

Extended analysis of a genome-wide association study in primary sclerosing cholangitis detects multiple novel risk loci

Trine Folseraas^{1,2,3,†}, Espen Melum^{1,2,3,†}, Philipp Rausch^{4,5}, Brian D. Juran⁶, Eva Ellinghaus⁷, Alexey Shiryayev^{1,2,3}, Jon K. Laerdahl^{8,9}, David Ellinghaus⁷, Christoph Schramm¹⁰, Tobias J. Weismüller^{11,12}, Daniel Nils Gotthardt¹³, Johannes Roksund Hov^{1,2,3}, Ole Petter Clausen^{3,14}, Rinse K. Weersma¹⁵, Marcel Janse¹⁵, Kirsten Muri Boberg¹, Einar Björnsson¹⁶, Hanns-Ulrich Marschall¹⁶, Isabelle Cleynen¹⁷, Philip Rosenstiel⁷, Kristian Holm¹, Andreas Teufel¹⁸, Christian Rust¹⁹, Christian Gieger²⁰, H-Erich Wichmann^{21,22,23}, Annika Bergquist²⁴, Euijung Ryu²⁵, Cyriel Y. Ponsioen²⁶, Heiko Runz²⁷, Martina Sterneck²⁸, Severine Vermeire¹⁷, Ulrich Beuers²⁶, Cisca Wijmenga²⁹, Erik Schruppf^{1,3}, Michael P. Manns^{11,12}, Konstantinos N. Lazaridis⁶, Stefan Schreiber^{7,30}, John F. Baines^{4,5}, Andre Franke⁷, Tom H. Karlsen^{1,2,31,*}

¹Norwegian PSC Research Center, Department of Transplantation Medicine, Oslo University Hospital, Rikshospitalet, Oslo, Norway; ²Research Institute for Internal Medicine, Oslo University Hospital, Rikshospitalet, Oslo, Norway; ³Faculty of Medicine, University of Oslo, Oslo, Norway; ⁴Institute for Experimental Medicine, Christian-Albrechts-University, Kiel, Germany; ⁵Max Planck Institute for Evolutionary Biology, Plön, Germany; ⁶Center for Basic Research in Digestive Diseases, Division of Gastroenterology and Hepatology, Mayo Clinic, College of Medicine, Rochester, Minnesota, United States; ⁷Institute of Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Germany; ⁸Centre for Molecular Biology and Neuroscience (CMBN) and Department of Microbiology, Oslo University Hospital, Rikshospitalet, Oslo, Norway; ⁹Bioinformatics Core Facility, Department of Informatics, University of Oslo, Oslo, Norway; ¹⁰1st Department of Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; ¹¹Department of Gastroenterology, Hepatology and Endocrinology, Hannover Medical School, Hannover, Germany; ¹²Integrated Research and Treatment Center-Transplantation (IFB-tx), Hannover Medical School, Hannover, Germany; ¹³Department of Medicine, University Hospital of Heidelberg, Heidelberg, Germany; ¹⁴Division of Pathology, Oslo University Hospital Rikshospitalet, Oslo, Norway; ¹⁵Department of Gastroenterology and Hepatology, University Medical Center Groningen and University of Groningen, The Netherlands; ¹⁶Department of Internal Medicine, Institute of Medicine, Sahlgrenska Academy and University Hospital, Gothenburg, Sweden; ¹⁷Department of Gastroenterology, University Hospital Gasthuisberg, Leuven, Belgium; ¹⁸1st Department of Medicine, University of Mainz, Mainz, Germany; ¹⁹Department of Medicine 2, Grosshadern, University of Munich, Munich, Germany; ²⁰Institute of Genetic Epidemiology, Helmholtz Center Munich, German Research Center for Environmental Health, Neuherberg, Germany; ²¹Institute of Epidemiology I, Helmholtz Center Munich, German Research Center for Environmental Health, Neuherberg, Germany; ²²Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany; ²³Klinikum Grosshadern, Munich, Germany; ²⁴Department of Gastroenterology and Hepatology, Karolinska University Hospital Huddinge, Stockholm, Sweden; ²⁵Division of Biomedical Statistics and Informatics, Mayo Clinic College of Medicine, Rochester, Minnesota, United States; ²⁶Department of Gastroenterology and Hepatology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands; ²⁷Department of Human Genetics, University Hospital of Heidelberg, Heidelberg, Germany; ²⁸Department of Hepatobiliary Surgery and Transplantation, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; ²⁹Department of Genetics, University Medical Center Groningen and University of Groningen, Groningen, The Netherlands; ³⁰Department for General Internal Medicine, Christian-Albrechts-University, Kiel, Germany; ³¹Division of Gastroenterology, Institute of Medicine, University of Bergen, Bergen, Norway

Keywords: Primary sclerosing cholangitis; Genome-wide association study; Single nucleotide polymorphism; Immunogenetics.

Received 6 October 2011; received in revised form 7 March 2012; accepted 26 March 2012; available online 18 April 2012

* Corresponding author. Address: Norwegian PSC Research Center, Department of Transplantation Medicine, Oslo University Hospital Rikshospitalet, Postboks 4950 Nydalen, N-0424 Oslo, Norway. Tel.: +47 23 07 2469; fax: +47 23 07 4869.

E-mail address: t.h.karlsen@medisin.uio.no (T.H. Karlsen).

† These authors contributed equally to this work.

Abbreviations: PSC, primary sclerosing cholangitis; SNP, single nucleotide polymorphism; GWAS, genome-wide association study; IBD, inflammatory bowel disease; CEPH, Centre d'Etude du Polymorphisme Humain; HWE, Hardy-Weinberg-Equilibrium; LD, linkage disequilibrium; CMH, Cochran-Mantel-Haenszel; BD, Breslow-Day; HET, heterogeneity; PBC, primary biliary cirrhosis; Chr, chromosome; Al, alleles; OR, odds ratio; CI, confidence interval; GRAIL, Gene Relationships Across Implicated Loci.



ELSEVIER

Background & Aims: A limited number of genetic risk factors have been reported in primary sclerosing cholangitis (PSC). To discover further genetic susceptibility factors for PSC, we followed up on a second tier of single nucleotide polymorphisms (SNPs) from a genome-wide association study (GWAS).

Methods: We analyzed 45 SNPs in 1221 PSC cases and 3508 controls. The association results from the replication analysis and the original GWAS (715 PSC cases and 2962 controls) were combined in a meta-analysis comprising 1936 PSC cases and 6470 controls. We performed an analysis of bile microbial community composition in 39 PSC patients by 16S rRNA sequencing.

Results: Seventeen SNPs representing 12 distinct genetic loci achieved nominal significance ($p_{\text{replication}} < 0.05$) in the replication. The most robust novel association was detected at chromosome 1p36 (rs3748816; $p_{\text{combined}} = 2.1 \times 10^{-8}$) where the *MMEL1* and *TNFRSF14* genes represent potential disease genes. Eight additional novel loci showed suggestive evidence of association ($p_{\text{rep}} < 0.05$). *FUT2* at chromosome 19q13 (rs602662; $p_{\text{comb}} = 1.9 \times 10^{-6}$, rs281377; $p_{\text{comb}} = 2.1 \times 10^{-6}$ and rs601338; $p_{\text{comb}} = 2.7 \times 10^{-6}$) is notable due to its implication in altered susceptibility to infectious agents. We found that *FUT2* secretor status and genotype defined by rs601338 significantly influence biliary microbial community composition in PSC patients.

Conclusions: We identify multiple new PSC risk loci by extended analysis of a PSC GWAS. *FUT2* genotype needs to be taken into account when assessing the influence of microbiota on biliary pathology in PSC.

© 2012 European Association for the Study of the Liver. Published by Elsevier B.V. All rights reserved.

Introduction

Primary sclerosing cholangitis (PSC) is a chronic cholestatic liver disease of unknown etiology, characterized by progressive inflammation and fibrosis of the bile ducts, leading to liver cirrhosis in many cases [1]. An important clinical feature of PSC is the frequent autoimmune manifestations in other organ systems, most commonly inflammatory bowel disease (IBD), which is reported in 62–83% of PSC patients of Northern European descent [1]. In addition, approximately 25% of PSC patients are affected with at least one autoimmune disease outside the liver and colon [2], most commonly, type 1 diabetes, thyroid disease, rheumatoid arthritis and psoriasis. Shared genetic susceptibility could potentially explain the frequent occurrence of these immune-related co-morbidities in PSC.

The importance of genetic risk factors in PSC is demonstrated by heritability studies estimating siblings of PSC patients to be 9–39 times more likely to develop PSC than the general population [1]. The contribution of genetic variants in the HLA complex on chromosome 6p21 to the risk of PSC is well established [1]. Previously, strong evidence for associated risk factors outside the HLA region has been reported at chromosome 3p21 in *MST1* and chromosome 2q13 near *BCL2L11* [3]. In addition, suggestive PSC associations have been reported at six additional loci harboring the likely susceptibility genes *GPBAR1*, *IL2RA*, *GPC5/GPC6*, *IL2/IL21*, *CARD9*, and *REL* [3–6]. In the most recent PSC GWAS [3], only the top 23 associated regions were selected for replication genotyping.

Replication attempts of promising, but lower ranked markers from GWAS have yielded valuable findings in several other

diseases, and have been particularly fruitful when combined with a semi-hypothesis-driven approach taking into consideration gene content and potential biological relevance [7]. In an attempt to identify novel susceptibility loci in PSC, we selected a second tier of promising associated markers from an available PSC GWAS [3] for replication in an independent cohort.

Materials and methods

Study subjects

The discovery panel included a total of 715 PSC cases and 2962 healthy controls, with 332 PSC cases and 262 controls from Scandinavia, and 383 PSC cases and 2700 controls from Germany [3]. The replication panel consisted of a total of 1221 PSC cases and 3508 controls, with 289 PSC cases and 820 controls from Scandinavia, 561 PSC patients and 2063 controls from Central Europe, and 371 PSC cases and 625 controls from the United States (US).

The recruitment of study subjects is described in detail in the Supplementary Methods section of the [Supplementary Material](#).

Written informed consent was obtained from all study participants. The study conforms to the ethical guidelines of the 1975 Declaration of Helsinki and was approved by the South-Eastern Norwegian Regional Ethics Committee (S-93178 and S-08872b).

Association analysis of genome-wide data

Genome-wide genotyping, imputation and quality control in the discovery panel are described in the [Supplementary Methods](#). Association analysis of the genotyped and imputed SNPs in the discovery panel was performed using a logistic regression procedure implemented in the R statistical package version 2.9.1 (<http://www.r-project.org>). The six first principal components generated by the EIGENSTRAT software [8] after removal of population outliers were included as co-variables in the regression to correct for residual population structure.

Selection of SNPs for extended replication

A total of 2,466,182 SNPs were included in the association analysis of the GWAS dataset. All nominally associated SNPs (p value < 0.05) outside the HLA complex (defined as position 25–35 million base pairs on chromosome 6p21), which had not been previously subjected to replication genotyping, were considered potential candidates for replication. We aimed at integrating a priori knowledge on disease mechanisms in PSC and genetic susceptibility factors for related conditions and traits by implementing two SNP prioritization strategies for replication genotyping.

Strategy I

All non-HLA SNPs with a GWAS p value (p_{GWAS}) < 0.05 ($n = 134,466$) were aligned with SNPs reported to be associated in immune-mediated and chronic inflammatory diseases and relevant biochemical parameters as listed in the Catalog of Genome-Wide Association Studies (<http://www.genome.gov/26525384>) (accessed 23.04.2010). [Supplementary Table 2](#) lists the phenotypes taken into consideration.

Strategy II

Genetic loci harboring SNPs with robust statistical evidence of association in the GWAS ($p_{\text{GWAS}} < 1.0 \times 10^{-4}$) underwent literature-based assessment taking potential pathogenetic mechanisms in PSC into account [1]. Before this assessment, the number of candidate SNPs with a $p_{\text{GWAS}} < 1.0 \times 10^{-4}$ was reduced by applying a linkage disequilibrium (LD) clumping threshold of $r^2 = 0.8$ and distance = 50 kb in PLINK version 1.06 [9] using LD measurements from the HapMap project.

To further avoid redundant genotyping, the total number of SNPs selected with strategies I and II was clumped using a LD clumping threshold of $r^2 = 0.8$ and distance = 2000 kb in PLINK version 1.06 [9]. In *FUT2*, two redundant coding SNPs (rs601338 and rs281377) were purposely retained. In addition, a tag (rs4143332, $r^2 = 1.0$, $D' = 1.0$) for the most strongly associated SNP (rs3134792) in *HLA-B* was included. If assay design was not possible for the index SNPs from the clumping, SNPs identified to be in LD with the SNP in the clumping procedure were used instead.

Research Article

Table 1. Clinical characteristics of the PSC patients included in the study.

	Genome-wide analysis			Replication analysis	
	Scandinavian	German	Scandinavian	Central Europe	United States
Number of PSC patients	332	383	289	561	371
Median age at diagnosis (range)	34 (10-76)	32 (9-76)	37 (12-74)	35 (2-81)	41 (8-77)
Male patients (%)	241 (73%)	259 (69%)	199 (70%)	354 (64%)	239 (64%)
IBD data available for (n) patients	331	364	249	544	329
IBD (%)	260 (79%)	245 (67%)	196 (79%)	347 (64%)	262 (80%)
IBD subtype					
Ulcerative colitis (%)	64%	51%	66%	47%	66%
Crohn's disease (%)	10%	10%	8%	13%	8%
IBD unclassified (%)	4%	7%	5%	4%	6%
CCA data available for (n) patients	330	375	243	461	329
CCA diagnosed (%)	35 (11%)	10 (3%)	27 (11%)	28 (6%)	27 (8%)

Clinical characteristics of the patients included in the study stratified according to study panel in the genome-wide analysis or the replication analysis. IBD, inflammatory bowel disease; CCA, cholangiocarcinoma.

Association analysis of replication data and meta-analysis

Replication genotyping and quality control are described in the [Supplementary Methods](#). Association analysis in the three replication panels was performed by the Cochran-Mantel-Haenszel (CMH) test as implemented in the PLINK version 1.06 [9]. Along with the CMH test, a Breslow-Day (BD) test for heterogeneity of odds ratios was performed. Meta-analysis of the summary statistics in the discovery and replication panels was performed using the Meta-Analysis Tool for genome-wide association scans, METAL (<http://www.sph.umich.edu/csg/abecasis/Metal>) (version released 2010-08-01) (see [Supplementary Methods](#) for details).

To adjust for multiple testing in the replication, a strict Bonferroni correction was applied assuming independence between the 45 SNPs analyzed (Bonferroni-adjusted p value threshold at $0.05/45 = 0.0011$).

Association signal plots, linkage disequilibrium calculations, and pathway analysis

Regional association plots were generated from the GWAS data using the LocusZoom software ([Supplementary Methods](#)) [10]. For all the other LD calculations, data from the HapMap project and PLINK version 1.06 [9] were used. To examine the functional relationship among genomic PSC risk regions, we performed a Gene Relationships Across Implicated Loci (GRAIL) pathway analysis (<http://www.broadinstitute.org/mpg/grail/>) ([Supplementary Methods](#)).

Biliary FUT2 phenotyping, FUT2 sequence alignment, and profiling of biliary microbiota composition

To demonstrate presence of FUT2 in the bile duct epithelium, immunostaining for the α (1,2) fucose-specific lectin *Ulex europaeus* agglutinin-I was performed. The biliary FUT2 phenotyping is described in detail in [Supplementary Methods](#). Sequence alignment of human *FUT2* with a range of different species was performed to evaluate evolutionary conservation (see figure legend of [Supplementary Fig. 2](#)). A fragment of the 16S rRNA gene spanning the V1 and V2 hypervariable regions was amplified from processed bile samples of PSC patients and sequenced using 454 GS-FLX Titanium sequencing chemistry as described [11] ([Supplementary Methods](#)).

Results

SNP selection and genotyping results

Seven hundred and fifteen PSC patients and 2962 controls were successfully genotyped and analyzed in the GWAS ([Table 1](#)).

Among the 2,466,182 genotyped and imputed SNPs analyzed, a subset of 59 SNPs were selected for follow-up; 37 SNPs based on SNP selection strategy I and 22 SNPs based on SNP selection strategy II. Following quality pruning of the replication dataset, a total of 45 SNPs were included in the association analysis.

Association results

A total of 17 SNPs achieved nominal significance in the replication ($p_{\text{repl}} < 0.05$) and demonstrated effect sizes in the same direction as observed in the discovery cohort ([Table 2](#)). Results for the additional 28 SNPs studied are provided in [Supplementary Table 1](#). The associated SNPs implicate 12 genetic loci, nine of which had not been previously reported to be associated with PSC. The association results for SNPs with nominally significant replication with the same direction of effect were homogenous ($P_{\text{HETEROGENEITY (HET)}} > 0.05$) when the GWAS discovery and replication cohorts were compared, except for one marker (rs11936230). Within the three panels constituting the replication cohort, the significant association results were also largely homogenous, however, two markers (rs11682163 and rs281377) had a significant Breslow-Day test, implying heterogeneity of odds ratios in between the three panels for these SNPs. Four SNPs at *MMEL1/TNFRSF14*, *IL2/IL21*, and *CARD9* demonstrated association results robust to correction for multiple testing using Bonferroni's method ($p < 0.0011$) in the replication analysis ([Table 2](#)).

The strongest finding in the replication panel was the missense SNP rs3748816 (M[ATG] \rightarrow T[ACG]) located in *MMEL1*, which achieved genome-wide significance [12] in the combined analysis ($p_{\text{comb}} = 2.1 \times 10^{-8}$) ([Table 2](#)). The association signal at this locus (1p36) encompasses *MMEL1* and extends into the tumor necrosis factor receptor superfamily member 14 (*TNFRSF14*) gene ([Fig. 1](#)).

FUT2 associations and bile microbial community composition

Eight novel loci demonstrated suggestive evidence for association in the replication panel ($p_{\text{repl}} < 0.05$), but were not robust to

Table 2. Allele frequencies and results from the association analyses for SNPs with association results reaching nominal significance with consistent effect sizes in the replication.

Chr	SNP	Position	Locus	AI	Genome-wide analysis				Replication analysis								
					Allele frequencies (Cases/Controls)				Allele frequencies (Cases/Controls)								
					Scandinavia (332/262)	Germany (383/2700)	<i>p</i> value*	OR (95% CI)*	Scandinavia (289/820)	Central Europe (561/2063)	United States (371/625)	<i>p</i> value†	OR (95% CI)†	BD <i>p</i> value	Combined <i>p</i> value	Het <i>p</i> value	Selec- tion‡
1	rs3748816	2,516,606	<i>TNFRSF14/</i> <i>MMEL1</i>	G/A	0.29/0.34	0.28/0.33	6.5 × 10 ⁻⁴	0.79 (0.69-0.90)	0.31/0.35	0.30/0.34	0.27/0.35	8.7 × 10 ⁻⁶	0.79 (0.71-0.88)	0.45	2.1 × 10 ⁻⁸	0.92	I
4	rs13132245	123,269,042	<i>KIAA1109/</i> <i>IL2/IL21</i>	G/A	0.16/0.21	0.12/0.17	7.3 × 10 ⁻⁵	0.69 (0.57-0.83)	0.13/0.18	0.12/0.14	0.12/0.15	3.0 × 10 ⁻⁴	0.77 (0.67-0.89)	0.65	1.2 × 10 ⁻⁷	0.40	II
4	rs13119723	123,437,763	<i>KIAA1109/</i> <i>IL2/IL21</i>	G/A	0.13/0.18	0.09/0.14	4.1 × 10 ⁻⁵	0.65 (0.53-0.80)	0.15/0.19	0.13/0.15	0.12/0.16	4.0 × 10 ⁻⁴	0.78 (0.68-0.90)	0.87	9.9 × 10 ⁻⁸	0.32	II
9	rs4077515	138,386,317	<i>CARD9</i>	T/C	0.48/0.45	0.43/0.40	0.045	1.14 (1.00-1.31)	0.49/0.45	0.47/0.42	0.45/0.42	5.3 × 10 ⁻⁴	1.18 (1.08-1.30)	0.95	7.6 × 10 ⁻⁵	0.55	I
16	rs2903692	11,146,284	<i>CLEC16A</i>	A/G	0.28/0.33	0.31/0.36	0.0014	0.80 (0.92-0.70)	0.27/0.33	0.32/0.35	0.32/0.36	0.0012	0.85 (0.76-0.94)	0.43	6.4 × 10 ⁻⁶	0.64	I
2	rs13017599	61,017,835	<i>REL</i>	A/G	0.40/0.35	0.40/0.37	0.027	1.16 (1.32-1.02)	0.40/0.36	0.42/0.38	0.38/0.37	0.0026	1.16 (1.05-1.28)	0.42	1.9 × 10 ⁻⁴	0.88	I
21	rs11203203	42,709,255	<i>UBASH3A</i>	A/G	0.39/0.35	0.42/0.37	0.0025	1.22 (1.39-1.07)	0.39/0.34	0.39/0.36	0.40/0.36	0.0033	1.16 (1.05-1.28)	0.68	2.9 × 10 ⁻⁵	0.60	I
19	rs602662	53,898,797	<i>FUT2</i>	A/G	0.56/0.50	0.53/0.46	7.7 × 10 ⁻⁵	1.30 (1.49-1.14)	0.53/0.47	0.48/0.45	0.51/0.50	0.0034	1.15 (1.05-1.27)	0.29	1.9 × 10 ⁻⁶	0.21	II
21	rs9976767	42,709,459	<i>UBASH3A</i>	G/A	0.48/0.46	0.48/0.44	0.028	1.15 (1.02-1.30)	0.45/0.43	0.47/0.44	0.49/0.43	0.0055	1.14 (1.04-1.26)	0.44	4.0 × 10 ⁻⁴	0.99	I
19	rs601338	53,898,486	<i>FUT2</i>	A/G	0.50/0.44	0.47/0.40	4.5 × 10 ⁻⁵	1.33 (1.52-1.16)	0.51/0.44	0.46/0.42	0.48/0.48	0.0061	1.14 (1.04-1.26)	0.062	2.7 × 10 ⁻⁶	0.14	II
2	rs11682163	3,727,421	<i>ALLC</i>	C/T	0.49/0.39	0.49/0.44	7.9 × 10 ⁻⁵	1.29 (1.47-1.14)	0.50/0.45	0.49/0.44	0.43/0.45	0.010	1.13 (1.03-1.24)	0.048	7.6 × 10 ⁻⁶	0.14	II
19	rs281377	53,898,415	<i>FUT2</i>	T/C	0.43/0.49	0.46/0.52	9.3 × 10 ⁻⁶	0.71 (0.61-0.83)	0.39/0.47	0.47/0.49	0.44/0.44	0.011	0.88 (0.80-0.97)	0.025	2.1 × 10 ⁻⁶	0.059	II
22	rs5771069	48,777,607	<i>IL17REL</i>	A/G	0.48/0.54	0.49/0.53	0.0057	0.82 (0.95-0.72)	0.45/0.49	0.47/0.50	0.48/0.50	0.028	0.90 (0.82-0.99)	0.89	5.8 × 10 ⁻⁴	0.43	I
4	rs11936230	123,236,205	<i>KIAA1109/</i> <i>IL2/IL21</i>	T/C	0.38/0.30	0.41/0.35	1.9 × 10 ⁻⁵	1.33 (1.17-1.52)	0.38/0.34	0.38/0.36	0.38/0.36	0.039	1.11 (1.01-1.22)	0.79	1.9 × 10 ⁻⁵	0.040	II
12	rs4149056	21,222,816	<i>SLCO1B1</i>	C/T	0.13/0.17	0.16/0.17	0.044	0.84 (1.00-0.70)	0.14/0.17	0.15/0.16	0.13/0.15	0.043	0.87 (0.76-1.00)	0.60	0.0045	0.76	I
22	rs4820599	23,320,213	<i>GGT1</i>	G/A	0.29/0.24	0.31/0.28	0.035	1.16 (1.01-1.33)	0.29/0.28	0.29/0.27	0.31/0.28	0.043	1.11 (1.00-1.23)	0.77	0.0038	0.69	I
12	rs3184504	110,368,991	<i>SH2B3</i>	T/C	0.54/0.52	0.57/0.51	0.0016	1.23 (1.08-1.40)	0.47/0.45	0.51/0.49	0.51/0.47	0.050	1.10 (1.00-1.21)	0.67	4.6 × 10 ⁻⁴	0.21	I

Complete association results for SNPs reaching nominal significance with consistent effect sizes in the replication. For the genome-wide analysis, the allele frequencies were calculated based on allele dosages and are listed separately for the German and Scandinavian discovery panels. For the replication analysis, allele frequencies are given for all three panels making up the combined replication panel. The combined *p* values for the genome-wide analysis and the replication analysis were generated using the METAL software [34]. The genomic positions refer to NCBI's build 36. SNP selection strategies I and II indicate SNPs selected based on the two SNP selection strategies. The HLA association was confirmed for rs4143332 ($p_{\text{repl}} = 4.1 \times 10^{-77}$, $p_{\text{comb}} = 4.4 \times 10^{-123}$) (not listed in the table).

Chr, chromosome; AI, alleles; OR, odds ratio; CI, confidence interval; BD, Breslow-Day; Het, heterogeneity.

*Odds ratios and *p* values derived from logistic regressions of allele dosages including the six first principal components from the principal components analysis as covariates.

†*p* values generated by using the Cochran-Mantel-Haenszel test.

‡SNP selection strategy.

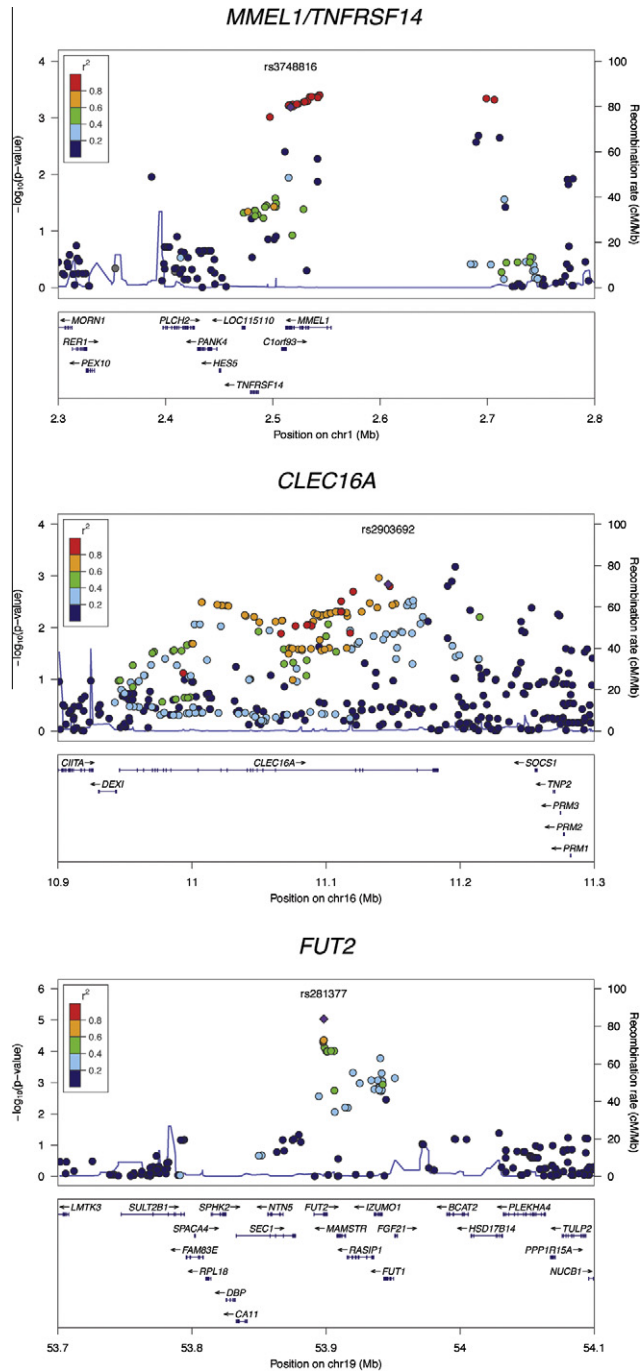


Fig. 1. Regional association plots for MMEL1/TNFRSF14, CLEC16A and FUT2. The association results for both the genotyped and imputed SNPs are represented by the $-\log_{10} p$ value plotted against the genomic position. The index SNP is indicated by a purple diamond while the colors of the remaining SNPs indicate the linkage disequilibrium with the index SNP. The recombination rates were derived from the HapMap project and are represented by the thin blue lines. The plots were generated using the LocusZoom software [10].

correction for multiple testing (Table 2). Of these, three SNPs in the Fucosyltransferase 2 gene (*FUT2*, Table 2 and Fig. 1) at 19q13 (a synonymous SNP, rs281377 ($p_{\text{comb}} = 2.1 \times 10^{-6}$), a non-sense SNP, rs601338 (W (TGG) \rightarrow *(TAG)) ($p_{\text{comb}} = 2.7 \times 10^{-6}$) and a

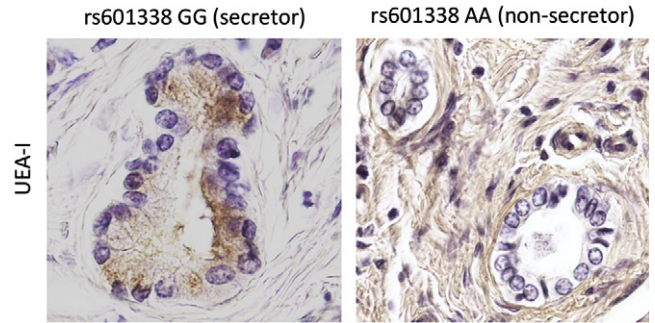


Fig. 2. Biliary FUT2 phenotyping. Lectin staining of the hilar liver biopsy specimens from PSC patients. Paraffin-embedded sections from individuals with AA (non-secretor) and GG (secretor) variants of the *FUT2* rs601338 SNP were used to evaluate the expression of $\alpha(1,2)$ fucosylated glycans in the bile duct epithelium. H antigen, detected with the $\alpha(1,2)$ fucose-specific lectin *Ulex europaeus* agglutinin-I (UEA-I) (brown staining), is expressed on the apical surface of the biliary epithelial layer of the secretor variant, while it is absent on non-secretor epithelia (Original magnification 400x).

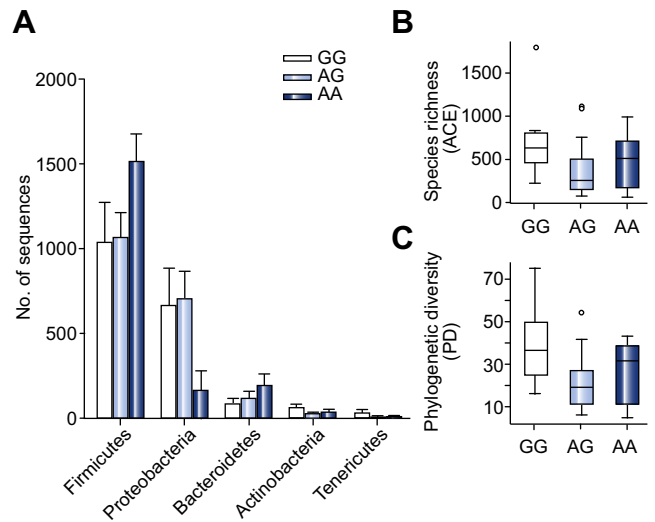


Fig. 3. Influence of FUT2 genotype on phyla abundances and alpha diversity. (A) Mean abundances (\pm SE) of the major phyla with respect to genotype at the non-sense SNP rs601338 (W (TGG) \rightarrow *(TAG)). (B) The Abundance based Coverage Estimator (ACE) as a measure of the approximated species richness [35]. (C) Alpha diversity measured as the sum of the total branch length in each sample (i.e. Phylogenetic Diversity) [36]. All values are based on the normalized dataset (2000 reads per individual).

missense SNP, rs602662 (G (GGT) \rightarrow S (ACT)) ($p_{\text{comb}} = 1.9 \times 10^{-6}$) are of particular interest, given the associations between *FUT2* genotype and several infectious diseases. Strong LD between these SNPs ($r_{\text{rs601338 vs. rs281377}}^2 = 0.84$, $r_{\text{rs601338 vs. rs602662}}^2 = 0.76$ and $r_{\text{rs281377 vs. rs602662}}^2 = 0.57$) implies that the putatively functional allele(s) cannot be directly determined by the association analysis. The rs601338 polymorphism generates a premature stop codon (W143X) and a truncated, dysfunctional *FUT2* protein [13]. By immunostaining for the $\alpha(1,2)$ fucose-specific lectin *Ulex europaeus* agglutinin-I (Fig. 2), we show that this stop codon appears to have a profound effect on expression of $\alpha(1,2)$ fucosylated glycans in bile duct epithelium. The rs602662 polymorphism results in an amino acid substitution (G247S) which may be detrimental to

FUT2 function, as shown by strong evolutionary conservation of this residue (Supplementary Fig. 2).

To evaluate a potential functional role of FUT2 secretor status and genotype as defined by rs601338 in the pathophysiology of PSC, we analyzed the bacterial community composition of bile samples from 39 PSC patients (8 homozygous for the functional allele "G", 21 heterozygous and 10 homozygous for the loss-of-function allele "A" (W (TGG) → *(TAG)), since we previously observed genotype-dependent changes in community structure of the colon in the context of Crohn's disease [11]. Intriguingly, the abundances of several phyla displayed significant differences with respect to FUT2 genotype and secretor status similar to those observed in the colon. In particular, the Firmicutes are significantly elevated and the Proteobacteria significantly decreased among non-secretors, in addition to differences observed in the Bacteroidetes, Actinobacteria, and Tenericutes (Fig. 3A and Supplementary Table 3). Similarly, alpha diversity measures display significant differences with respect to genotype similarly to what observed in the colon, i.e. a decrease of diversity in the heterozygous state compared to both homozygous genotypes (Fig. 3B and C and Supplementary Table 4). Finally, differences in inter-individual variability, i.e. beta diversity, are also apparent (*adonis*: unweighted UniFrac- $r^2 = 0.074$, $p = 0.055$ (genotype); normalized weighted UniFrac- $r^2 = 0.057$, $p = 0.049$ (secretor status)) (Supplementary Fig. 4A and B). Using the full species assemblage to test for differentiation with respect to FUT2 genotype also revealed marginal results (Redundancy Analysis: $F = 1.140$, $p = 0.085$ (Supplementary Fig. 4C).

Additional suggestive associations

Among the other associated SNPs, rs2903692 ($p_{\text{comb}} = 6.4 \times 10^{-6}$) is located in the gene C-type lectin domain family 16, member A (*CLEC16A*, Table 2 and Fig. 1). Two SNPs, rs11203203 ($p_{\text{comb}} = 2.9 \times 10^{-5}$) and rs9976767 ($p_{\text{comb}} = 4.0 \times 10^{-4}$), localize in an intron of ubiquitin associated and SH3 domain containing A gene (*UBASH3A*, Table 2) and are in LD ($r^2 = 0.63$). The rs11682163 SNP ($p_{\text{comb}} = 7.6 \times 10^{-6}$) is localized in allantoicase (*ALLC*, Table 2) at 2p25. At 22q13, rs5771069 ($p_{\text{comb}} = 5.8 \times 10^{-4}$) is a missense SNP (L(CTT) → P(CCT)) in interleukin 17 receptor E-like (*IL17REL*, see Table 2). At 12p12, the rs4149056 SNP ($p_{\text{comb}} = 0.0045$) is a missense SNP (V(GTG) → A(GCG) in an exon of the solute carrier organic anion transporter family gene (*SLCO1B1*). The rs4820599 SNP ($p_{\text{comb}} = 0.0038$) is localized in an intron of the gamma-glutamyltransferase 1 gene (*GGT1*) at 22q11. The rs3184504 SNP ($p_{\text{comb}} = 4.6 \times 10^{-4}$) represents a missense mutation (W (TGG) → R (CGG) in an exon of the SH2B adaptor protein 3 gene (*SH2B3*, Table 2) at 12q24.

The associated SNPs at 4q27 (rs13132245; $p_{\text{comb}} = 1.2 \times 10^{-7}$ and rs13119723; $p_{\text{comb}} = 9.9 \times 10^{-8}$) are localized 42 kb upstream of and in an intron of *KIAA1109*, respectively. Both rs13119723 and another genetic variant residing in this region, rs6822844, which is in LD with the replicated SNPs at rs13132245 ($r^2 = 0.85$) and rs13119723 ($r^2 = 0.66$), have been reported nominally associated with PSC in two candidate gene studies [5,6]. The replicated SNPs at 9q34 (rs4077515; $p_{\text{comb}} = 7.6 \times 10^{-5}$) in the caspase-recruitment domain family, member 9 gene (*CARD9*) and at 2p16 (rs13017599; $p_{\text{comb}} = 1.9 \times 10^{-4}$) near v-rel reticuloendotheliosis viral oncogene (*REL*) (Table 2 and Supplementary Fig. 1) also belong to loci which previously have been reported nominally associated with

PSC [5]. A GRAIL analysis of previously published and hereby detected PSC risk loci demonstrated that several of the loci are functionally related (Supplementary Fig. 3A and B).

Discussion

In the largest PSC cohort presented, we performed replication genotyping and a combined analysis of 45 SNPs not followed up in a previous GWAS in PSC [3]. We identified one novel PSC risk locus with association results below the threshold for genome-wide significance along with suggestive evidence for replication of 8 novel additional loci. The PSC-associated *FUT2* variant was shown to significantly influence the bile microbial community composition in PSC patients.

The replicated SNP at 1p36 (*MMEL1-TNFRSF14*) represents the first genetic overlap demonstrated between PSC and primary biliary cirrhosis (PBC) [14]. Interestingly, the SNP demonstrates opposite effect sizes in PSC and PBC [14]. Considering the peak association signal of the observed non-synonymous SNP at *MMEL1* (Fig. 1), the suggestive presence of *MMEL1* protein expression in bile duct cells and glandular cells of the gall bladder (<http://www.proteinatlas.org>) and the fundamental role of other membrane metallo-endopeptidase family members in processes of metabolism, *MMEL1* has a relevant candidate role at this locus. For PSC, *TNFRSF14* is also an intriguing candidate. The *TNFRSF14* protein is expressed on T lymphocytes and the mucosal epithelium and acts as a receptor for the proinflammatory cytokine LIGHT [15]. *TNFRSF14* signaling has also inhibitory effects on B and T cells, and this dual role is thought to regulate immune tolerance [16]. Mice lacking *Btla* (*Btla*^{-/-} mice), the mediator responsible for the inhibitory actions of *TNFRSF14*, spontaneously develop an autoimmune hepatitis-like disease accompanied by severe bile duct epithelium inflammation with nuclear pleomorphism and irregularity of duct outlines [17], paralleling features of histopathological liver biopsy specimens in human PSC [18].

FUT2 encodes an enzyme (Galactoside 2- α -L-fucosyltransferase 2) involved in protein glycosylation, including the ABH blood-antigen synthesis pathway [13]. Genetic variants giving rise to a non-functional, truncated *FUT2* enzyme result in an inability to synthesize ABH antigens on mucosal surfaces and in salivary glands, which is referred to as non-secretor status [13]. *FUT2*-determined secretor status has been observed to affect susceptibility to a number of infectious agents [19], possibly by altering the recognition and binding of pathogen adhesins to their preferred carbohydrate receptors on mucosal surfaces [20]. Homozygosity for the PSC-associated non-sense variant rs601338 has been reported to account for >95% of the prevalence of the non-secretor status in Caucasians [21]. The non-sense SNP rs601338 is also a risk factor for Crohn's disease [19]. Although the bile duct is an environment distinct from the portions of the GI tract most frequently afflicted by Crohn's disease, we observe strikingly similar changes with respect to the abundance of major bacterial phyla and the level and pattern of bacterial diversity compared to previous observations in the colon [11]. Thus, these general patterns appear to be consistent between different habitats throughout the GI tract, which may underlie the overlapping association of *FUT2* to different chronic inflammatory disorders. Importantly, based on the present data, further enquiries into this topic need to take into account both biliary *FUT2* expression and the biliary microbiome in the contaminated bile

Research Article

Table 3. Associations in other autoimmune diseases for PSC loci with robust or suggestive association in the current study.

Locus	Candidate gene(s) and SNP(s) associated in the current study	Autoimmune phenotypes with a reported association at the locus	Associated SNP	References		
1p36	<i>MMEL1-TNFRSF14</i> rs3748816	Celiac disease	rs3748816	Dubois <i>et al.</i> , <i>Nat Genet</i> 42, 295 (2010)		
		Multiple sclerosis	rs6684864	Blanco-Kelly <i>et al.</i> , <i>Genes Immun</i> 12, 145 (2011)		
			rs3748816	Ban <i>et al.</i> , <i>Genes Immun</i> 11, 660 (2010)		
		Primary biliary cirrhosis	rs3748816	Hirschfield <i>et al.</i> , <i>Nat Genet</i> 42, 655 (2010)		
		Rheumatoid arthritis	rs3890745	Raychaudhuri <i>et al.</i> , <i>Nat Genet</i> 40, 1216 (2008)		
		Ulcerative colitis	rs734999	Anderson <i>et al.</i> , <i>Nat Genet</i> 43, 246 (2011)		
2p15	<i>REL</i> rs13017599	Celiac disease	rs13003464	Dubois <i>et al.</i> , <i>Nat Genet</i> 42, 295 (2010)		
		Crohn's disease	rs10181042	Franke <i>et al.</i> , <i>Nat Genet</i> 42, 1118 (2010)		
		Psoriasis	rs702873	Strange <i>et al.</i> , <i>Nat Genet</i> 42, 985 (2010)		
		Psoriatic arthritis	rs13017599	Ellinghaus <i>et al.</i> , <i>J Invest Dermatol</i> 132, 1133 (2012)		
		Rheumatoid arthritis	rs13017599	Gregersen <i>et al.</i> , <i>Nat Genet</i> 41, 820 (2009)		
		Ulcerative colitis	rs13031237 rs13003464	Stahl <i>et al.</i> , <i>Nat Genet</i> 42, 508 (2010) McGovern <i>et al.</i> , <i>Nat Genet</i> 42, 332 (2010)		
2p25	<i>ALLC</i> ra11682163	n.a.				
4q27	<i>IL2-IL21</i> rs13132245 rs13119723 rs11936230	Celiac disease	rs13119723	van Heel <i>et al.</i> , <i>Nat Genet</i> 39, 827 (2007)		
			rs6822844	Hunt <i>et al.</i> , <i>Nat Genet</i> 40, 395 (2008)		
			rs13151961	Dubois <i>et al.</i> , <i>Nat Genet</i> 42, 295 (2010)		
		Graves' disease	rs17388568	Todd <i>et al.</i> , <i>Nat Genet</i> 39, 857 (2007)		
		Psoriatic arthritis	rs13119723	Liu <i>et al.</i> , <i>PLoS Genet</i> 4, e1000041 (2008)		
			rs13151961	Liu <i>et al.</i> , <i>PLoS Genet</i> 4, e1000041 (2008)		
		Rheumatoid arthritis	rs13119723	Stahl <i>et al.</i> , <i>Nat Genet</i> 42, 508 (2010)		
			rs6822844	Maiti <i>et al.</i> , <i>Arthritis Rheum</i> 62, 323 (2010)		
		Systemic lupus erythematosus	rs6822844	Maiti <i>et al.</i> , <i>Arthritis Rheum</i> 62, 323 (2010)		
		Sjögren's syndrome	rs6822845	Maiti <i>et al.</i> , <i>Arthritis Rheum</i> 62, 323 (2010)		
		Type 1 diabetes	rs4505848	Barrett <i>et al.</i> , <i>Nat Genet</i> 41, 703 (2009)		
			rs6534347	WTCCC, <i>Nature</i> 447, 661 (2007)		
		Ulcerative colitis	rs17388568	Anderson <i>et al.</i> , <i>Nat Genet</i> 43, 246 (2011)		
		9q35	<i>CARD9</i> rs4077515	Ankylosing spondylitis	rs4077515	Burton <i>et al.</i> , <i>Nat Genet</i> 39, 1329 (2007)
				rs4077515	Pointon <i>et al.</i> , <i>Genes Immun</i> 11, 490 (2010)	
	rs10781500	Evans <i>et al.</i> , <i>Nat Genet</i> 43, 761 (2011)				
Crohn's disease	rs4077515	Franke <i>et al.</i> , <i>Nat Genet</i> 42, 1118 (2010)				
	rs10870077	Zhemakova <i>et al.</i> , <i>Am J Hum Genet</i> 82, 1202 (2008)				
		Ulcerative colitis	rs4077515	McGovern <i>et al.</i> , <i>Nat Genet</i> 42, 332 (2010)		
			rs10781500	Barrett <i>et al.</i> , <i>Nat Genet</i> 41, 1330 (2009)		
			rs10781499	Anderson <i>et al.</i> , <i>Nat Genet</i> 43, 246 (2011)		
			rs10870077	Zhemakova <i>et al.</i> , <i>Am J Hum Genet</i> 82, 1202 (2008)		
12p12	<i>SLCO1B1</i> rs4149056	n.a.				

(continued on next page)

Table 3 (continued)

Locus	Candidate gene(s) and SNP(s) associated in the current study	Autoimmune phenotypes with a reported association at the locus	Associated SNP	References	
12q31	<i>SH2B3</i> rs3184504	Celiac disease	rs653178	Hunt <i>et al.</i> , <i>Nat Genet</i> 40, 395 (2008)	
			rs653178	Dubois <i>et al.</i> , <i>Nat Genet</i> 42, 295 (2010)	
		Multiple sclerosis	rs653178	Zhernakova <i>et al.</i> , <i>PLoS Genet</i> 7, e1002004 (2011)	
			rs3184504	Alcina <i>et al.</i> , <i>Genes Immun</i> 11, 439 (2010)	
		Rheumatoid arthritis	rs3184504	Stahl <i>et al.</i> , <i>Nat Genet</i> 42, 508 (2010)	
			rs653178	Zhernakova <i>et al.</i> , <i>PLoS Genet</i> 7, e1002004 (2011)	
		Systemic lupus erythematosus	rs17696736	Gateva <i>et al.</i> , <i>Nat Genet</i> 41, 1228 (2009)	
			rs3184504	Barrett <i>et al.</i> , <i>Nat Genet</i> 41, 703 (2009)	
		Type 1 diabetes	rs17696736	WTCCC, <i>Nature</i> 7, 661 (2007)	
16p14	<i>CLEC16A</i> rs2903692	Celiac disease	rs12928822	Dubois <i>et al.</i> , <i>Nat Genet</i> 42, 295 (2010)	
		Multiple sclerosis	rs11865121	De Jager <i>et al.</i> , <i>Nat Genet</i> 41, 776 (2009)	
		Primary biliary cirrhosis	rs2903692	Martinez <i>et al.</i> , <i>Ann Rheum Dis</i> 69, 309 (2010)	
			rs12924729	Mells <i>et al.</i> , <i>Nat Genet</i> 43, 329 (2011)	
		Rheumatoid arthritis	rs6498169	Martinez <i>et al.</i> , <i>Ann Rheum Dis</i> 69, 309 (2010)	
			rs12708716	Gateva <i>et al.</i> , <i>Nat Genet</i> 41, 1228 (2009)	
		Systemic lupus erythematosus	rs12708716	Cooper <i>et al.</i> , <i>Nat Genet</i> 40, 1399 (2008)	
			rs2903692	Martinez <i>et al.</i> , <i>Ann Rheum Dis</i> 69, 309 (2010)	
19q13	<i>FUT2</i> rs602662 rs601338 rs281377	Crohn's disease	rs504963	McGovern <i>et al.</i> , <i>Hum Mol Genet</i> 19, 3468 (2010)	
			rs602662	McGovern <i>et al.</i> , <i>Hum Mol Genet</i> 19, 3468 (2010)	
			rs601338	McGovern <i>et al.</i> , <i>Hum Mol Genet</i> 19, 3468 (2010)	
			rs281379	Franke <i>et al.</i> , <i>Nat Genet</i> 42, 1118 (2010)	
22q11	<i>GGT1</i> rs4820599	n.a.			
21q22	<i>UBASH3A</i> rs11203203 rs9976767	Celiac disease	rs11203203	Zhernakova <i>et al.</i> , <i>PLoS Genet</i> 7, e1002004 (2011)	
		Rheumatoid arthritis	rs11203203	Stahl <i>et al.</i> , <i>Nat Genet</i> 42, 508 (2010)	
			Type 1 diabetes	rs11203203	Zhernakova <i>et al.</i> , <i>PLoS Genet</i> 7, e1002004 (2011)
				rs9976767	Grant <i>et al.</i> , <i>Diabetes</i> 58, 290 (2009)
			Vitiligo	rs11203203	Barrett <i>et al.</i> , <i>Nat Genet</i> 41, 703 (2009)
22q13	<i>IL17REL</i> rs5771069	Ulcerative colitis	rs5771069	Anderson <i>et al.</i> , <i>Nat Genet</i> 43, 246 (2011)	
			rs5771069	Franke <i>et al.</i> , <i>Nat Genet</i> 42, 292 (2010)	

All currently reported associations in autoimmune diseases for the nominally replicated candidate genes studied in the present paper are listed in the table. The index SNP from the current study is indicated in bold. In studies where multiple SNPs at the locus are reported associated to the respective disease, the SNP with the lowest *p* value is presented in the table. Significantly associated index SNPs from the current study are additionally presented in the table. WTCCC, Wellcome Trust Case Control Consortium; n.a., no associations to autoimmune diseases reported for genetic variants in these genes.

ducts of these patients, as well as the potential effects of rs602662 genotype on *FUT2* function [22,23].

Among the additional novel suggestive loci, the *CLEC16A* association was most robust. *CLEC16A* has a predicted C-type lectin domain structure [24] and C-type lectins serve as central players in the immune response by modulating toll receptor function [25]. A role for toll-like receptor signaling in PSC pathogenesis has been suggested [26], but whether *CLEC16A* is implicated can only be speculated. Like for *CLEC16A*, the suggestive associations observed at *SLCO1B1* and *GGT* might also involve immunological mechanisms since *SLCO1B1* and *GGT* are involved in the hepatic uptake and interconversion of the inflammatory mediator leukotriene C4, respectively [27,28]. The associations at *SLCO1B1*

and *GGT* can also reflect other roles of these genes as the PSC associated variants have been shown to alter blood levels of bilirubin and gamma glutamyltransferase, respectively, in previous GWAS [29,30].

The potential roles of the PSC associations at *IL2/IL21* and *REL/PUS10* that are strengthened by the current study have previously been extensively reviewed [31]. In a disease like PSC, in which no effective medical therapy is available, the *IL2* and *IL2RA* genetic findings could prove to be of unique translational value (e.g. an IL-2 monoclonal antibody (Daclizumab/Zenapax) is under clinical trials in other conditions). Like for *FUT2*, the associations observed at *CARD9* point towards the interaction of host genetics and microbiota. When stimulated by intracellular pathogens,

Research Article

CARD9 interacts with NOD2 and induces cytokine production or activates the NF- κ B pathway [32].

It has been established that susceptibility loci in related conditions are typically implicated in several disorders [33]. What is increasingly evident, and strengthened by our study is that PSC belongs to the family of autoimmune diseases, and that the overlapping genes are not restricted to IBD susceptibility genes (Table 3). For a rare disease, a clear understanding of the overlap with other immune mediated diseases may provide therapeutic opportunities otherwise not within scope. There is also a strong need to expand on the genetic susceptibility to PSC in study populations recruited outside Northern Europe and the United States.

In conclusion, we identify multiple PSC risk loci involved in the regulation of the immune response and the present study demonstrates a considerable overlap in the genetic susceptibility with other autoimmune diseases not restricted to IBD. The *FUT2* association potentially represents other components of the pathogenesis, and the significant influence of *FUT2* genotype on biliary microbial composition demonstrated by our study should be taken into account when further investigating the influence of microbiota on biliary pathology.

Financial support

The study was supported by The Norwegian PSC Research Center (<http://ous-research.no/nopsc/>) and the German Ministry of Education and Research (BMBF) through the National Genome Research Network (NGFN) and the Integrated Research and Treatment Center – Transplantation (reference number: 01EO0802) and the PopGen biobank (<http://www.popgen.de>). The US part of the study was supported by the NIH (DK 84960). The project received infrastructure support through the Norwegian Functional Genomics Programme (FUGE) via the “CIGENE” platform (<http://www.cigene.no/>), the Research Computing Services at the University of Oslo and the DFG excellence cluster “Inflammation at Interfaces” (<http://www.inflammation-at-interfaces.de/>).

Conflict of interest

The authors who have taken part in this study declared that they do not have anything to disclose regarding funding or conflict of interest with respect to this manuscript.

Acknowledgements

The authors are indebted to all PSC patients and healthy controls for their participation. We thank Katja Cloppenborg-Schmidt, Ilo-na Urbach, Irene Pauselis, Tanja Wesse, Tanja Henke, Rainer Vogler, Hege Dahlen Sollid, Bente Woldseth, and Liv Wenche Torbjørnsen for expert technical help. Paul R. Berg and Silje Karoliussen are specially thanked for providing expert assistance in the genotyping process. We are grateful to Arthur Kaser and Michael Nothnagel for helpful discussions. Benedicte A. Lie and The Norwegian Bone Marrow Donor Registry at Oslo University Hospital, Rikshospitalet in Oslo are acknowledged for sharing the healthy Norwegian controls. Dr. Felix Braun, Dr. Wolfgang Kreisler, Dr. Thomas Berg and Dr. Rainer Günther are acknowledged for contributing German PSC patients.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jhep.2012.03.031>.

References

- [1] Karlsen TH, Schrumpf E, Boberg KM. Update on primary sclerosing cholangitis. *Dig Liver Dis* 2010;42:390–400.
- [2] Saareinen S, Olerup O, Broome U. Increased frequency of autoimmune diseases in patients with primary sclerosing cholangitis. *Am J Gastroenterol* 2000;95:3195–3199.
- [3] Melum E, Franke A, Schramm C, Weismuller TJ, Gotthardt DN, Offner FA, et al. Genome-wide association analysis in primary sclerosing cholangitis identifies two non-HLA susceptibility loci. *Nat Genet* 2011;43:17–19.
- [4] Karlsen TH, Franke A, Melum E, Kaser A, Hov JR, Balschun T, et al. Genome-wide association analysis in primary sclerosing cholangitis. *Gastroenterology* 2010;138:1102–1111.
- [5] Janse M, Lamberts LE, Franke L, Raychaudhuri S, Ellinghaus E, Muri Boberg K, et al. Three ulcerative colitis susceptibility loci are associated with primary sclerosing cholangitis and indicate a role for IL2, REL, and CARD9. *Hepatology* 2011;53:1977–1985.
- [6] Stallhofer J, Denk GU, Glas J, Laubender RP, Goke B, Rust C, et al. Analysis of IL2/IL21 Gene Variants in Cholestatic Liver Diseases Reveals an Association with Primary Sclerosing Cholangitis. *Digestion* 2011;84:29–35.
- [7] Trynka G, Zhernakova A, Romanos J, Franke L, Hunt KA, Turner G, et al. Coeliac disease-associated risk variants in TNFAIP3 and REL implicate altered NF- κ B signalling. *Gut* 2009;58:1078–1083.
- [8] Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38:904–909.
- [9] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–575.
- [10] Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Glied TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 2010;26:2336–2337.
- [11] Rausch P, Rehman A, Kunzel S, Hasler R, Ott SJ, Schreiber S, et al. Colonic mucosa-associated microbiota is influenced by an interaction of Crohn disease and *FUT2* (Secretor) genotype. *Proc Natl Acad Sci U S A* 2011;108:19030–19035.
- [12] Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 2005;6:95–108.
- [13] Kelly RJ, Rouquier S, Giorgi D, Lennon GG, Lowe JB. Sequence and expression of a candidate for the human Secretor blood group alpha(1,2)fucosyltransferase gene (*FUT2*). Homozygosity for an enzyme-inactivating nonsense mutation commonly correlates with the non-secretor phenotype. *J Biol Chem* 1995;270:4640–4649.
- [14] Hirschfield GM, Liu X, Han Y, Gorlov IP, Lu Y, Xu C, et al. Variants at *IRF5-TNPO3*, *17q12-21* and *MMEL1* are associated with primary biliary cirrhosis. *Nat Genet* 2010;42:655–657.
- [15] Steinberg MW, Shui JW, Ware CF, Kronenberg M. Regulating the mucosal immune system: the contrasting roles of LIGHT, HVEM, and their various partners. *Semin Immunopathol* 2009;31:207–221.
- [16] Sedy JR, Gavrieli M, Potter KG, Hurchla MA, Lindsley RC, Hildner K, et al. B and T lymphocyte attenuator regulates T cell activation through interaction with herpesvirus entry mediator. *Nat Immunol* 2005;6:90–98.
- [17] Oya Y, Watanabe N, Owada T, Oki M, Hirose K, Suto A, et al. Development of autoimmune hepatitis-like disease and production of autoantibodies to nuclear antigens in mice lacking B and T lymphocyte attenuator. *Arthritis Rheum* 2008;58:2498–2510.
- [18] Chapman RW, Arborgh BA, Rhodes JM, Summerfield JA, Dick R, Scheuer PJ, et al. Primary sclerosing cholangitis: a review of its clinical features, cholangiography, and hepatic histology. *Gut* 1980;21:870–877.
- [19] McGovern DP, Jones MR, Taylor KD, Marciani K, Yan X, Dubinsky M, et al. Fucosyltransferase 2 (*FUT2*) non-secretor status is associated with Crohn's disease. *Hum Mol Genet* 2010;19:3468–3476.
- [20] Mentis A, Blackwell CC, Weir DM, Spiliadis C, Dailianas A, Skandalis N. ABO blood group, secretor status and detection of *Helicobacter pylori* among patients with gastric or duodenal ulcers. *Epidemiol Infect* 1991;106:221–229.

- [21] Svensson L, Petersson A, Henry SM. Secretor genotyping for A385T, G428A, C571T, C628T, 685delTGG, G849A, and other mutations from a single PCR. *Transfusion (Paris)* 2000;40:856–860.
- [22] Serpa J, Mendes N, Reis CA, Santos Silva LF, Almeida R, Le Pendu J, et al. Two new FUT2 (fucosyltransferase 2 gene) missense polymorphisms, 739G → A and 839T → C, are partly responsible for non-secretor status in a Caucasian population from Northern Portugal. *Biochem J* 2004;383:469–474.
- [23] Silva LM, Carvalho AS, Guillon P, Seixas S, Azevedo M, Almeida R, et al. Infection-associated FUT2 (Fucosyltransferase 2) genetic variation and impact on functionality assessed by in vivo studies. *Glycoconj J* 2010;27:61–68.
- [24] Hakonarson H, Grant SF, Bradfield JP, Marchand L, Kim CE, Glessner JT, et al. A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. *Nature* 2007;448:591–594.
- [25] Svajcer U, Anderluh M, Jeras M, Obermajer N. C-type lectin DC-SIGN: an adhesion, signalling and antigen-uptake molecule that guides dendritic cells in immunity. *Cell Signal* 2010;22:1397–1405.
- [26] Karrar A, Broome U, Sodergren T, Jaksch M, Bergquist A, Bjornstedt M, et al. Biliary epithelial cell antibodies link adaptive and innate immune responses in primary sclerosing cholangitis. *Gastroenterology* 2007;132:1504–1514.
- [27] Jedlitschky G, Keppler D. Transport of leukotriene C4 and structurally related conjugates. *Vitam Horm* 2002;64:153–184.
- [28] Mistry D, Stockley RA. Gamma-glutamyl transferase: the silent partner? *Copd* 2010;7:285–290.
- [29] Johnson AD, Kavousi M, Smith AV, Chen MH, Dehghan A, Aspelund T, et al. Genome-wide association meta-analysis for total serum bilirubin levels. *Hum Mol Genet* 2009;18:2700–2710.
- [30] Yuan X, Waterworth D, Perry JR, Lim N, Song K, Chambers JC, et al. Population-based genome-wide association studies reveal six loci influencing plasma levels of liver enzymes. *Am J Hum Genet* 2008;83:520–528.
- [31] Karlsen TH, Kaser A. Deciphering the genetic predisposition to primary sclerosing cholangitis. *Semin Liver Dis* 2011;31:188–207.
- [32] Hara H, Saito T. CARD9 versus CARMA1 in innate and adaptive immunity. *Trends Immunol* 2009;30:234–242.
- [33] 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.
- [34] Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010;26:2190–2191.
- [35] Chao A. Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* 1987;43:783–791.
- [36] Faith DP. Conservation evaluation and phylogenetic diversity. *Biol Conserv* 1992;61:1–10.