nature genetics

Genome-wide association analysis in primary sclerosing cholangitis identifies two non-HLA susceptibility loci

Espen Melum[1,2,27], Andre Franke[3,27], Christoph Schramm[4], Tobias J Weismüller[5,6], Daniel Nils Gotthardt[7], Felix A Offner[8], Brian D Juran[9], Jon K Laerdahl[10], Verena Labi[11], Einar Björnsson[12], Rinse K Weersma[13], Liesbet Henckaerts[14], Andreas Teufel[15], Christian Rust[16], Eva Ellinghaus[3], Tobias Balschun[3], Kirsten Muri Boberg[1], David Ellinghaus[3], Annika Bergquist[17], Peter Sauer[7], Euijung Ryu[18], Johannes Roksund Hov[1,2], Jochen Wedemeyer[5,6], Björn Lindkvist[12], Michael Wittig[3], Robert J Porte[19], Kristian Holm[1], Christian Gieger[20], H-Erich Wichmann[20–22], Pieter Stokkers[23], Cyriel Y Ponsioen[23], Heiko Runz[24], Adolf Stiehl[7], Cisca Wijmenga[25], Martina Sterneck[4], Severine Vermeire[14], Ulrich Beuers[23], Andreas Villunger[11], Erik Schrumpf[1], Konstantinos N Lazaridis[9], Michael P Manns[5,6], Stefan Schreiber[3,26,28] & Tom H Karlsen[1,28]

Primary sclerosing cholangitis (PSC) is a chronic bile duct disease affecting 2.4–7.5% of individuals with inflammatory bowel disease. We performed a genome-wide association analysis of 2,466,182 SNPs in 715 individuals with PSC and 2,962 controls, followed by replication in 1,025 PSC cases and 2,174 controls. We detected non-HLA associations at rs3197999 in *MST1* and rs6720394 near *BCL2L11* (combined $P = 1.1 \times 10^{-16}$ and $P = 4.1 \times 10^{-8}$, respectively).

Genetic associations at a genome-wide significance level in PSC[1] have only previously been detected within the human leukocyte antigen (HLA) complex on chromosome 6p21 (ref. 2). In the majority of HLA-associated diseases, additional susceptibility genes have been identified at other chromosomal loci[3]. The heritability in PSC is estimated to be in the same range as most of these other HLA-associated conditions (with a relative sibling risk of approximately 10)[4], suggesting a similar genetic architecture. To identify non-HLA susceptibility loci for PSC, we analyzed 332 Scandinavian PSC cases and 383 German PSC cases, along with 262 Scandinavian controls and 2,700 German controls genotyped with the Affymetrix Genome-Wide Human SNP Array 6.0 (Affymetrix). Two hundred sixty of the cases and 262 of the controls were included in a previous genome-wide assessment of 375,487 SNPs[2]. The power to detect association at a genome-wide significance level with a log-additive model in the discovery panel was 80% for a SNP with a frequency of 40% in the controls and an odds ratio of 1.42 (**Supplementary Fig. 1**). For a detailed description of the study populations and the experimental protocols, see the **Supplementary Methods**.

We applied extensive quality control measures[5] and excluded samples with a low genotyping success rate (<95%), heterozygosity outliers and samples with evidence for cryptic relatedness. We assessed study population heterogeneity by means of a principal components analysis[6] (**Supplementary Fig. 1**), and we removed ethnic outliers before proceeding with further analyses. We also excluded SNPs with a minor allele frequency <1%, a genotyping success rate <95% or a deviation of the genotype distribution from Hardy-Weinberg equilibrium in the controls ($P < 10^{-4}$). To increase the genomic coverage of the dataset, we imputed missing genotypes and SNPs using the phased European CEU HapMap data release 22 reference dataset. We subjected all imputed markers to the same quality criteria described above and added the requirement of good imputation quality, leaving a total of 2,466,182 SNPs for association analysis. We used a logistic regression procedure to test both genotyped and imputed SNPs for association. We used allele dosages from the imputation to account for uncertainty in the imputation procedure,

[1]Norwegian PSC Research Center, Clinic for Specialized Medicine and Surgery, Oslo University Hospital, Rikshospitalet, Oslo, Norway. [2]Research Institute for Internal Medicine, Oslo University Hospital, Rikshospitalet, Oslo, Norway. [3]Institute of Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Germany. [4]1st Department of Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. [5]Department of Gastroenterology, Hepatology and Endocrinology, Hannover Medical School, Hannover, Germany. [6]Integrated Research and Treatment Center-Transplantation (IFB-Tx), Hannover Medical School, Hannover, Germany. [7]Department of Medicine, University Hospital of Heidelberg, Heidelberg, Germany. [8]Academic Teaching Hospital Feldkirch, Feldkirch, Austria. [9]Center for Basic Research in Digestive Diseases, Division of Gastroenterology and Hepatology, Mayo Clinic College of Medicine, Rochester, Minnesota, USA. [10]Centre for Molecular Biology and Neuroscience (CMBN) and Institute of Medical Microbiology, Oslo University Hospital, Rikshospitalet, Oslo, Norway. [11]Division of Developmental Immunology, Biocenter, Innsbruck Medical University, Innsbruck, Austria. [12]Section of Gastroenterology and Hepatology, Department of Internal Medicine, Sahlgrenska University Hospital, Gothenburg, Sweden. [13]Department of Gastroenterology and Hepatology, University Medical Center Groningen and University of Groningen, Groningen, The Netherlands. [14]Department of Gastroenterology, University Hospital Gasthuisberg, Leuven, Belgium. [15]1st Department of Medicine, University of Mainz, Mainz, Germany. [16]Department of Medicine 2, Grosshadern, University of Munich, Munich, Germany. [17]Department of Gastroenterology and Hepatology, Karolinska University Hospital Huddinge, Stockholm, Sweden. [18]Division of Biomedical Statistics and Informatics, Mayo Clinic College of Medicine, Rochester, Minnesota, USA. [19]Hepatobiliary Surgery and Liver Transplantation, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. [20]Institute of Epidemiology, Helmholtz Centre Munich, German Research Center for Environmental Health, Neuherberg, Germany. [21]Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-University, Munich, Germany. [22]Klinikum Grosshadern, Munich, Germany. [23]Department of Gastroenterology and Hepatology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands. [24]Department of Human Genetics, University Hospital of Heidelberg, Heidelberg, Germany. [25]Department of Genetics, University Medical Center Groningen and University of Groningen, Groningen, The Netherlands. [26]Department for General Internal Medicine, Christian-Albrechts-University, Kiel, Germany. [27]These authors contributed equally to this work. [28]These authors jointly directed this work. Correspondence should be addressed to T.H.K. (t.h.karlsen@medisin.uio.no).

Received 20 July; accepted 16 November; published online 12 December 2010; doi:10.1038/ng.728

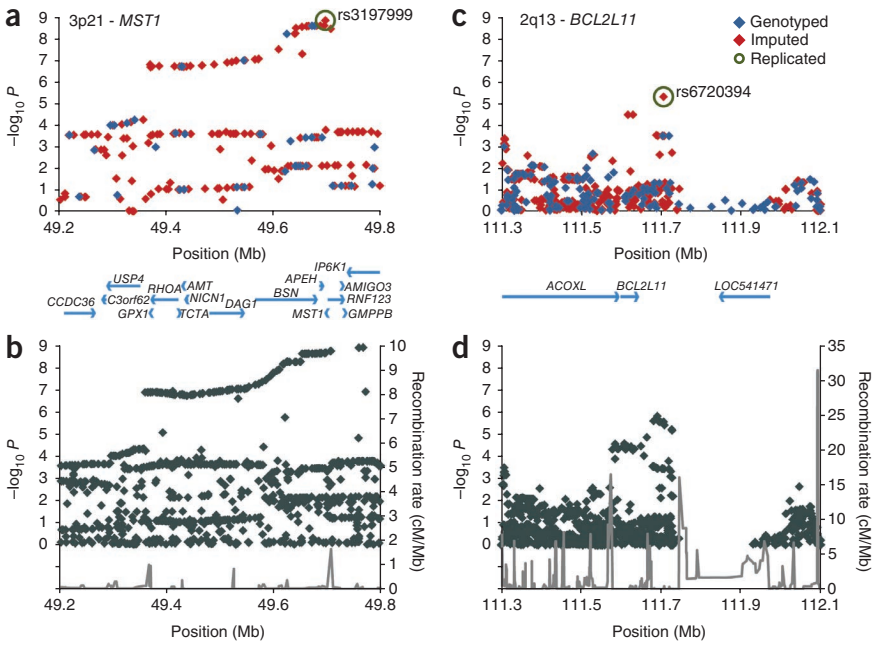

Figure 1 Association results at the *MST1* and *BCL2L11* loci. (**a**,**c**) The association results from the genotyped and imputed markers at the *MST1* and *BCL2L11* loci are shown as the $-\log_{10}$ of the *P* values plotted against the genomic position (NCBI build 36). The *MST1* and *BCL2L11* SNPs with robust evidence for replication are highlighted with green circles. (**b**,**d**) The recombination rates (gray lines) derived from the HapMap project, along with the association results based on an imputed dataset using reference data from the 1000 Genomes Project. None of the imputed 1000 Genomes Project SNPs in the region from 111,745,000 bp to 111,929,500 bp on chromosome 2 passed imputation quality thresholds (**Supplementary Methods**). The *MST1* and *BCL2L11* associations (represented by the SNPs given in **Table 1**) were independent of the HLA association in a logistic regression analysis with inclusion of rs3134792 (the most strongly associated HLA SNP) as a covariate and in a stratified analysis according to rs3134792 risk allele carriership (**Supplementary Table 2**). Association results at the *MST1* and *BCL2L11* loci in ulcerative colitis compared to PSC are shown in **Supplementary Figure 3**.

and we included the first six principal components as covariates to adjust for population structure (**Supplementary Fig. 1**).

In line with our previous study[2], we detected the strongest associations with PSC at SNPs in the HLA complex at chromosome 6p21, peaking at rs3134792 in *HLA-B* ($P = 6.8 \times 10^{-49}$) (**Supplementary Fig. 2**). Carriers of the associated G allele at rs3134792 were HLA-B*08 carriers in 99% of the cases and were HLA-DRB1*03 carriers in 90% of the cases (**Supplementary Methods**). Inclusion of rs3134792 as a covariate in our model showed a complex residual association signal in the vicinity of the class II region (lowest $P = 7.6 \times 10^{-17}$ for rs9272723), suggesting the presence of multiple causative loci within the region (**Supplementary Fig. 2**). In ulcerative colitis, the association signal in the HLA complex is less extensive (**Supplementary Fig. 3**), with associated SNPs predominantly being observed near the HLA class II genes, and dedicated studies will be needed to differentiate between disease-specific and shared risk variants for PSC and ulcerative colitis in this region. In addition to SNPs in the HLA complex, multiple SNPs in strong linkage disequilibrium (LD) at chromosome 3p21 were also associated at a genome-wide significance level. The 3p21 signal stretches over a 0.34-Mb interval and peaks at rs3197999 in *MST1*, the macrophage stimulating 1 gene ($P = 1.4 \times 10^{-9}$) (**Fig. 1a**). Three hundred seventy-nine non-HLA SNPs (that is, excluding markers in the region between 25 Mb and 35 Mb on chromosome 6) with $P < 10^{-4}$ were subsequently evaluated for replication genotyping. To exclude technical artifacts, we visually inspected raw intensity cluster plots for the genotyped SNPs. By grouping correlated SNPs based on LD (**Supplementary Methods**), we defined the top 23 associated regions for follow up. We performed replication genotyping using Sequenom mass spectrometry–based technology and a total of 1,025 PSC cases along with 2,174 controls from Scandinavia, Central Europe and the United States (see **Supplementary Table 1** for a complete listing and for allele frequencies).

In the replication analysis, we detected the most prominent association for the non-synonymous (p.Arg689Cys) SNP rs3197999 at *MST1* (**Table 1**). This PSC-associated amino acid change has previously been proposed to influence MST1 receptor interaction, as well as the risk for ulcerative colitis and Crohn's disease[7]. The present finding suggests that the *MST1* locus represents an important overlapping susceptibility locus for PSC and inflammatory bowel disease and that the influence from the disease-associated variant on biliary and intestinal inflammation needs to be further studied. The fact that the MST1 protein is expressed at high levels in gallbladder epithelium

Table 1 SNPs with significant association results in the replication cohort

					Genome-wide analysis				Replication analysis					
					Allele frequencies (cases/controls)				Allele frequencies (cases/controls)					
Locus	SNP	Chr.	Position	Alleles	Scandinavia (332/262)	Germany (383/2700)	P^a	OR (95% CI)[a]	Scandinavia (259/729)	Central Europe (498/891)	United States (268/554)	$P_{CMH}{}^b$	OR (95% CI)	Combined *P*
MST1	rs3197999	3	49,696,536	A/G	0.36/0.32	0.40/0.28	1.4×10^{-9}	1.51 (1.32–1.72)	0.35/0.31	0.33/0.25	0.38/0.30	1.5×10^{-8}	1.39 (1.24–1.56)	1.1×10^{-16}
BCL2L11	rs6720394	2	111,705,843	G/T	0.17/0.13	0.15/0.11	5.2×10^{-6}	1.60 (1.31–1.96)	0.14/0.12	0.13/0.11	0.16/0.10	0.0016	1.29 (1.10–1.51)	4.1×10^{-8}

Association results for the genome-wide analysis and the replication analysis for the two SNPs with replication results robust to correction for multiple testing using Bonferroni's method. We performed association testing with logistic regression including correction for population structure in the genome-wide discovery analysis, and the Cochran-Mantel-Haenszel test was used for the replication analysis. Allele frequencies are given for each of the five study panels separately. Positions refer to NCBI's build 36. CMH, Cochran-Mantel-Haenszel; OR, odds ratio; Chr., chromosome.

[a]Odds ratios and *P* values derived from logistic regressions of allele dosages including the first six principal components from the principal components analysis as covariates. [b]Cochran-Mantel-Haenszel test; Breslow-Day test, $P = 0.27$ for rs3197999 and $P = 0.10$ for rs6720394.

(antibody ID HPA024036; Human Protein Atlas) further supports this notion. In addition, the strong LD present in this region (**Fig. 1b**) means that further genetic characterization is necessary to exclude the presence of several susceptibility variants at the 3p21 locus.

The replicated association signal at chromosome 2q13 (**Table 1**) was supported by multiple SNPs (**Fig. 1c**,**d**) encompassing the *BCL2L11* (encoding the BCL2-like 11 protein) locus and extending into a duplicated region with a transcript of unknown function (*LOC541471*). *BCL2L11* encodes the Bcl-2 interacting protein (Bim), which is crucial for maintaining immunological tolerance through induction of apoptosis of autoreactive T cells, as well as the deletion of activated T cells after an immune response[8]. The neighboring putative gene, *LOC541471*, is very unlikely to give rise to protein-coding transcripts, as these transcripts contain only short open reading frames that do not match any known protein homolog (**Supplementary Fig. 4**). Although it can not be completely ruled out that the *LOC541471* transcripts function as non–protein-coding RNA, we conclude that the lead SNP, rs6720394, is more likely to represent genetic variation that affects *BCL2L11* rather than *LOC541471*. To gain potential insight into a role of Bim in biliary physiology, we assessed hepatic hematoxylin and eosin stainings of 8-week-old *Bcl2l11*$^{-/-}$ mice and matched wildtype controls (**Supplementary Methods**). Although both genotypes presented with histologically normal livers, it was remarkable that in four out of four *Bcl2l11*$^{-/-}$ livers, mononuclear cells were present surrounding several intrahepatic bile ducts, whereas none of the portal fields in four wildtype livers showed similar subtle alterations ($P = 0.029$) (**Supplementary Fig. 5**). Further mechanistic studies aimed at characterizing Bim's functional effects on liver and biliary physiology appear warranted.

At the *IL2RA* (encoding interleukin 2 receptor alpha) locus, several SNPs showed highly significant association in the genome-wide analysis (lowest $P = 2.4 \times 10^{-7}$, for rs10905718) (**Supplementary Fig. 2**). This locus is of particular interest in PSC because *Il2ra*$^{-/-}$ mice spontaneously develop both intestinal and biliary inflammation[9]. Notably, although the *Il2ra*$^{-/-}$ model has been proposed to mimic pathogenesis in another disease of the bile ducts (primary biliary cirrhosis), no association at *IL2RA* was reported in a genome-wide analysis of this condition[10]. The SNPs at *IL2RA* selected for replication based on the genome-wide analysis demonstrated non-uniform effect sizes in the replication analysis (Breslow-Day test, $P < 0.05$) and did not achieve nominal significance when analyzed with a random effects model (**Supplementary Table 1**). The lack of formal replication at the *IL2RA* locus could be due to population differences, which we could not correct for because genome-wide data were not available for the replication panels. For *IL2RA* SNPs previously shown to influence risk of multiple sclerosis and type 1 diabetes[3], we observed significant associations in the genome-wide analysis ($P = 8.7 \times 10^{-4}$ for rs2104286 and $P = 0.0025$ for rs12251307, respectively), as well as similar trends in the replication analysis (Cochran-Mantel-Haenszel (CMH), $P = 0.064$ for rs2104286 and CMH, $P = 0.033$ for rs12251307; Breslow-Day, $P = 0.30$ and $P = 0.39$, respectively). The heterogeneous association signal at this locus makes it impossible to conclude genetically regarding an involvement of *IL2RA* in PSC pathogenesis, yet the combined evidence from the statistical association analysis and the results from the *Il2ra*$^{-/-}$ mouse makes such an involvement possible.

In conclusion, to our knowledge, we are able to provide the first evidence at genome-wide significance for involvement of a non-HLA gene in PSC susceptibility. Furthermore, we are able to provide suggestive evidence for at least two additional loci involved in T cell activation and immunological tolerance, supporting the possibility raised by the strong HLA associations that PSC pathogenesis has an autoreactive component.

URLs. HapMap, http://www.hapmap.org/; Human Protein Atlas, http://www.proteinatlas.org/; 1000 Genomes Project, http://www.1000genomes.org/.

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

ACKNOWLEDGMENTS

The authors wish to thank all PSC cases and healthy controls for their participation. We also thank K. Cloppenborg-Schmidt, I. Urbach, I. Pauselis, T. Wesse, T. Henke, R. Vogler, B. Stade, T. Vennegerts, P.R. Berg, H.D. Sollid and B. Woldseth for expert technical help. We are grateful to M.K. Viken and M. Nothnagel for helpful discussions. We acknowledge B.A. Lie and the Norwegian Bone Marrow Donor Registry at Oslo University Hospital, Rikshospitalet for contributing the healthy Norwegian control population. We acknowledge F. Braun, W. Kreisel, T. Berg and R. Günther for contributing German PSC cases. We acknowledge A. Strasser for generating and kindly providing the *Bcl2l11*$^{-/-}$ mouse model. We greatly acknowledge A. Kaser for managing the *Bcl2l11*$^{-/-}$ liver histology assessment and for helpful discussions on the functional implications of all findings. The study was supported by The Norwegian PSC research center, the German Federal Ministry of Education and Research (BMBF) through the National Genome Research Network (NGFN), the PopGen biobank, the Integrated Research and Treatment Center–Transplantation (reference number: 01EO0802), the Palumbo Charitable Trust, the Musette and Allen Morgan Jr. Foundation for the Study of PSC, PSC Partners Seeking a Cure and the Mayo Clinic College of Medicine. The project received infrastructure support through the Norwegian Functional Genomics Programme (FUGE) through the 'CIGENE' platform, the Research Computing Services at the University of Oslo and the Deutsche Forschungsgemeinschaft (DFG) excellence cluster 'Inflammation at Interfaces'. The Kooperative Gesundheitsforschung in der Region Augsburg (KORA) research platform was initiated and financed by the Helmholtz Center Munich, German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Part of this work was financed by the German National Genome Research Network (NGFN-2 and NGFNPlus: 01GS0823). This research was also supported within the Munich Center of Health Sciences (MC Health) as part of Ludwig-Maximilians-Universität (LMU) innovativ.

AUTHOR CONTRIBUTIONS

E.M. performed data analysis. A.F. and T.H.K. supervised data analysis and coordinated project contributions. E.E., T.B., D.E., J.R.H. and E.R. helped with data analysis. F.A.O., V.L. and A.V. performed the *Bcl2l11*$^{-/-}$ animal work and liver histology assessments. J.K.L. performed *in silico* analysis of chromosome 2q13 transcripts. M.W. and K.H. were responsible for in-house conversion and database management of genome-wide association study data. C.S., T.J.W., D.N.G., B.D.J., E.B., R.K.W., L.H., A.T., C.R., K.M.B., C.G., H.-E.W., A.B., P. Sauer, J.W., B.L., R.J.P., P. Stokkers, C.Y.P., H.R., A.S., C.W., M.S., S.V., U.B., E.S., K.N.L., M.P.M. and S.S. provided the case populations and healthy controls. S.S., A.F., T.H.K. and E.M. designed the experiment. E.M. and T.H.K. drafted the manuscript. All authors revised the manuscript and approved of the final version.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

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

1. Broomé, U. & Bergquist, A. *Semin. Liver Dis.* **26**, 31–41 (2006).
2. Karlsen, T.H. *et al. Gastroenterology* **138**, 1102–1111 (2010).
3. Zhernakova, A., van Diemen, C.C. & Wijmenga, C. *Nat. Rev. Genet.* **10**, 43–55 (2009).
4. Bergquist, A. *et al. Clin. Gastroenterol. Hepatol.* **6**, 939–943 (2008).
5. Karlsen, T.H., Melum, E. & Franke, A. *Hepatology* **51**, 1833–1842 (2010).
6. Price, A.L. *et al. Nat. Genet.* **38**, 904–909 (2006).
7. Goyette, P. *et al. Mucosal Immunol.* **1**, 131–138 (2008).
8. Strasser, A. *Nat. Rev. Immunol.* **5**, 189–200 (2005).
9. Hsu, W. *et al. Hepatology* **49**, 133–140 (2009).
10. Hirschfield, G.M. *et al. N. Engl. J. Med.* **360**, 2544–2555 (2009).