Gut

Genetic association analysis identifies variants associated with disease progression in primary sclerosing cholangitis

Journal:	*Gut*
Manuscript ID	gutjnl-2016-313598.R1
Article Type:	Original Article
Date Submitted by the Author:	28-Apr-2017
Complete List of Authors:	Alberts, Rudi; University of Groningen and University Medical Center Groningen, Department of Gastroenterology and Hepatology de Vries, Elisabeth; Academic Medical Center, Department of Gastroenterology and Hepatology Goode, Elizabeth; University of East Anglia, Norwich Medical School, Faculty of Medicine and Health Sciences; University of Cambridge, Academic Department of Medical Genetics Jiang, Xiaojun; Oslo University Hospital, Rikshospitalet, Norwegian PSC Research Center, Division of Cancer Medicine, Surgery and Transplantation; Oslo University Hospital, Rikshospitalet, Research Institute of Internal Medicine Sampaziotis, Fotios; University of Cambridge, Wellcome Trust-Medical Research Council Stem Cell Institute, Anne McLaren Laboratory, Department of Surgery; University of Cambridge and NIHR Cambridge Biomedical Research Centre, Department of Surgery Rombouts, Krista; University College London, Royal Free Hospital, Institute for Liver and Digestive Health Böttcher, Katrin; University College London, Royal Free Hospital, Institute for Liver and Digestive Health Folseraas, Trine; Oslo University Hospital, Rikshospitalet, Norwegian PSC Research Center, Division of Cancer Medicine, Surgery and Transplantation; Oslo University Hospital, Rikshospitalet, Research Institute of Internal Medicine Weismüller, Tobias; Hannover Medical School, Department of Gastroenterology, Hepatology and Endocrinology; Hannover Medical School, Integrated Research and Treatment Center-Transplantation (IFB-tx) Mason, Andrew; University of Alberta Hospital, Division of Gastroenterology and Hepatology Wang, Weiwei; University of Alberta Hospital, Division of Gastroenterology and Hepatology Alexander, Graeme; University of Cambridge, Department of Medicine, Division of Hepatology Alvaro, D; Sapienza University of Rome, Department of Clinical Medicine, Division of Gastroenterology Bergquist, Annika; Karolinska University Hospital, Karolinska Institutet, Center for Digestive Diseases Björkström, Niklas; Karolinska University Hospital, Karolinska Institutet,

	Center for Infectious Medicine Beuers, Ulrich; Academic Medical Center, Department of Gastroenterology and Hepatology Bj&#65533;rnsson, Einar; Landspitali University Hospital, Department of Internal Medicine, Division of Gastroenterology and Hepatology Boberg, Kirsten; Oslo University Hospital, Rikshospitalet, Norwegian PSC Research Center, Division of Cancer Medicine, Surgery and Transplantation; University of Oslo, K.G. Jebsen Inflammation Research Centre and Institute of Clinical Medicine Bowlus, Christopher; University of California Davis, Division of Gastroenterology and Hepatology Bragazzi, Maria Consiglia; Sapienza University of Rome, Medico-Surgical Sciences and Biotechnologies Carbone, Marco; University of Milan-Bicocca, Program for Autoimmune Liver Diseases, International Center for Digestive Health, Department of Medicine and Surgery Chazouilleres, Olivier; AP-HP, Hôpital Saint Antoine, Department of Hepatology Cheung, Angela; Toronto General Hospital, Gen. Internal Medicine, University Health Network Dalekos, George; University of Thessaly, Department of Medicine and Research Laboratory of Internal Medicine, Medical School Eaton, John; Mayo Clinic Minnesota, Division of Gastroenterology and Hepatology Eksteen, Bertus; University of Calgary, Snyder Institute of Chronic Diseases, Department of Medicine Ellinghaus, David; Christian-Albrechts-University, Institute of Clinical Molecular Biology Färkkilä, Martti; Helsinki University Central Hospital, Division of Gastroenterology, Department of Medicine Festen, Noortje; University of Groningen and University Medical Center Groningen, Department of Gastroenterology and Hepatology Floreani, Annarosa; University of Padova, Department of Surgery, Oncology and Gastroenterology Franceschet, Irene; University of Padova, Department of Surgery, Oncology and Gastroenterology Gotthardt, Daniel; University Hospital of Heidelberg, Department of Medicine Hirschfield, Gideon; University of Birmingham, Centre for Liver Research, NIHR Biomedical Research Unit van Hoek, Bart; Leiden University Medical Centre, Department of Gastroenterology and Hepatology Holm, Kristian; Oslo University Hospital, Rikshospitalet, Norwegian PSC Research Center, Division of Cancer Medicine, Surgery and Transplantation; Oslo University Hospital, Rikshospitalet, Research Institute of Internal Medicine Hohenester, Simon; University of Munich, Liver Center Munich, Department of Medicine II Hov, Johannes; Oslo University Hospital, Rikshospitalet, Norwegian PSC Research Center, Division of Cancer Medicine, Surgery and Transplantation; Oslo University Hospital, Rikshospitalet, Research Institute of Internal Medicine Imhann, Floris; University of Groningen and University Medical Center Groningen, Department of Gastroenterology and Hepatology Invernizzi, Pietro; University of Milan-Bicocca, Program for Autoimmune Liver Diseases, International Center for Digestive Health, Department of Medicine and Surgery Juran, Brian; Mayo Clinic Minnesota, Division of Gastroenterology and Hepatology Lenzen, Henrike; Hannover Medical School, Department of Gastroenterology, Hepatology and Endocrinology

Lieb, Wolfgang; Christian-Albrechts-University, Popgen Biobank, University Hospital Schleswig-Holstein; Christian-Albrechts University, Institute for Epidemiology
Liu, Jimmy; Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus
Marschall, Hanns-Ulrich; Sahlgrenska Academy, Department of Molecular and Clinical Medicine, Institute of Medicine
Marzioni, Marco; Università Politecnica delle Marche, Ospedali Riuniti University Hospital, Department of Gastroenterology
Melum, Espen; Oslo University Hospital, Rikshospitalet, Norwegian PSC Research Center, Division of Cancer Medicine, Surgery and Transplantation; Oslo University Hospital, Rikshospitalet, Research Institute of Internal Medicine
Milkiewicz, Piotr; Medical University of Warsaw, Liver and Internal Medicine Unit
Müller, Tobias; Charité Universitätsmedizin Berlin, Campus Virchow Klinikum, Department of Internal Medicine, Hepatology and Gastroenterology
Pares, Albert; Hospital Clinic, IDIBAPS, CIBERehd, University of Barcelona, Liver Unit
Rupp, Christian; University Hospital of Heidelberg, Department of Internal Medicine IV
Rust, Christian; Krankenhaus Barmherzige Brüder, Department of Medicine I
Sandford, Richard; University of Cambridge, Academic Department of Medical Genetics
Christoph, Schramm; University Medical Center Hamburg-Eppendorf, 1st Department of Medicine
Schreiber, Stefan; Christian-Albrechts-University, Institute of Clinical Molecular Biology; Christian-Albrechts-University, Department for General Internal Medicine
Schrumpf, Erik; Oslo University Hospital, Rikshospitalet, Norwegian PSC Research Center, Division of Cancer Medicine, Surgery and Transplantation; Oslo University Hospital, Rikshospitalet, Section of Gastroenterology, Department of Transplantation Medicine, Division of Cancer, Surgery and Transplantation
Silverberg, Mark; Mount Sinai Hospital Toronto, Inflammatory Bowel Disease (IBD) Group, Zane Cohen Centre for Digestive Diseases
Srivastava, Brijesh; University of Cambridge, Academic Department of Medical Genetics
Sterneck, Martina; University Medical Center Hamburg-Eppendorf, Department of Hepatobiliary Surgery and Transplantation
Teufel, Andreas; University of Mainz, 1st Department of Medicine
Vallier, Ludovic; University of Cambridge, Wellcome Trust-Medical Research Council Stem Cell Institute, Anne McLaren Laboratory, Department of Surgery; Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus
Verheij, Joanne; Academic Medical Center, Department of Pathology
Vich Vila, Arnau; University of Groningen and University Medical Center Groningen, Department of Gastroenterology and Hepatology
de Vries, Boudewijn; University of Groningen and University Medical Center Groningen, Department of Gastroenterology and Hepatology
Zachou, Kalliopi; University of Thessaly, Department of Internal Medicine
International PSC Study Group, The
UK PSC Consortium, The
CHAPMAN, ROGER WILLIAM; John Radcliffe University Hospitals NHS Trust, Department of Hepatology
Manns, Michael; Hannover Medical School, Department of Gastroenterology, Hepatology, and Endocrinology; Hannover Medical School, Integrated Research and Treatment Center-Transplantation (IFB-tx)

	Pinzani, Massimo; University College London, Royal Free Hospital, Institute for Liver and Digestive Health Rushbrook, Simon; University of East Anglia, Norwich Medical School, Faculty of Medicine and Health Sciences Lazaridis, Konstantinos; Mayo Clinic Minnesota, Division of Gastroenterology and Hepatology Franke, Andre; Christian-Albrechts-University, Institute of Clinical Molecular Biology Anderson, Carl; Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus Karlsen, Tom; Oslo University Hospital, Rikshospitalet, Norwegian PSC Research Center, Division of Cancer Medicine, Surgery and Transplantation; Oslo University Hospital, Rikshospitalet, Research Institute of Internal Medicine Ponsioen, Cyriel; Academic Medical Center, Department of Gastroenterology and Hepatology Weersma, Rinse; University of Groningen and University Medical Centre Groningen, Department of Gastroenterology and Hepatology
Keywords:	PRIMARY SCLEROSING CHOLANGITIS, GENETICS, LIVER TRANSPLANTATION

Genetic association analysis identifies variants associated with disease progression in primary sclerosing cholangitis

Rudi Alberts[1*], Elisabeth M.G. de Vries[2*], Elizabeth C. Goode[3,4], Xiaojun Jiang[5,6], Fotis Sampaziotis[7,8], Krista Rombouts[9], Katrin Böttcher[9], Trine Folseraas[5,6], Tobias J. Weismüller[10,11], Andrew L. Mason[12], Weiwei Wang[12], Graeme Alexander[13], Domenico Alvaro[14], Annika Bergquist[15], Niklas K. Björkström[16], Ulrich Beuers[2], Einar Björnsson[17], Kirsten Muri Boberg[5,18], Christopher L. Bowlus[19], Maria C. Bragazzi[20], Marco Carbone[21], Olivier Chazouillères[22], Angela Cheung[23], Georgios Dalekos[24], John Eaton[25], Bertus Eksteen[26], David Ellinghaus[27], Martti Färkkilä[28], Eleonora A.M. Festen[1], Annarosa Floreani[29], Irene Franceschet[30], Daniel Nils Gotthardt[31], Gideon M. Hirschfield[32], Bart van Hoek[33], Kristian Holm[5,6], Simon Hohenester[34], Johannes Roksund Hov[5,6], Floris Imhann[1], Pietro Invernizzi[21], Brian D. Juran[25], Henrike Lenzen[35], Wolfgang Lieb[36,37], Jimmy Z. Liu[38], Hanns-Ulrich Marschall[39], Marco Marzioni[40], Espen Melum[5,6], Piotr Milkiewicz[41], Tobias Müller[42], Albert Pares[43], Christian Rupp[44], Christian Rust[45], Richard N. Sandford[4], Christoph Schramm[46], Stefan Schreiber[27,47], Erik Schrumpf[6,48], Mark S. Silverberg[49], Brijesh Srivastava[4], Martina Sterneck[50], Andreas Teufel[51], Ludovic Vallier[7,38], Joanne Verheij[52], Arnau Vich Vila[1], Boudewijn de Vries[1], Kalliopi Zachou[53], The International PSC Study Group, The UK PSC Consortium, Roger W. Chapman[54], Michael P. Manns[10,11], Massimo Pinzani[9], Simon M. Rushbrook[3], Konstantinos N. Lazaridis[25], Andre Franke[27], Carl A. Anderson[38], Tom H. Karlsen[5,6], Cyriel Y. Ponsioen[2#], Rinse K. Weersma[1#]

[*]equal contribution

[#]shared last author

1 Department of Gastroenterology and Hepatology, University of Groningen and University Medical Centre Groningen, Groningen, NL

2 Department of Gastroenterology and Hepatology, Academic Medical Center, Amsterdam, NL

3 Norwich Medical School, Faculty of Medicine and Health Sciences, University of East Anglia, Norwich, UK

4 Academic Department of Medical Genetics, University of Cambridge, Cambridge, UK

5 Norwegian PSC Research Center, Division of Cancer Medicine, Surgery and Transplantation, Oslo University Hospital, Rikshospitalet, Oslo, Norway

6 Research Institute of Internal Medicine, Oslo University Hospital, Rikshospitalet, Oslo, Norway

7 Wellcome Trust-Medical Research Council Stem Cell Institute, Anne McLaren Laboratory, Department of Surgery, University of Cambridge, Cambridge, UK

8 Department of Surgery, University of Cambridge and NIHR Cambridge Biomedical Research Centre, Cambridge, UK

9 Institute for Liver and Digestive Health, University College London, Royal Free Hospital, London, UK

10 Department of Gastroenterology, Hepatology and Endocrinology, Hannover Medical School, Hannover, Germany

11 Integrated Research and Treatment Center-Transplantation (IFB-tx), Hannover Medical School, Hannover, Germany

12 Division of Gastroenterology and Hepatology, University of Alberta, Edmonton, Alberta, Canada

13 Department of Medicine, Division of Hepatology, University of Cambridge, Cambridge, UK

14 Department of Clinical Medicine, Division of Gastroenterology, Sapienza University of Rome, Rome, Italy

15 Center for Digestive Diseases, Karolinska University Hospital, Karolinska Institutet, Stockholm, Sweden

16 Center for Infectious Medicine, Department of Medicine Huddinge, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden

17 Department of Internal Medicine, Division of Gastroenterology and Hepatology, Landspitali University Hospital, Reykjavik, Iceland

18 K.G. Jebsen Inflammation Research Centre and Institute of Clinical Medicine, University of Oslo, Oslo, Norway

19 Division of Gastroenterology and Hepatology, University of California Davis, Davis, CA, USA

20 Sapienza University of Rome, Medico-Surgical Sciences and Biotechnologies, Rome, Italy

21 Program for Autoimmune Liver Diseases, International Center for Digestive Health, Department of Medicine and Surgery, University of Milan-Bicocca, Milan, Italy

22 AP-HP, Hôpital Saint Antoine, Department of Hepatology, Paris, France

23 Gen. Internal Medicine, University Health Network, Toronto General Hospital, Toronto, Canada

24 Department of Medicine and Research Laboratory of Internal Medicine, Medical School, University of Thessaly, Larissa, Greece

25 Division of Gastroenterology and Hepatology, Mayo Clinic Minnesota, Rochester, Minnesota, United States

26 Snyder Institute of Chronic Diseases, Department of Medicine, University of Calgary, Calgary, Canada

27 Institute of Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Germany

28 Division of Gastroenterology, Department of Medicine, Helsinki University Hospital, Finland

29 Dept. of Surgical, Oncological and Gastroenterological Sciences, University of Padova, Padova, Italy

30 Department of Surgery, Oncology and Gastroenterology, University of Padova, Padova, Italy

31 Department of Medicine, University Hospital of Heidelberg, Heidelberg, Germany

32 Centre for Liver Research, NIHR Biomedical Research Unit, University of Birmingham, Birmingham, UK

33 Department of Gastroenterology and Hepatology, Leiden University Medical Centre, Leiden, the Netherlands

34 Liver Center Munich, Department of Medicine II, University of Munich, Munich, Germany

35 Department of Gastroenterology, Hepatology and Endocrinology, Hannover Medical School, Hannover, Germany

36 Popgen Biobank, University Hospital Schleswig-Holstein, Christian-Albrechts-University, Kiel, Germany

37 Institute for Epidemiology, Christian-Albrechts University, Kiel, Germany

38 Wellcome Trust Genome Campus, Wellcome Trust Sanger Institute, Cambridge, UK

39 Department of Molecular and Clinical Medicine, Institute of Medicine, Sahlgrenska Academy, Gothenburg, Sweden

40 Department of Gastroenterology, Università Politecnica delle Marche, Ospedali Riuniti University Hospital, Ancona, Italy

41 Liver and Internal Medicine Unit, Medical University of Warsaw, Warsaw, Poland

42 Department of Internal Medicine, Hepatology and Gastroenterology, Charité Universitätsmedizin Berlin, Campus Virchow Klinikum, Berlin, Germany

43 Liver Unit, Hospital Clinic, IDIBAPS, CIBERehd, University of Barcelona, Barcelona, Spain

44 Department of Internal Medicine IV, University Hospital of Heidelberg, Heidelberg,Germany

45 Department of Medicine I, Krankenhaus Barmherzige Brüder, Munich, Germany

46 1st Department of Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

47 Department for General Internal Medicine, Christian-Albrechts-University, Kiel, Germany

48 Section of Gastroenterology, Department of Transplantation Medicine, Division of Cancer, Surgery and Transplantation, Oslo University Hospital, Rikshospitalet, Oslo, Norway

49 Inflammatory Bowel Disease (IBD) Group, Zane Cohen Centre for Digestive Diseases, Mount Sinai Hospital Toronto, Ontario, Canada

50 Department of Hepatobiliary Surgery and Transplantation, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

51 1st Department of Medicine, University of Mainz, Mainz, Germany

52 Department of Pathology, Academic Medical Center, Amsterdam, the Netherlands

53 Department of Internal Medicine, University of Thessaly, Larissa, Greece

54 Department of Hepatology, John Radcliffe University Hospitals NHS Trust, Oxford, UK

Abbreviations

CCA: cholangiocarcinoma; CLC: cholangiocyte-like-cell; CRC: colorectal carcinoma; HCC: hepatocellular carcinoma; hiPSC: human induced pluripotent stem cell; HSC: hepatic stellate cell; IBD: inflammatory bowel disease; MHC: major histocompatibility complex; PBC: primary biliary cholangitis; PBD: primary bile duct; PSC: primary sclerosing cholangitis; SNP: single nucleotide polymorphism;

Keywords

Primary sclerosing cholangitis, genetics, liver transplantation

Correspondence

Rinse K Weersma, Department of Gastroenterology and Hepatology, University of Groningen and University Medical Center Groningen, PO Box 30.001, 9700RB, Groningen, the Netherlands

Tel: +31503610426, Fax: +31503619306, Email: r.k.weersma@umcg.nl

Word count: 3655

ABSTRACT

Objective Primary sclerosing cholangitis (PSC) is a genetically complex, inflammatory bile duct disease of largely unknown etiology often leading to liver transplantation or death. Little is known about the genetic contribution to the severity and progression of PSC. The aim of this study is to identify genetic variants associated with PSC disease progression and development of complications.

Design We collected standardized PSC subphenotypes in a large cohort of 3,402 PSC patients. After quality control we combined 130,422 single nucleotide polymorphisms of all patients – obtained using the Illumina Immunochip – with their disease subphenotypes. Using logistic regression and Cox proportional hazards models we identified genetic variants associated with binary and time-to-event PSC subphenotypes.

Results We identified genetic variant rs853974 to be associated with liver transplant-free survival ($P = 6.07 \times 10^{-9}$). Kaplan-Meier survival analysis showed a 50.9% (95% CI 41.5-59.5%) transplant-free survival for homozygous AA allele carriers of rs853974 compared with 72.8% (95% CI 69.6-75.7%) for GG carriers at ten years after PSC diagnosis. For the candidate gene in the region, *RSPO3*, we demonstrated expression in key liver-resident effector cells, such as human and murine cholangiocytes and human hepatic stellate cells.

Conclusion We present a large international PSC patient cohort, and report genetic loci associated with PSC disease progression. For liver transplant-free survival, we identified a genome wide significant signal and demonstrated expression of the candidate gene *RSPO3* in key liver-resident effector cells. This warrants further assessments of the role of this potential key PSC modifier gene.

SUMMARY BOX

What is already known about this subject:

- Several case-control genome-wide association studies have revealed 20 susceptibility loci for primary sclerosing cholangitis.
- Little is known about the genetic contribution to the severity and progression of complex diseases in general and primary sclerosing cholangitis in particular.
- *RSPO3* plays a role in the activation of the canonical Wnt signaling pathway, which is involved in liver fibrosis.

What are the new findings:

- The genetic variant rs853974 is genome-wide significantly associated with liver transplant-free survival in primary sclerosing cholangitis.
- Candidate gene *RSPO3* is expressed in both murine and human cholangiocytes, and in human hepatic stellate cells.
- Three new loci were found to be associated with time to cholangiocarcinoma in patients with primary sclerosing cholangitis.

How might it impact on clinical practice in the foreseeable future?

- Through its effect on liver fibrosis, *RSPO3* could play an important role in PSC disease progression, and insight in its mechanism could lead to new therapeutic targets.
- Furthermore, since we demonstrated that genetic variants are associated with PSC disease progression, genetics could provide a tool for risk stratification of patients with PSC in the future.

INTRODUCTION

Primary sclerosing cholangitis (PSC) is a complex, cholestatic liver disease, in which chronic biliary inflammation and bile duct destruction leads to biliary fibrosis and liver cirrhosis, often in a slowly progressive manner.[1] PSC is characterized by a cholangiographic image of strictures interchanged with dilatations throughout the biliary tract. Reported incidence rates of PSC vary widely, with incidence rates of 0.91, 1.31, and 0.5 per 100,000 inhabitants per year for North America, Norway, and the Netherlands, respectively.[2,3] There is a male to female ratio of 2:1, and the disease can occur at any age, with a peak incidence around 40 years.[3] There is a close association between PSC and inflammatory bowel disease (IBD), and PSC patients are subject to a 5-fold increased risk of developing colorectal carcinoma (CRC) when compared with the general population.[3] In addition, PSC carries an excess risk of cholangiocarcinoma (CCA) which seems to be unrelated to the disease duration and the presence of liver cirrhosis.[4] There is no effective medical therapy that can halt disease progression in PSC. The only curative option to date is liver transplantation.

The etiology of PSC is still largely unknown. The etiology is most likely to be multifactorial, in which the occurrence of PSC could be triggered by environmental factors in a genetic susceptible host.[5] The relationship between susceptibility to PSC and environmental factors has been studied for several risk factors, of which smoking has repeatedly been shown to be associated with a decreased risk of developing PSC, independent of its protective effect in ulcerative colitis.[6]

Already in 1983, the identification of associations between PSC and HLA-B8 of the human leukocyte antigen (HLA) complex located on chromosome 6 - harboring several genes that are involved in antigen presentation and are important in immunity - raised interest in the role of genetics in PSC.[7] This was amplified by a large Swedish study on PSC heritability demonstrated a nearly 4 to 17 times increased risk for first degree relatives of PSC patients to develop PSC, when compared with the general population.[8] The additional 3.3 times increased risk of ulcerative colitis, and the presence of at least one

concomitant immune mediated disease outside the liver and bowel in approximately 20 to 25% of PSC patients, suggests a shared genetic component between these diseases.[8–10] Over the last 5 years, the application of genome-wide association studies has resulted in an increasing insight in the genetic architecture of PSC, with the identification of 19 non-HLA risk loci at the time of writing.[11–14]

Little is known about the genetic contribution to the severity and progression of complex diseases in general and PSC specifically. In Mendelian traits like cystic fibrosis and hemochromatosis, consortia efforts have led to the identification of robust and important modifier genes.[15,16] If genetic variants would be associated with PSC phenotypes, this could enable risk stratification of PSC patients according to disease behavior and would lead to insight into pathogenetic mechanisms associated to disease progression. Whilst translational research from susceptibility genes has yet to prove useful for the development of new drugs in complex disease, modifier genes may point toward pathways involved in disease progression amenable by pharmacological interventions.

The aim of this study is to identify genetic variants associated with PSC disease progression and development of complications, in a large, international, multicenter PSC cohort.

METHODS

Study design and patients

PSC patients previously recruited throughout Europe, the USA and Canada, and genotyped using the Immunochip by Liu et al.[13] were included. Subject recruitment was approved by the ethics committees or institutional review boards of all participating centers. Written informed consent was obtained from all participants. Patients of whom PSC diagnosis was revised after they were genotyped were excluded.

The following phenotypic data were collected for patient and disease characteristics: sex, date of birth, PSC subtype (small or large duct), date of PSC diagnosis, intra and-/or extra hepatic disease, dominant strictures, concomitant IBD and type of IBD, date of IBD diagnosis, and smoking status. Furthermore, follow-up data were collected for: date and cause of death, date and indication of liver transplantation, the occurrence and date of diagnosis of hepatocellular carcinoma (HCC), CCA, CRC, gallbladder carcinoma, and the occurrence and date of a colectomy.

PSC diagnosis was based on clinical, biochemical, cholangiographic and histologic criteria, as formulated by the EASL guidelines.[17] IBD diagnosis was scored based on accepted endoscopic, radiologic and histologic criteria.[18]

PSC related death was defined as death from liver failure, death from cholangiosepsis, death from cholangiocarcinoma or death from gallbladder carcinoma. The time-to-event phenotype liver transplant-free survival was defined as the time between PSC diagnosis and the composite endpoint of either liver transplantation or PSC related death.

We used genotype data of PSC patients as previously described.[13] Appendix A describes the quality control applied to this dataset. A total of 130,422 single nucleotide polymorphisms (SNPs) for 3,402 PSC samples remained after quality control and were used in the analysis.

Statistical analysis

The age at PSC diagnosis was expressed as median value and interquartile range. Categorical variables were expressed as numbers and percentages based on non-missing values.

Binary associations

Binary associations were calculated using multiple logistic regression. We corrected for clinical covariates by adding them to the regression model. To determine which clinical covariates to correct for, we performed a backwards elimination procedure per binary phenotype. We started with the full model including sex, country, date of PSC diagnosis, established IBD diagnosis, and smoking status, and removed covariates from the model until the AIC (Akaike information criterion) stabilized.

Time-to-event associations

Cox proportional hazards regression was used to estimate the effect of genetic variants on time-to-event PSC subphenotypes. Clinical variables that were significantly associated with the time-to-event phenotype in univariable Cox regression analyses ($P < 0.05$) were entered into a multivariable Cox model alongside the genotype. To visualize the effect of genotype on time-to-event phenotypes, Kaplan-Meier survival estimates were plotted. (Methods described in supplementary appendix A).

We used the SNP2HLA software[19] to impute classical HLA alleles from genotype data for *HLA-A, HLA-B, HLA-C, HLA-DRB1, HLA-DQA1, HLA-DQB1, HLA-DPA1* and *HLA-DPB1* and their corresponding amino acid polymorphisms from the genotype data (Methods described in supplementary appendix A).

Mouse experiments and in vitro experiments on the role of *RSPO3* in PSC

Mouse experiments

C57BL/6 (B6) mice were purchased from Charles River (Milan, Italy). Normal C57BL/6 mice were sacrificed at the age of 8-10 weeks. Organs were harvested and washed by cold phosphate buffered saline (PBS). Cholangiocytes were isolated both from normal mice (n=3) and from mice (n=3) fed 0.1%3,5-diethoxycarbonyl-1,4-dihydrocollidine (DDC) for 4 weeks, as a model of sclerosing cholangitis.[20] Total RNA was extracted and sequenced on an Illumina HiSeq 2000 machine. See Appendix A for more details.

In vitro experiments on human primary biliary tissue, cholangiocyte-like-cells and hepatic stellate cells

Human primary biliary tissue was obtained from a liver and pancreas organ donor after obtaining informed consent from the donor's family. A section of the bile duct was excised and homogenized and RNA was extracted. Cholangiocyte-like-cells were generated from human induced pluripotent stem cells and cultured. *RSPO3* expression was determined using qPCR and compared with the housekeeping gene using the $2^{-\Delta Ct}$ method.[21] Next to that, we used previously published microarray data to verify *RSPO3* expression.[22] The R/Bioconductor package limma [23] was used to evaluate differential expression between pairs of conditions (hIPSCs and CLCs and hIPSCs and PBD). A linear model fit was applied and p-values were corrected using the method of Benjamini and Hochberg.[24] Methods are further described in Appendix A.

Primary human hepatic stellate cells were isolated and cultured from wedge sections of liver tissue, obtained from patients undergoing surgery at the Royal Free Hospital in London. Total RNA was extracted and retro-transcribed into cDNA, which was used for gene expression assessment with qPCR. Gene expression was compared with the housekeeping gene using the $2^{-\Delta Ct}$ method.

RESULTS

Patient characteristics and natural history

Clinical characteristics of the PSC cohort are described in table 1. The cohort consisted of 2881 patients from Europe and 521 patients from the United States and Canada (Supplementary table 1). A total of 2,185 (65%) patients were male, and the median age at PSC diagnosis was 38.6 years (IQR 28.0 - 50.1). Concomitant IBD was diagnosed in 2,390 (75%) patients. The median follow-up was 8.7 years. In total 874 (26%) patients underwent liver transplantation and 181 (5%) patients died of PSC related causes. Over 11% of patients developed a malignancy, most often CCA (5.6%) or CRC (4.3%).

Genetic associations with binary subphenotypes

Genome-wide association analyses focusing on the occurrence of malignancy in PSC patients revealed several suggestive associations (Supplementary table 2). When comparing 107 PSC-AIH patients with 3,159 PSC patients without AIH overlap, a strong genetic association in the *HLA-DQB1* gene was identified (top SNP rs3891175, $P = 4.6 \times 10^{-11}$, OR = 2.41). After imputation of the classical HLA alleles, we found that the alleles DQA1*05:01 and DQB1*02:01 were most significantly associated with PSC/AIH overlap (P-values of 3.8×10^{-11} and 1.8×10^{-07}). For other binary subphenotypes - small duct PSC, the occurrence of HCC, gallbladder carcinoma, and proctocolectomy – no genetic associations were found.

Table 1 Clinical characteristics of the PSC cohort consisting of 3,402 patients

Variable	Groups	Number (%)
Age at PSC diagnosis*		38.6y (28.0y - 50.1y)
Sex	male	2185 (64.7%)
	female	1193 (35.3%)
	missing	24 (0.7%)
Main diagnosis	PSC	3159 (94.6%)
	small duct PSC	75 (2.2%)
	PSC with AIH overlap	107 (3.2%)
	missing	61 (1.8%)
Liver transplantation	Yes	874 (26.3%)
	No	2444 (73.7%)
	missing	84 (2.5%)
Colectomy	Yes	419 (12.6%)
	No	2897 (87.4%)
	missing	86 (2.5%)
IBD	No IBD	816 (25.5%)
	Ulcerative colitis	1940 (60.5%)
	Crohn's disease	357 (11.1%)
	IBD-U	93 (2.9%)
	missing	196 (5.8%)
Cholangiocarcinoma	Yes	188 (5.6%)
	No	3147 (94.4%)
	missing	67 (2.0%)
Colorectal carcinoma	Yes	127 (4.3%)
	No	2822 (95.7%)
	missing	453 (13.3%)
Gall bladder carcinoma	Yes	30 (1.0%)
	No	2977 (99.0%)
	missing	395 (11.6%)
Hepatocellular carcinoma	Yes	22 (0.7%)
	No	2984 (99.3%)
	missing	396 (11.6%)
Smoking status	Smoker	140 (6.0%)
	Ex-Smoker	529 (22.7%)
	Non-smoker	1657 (71.2%)
	missing	1076 (31.6%)
Death	Non PSC related	47 (1.5%)
	Liver failure	66 (2.1%)
	Cholangiosepsis	18 (0.6%)
	Gallbladder carcinoma	12 (0.4%)
	Cholangiocarcinoma	85 (2.6%)
	Hepatocellular carcinoma	6 (0.2%)
	Colorectal carcinoma in case of coexisting IBD	3 (0.1%)
	Alive	2977 (92.6%)
	missing	188 (5.5%)

*Values shown as median (IQR).
Quantitative data are expressed as counts and percentages excluding missing data.

Genetic associations with time-to-event subphenotypes

Next, we aimed to determine whether genetic variants are associated with important time-to-event variables reflecting the PSC disease course, e.g. time between PSC diagnosis and the development of a carcinoma. For this, we developed a framework to perform Immunochip-wide Cox proportional hazards analyses. We defined the liver transplant-free survival subphenotype as time from PSC diagnosis until liver transplantation or PSC related death. Univariable Cox regression analyses including clinical parameters showed statistically significant associations with the time to event endpoint transplant-free survival for sex, country, date of PSC diagnosis, established IBD diagnosis, and smoking status. Next, we tested 130,422 SNPs for association with liver transplant-free survival using multivariable Cox proportional hazards regression, including the genotype effect alongside the significant clinical co-variables. We found SNP rs853974 to be associated with liver transplant-free survival of PSC patients at genome wide significance ($P = 6.07 \times 10^{-9}$). Kaplan-Meier survival analysis showed a 50.9% (95% CI 41.5-59.5%) transplant-free survival for homozygous AA allele carriers of rs853974 compared with 72.8% (95% CI 69.6-75.7%) for GG carriers at ten years after PSC diagnosis (figure 1A). AA homozygotes had a 2.14 (95% CI = 1.66-2.76) increased hazard, indicating a 2.14 larger relative risk for need for liver transplantation or for PSC related death compared to GG homozygotes. Figure 1B shows a regional plot of this observed association.

SNP rs853974 is located on chromosome 6. We did not identify a direct functional effect of this SNP on gene expression or regulatory features (Appendix A). We hypothesized that neighbouring gene *R-spondin 3 (RSPO3)* would be the most likely positional candidate gene. The other neighboring gene, *CENPW*, has a fundamental role in kinetochore assembly and is required for normal chromosome organization and progress through mitosis and therefore not a good candidate. In addition to SNP rs853974, additional suggestive genetic associations with time-to-event phenotypes liver transplant-free survival and time to CCA were found (Supplementary table 3).

Expression of *RSPO3* in key liver-resident effector cells

To assess whether *RSPO3* is expressed in disease relevant cells (cholangiocytes and hepatic stellate cells), we performed RNA-sequencing on healthy and cholestatic cholangiocytes and multiple organs derived from normal C57BL/6 mice. *RSPO3* expression was 7 to 20 folds higher in cholangiocytes as compared with any of the organs. Furthermore, *RSPO3* expression was higher in healthy cholangiocytes than in cholestatic cholangiocytes (figure 2A).

Next, using microarrays, we assessed expression of *RSPO3* in human induced pluripotent stem cells, human induced pluripotent stem cell-derived cholangiocyte-like cells, and human primary bile duct samples. *RSPO3* expression was significantly higher in cholangiocyte-like cells and primary bile duct cells compared with human induced pluripotent stem cells (figure 2B). This finding was confirmed by qPCR (figure 2C).

Since activated hepatic stellate cells are the main cells involved in liver fibrosis,[25] we also investigated expression of *RSPO3* in human culture-activated hepatic stellate cells. We isolated, cultured and activated human hepatic stellate cells of three patients without PSC. Using qPCR we observed expression of the hepatic stellate cell marker gene Cytoglobin B, as well as expression of *RSPO3* in all three subjects (figure 2D). We did not observe *RSPO3* expression in human CD4 and CD8 T lymphocytes (data not shown).

DISCUSSION

To date, very few disease-modifying genes have been identified in rare complex diseases. Collaboration within the international PSC Study group enabled the establishment of a cohort of unprecedented size, for an orphan disease such as PSC, enabling the investigation of genetic variation underlying the progression of PSC through time. Overall, it is a major challenge to determine genetic variants associated with survival, and only few genetic studies investigating this have been published.[26] We present a conceptually new method to determine associations between genetic variants and disease course, using genome wide multivariable Cox proportional hazards regression analyses. Here we identify a genome wide significant association between SNP rs853974 – located close to the *RSPO3* gene – and liver transplant-free survival in PSC. Interestingly, this locus is not associated with PSC susceptibility and thus exemplifies different genetic regulation of disease susceptibility and disease progression.

This study is based on genotype data obtained using Illumina immunochip, a genotyping platform that densely covers genetic regions associated to immune mediated diseases. Use of GWAS arrays, that more uniformly cover genetic variants all over the genome, would have been ideal. However, a complete GWAS dataset for the entire international cohort was not available at the time of study. For that reason we started with the available immunochip data. Given the positive findings in this study, a similar study on GWAS arrays data could very well be of additional value.

For the binary phenotype of developing cholangiocarcinoma or not, we found an association at chromosome 5 at 150 Mb (Supplementary table 2). Of interest, this locus contains an established genetic association with Crohn's disease susceptibility in the autophagy gene *IRGM*. [27] For the phenotype of developing colorectal carcinoma or not we found an association at chromosome 14 at 35 Mb. This locus appeared not to be associated with sporadic colorectal cancer.

Since transplant-free survival is a combined and heterogeneous phenotype we

assessed to which extent the following three subgroups contribute to the association: 1) transplanted patients with indication for transplantation "end stage liver disease" and patients died because of "liver failure"; 2) transplanted patients with indication for transplantation "cca/high grade dysplasia" and patients died because of "cca or gallbladder carcinoma"; and 3) transplanted patients with indication for transplantation "intolerable complaints/pruritus/recurrent cholangitis" and patients died of "cholangiosepsis". We observed a stronger contribution of subgroups 1 and 3 to the association, indicating that the underlying biological mechanism is more likely one involved in causing progression of liver disease and/or cholangitis or cholangiosepsis rather than a mechanism involved in cancer development.

R-spondin 3 is a member of the R-spondin protein family (R-spondin 1-4).[28] These proteins are secreted agonists of the canonical Wnt/β-catenin signaling pathway.[28] They activate the pathway leading to induced transcription of Wnt target genes. Wnt/β-catenin signaling plays a central role in embryogenesis, organogenesis and adult homeostasis, and is a critical regulator of stem cell maintenance.[29,30] RSPO3 is a ligand of the Frizzled 8 and LRP5/6 receptors.[28] In the canonical form of the Wnt pathway, binding of ligands to the Frizzled (Fzd) receptor and LRP5 or 6 co-receptors causes β-catenin to dephosphorylate in the cytoplasm. Accumulated β-catenin translocates to the nucleus where it binds to T cell factor (TCF)/Lymphoid enhancer-binding factor (LEF), causing transcription of Wnt target genes – such as *Fibronectin*, *MMP-7, Twist,* and *Snail*. These factors activate hepatic stellate cells and induce liver fibrosis. Blocking the Wnt signaling pathway using Dickkopf-1 (Dkk-1), a Wnt co-receptor antagonist, restores hepatic stellate cells quiescence in culture.[31] Hence, Wnt signaling is involved in both progression and regression of liver fibrosis, either by inhibiting or promoting activation and survival of hepatic stellate cells.[31,32] Also, RSPOs have been shown to facilitate hepatic stellate cell activation and promote hepatic fibrogenesis.[33] Here, we demonstrate that *RSPO3* is expressed in key effector cells involved in the pathogenesis of PSC. Since we have shown that PSC patients that are homozygote AA carriers of rs853974 progress more rapidly towards PSC related death or

liver transplantation, *RSPO3* can be regarded a plausible candidate gene to be involved in PSC disease progression. Hypothetically, PSC patients might benefit from reduction of *RSPO3* or generally canonical Wnt signaling.

In an Immunochip analysis of the International IBD Genetics Consortium including over 75,000 individuals[34], an intronic SNP rs9491697 in *RSPO3* (which is not in linkage disequilibrium with rs853974, $r^2 = 0.014$) was identified to be associated with Crohn's disease ($P = 3.79 \times 10^{-10}$, OR = 1.077) but not with ulcerative colitis. Given the small number of Crohn's disease patients (n = 357) within the present study, the lack of linkage disequilibrium between the two “hit SNPs”, and the fact that our multivariate Cox model corrected for IBD-status, the identified association signal does not seem to be driven by the co-occurrence of Crohn's disease in our cohort.

For several binary and time-to-event subphenotypes we found suggestive genetic associations. Two additional SNPs, rs1532244 on chromosome 3 and rs17649817 on chromosome 5, were suggestively associated with transplant-free survival. Furthermore, one SNP, rs7731017, was suggestively associated with the presence of CCA. We investigated whether any of the candidate genes in the locus overlapped with genes identified in tumour sequencing studies of cholangiocarcinoma. We did not find an overlap with the 32 genes reported to be significantly altered in intrahepatic, extrahepatic, and gallbladder cancer by Nakamura et al.[35] When comparing the genes in our CCA locus with 1146 genes containing non-synonymous somatic mutations in intrahepatic cholanciocarcinoma[36], we found that the *SYNPO* gene was both in the list of 1146 genes of the sequencing study as well as in the locus that we identified to be associated with the presence of CCA. There is little known about this gene and there is no connection with oncogenesis. Another gene, *SP100,* was found in this study to be in the locus associated with time to CCA and is also in the list of 1146 genes. *SP100* is associated with autoimmune disease of the urogenital tract and also with PBC. Interestingly, anti-sp100 autoantibodies have been described for PBC[37]. The genetic association of *SP100* with both PBC and the time to CCA within PSC patients as well as the existence of anti-sp100 autoantibodies makes this an interesting gene

for future follow-up studies.

When comparing PSC-AIH patients with PSC patients without AIH, we found a strong genetic association with PSC-AIH in the HLA-DQB1 gene. The identified variant was tagging the classical HLA haplotypes DQA1*05:01 and DQB1*02:01. These associations overlap the associations found by a previous genome-wide association study of AIH type 1 in The Netherlands,[38] suggesting that the genetic basis for AIH type 1 pathogenesis is similar for isolated AIH type 1 patients, compared with PSC-AIH patients.

This study is limited by the relatively small cohort size, when compared with other GWAS studies that incorporate tens of thousands of samples. The resulting lack of statistical power may have played a role in the binary analyses, in which suggestive hits were found for CCA and CRC but genome wide significance was not reached. However, PSC is a rare disease, and the present study has included patients recruited throughout the world in a joined effort. It is therefore not expected that a larger cohort of PSC cases will become available soon.

In conclusion, we present the largest association study of PSC genotypes with disease phenotypes to date. We identified several genetic variants associated with PSC disease course. Specifically, we report rs853974 to be genome-wide significantly associated with liver transplant-free survival in PSC. Findings of candidate gene *RSPO3* being expressed in both mouse and human cholangiocytes and human activated hepatic stellate cells warrant further assessments of the role of this potential key PSC modifier gene.

Acknowledgments

We thank all PSC patients for participation in this study. We acknowledge the members of the International PSC Study Group and the UK-PSC Consortium for their participation. We acknowledge Lukas Tittman for providing samples.

RA is supported by a PSC Partners Seeking a Cure grant "Unraveling genetics driving PSC subphenotypes; An IPSCSG study". LV and FS are supported by the ERC grant Relieve IMDs and the Cambridge Hospitals National Institute for Health Research Biomedical Research Center. SH is supported by a grant from the German Research Community (DFG) grant HO 4460/2-1. TM is supported by the German Research Community (DFG), grants MU 2864/1-1 and MU 2864/1-3. KNL is supported by the NIH RO1 DK 084960 and Sigismunda Palumbo Charitable Trust. EAMF is supported by a Career Development Grant from Dutch Digestive Foundation (Maag Lever Darm Stichting, MLDS). RKW is supported by a VIDI grant (016.136.308) from the Netherlands Organization for Scientific Research (NWO) and a PSC Partners Seeking a Cure grant "The Exome in PSC".

Contributors

RA, EMGdV, XJ, FS, KR, KS, ALM, and WW: statistical analysis and interpretation of data. CYP and RKW: study supervision. KR, KS, XJ, FS, and MP: performed experiments. RA, EMGdV, THK, SH, CS, TF, JRH, EM, FS, CYP, and RKW wrote the manuscript. JZL, AFranke, DE, and CAA performed genotyping, calling and QC. RA, EMGdV, ECG, XJ, FS, KR, KS, TF, TJW, ALM, WW, GA, DA, AB, NKB, UB, EB, KMB, CLB, MCB, MC, OC, AC, GD, JE, BE, DE, MF, EAMF, AFloreani, IF, DNG, GMH, BvH, KH, SH, JRH, FI, PI,BDJ, HL,WL, JZL, H-UM, MM, EM, PM, TM, AP, CRupp, CRust, RNS, CS, SS, ES, MSilverberg, BS, MSterneck, AT, LV, JV, AVV, BdV, KZ, RWC, MPM, MP, SMR, KNL, AFranke, CAA, THK, CYP, RKW, The UK-PSC Consortium, and The International PSC Study Group contributed to sample and clinical data collection. All authors revised the manuscript for critical content and approved the final version.

Competing Interests

None declared.

Transcript profiling

Previously published microarray data characterizing the transcriptomic profile of PBD and CLCs were used. This data is available on ArrayExpress, accession number: E-MTAB-2965.

References

1 Hirschfield GM, Karlsen TH, Lindor KD, *et al.* Primary sclerosing cholangitis. *Lancet* 2013;**382**:1587–99.

2 Boonstra K, Beuers U, Ponsioen CY. Epidemiology of primary sclerosing cholangitis and primary biliary cirrhosis: A systematic review. *J Hepatol* 2012;**56**:1181–8.

3 Boonstra K, Weersma RK, van Erpecum KJ, *et al.* Population-based epidemiology, malignancy risk, and outcome of primary sclerosing cholangitis. *Hepatology* 2013;**58**:2045–55.

4 Boberg KM, Lind GE. Primary sclerosing cholangitis and malignancy. Best Pract. Res. Clin. Gastroenterol. 2011;**25**:753–64.

5 Eaton JE, Talwalkar J a, Lazaridis KN, *et al.* Pathogenesis of primary sclerosing cholangitis and advances in diagnosis and management. *Gastroenterology* 2013;**145**:521–36.

6 Mitchell S, Thyssen M, Orchard T, *et al.* Cigarette smoking, appendectomy, and tonsillectomy as risk factors for the development of primary sclerosing cholangitis: a case control study. *Gut* 2002;**51**:567–73.

7 Chapman R, Varghese Z, Gaul R, *et al.* Association of primary sclerosing cholangitis with HLA-B8. *Gut* 1983;**24**:38–41.

8 Bergquist A, Montgomery SM, Bahmanyar S, *et al.* Increased risk of primary sclerosing cholangitis and ulcerative colitis in first-degree relatives of patients with

primary sclerosing cholangitis. *Clin Gastroenterol Hepatol* 2008;**6**:939–43.

9 Lamberts LE, Janse M, Haagsma EB, *et al.* Immune-mediated diseases in primary sclerosing cholangitis. *Dig liver Dis* 2011;**43**:802–6.

10 Saarinen S, Olerup O, Broomé U. Increased frequency of autoimmune diseases in patients with primary sclerosing cholangitis. *Am J Gastroenterol* 2000;**95**:3195–9.

11 Folseraas T, Melum E, Rausch P, *et al.* Extended analysis of a genome-wide association study in primary sclerosing cholangitis detects multiple novel risk loci. *J Hepatol* 2012;**57**:366–75.

12 Ellinghaus D, Folseraas T, Holm K, *et al.* Genome-wide association analysis in primary sclerosing cholangitis and ulcerative colitis identifies risk loci at GPR35 and TCF4. *Hepatology* 2013;**58**:1074–83.

13 Liu JZ, Hov JR, Folseraas T, *et al.* Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis. *Nat Genet* 2013;**45**:670–5.

14 Ellinghaus D, Jostins L, Spain SL, *et al.* Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat Genet* 2016;**48**:510–8.

15 Bartlett JR, Friedman KJ, Ling SC, *et al.* Genetic modifiers of liver disease in cystic fibrosis. *JAMA* 2009;**302**:1076–83.

16 Stickel F, Buch S, Zoller H, *et al.* Evaluation of genome-wide loci of iron metabolism in hereditary hemochromatosis identifies PCSK7 as a host risk factor of liver cirrhosis. *Hum Mol Genet* 2014;**23**:3883–90.

17 Lennard-Jones JE. Classification of inflammatory bowel disease. *Scand J Gastroenterol Suppl* 1989;**170**:2-6-19.

18 'European Association for the Study of the Liver'. EASL Clinical Practice Guidelines: management of cholestatic liver diseases. *J Hepatol* 2009;**51**:237–67.

19 Jia X, Han B, Onengut-Gumuscu S, *et al.* Imputing Amino Acid Polymorphisms in Human Leukocyte Antigens. *PLoS One* 2013;**8**.

20 Fickert P, Stöger U, Fuchsbichler A, *et al.* A new xenobiotic-induced mouse model of sclerosing cholangitis and biliary fibrosis. *Am J Pathol* 2007;**171**:525–36.

21 Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative CT method. *Nat Protoc* 2008;**3**:1101–8.

22 Sampaziotis F, Cardoso de Brito M, Madrigal P, *et al.* Cholangiocytes derived from human induced pluripotent stem cells for disease modeling and drug validation. *Nat Biotechnol* 2015;**33**:845–52.

23 Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 2004;**3**:Article3.

24 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B* 1995;**57**:289–300.

25 Mederacke I, Hsu CC, Troeger JS, *et al.* Fate tracing reveals hepatic stellate cells as dominant contributors to liver fibrosis independent of its aetiology. *Nat Commun* 2013;**4**:2823.

26 Wu C, Li D, Jia W, *et al.* Genome-wide association study identifies common variants in SLC39A6 associated with length of survival in esophageal squamous-cell carcinoma. *Nat Genet* 2013;**45**:632–8.

27 Parkes M, Barrett JC, Prescott NJ, *et al.* Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nat Genet* 2007;**39**:830–2.

28 Nam JS, Turcotte TJ, Smith PF, *et al.* Mouse cristin/R-spondin family proteins are novel ligands for the frizzled 8 and LRP6 receptors and activate ??-catenin-dependent gene expression. *J Biol Chem* 2006;**281**:13247–57.

29 van Amerongen R, Nusse R. Towards an integrated view of Wnt signaling in development. *Development* 2009;**136**:3205–14.

30 de Lau WB, Snel B, Clevers HC. The R-spondin protein family. *Genome Biol* 2012;**13**:242.

31 Cheng JH, She H, Han Y-P, *et al.* Wnt antagonism inhibits hepatic stellate cell

activation and liver fibrosis. *Am J Physiol Gastrointest Liver Physiol* 2008;**294**:G39–49.

32 Myung SJ, Yoon J-H, Gwak G-Y, *et al.* Wnt signaling enhances the activation and survival of human hepatic stellate cells. *FEBS Lett* 2007;**581**:2954–8.

33 Xinguang Y, Huixing Y, Linlin W, *et al.* RSPOs facilitated HSC activation and promoted hepatic fibrogenesis. *Oncotarget* 2016;**5**.

34 Jostins L, Ripke S, Weersma RK, *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 2012;**491**:119–24.

35 Nakamura H, Arai Y, Totoki Y, *et al.* Genomic spectra of biliary tract cancer. *Nat Genet* 2015;**47**:1003–10.

36 Gao Q, Zhao YJ, Wang XY, *et al.* Activating mutations in PTPN3 promote cholangiocarcinoma cell proliferation and migration and are associated with tumor recurrence in patients. *Gastroenterology* 2014;**146**:1397–407.

37 Norman GL, Bialek A, Encabo S, *et al.* Is prevalence of PBC underestimated in patients with systemic sclerosis? *Dig Liver Dis* 2009;**41**:762–4.

38 De Boer YS, Van Gerven NMF, Zwiers A, *et al.* Genome-wide association study identifies variants associated with autoimmune hepatitis type 1. *Gastroenterology* 2014;**147**.

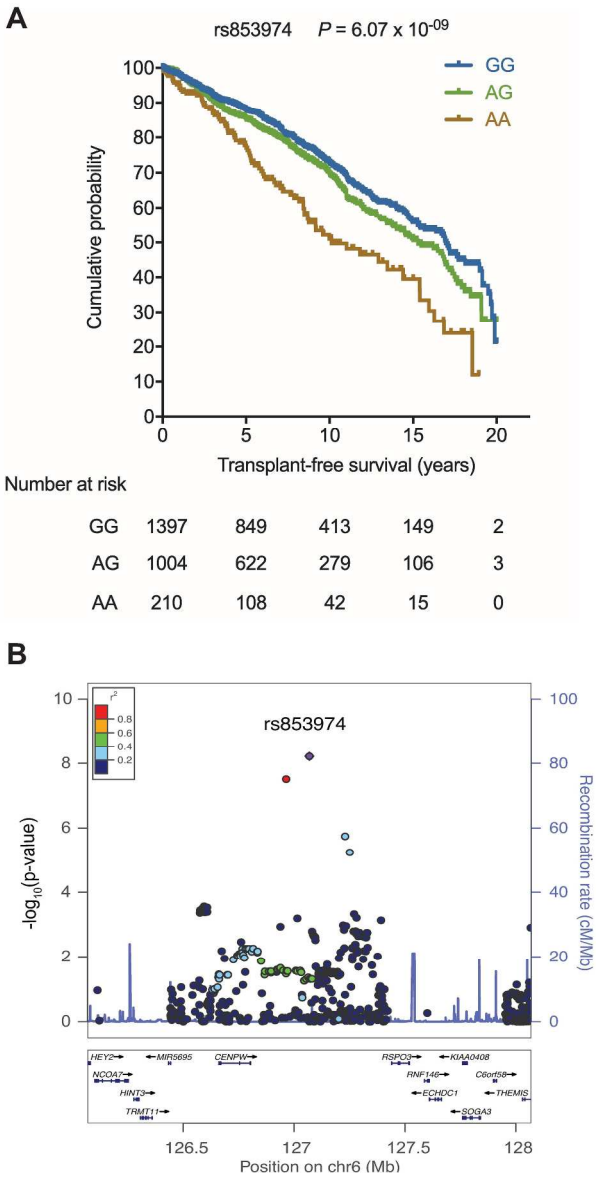
Figure legends

Figure 1: Association of genetic variants on chromosome 6 with transplant-free survival of PSC patients.

(A) Kaplan-Meier curves of transplant-free survival. Patients are stratified according to their genotype for SNP rs853974. The *P*-value for genotype effect in the Cox proportional hazards model is $P = 6.07 \times 10^{-09}$. (B) Regional association plot for transplant-free survival. The Y-axis shows the $-\log_{10}$(*P*-value) for genotype effect in the Cox proportional hazards model.

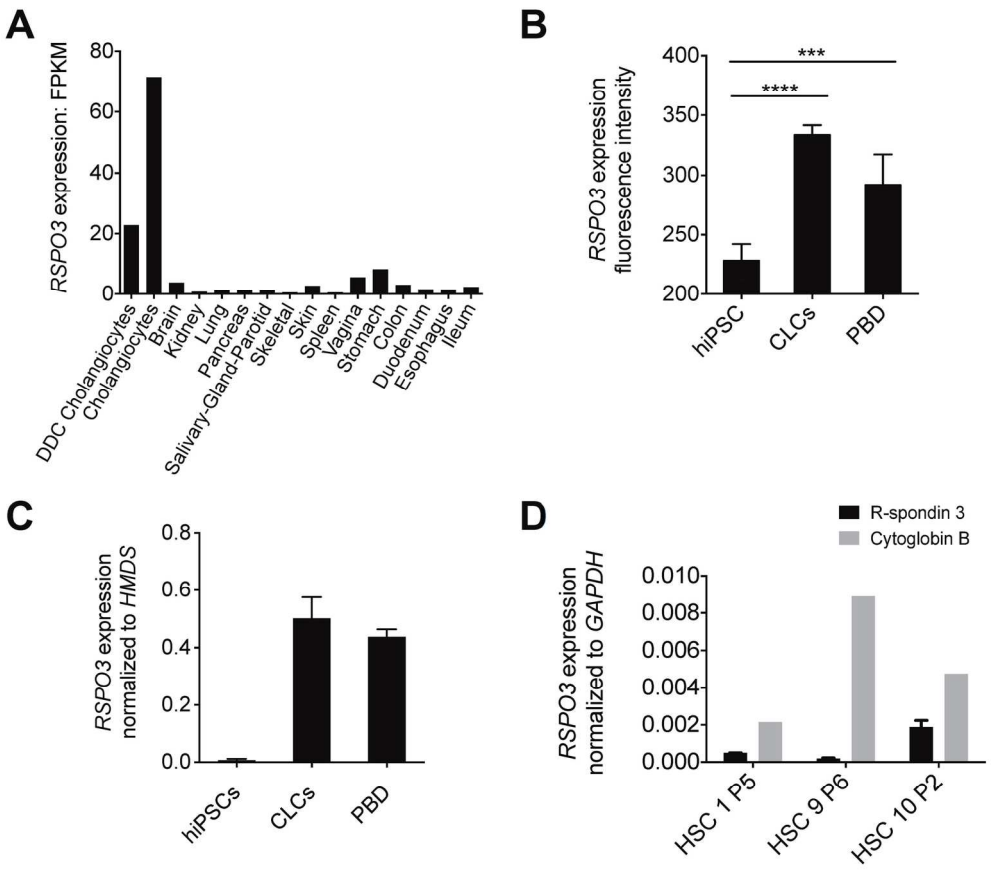
Figure 2: *RSPO3* expression in mouse cholangiocytes and in human cholangiocyte-like-cells, primary bile duct and hepatic stellate cells

(A) RNAseq analysis of *RSPO3* expression in DDC induced cholestatic cholangiocytes, healthy cholangiocytes and multiple organs of normal C57BL/6 mice. FPKM: Fragments Per Kilobase of exon per Million mapped reads. (B) Microarray *RSPO3* expression in human induced pluripotent stem cells (hiPSCs), cholangiocyte-like-cells (CLCs) and primary bile duct (PBD). *RSPO3* expression is significantly increased in CLCs and in PBD compared to hiPSCs. n=3; error bars, standard deviation. Asterisks represent statistical significance (****adjusted P<0.0001, ***adjusted P<0.001, Benjamini and Hochberg corrected P-values). (C) Quantitative real time PCR (qPCR) analysis demonstrating the expression of *RSPO3* in hiPSC-derived CLCs and PBD samples compared to expression in hiPSCs. Expression levels are fold changes compared to housekeeping gene HMDS calculated using the $2^{-\Delta Ct}$ method. (D) Quantitative real time PCR (qPCR) analysis showing expression of *RSPO3* and Cytoglobin B in three patients without PSC. Cytoglobin B mRNA expression was evaluated as specific HSC marker. Target genes were normalized using *GAPDH* as endogenous control and their relative expression was calculated with the $2^{-\Delta Ct}$ method.



Association of genetic variants on chromosome 6 with transplant-free survival of PSC patients

254x297mm (300 x 300 DPI)



RSPO3 expression in mouse cholangiocytes and in human cholangiocyte-like-cells, primary bile duct and hepatic stellate cells

158x141mm (300 x 300 DPI)

Supplementary appendix

Genetic association analysis identifies variants associated with disease progression in primary sclerosing cholangitis

Alberts R*, de Vries EMG* et al

Appendix

Contents

1. Supplementary Methods

1.1 Genotype data

Immunochip genotyping and quality control

Here, we used genotype data of 3,789 PSC patients previously described.[1] As quality control, SNPs with call rate below 80% were removed. Next, using PLINK v1.07 [2], per sample genotype call rate and heterozygosity rate were determined. Outlying samples were identified using Aberrant.[3] Also, for each pair of individuals with estimated identity by descent ≥ 0.9, the sample with the lower call rate was removed. Related individuals with identity by descent > 0.1875 and < 0.9 were kept. Samples of non-European ancestry were identified using PCA analysis and removed.[4] Using samples from the 1000 Genomes Project[5] principal components were calculated and PSC cases were projected onto them.[1,6,7] 3,402 PSC cases remained after quality control. As SNP quality control, SNPs were excluded that (i) had a MAF below 0.1%, (ii) had a Hardy-Weinberg equilibrium $P < 1 \times 10^{-5}$, (iii) had a call rate below 98% or (iv) failed the non-random differential missing data rate test of PLINK between cases and controls at a threshold of $P < 1 \times 10^{-5}$. After cluster plot inspection, a total of 130,422 SNPs remained for analysis.

1.2 Statistical and computational analyses

Statistical analyses

Binary associations were calculated using multiple logistic regression. We corrected for clinical covariates by adding them to the logistic regression model. To determine which clinical covariates to correct for, we performed a backwards elimination procedure for each binary phenotype. We started with the full model and removed covariates from the model until the AIC (Akaike information criterion) stabilized.

The time-to-event phenotype liver transplant-free survival was defined as the composite endpoint of liver transplantation or PSC related death (death from liver failure, death from cholangiosepsis, death from cholangiocarcinoma or death from gallbladder carcinoma). To determine whether clinical parameters have an effect on liver transplant-free survival, univariable Cox proportional hazards models were fitted including clinical parameters as covariates. Next, to determine associations between genotype and liver transplant-free survival, for all SNPs on the Immunochip after quality control, we fitted a Cox regression model including the significant clinical variables as well as genotype. Genotype was modeled as nominal variable. Associations with a *P*-value < 5.0×10^{-08} were considered significant, whereas associations with a *P*-value between 5.0×10^{-06} and 5.0×10^{-08} were considered suggestive. Proportional hazard assumptions were judged using Therneau and Grambsch tests. After stratifying the IBD type (No IBD, CD, UC or IBD-U) and categorizing time of IBD diagnosis into five groups, the proportional hazard assumptions were met. To visualize the effect of genotype on time-to-event phenotypes, Kaplan-Meier survival estimates were plotted. Cox regression analyses were also performed for time to death, time to liver transplantation, time to CCA, time to CRC or time to HCC, time to gall bladder carcinoma and time to proctocolectomy. All time intervals started at the time of PSC diagnosis.

Association tests were performed in PLINK version 1.07[2]. Survival analyses were done in R software version 3.1.3 (http://www.r-project.org). Genetic loci were defined as 250kb around the lead SNP and eventually extended to include all SNPs in LD with the lead SNP ($r^2 > 0.8$). In case a locus contained no genes we investigated the closest neighbouring ones.

Imputation of the HLA region

Based on the Immunochip genotypes, we imputed classical HLA alleles for *HLA-A, HLA-B, HLA-C, HLA-DRB1, HLA-DQA1, HLA-DQB1, HLA-DPA1* and *HLA-DPB1* and their corresponding amino acid polymorphisms, using a reference panel of 5,225 individuals of European descent, collected by the Type 1 Diabetes Genetics Consortium and using SNP2HLA with default settings.[8] This reference panel showed high imputation quality for the HLA region.[8–10]

In silico functional analyses

The Encyclopedia of DNA Elements (ENCODE)[11] was searched using the UCSC Genome Browser.[12] Specifically, overlaps of SNPs with the following regulatory features were searched: DNaseI hypersensitivity sites, transcription factor binding sites, histone modification and DNA-polymerase sites. We tested whether associated variants showed an effect on gene expression levels of genes. For this we used an expression

quantitative trait loci (eQTL) study of non-transformed peripheral blood in 5,311 individuals (http://www.genenetwork.nl/bloodeqtlbrowser/)[13] and we queried Single Tissue and Multi-tissue eQTL for all available tissues on the GTEx Portal (http://gtexportal.org), an online resource and associated tissue bank for the scientific community for studying the relationship between genetic variation and gene expression in human tissues.[14]

1.3 *RSPO3* expression in murine cholangiocytes and normal tissue

Mice

C57BL/6 (B6) mice were purchased from Charles River (Milan, Italy). The mice were housed in a Minimal Disease Unit at the animal facility at the Università Politecnica delle Marche, Ancona, Italy. All animal experiments were performed in compliance with local institution guidelines (Italian Ministry of Health).

Tissues and cholangiocytes preparation

Normal C57BL/6 mice were sacrificed at the age of 8-10 weeks. Brain, kidney, lung, pancreas, salivary-gland-parotid, skin, skeletal, spleen, stomach, vagina, colon, duodenum, esophagus and ileum were harvested and washed by cold phosphate buffered saline (PBS). Cholangiocytes were isolated both from normal mice (n=3) and from mice (n=3) fed 0.1%3,5-diethoxycarbonyl-1,4-dihydrocollidine (DDC) for 4 weeks, as a model of sclerosing cholangitis.[15] Purification of cholangiocytes from mice was performed by immune-bead purification, as previously reported.[16]

RNA sequencing

Total RNA was extracted by TRIzol Reagent (Life Technologies Corporation, Woburn, MA). Libraries were prepared using TruSeq RNA (Illumina) reagents, and sequenced on 4 lanes of an Illumina HiSeq 2000 machine, using SBS v3 reagents (Illumina) to generate 50 bp single end reads. The reads then were aligned to the mouse genome reference sequence and annotations from Illumina built for UCSC mm10 (version of 2014_05_23) with the Tophat software (version 2.1.0) using default parameters without gene/transcript discovery. Gene expression levels were quantified and normalized using cuffquant and cuffnorm with default parameters from the software cufflinks (version 2.2.1). Fragments Per Kilobase of exon per Million mapped reads (FPKM) values were used as gene expression levels.

1.4 *RSPO3* expression in human primary biliary tissue and cholangiocyte-like-cells

Human primary biliary tissue and cholangiocyte-like-cells culture

Primary biliary tissue (primary bile duct, PBD) was obtained from a liver and pancreas organ donor after obtaining informed consent from the donor's family (REC reference number: 09/H0306/73). A section of the bile duct was excised and homogenized using a tissue homogenizer. RNA was extracted using a kit (Sigma-Aldrich), according to the manufacturer's instructions. Cholangiocyte-like-cells (CLCs) were generated from human induced Pluripotent Stem Cells (hPSCs) and cultured as previously described.[17]

Real-time PCR (qPCR)

500ng of cellular RNA was reverse transcribed using Superscript II Reverse Transcriptase (Invitrogen). qPCR reaction mixtures were prepared using the SensiMix™ SYBR® & Fluorescein Kit (Bioline), according to the manufacturer's instructions. The cDNA was denatured at 94°C for 5 minutes, followed by 40 cycles at 94°C for 30 seconds, 60°C for 30 seconds, and 72°C for 30 seconds, and a final extension step at 72°C for 10 minutes. A Stratagene Mx3005P was used for all qPCR reactions.

Microarray data

Previously published raw and processed microarray data characterizing the transcriptomic profile of primary bile duct cells and cholangiocyte-like-cells were used.[17] This data is available on ArrayExpress (Accession number: E-MTAB-2965).

1.5 In vitro studies on human hepatic stellate cells

Isolation, culture and experimental treatment of human hepatic stellate cells

Primary human hepatic stellate cells were isolated from wedge sections of liver tissue, obtained from three patients undergoing surgery at the Royal Free Hospital after giving informed consent (EC01.14-RF). Cells were isolated according to Mederacke et al.[18], with modifications for human liver.[19,20] Briefly, 10 g of total human liver tissue was digested with 0.01% Collagenase, 0.05% Pronase and 0.001% DNase I without performing perfusion. The homogenate was filtered through a 100µm cell strainer (BD Falcon) and the flow-through was centrifuged at 50xg for 2 minutes at 4°C. After washing the supernatant, gradient centrifugation was performed at 1400xg for 17 minutes at 4°C using an 11.5% Optiprep gradient (Sigma). Finally, the interface was collected and washed. Purity of human hepatic stellate cells was established by detection of CD140b (PDGFRbeta), CD29 (Integrin beta 1) and Cytoglobin B (CYGB).

The obtained human hepatic stellate cells were cultured in IMDM supplemented with 20% foetal bovine serum (FBS), Glutamine, nonessential amino acids 1X, 1.0 mM sodium pyruvate, 1X antibiotic-antimycotic (all Life Technologies), referred to as complete HSC medium hereinafter. Experiments described in this study were performed on human hepatic stellate cells of at least three independent cell preparations, between passages 2 and 8.

Real-time qPCR

Total cellular RNA was extracted and then cleaned up by using QIAzol Lysis Reagent and RNeasy Kit (Qiagen, CA, USA), respectively, according to the manufacturer's protocol. After quantification using the NanoDrop1000 System (Thermo Scientific, USA), RNA was retro-transcribed into cDNA using the QuantiTect Reverse Transcription Kit (Qiagen, CA, USA) and 6.7 µg of the cDNA sample was used to set up real-time quantitative PCR reactions using TaqMan gene expression assays RSPO3_TaqMan® Gene Expression Assays_Hs01072567_m1, GAPDH_TaqMan® Gene Expression Assays_Hs02758991_g1 (Life technologies, CA, USA) and 7500 Fast Real-Time PCR System following the manufacturer's protocol. Each sample was tested in duplicate. Target genes were normalized using *GAPDH* as endogenous control and their relative expression was carried out with the 2−ΔCt method (where Ct represents the threshold cycle)[21]. The amplification efficiency of target and reference genes was approximately the same (slope < 0.1).

2. Supplementary Tables

Supplementary table 1 Numbers of patients per country included in this study

Country	No. of patients
Canada	158
Finland	290
France	34
Germany	558
Greece	10
Italy	51
Norway	397
Poland	42
Spain	18
Sweden	220
Netherlands	228
United Kingdom	1033
United States	363
total	3402

Supplementary table 2 Association statistics of binary phenotypes

Phenotype	SNP	Chr.	BP	Minor allele	MAF	OR (95% CI)	P-value	Candidate genes
Cholangiocarcinoma	rs7731017	5	150111618	G	0.011	4.81 (2.57-9.02)	9.62×10^{-07}	*LOC102546298, NDST1, SYNPO, MYOZ3, RBM22, DCTN4, SMIM3, IRGM, ZNF300, ZNF300P1*
Colorectal carcinoma	rs17102823	14	35363904	C	0.139	2.29 (1.65-3.17)	7.23×10^{-07}	*CFL2, BAZ1A, IGBP1P1, SRP54, FAM177A1, LOC101927178, PPP2R3C*

For each phenotype, PSC patients were divided into a group with the phenotype and a group without. Genetic associations were identified for each phenotype by Immunochip-wide association analysis.
All associations are suggestive ($P < 5.0 \times 10^{-06}$ and $P > 5.0 \times 10^{-08}$).
SNP, single nucleotide polymorphism; Chr, chromosome; BP, basepair position on genome build hg19/GRCh37; MAF, minor allele frequency; OR, odds ratio;

Supplementary table 3 Association statistics of time-to-event phenotypes

Phenotype	SNP	Chr	BP	Minor allele	Overall P-value genotype effect	Comparison	P-value	HR (95% CI)	Candidate genes
Liver transplant-free survival	**rs853974**	**6**	**127068983**	**A**	**6.07×10^{-09}**	**GG vs. AA**	**8.0×10^{-10}**	**0.46 (0.36-0.59)**	***CENPW, RSPO3***
						AG vs. AA	**2.3×10^{-06}**	**0.55 (0.43-0.70)**	
	rs1532244	3	28057905	G	9.24×10^{-07}	GG vs. AA	1.6×10^{-07}	4.77 (2.66-8.55)	*CMC1*
						GA vs. AA	3.3×10^{-01}	1.10 (0.91-1.35)	
	rs17649817	5	169956579	A	7.0×10^{-08}	CC vs. AA	1.1×10^{-08}	0.27 (0.17-0.42)	*LCP2, LINC01366, KCNIP1, KCNMB1, CTD-2270F17.1, LOC105377716*
						CA vs. AA	3.4×10^{-07}	0.29 (0.18-0.47)	
Time to CCA	rs3769839	2	231076625	G	9.29×10^{-07}	GG vs. AA	2.3×10^{-06}	9.94 (3.83-25.8)	*FBXO36, SLC16A14, SP110, SP140, SP140L, SP100*
						GA vs. AA	4.0×10^{-02}	0.54 (0.30-0.97)	
	rs2675647	10	63518620	C	2.72×10^{-06}	CC vs. AA	4.9×10^{-07}	10.40 (4.17-25.8)	*C10orf107, ARID5B, MIR548AV*
						CA vs. AA	9.6×10^{-01}	0.99 (0.61-1.61)	
	rs34985176	3	50711001	G	4.3×10^{-06}	GG vs. AA	1.8×10^{-06}	3.80 (2.20-6.58)	*CACNA2D2, C3orf18, HEMK1, CISH, MAPKAPK3, MIR4787, DOCK3*
						GA vs. AA	7.8×10^{-01}	1.06 (0.72-1.56)	

A Cox proportional hazards model was fitted Immunochip-wide. Sex, country, time point of diagnosis, smoking status and IBD status were included as covariates. P-values represent the genotype effect on the time-to-event phenotypes. All time intervals start at the time of PSC diagnosis. Both genome-wide significant ($P < 5.0 \times 10^{-08}$, in bold) and suggestive ($P < 5.0 \times 10^{-06}$ and $P > 5.0 \times 10^{-08}$) associations are given.
SNP, single nucleotide polymorphism; Chr, chromosome; BP, basepair position on genome build hg19/GRCh37; HR, hazard ratio; CCA, cholangiocarcinoma

References

1 Liu JZ, Hov JR, Folseraas T, *et al.* Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis. *Nat Genet* 2013; **45**: 670–5.

2 Purcell S, Neale B, Todd-Brown K, *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–75.

3 Bellenguez C, Strange A, Freeman C, Consortium WTCC, Donnelly P, Spencer CCA. A robust clustering algorithm for identifying problematic samples in genome-wide association studies. *Bioinformatics* 2012; **28**: 134–5.

4 Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet* 2006; **2**: e190.

5 Abecasis GR, Altshuler D, Auton A, *et al.* A map of human genome variation from population-scale sequencing. *Nature* 2010; **467**: 1061–73.

6 Trynka G, Hunt KA, Bockett NA, *et al.* Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat Genet* 2012; **43**: 1193–201.

7 Jostins L, Ripke S, Weersma RK, *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 2012; **491**: 119–24.

8 Jia X, Han B, Onengut-Gumuscu S, *et al.* Imputing Amino Acid Polymorphisms in Human Leukocyte Antigens. *PLoS One* 2013; **8**. DOI:10.1371/journal.pone.0064683.

9 Han B, Diogo D, Eyre S, *et al.* Fine Mapping Seronegative and Seropositive Rheumatoid Arthritis to Shared and Distinct HLA Alleles by Adjusting for the Effects of Heterogeneity. *Am J Hum Genet* 2014; **94**: 522–32.

10 Patsopoulos NA, Barcellos LF, Hintzen RQ, *et al.* Fine-Mapping the Genetic Association of the Major Histocompatibility Complex in Multiple Sclerosis: HLA and Non-HLA Effects. *PLoS Genet* 2013; **9**: e1003926.

11 Material SO, Web S, Press H, York N, Nw A. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* 2004; **306**: 636–40.

12 Rosenbloom KR, Dreszer TR, Pheasant M, *et al.* ENCODE whole-genome data in the UCSC Genome Browser. *Nucleic Acids Res* 2010; **38**: D620-5.

13 Westra H-J, Peters MJ, Esko T, *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* 2013; **45**: 1238–43.

14 The GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 2013; **45**: 580–5.

15 Fickert P, Stöger U, Fuchsbichler A, *et al.* A new xenobiotic-induced mouse model of sclerosing cholangitis and biliary fibrosis. *Am J Pathol* 2007; **171**: 525–36.

16 Marzioni M, Saccomanno S, Agostinelli L, *et al.* PDX-1/Hes-1 interactions determine cholangiocyte proliferative response to injury in rodents: Possible implications for sclerosing cholangitis. *J Hepatol* 2013; **58**: 750–6.

17 Sampaziotis F, Cardoso de Brito M, Madrigal P, *et al.* Cholangiocytes derived from human induced pluripotent stem cells for disease modeling and drug validation. *Nat Biotechnol* 2015; **33**: 845–52.

18 Mederacke I, Dapito DH, Affò S, Uchinami H, Schwabe RF. High-yield and high-purity isolation of hepatic stellate cells from normal and fibrotic mouse livers. *Nat Protoc* 2015; **10**: 305–15.

19 Rombouts K, Carloni V. Lipid Signaling Protocols. *Methods Mol Biol* 2016; **1376**: 203–12.

20 Jalan R, De Chiara F, Balasubramaniyan V, *et al.* Ammonia produces pathological changes in human hepatic stellate cells and is a target of therapy of portal hypertension. *J Hepatol* 2016; **64**: 823–33.

21 Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative CT method. *Nat Protoc* 2008; **3**: 1101–8.

3. International consortia

4.1 Members of the UK-PSC consortium not listed as co-authors

Abertawe Bro Morgannwg University NHS Trust (**Dr Chin Lye Ch'ng, Dr Clement Lai, Dr Tom Yapp**), Aintree University Hospitals NHS Foundation Trust (**Dr Richard Sturgess**), Airedale NHS Trust (**Dr Chris Healey**), Barnsley Hospital NHS Foundation Trust (**Dr Kapil Kapur**), Ashford and St Peter's Hospitals NHS Trust (**DrJohn Thornton**), Barking, Havering and Redbridge University Hospitals NHS Trust (**Dr Stephen Grainger, Dr Purushothaman Premchand**), Barnet and Chase Farm Hospitals NHS Trust (**Dr Stephen Mann, Dr Kalpesh Besherdas**), Barts and The London NHS Trust (**Dr Richard Marley, Prof Graham Foster**), Basingstoke and North Hampshire NHS Foundation Trust (**Dr John Ramage**), Bedford Hospital NHS Trust (**Dr Rory Harvey**), Blackpool Teaching Hospitals NHS Foundation Trust (**Dr Peter Isaacs**), Bolton NHS Foundation Trust (**Dr George Lipscomb**), Bradford Teaching Hospitals NHS Foundation Trust (**Dr Sulleman Moreea**), Brighton and Sussex University Hospitals NHS Trust (**Dr Jeremy Tibble, Dr Nick Parnell**), Bromley Hospitals NHS Trust (**Dr Anthony Jenkins**), Buckinghamshire Healthcare NHS Trust (**Dr Sue Cullen, Dr David Gorard**), Burton Hospitals NHS Foundation Trust (**Dr Altaf Palejwala**), Cambridge University Hospitals NHS Foundation Trust (**Dr George Mells, Dr Muhammad F Dawwas**), Calderdale And Huddersfield NHS Foundation Trust (**Dr Susan Jones, Dr Ashwin Verma**), Cardiff and Vale NHS Trust (**Dr Richard Aspinall, Dr Sunil Dolwani**), Central Manchester University Hospitals NHS Foundation Trust (**Dr Martin Prince**), Chelsea and Westminster Hospital NHS Foundation Trust (**Dr Karen Hawkins, Dr Priyajit Prasad**), Chesterfield Royal Hospitals NHS Foundation Trust (**Dr David Elphick**), City Hospitals Sunderland NHS Foundation Trust (**Dr Harriet Mitchison**), Colchester Hospital University NHS Foundation Trust (**Dr Ian Gooding**), Countess of Chester Hospital NHS Foundation Trust (**Dr Mazn Karmo**), County Durham and Darlington NHS Foundation Trust (**Dr Anjan Dhar, Dr Stephen Mitchell, Dr Sushma Saksena**), Cwm Taf NHS Trust (**Dr Minesh Patel**), Dartford And Gravesham NHS Trust (**Dr Roland Ede**), Derby Hospitals NHS Foundation Trust (**Dr Andrew Austin**), Doncaster and Bassetlaw Hospitals NHS Foundation Trust (**Dr Joanne Sayer**), Dorset County Hospitals NHS Foundation Trust (**Dr Chris Hovell**), East and North Hertfordshire NHS Trust (**Dr Martyn Carter, Dr Peter McIntyre**), East Cheshire NHS Trust (**Dr Konrad Koss**), Medway NHS Foundation Trust (**Dr Gray Smith-Laing**), Maidstone and Tunbridge Wells NHS Trust (**Dr George Bird**), East Kent Hospitals University NHS Foundation Trust (**Dr Kate Hill**), East Sussex NHS Trust (**Dr David Neal**), East Lancashire Hospitals NHS Trust (**Dr Vishal Kaushik**), Epsom and St Helier University Hospitals NHS Trust (**Dr Guan Lim**), Frimley Park NHS Foundation Trust (**Dr Aftab Ala**), Gateshead Health NHS Foundation Trust (**Dr Athar Saeed**), George Eliot Hospital NHS Trust (**Dr Sankara Raman**), Gloucestershire Hospitals NHS Foundation Trust (**Professor Jonathan Brown**), Great Western Hospitals NHS Foundation Trust (**Dr Ajeya Shetty**), Guy's and St Thomas' NHS Trust (**Dr Mark Wilkinson**), Gwent Healthcare NHS Trust (**Dr Miles Allison, Dr Marek Czajkowfki, Dr Peter Neville**), Harrogate and District NHS Foundation Trust (**Dr Butt,Dr Gareth Davies**), Heart of England NHS Foundation Trust (**Dr Theodore Ngatchu**), Hereford Hospitals NHS Trust (**Dr Rupert Ransford**), Hinchingbrooke Health Care NHS Trust (**Dr Richard Dickinson**), Homerton University Hospital NHS Foundation Trust (**Dr Ray Shidrawi**), Hull And East Yorkshire Hospitals NHS Trust (**Dr George Abouda**), Hywel Dda Health Board NHS Trust (**Dr Faiz Ali, Dr Mark Narain, Dr Ian Rees, Dr Imroz Salam**),Imperial College Healthcare NHS Trust (**Dr Ashley Brown**), Ipswich Hospital NHS Trust (**Dr Simon Williams**), NHS Isle Of Wight (**Dr Christopher Sheen**), James Paget University Hospitals NHS Foundation Trust (**Dr Matthew Williams**), Kings College Hospital NHS Foundation Trust (**Dr Michael Heneghan**), Kingston Hospital NHS Trust (**Dr Chris Rodrigues**), Lancashire Teaching Hospitals NHS Foundation Trust (**Dr Phillip Shields**), Leeds Teaching Hospitals NHS Trust (**Dr Mark Aldersley, Dr Mervyn Davies, Dr Charles Millson**), Luton and Dunstable Hospital NHS Foundation Trust (**Dr Sambit Sen**), Mid Cheshire Hospitals NHS Foundation Trust (**Dr Kevin Yoong**), Mid Essex Hospital Services NHS Trust (**Dr Cho Cho Khin**), Mid Staffordshire NHS Foundation Trust (**Dr Pradip Singh**), Mid Yorkshire Hospitals NHS Trust (**Dr Nurani Sivaramakrishnan**), Milton Keynes Hospital NHS Foundation Trust (**Dr George MacFaul**), Newham University Hospital NHS Trust (**Dr Matthew Guinane**), NHS Ayrshire & Arran (**Dr Chris Gillen, Dr James Rose, Dr Amir Shah**), NHS Borders (**Dr Chris Evans**), NHS Dumfries & Galloway (**Dr Subrata Saha**), NHS Fife (**Dr Sherzad Balta, Dr John Wilson**), NHS Forth Valley (**Dr Peter Bramley**), NHS Grampian (**Dr Andrew Fraser**), NHS Greater Glasgow and Clyde (**Dr Aidan Cahill, Dr G Curry, Dr Shouren Datta, Dr Ewan Forrest, Dr Neil Jamieson, Dr James McPeake, Professor Peter Mills, Dr Judith Morris, Dr Adrian Stanley**), NHS Highland (**Dr Dara De Las Heras, Dr Tim Shallcross**), NHS Lanarkshire (**Dr Stuart Campbell, Dr Richard Crofton, Dr Andrzej Prach**), NHS Lothian (**Dr Andrew Bathgate, Dr Kevin Palmer**), NHS Tayside (**Dr Alan Shepard, Dr John Dillon**), Norfolk and Norwich University Hospitals NHS Foundation Trust (**Dr Martin Phillips**), North Bristol NHS Trust (**Dr Robert Przemioslo**), North Cumbria University Hospitals NHS Trust (**Dr Babur Javaid, Dr Chris McDonald**), North Tees And Hartlepool NHS Foundation Trust (**Dr Jane Metcalf**), North Wales NHS Trust (**Dr Paulose George, Dr David Ramanaden**),

The North West London Hospitals NHS Trust (**Dr Maxton Pitcher**), North West Wales NHS Trust (**Dr Jaber Gasem Dr Richard Evans**), Northampton General Hospital NHS Trust (**Dr Udi Shmueli**), Northern Devon Healthcare NHS Trust (**Dr Andrew Davis**), Northern Lincolnshire and Goole Hospitals NHS Foundation Trust (**Dr Prabhakar Mysore, Dr Asifabbas Naqvi**), Northumbria Healthcare NHS Trust (**Dr Mark Welfare**), Nottingham University Hospitals NHS Trust (**Dr Steve Ryder**), Pennine Acute Hospitals NHS Trust (**Dr Howard Klass, Dr Jimmy Limdi, Dr Bashir Rameh**), Peterborough Stamford Hospitals NHS Foundation Trust (**Dr Mary Ninkovic**), Plymouth Hospitals NHS Trust (**Dr Matthew Cramp**), Poole Hospital NHS Foundation Trust (**Dr Nicholas Sharer**), Portsmouth Hospitals NHS Trust (**Dr Patrick Goggin**), Queen Mary's Sidcup NHS Trust (**Dr Howard Curtis**), Royal Berkshire NHS Foundation Trust (**Dr Jonathan Booth**), Royal Cornwall Hospitals NHS Trust (**Dr Hyder Hussaini**), Royal Devon and Exeter NHS Foundation Trust (**Dr Reuben Ayres),** Royal Free Hampstead NHS Trust (**Dr Douglas Thorburn**), Royal Liverpool and Broadgreen University Hospitals NHS Trust (**Dr Martin Lombard**), Royal Surrey County Hospital NHS Trust (**Dr Michelle Gallagher**), Royal United Hospital Bath NHS Trust (**Dr Duncan Robertson**), Salisbury NHS Foundation Trust (**Dr Sam Vyas**), Sandwell and West Birmingham Hospitals NHS Trust (**Dr Saket Singhal**), Scarborough And North East Yorkshire Health Care NHS Trust (**Dr Sathish Babu**), Sheffield Teaching Hospitals NHS Foundation Trust (**Dr Dermot Gleeson**), Sherwood Forest Hospitals NHS Foundation Trust (**Dr Sharat Misra**), Shrewsbury and Telford Hospital NHS Trust (**Dr Jeff Butterworth**), South Devon Healthcare NHS Trust (**Dr Keith George**), South London Healthcare NHS Trust (**Dr Alastair McNair**), South Tees Hospitals NHS Trust (**Dr Andrew Douglas**), South Tyneside NHS Foundation Trust (**Dr Colin Rees**), South Warwickshire General Hospitals NHS Trust (**Dr Jeremy Shearman**), Southampton University Hospitals NHS Trust (**Dr Kate Nash, Dr Mark Wright**), Southend University Hospital NHS Foundation Trust (**Dr Gary Bray**), Southport And Ormskirk Hospital NHS Trust (**Dr Graham Butcher**), St George's Healthcare NHS Trust (**Dr Daniel Forton**), St Helens And Knowsley Hospitals NHS Trust (**Dr John McLindon**), Stockport NHS Foundation Trust (**Dr Debashis Das**), Surrey and Sussex Healthcare NHS Trust (**Dr Gary Mackenzie, Dr Azhar Ansari, Dr Gregory Whatley**), Taunton and Somerset NHS Foundation Trust (**Dr Stirling Pugh**), Dudley Group of Hospitals NHS Trust (**Dr Neil Fisher**), The Hillingdon Hospital NHS Trust (**Dr Deb Datta**), The Lewisham Hospital NHS Trust (**Dr John Odonohue**), The Newcastle upon Tyne Hospitals NHS Foundation Trust (**Dr Mark Hudson**), The Princess Alexandra Hospital NHS Trust (**Dr Rosemary Phillips**), The Queen Elizabeth Hospital King's Lynn NHS Trust (**Dr Andrew Douds**), Rotherham NHS Foundation Trust (**Dr Barbara Hoeroldt**), Royal Bournemouth And Christchurch Hospitals NHS Foundation Trust (**Dr Earl Williams**), The Royal Wolverhampton Hospitals NHS Trust (**Dr Matthew Brookes**), Trafford Healthcare NHS Trust (**Dr Chris Summerton**), United Lincolnshire Hospitals NHS Trust (**Dr Aravamuthan Sreedharan, Dr Sanjiv Jain, Dr Martin James**), University College London Hospitals NHS Foundation Trust (**Dr Stephen Pereira**), University Hospital Birmingham NHS Foundation Trust (**Professor David Adams**), University Hospital of North Staffordshire NHS Trust (**Dr Alison Brind**), University Hospital of South Manchester NHS Foundation Trust (**Dr Gill Watt**), University Hospitals Bristol NHS Foundation Trust (**Dr Fiona Gordon, Dr Jim Portal**), University Hospitals Coventry and Warwickshire NHS Trust (**Dr Esther Unitt**), University Hospitals of Leicester NHS Trust (**Dr Allister Grant**), Walsall Hospitals NHS Trust (**Dr Mark Cox**), Warrington and Halton Hospitals NHS Foundation Trust (**Dr Subramaniam Ramakrishnan**), West Hertfordshire Hospitals NHS Trust (**Dr Alistair King**), West Suffolk Hospitals NHS Trust (**Dr Simon Whalley**), Western Sussex Hospitals NHS Trust (**Dr Andy Li, Dr Mohammed Rashid**), Weston Area Health NHS Trust (**Dr Andrew Bell**), Winchester And Eastleigh Healthcare NHS Trust (**Dr Harriet Gordon**), Wirral University Teaching Hospital NHS Foundation Trust (**Dr Riyaz Faizallah**), Worcestershire Acute Hospitals NHS Trust (**Dr Ishfaq Ahmad, Dr Ian Gee**), Wrightington, Wigan and Leigh NHS Trust (**Dr Gurvinder Banait**), Yeovil District Hospital NHS Foundation Trust (**Dr Steve Gore, Dr James Gotto**), York Hospitals NHS Foundation Trust (**Dr Alastair Turnbull**).

Genetic association analysis identifies variants associated with disease progression in primary sclerosing cholangitis

Rudi Alberts[1*], Elisabeth M.G. de Vries[2*], Elizabeth C. Goode[3,4], Xiaojun Jiang[5,6], Fotis Sampaziotis[7,8], Krista Rombouts[9], Katrin Böttcher[9], Trine Folseraas[5,6], Tobias J. Weismüller[10,11], Andrew L. Mason[12], Weiwei Wang[12], Graeme Alexander[13], Domenico Alvaro[14], Annika Bergquist[15], Niklas K. Björkström[16], Ulrich Beuers[2], Einar Björnsson[17], Kirsten Muri Boberg[5,18], Christopher L. Bowlus[19], Maria C. Bragazzi[20], Marco Carbone[21], Olivier Chazouillères[22], Angela Cheung[23], Georgios Dalekos[24], John Eaton[25], Bertus Eksteen[26], David Ellinghaus[27], Martti Färkkilä[28], Eleonora A.M. Festen[1], Annarosa Floreani[29], Irene Franceschet[30], Daniel Nils Gotthardt[31], Gideon M. Hirschfield[32], Bart van Hoek[33], Kristian Holm[5,6], Simon Hohenester[34], Johannes Roksund Hov[5,6], Floris Imhann[1], Pietro Invernizzi[21], Brian D. Juran[25], Henrike Lenzen[35], Wolfgang Lieb[36,37], Jimmy Z. Liu[38], Hanns-Ulrich Marschall[39], Marco Marzioni[40], Espen Melum[5,6], Piotr Milkiewicz[41], Tobias Müller[42], Albert Pares[43], Christian Rupp[44], Christian Rust[45], Richard N. Sandford[4], Christoph Schramm[46], Stefan Schreiber[27,47], Erik Schrumpf[6,48], Mark S. Silverberg[49], Brijesh Srivastava[4], Martina Sterneck[50], Andreas Teufel[51], Ludovic Vallier[7,38], Joanne Verheij[52], Arnau Vich Vila[1], Boudewijn de Vries[1], Kalliopi Zachou[53], The International PSC Study Group, The UK PSC Consortium, Roger W. Chapman[54], Michael P. Manns[10,11], Massimo Pinzani[9], Simon M. Rushbrook[3], Konstantinos N. Lazaridis[25], Andre Franke[27], Carl A. Anderson[38], Tom H. Karlsen[5,6], Cyriel Y. Ponsioen[2#], Rinse K. Weersma[1#]

[*]equal contribution

[#]shared last author

1 Department of Gastroenterology and Hepatology, University of Groningen and University Medical Centre Groningen, Groningen, NL

2 Department of Gastroenterology and Hepatology, Academic Medical Center, Amsterdam, NL

3 Norwich Medical School, Faculty of Medicine and Health Sciences, University of East Anglia, Norwich, UK

4 Academic Department of Medical Genetics, University of Cambridge, Cambridge, UK

5 Norwegian PSC Research Center, Division of Cancer Medicine, Surgery and Transplantation, Oslo University Hospital, Rikshospitalet, Oslo, Norway

6 Research Institute of Internal Medicine, Oslo University Hospital, Rikshospitalet, Oslo, Norway

7 Wellcome Trust-Medical Research Council Stem Cell Institute, Anne McLaren Laboratory, Department of Surgery, University of Cambridge, Cambridge, UK

8 Department of Surgery, University of Cambridge and NIHR Cambridge Biomedical Research Centre, Cambridge, UK

9 Institute for Liver and Digestive Health, University College London, Royal Free Hospital, London, UK

10 Department of Gastroenterology, Hepatology and Endocrinology, Hannover Medical School, Hannover, Germany

11 Integrated Research and Treatment Center-Transplantation (IFB-tx), Hannover Medical School, Hannover, Germany

12 Division of Gastroenterology and Hepatology, University of Alberta, Edmonton, Alberta, Canada

13 Department of Medicine, Division of Hepatology, University of Cambridge, Cambridge, UK

14 Department of Clinical Medicine, Division of Gastroenterology, Sapienza University of Rome, Rome, Italy

15 Center for Digestive Diseases, Karolinska University Hospital, Karolinska Institutet, Stockholm, Sweden

16 Center for Infectious Medicine, Department of Medicine Huddinge, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden

17 Department of Internal Medicine, Division of Gastroenterology and Hepatology, Landspitali University Hospital, Reykjavik, Iceland

18 K.G. Jebsen Inflammation Research Centre and Institute of Clinical Medicine, University of Oslo, Oslo, Norway

19 Division of Gastroenterology and Hepatology, University of California Davis, Davis, CA, USA

20 Sapienza University of Rome, Medico-Surgical Sciences and Biotechnologies, Rome, Italy

21 Program for Autoimmune Liver Diseases, International Center for Digestive Health, Department of Medicine and Surgery, University of Milan-Bicocca, Milan, Italy

22 AP-HP, Hôpital Saint Antoine, Department of Hepatology, Paris, France

23 Gen. Internal Medicine, University Health Network, Toronto General Hospital, Toronto, Canada

24 Department of Medicine and Research Laboratory of Internal Medicine, Medical School, University of Thessaly, Larissa, Greece

25 Division of Gastroenterology and Hepatology, Mayo Clinic Minnesota, Rochester, Minnesota, United States

26 Snyder Institute of Chronic Diseases, Department of Medicine, University of Calgary, Calgary, Canada

27 Institute of Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Germany

28 Division of Gastroenterology, Department of Medicine, Helsinki University Hospital, Finland

29 Dept. of Surgical, Oncological and Gastroenterological Sciences, University of Padova, Padova, Italy

30 Department of Surgery, Oncology and Gastroenterology, University of Padova, Padova, Italy

31 Department of Medicine, University Hospital of Heidelberg, Heidelberg, Germany

32 Centre for Liver Research, NIHR Biomedical Research Unit, University of Birmingham, Birmingham, UK

33 Department of Gastroenterology and Hepatology, Leiden University Medical Centre, Leiden, the Netherlands

34 Liver Center Munich, Department of Medicine II, University of Munich, Munich, Germany

35 Department of Gastroenterology, Hepatology and Endocrinology, Hannover Medical School, Hannover, Germany

36 Popgen Biobank, University Hospital Schleswig-Holstein, Christian-Albrechts-University, Kiel, Germany

37 Institute for Epidemiology, Christian-Albrechts University, Kiel, Germany

38 Wellcome Trust Genome Campus, Wellcome Trust Sanger Institute, Cambridge, UK

39 Department of Molecular and Clinical Medicine, Institute of Medicine, Sahlgrenska Academy, Gothenburg, Sweden

40 Department of Gastroenterology, Università Politecnica delle Marche, Ospedali Riuniti University Hospital, Ancona, Italy

41 Liver and Internal Medicine Unit, Medical University of Warsaw, Warsaw, Poland

42 Department of Internal Medicine, Hepatology and Gastroenterology, Charité Universitätsmedizin Berlin, Campus Virchow Klinikum, Berlin, Germany

43 Liver Unit, Hospital Clinic, IDIBAPS, CIBERehd, University of Barcelona, Barcelona, Spain

44 Department of Internal Medicine IV, University Hospital of Heidelberg, Heidelberg,Germany

45 Department of Medicine I, Krankenhaus Barmherzige Brüder, Munich, Germany

46 1st Department of Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

47 Department for General Internal Medicine, Christian-Albrechts-University, Kiel, Germany

48 Section of Gastroenterology, Department of Transplantation Medicine, Division of Cancer, Surgery and Transplantation, Oslo University Hospital, Rikshospitalet, Oslo, Norway

49 Inflammatory Bowel Disease (IBD) Group, Zane Cohen Centre for Digestive Diseases, Mount Sinai Hospital Toronto, Ontario, Canada

50 Department of Hepatobiliary Surgery and Transplantation, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

51 1st Department of Medicine, University of Mainz, Mainz, Germany

52 Department of Pathology, Academic Medical Center, Amsterdam, the Netherlands

53 Department of Internal Medicine, University of Thessaly, Larissa, Greece

54 Department of Hepatology, John Radcliffe University Hospitals NHS Trust, Oxford, UK

Abbreviations

CCA: cholangiocarcinoma; CLC: cholangiocyte-like-cell; CRC: colorectal carcinoma; HCC: hepatocellular carcinoma; hiPSC: human induced pluripotent stem cell; HSC: hepatic stellate cell; IBD: inflammatory bowel disease; MHC: major histocompatibility complex; PBC: primary biliary cholangitis; PBD: primary bile duct; PSC: primary sclerosing cholangitis; SNP: single nucleotide polymorphism;

Keywords

Primary sclerosing cholangitis, genetics, liver transplantation

Correspondence

Rinse K Weersma, Department of Gastroenterology and Hepatology, University of Groningen and University Medical Center Groningen, PO Box 30.001, 9700RB, Groningen, the Netherlands

Tel: +31503610426, Fax: +31503619306, Email: r.k.weersma@umcg.nl

Word count: 3655

ABSTRACT

Objective Primary sclerosing cholangitis (PSC) is a genetically complex, inflammatory bile duct disease of largely unknown etiology often leading to liver transplantation or death. Little is known about the genetic contribution to the severity and progression of PSC. The aim of this study is to identify genetic variants associated with PSC disease progression and development of complications.

Design We collected standardized PSC subphenotypes in a large cohort of 3,402 PSC patients. After quality control we combined 130,422 single nucleotide polymorphisms of all patients – obtained using the Illumina Immunochip – with their disease subphenotypes. Using logistic regression and Cox proportional hazards models we identified genetic variants associated with binary and time-to-event PSC subphenotypes.

Results We identified genetic variant rs853974 to be associated with liver transplant-free survival ($P = 6.07 \times 10^{-9}$). Kaplan-Meier survival analysis showed a 50.9% (95% CI 41.5-59.5%) transplant-free survival for homozygous AA allele carriers of rs853974 compared with 72.8% (95% CI 69.6-75.7%) for GG carriers at ten years after PSC diagnosis. For the candidate gene in the region, *RSPO3*, we demonstrated expression in key liver-resident effector cells, such as human and murine cholangiocytes and human hepatic stellate cells.

Conclusion We present a large international PSC patient cohort, and report genetic loci associated with PSC disease progression. For liver transplant-free survival, we identified a genome wide significant signal and demonstrated expression of the candidate gene *RSPO3* in key liver-resident effector cells. This warrants further assessments of the role of this potential key PSC modifier gene.

SUMMARY BOX

What is already known about this subject:

- Several case-control genome-wide association studies have revealed 20 susceptibility loci for primary sclerosing cholangitis.
- Little is known about the genetic contribution to the severity and progression of complex diseases in general and primary sclerosing cholangitis in particular.
- *RSPO3* plays a role in the activation of the canonical Wnt signaling pathway, which is involved in liver fibrosis.

What are the new findings:

- The genetic variant rs853974 is genome-wide significantly associated with liver transplant-free survival in primary sclerosing cholangitis.
- Candidate gene *RSPO3* is expressed in both murine and human cholangiocytes, and in human hepatic stellate cells.
- Three new loci were found to be associated with time to cholangiocarcinoma in patients with primary sclerosing cholangitis.

How might it impact on clinical practice in the foreseeable future?

- Through its effect on liver fibrosis, *RSPO3* could play an important role in PSC disease progression, and insight in its mechanism could lead to new therapeutic targets.
- Furthermore, since we demonstrated that genetic variants are associated with PSC disease progression, genetics could provide a tool for risk stratification of patients with PSC in the future.

INTRODUCTION

Primary sclerosing cholangitis (PSC) is a complex, cholestatic liver disease, in which chronic biliary inflammation and bile duct destruction leads to biliary fibrosis and liver cirrhosis, often in a slowly progressive manner.[1] PSC is characterized by a cholangiographic image of strictures interchanged with dilatations throughout the biliary tract. Reported incidence rates of PSC vary widely, with incidence rates of 0.91, 1.31, and 0.5 per 100,000 inhabitants per year for North America, Norway, and the Netherlands, respectively.[2,3] There is a male to female ratio of 2:1, and the disease can occur at any age, with a peak incidence around 40 years.[3] There is a close association between PSC and inflammatory bowel disease (IBD), and PSC patients are subject to a 5-fold increased risk of developing colorectal carcinoma (CRC) when compared with the general population.[3] In addition, PSC carries an excess risk of cholangiocarcinoma (CCA) which seems to be unrelated to the disease duration and the presence of liver cirrhosis.[4] There is no effective medical therapy that can halt disease progression in PSC. The only curative option to date is liver transplantation.

The etiology of PSC is still largely unknown. The etiology is most likely to be multifactorial, in which the occurrence of PSC could be triggered by environmental factors in a genetic susceptible host.[5] The relationship between susceptibility to PSC and environmental factors has been studied for several risk factors, of which smoking has repeatedly been shown to be associated with a decreased risk of developing PSC, independent of its protective effect in ulcerative colitis.[6]

Already in 1983, the identification of associations between PSC and HLA-B8 of the human leukocyte antigen (HLA) complex located on chromosome 6 - harboring several genes that are involved in antigen presentation and are important in immunity - raised interest in the role of genetics in PSC.[7] This was amplified by a large Swedish study on PSC heritability demonstrated a nearly 4 to 17 times increased risk for first degree relatives of PSC patients to develop PSC, when compared with the general population.[8] The additional 3.3 times increased risk of ulcerative colitis, and the presence of at least one

concomitant immune mediated disease outside the liver and bowel in approximately 20 to 25% of PSC patients, suggests a shared genetic component between these diseases.[8–10] Over the last 5 years, the application of genome-wide association studies has resulted in an increasing insight in the genetic architecture of PSC, with the identification of 19 non-HLA risk loci at the time of writing.[11–14]

Little is known about the genetic contribution to the severity and progression of complex diseases in general and PSC specifically. In Mendelian traits like cystic fibrosis and hemochromatosis, consortia efforts have led to the identification of robust and important modifier genes.[15,16] If genetic variants would be associated with PSC phenotypes, this could enable risk stratification of PSC patients according to disease behavior and would lead to insight into pathogenetic mechanisms associated to disease progression. Whilst translational research from susceptibility genes has yet to prove useful for the development of new drugs in complex disease, modifier genes may point toward pathways involved in disease progression amenable by pharmacological interventions.

The aim of this study is to identify genetic variants associated with PSC disease progression and development of complications, in a large, international, multicenter PSC cohort.

METHODS

Study design and patients

PSC patients previously recruited throughout Europe, the USA and Canada, and genotyped using the Immunochip by Liu et al.[13] were included. Subject recruitment was approved by the ethics committees or institutional review boards of all participating centers. Written informed consent was obtained from all participants. Patients of whom PSC diagnosis was revised after they were genotyped were excluded.

The following phenotypic data were collected for patient and disease characteristics: sex, date of birth, PSC subtype (small or large duct), date of PSC diagnosis, intra and-/or extra hepatic disease, dominant strictures, concomitant IBD and type of IBD, date of IBD diagnosis, and smoking status. Furthermore, follow-up data were collected for: date and cause of death, date and indication of liver transplantation, the occurrence and date of diagnosis of hepatocellular carcinoma (HCC), CCA, CRC, gallbladder carcinoma, and the occurrence and date of a colectomy.

PSC diagnosis was based on clinical, biochemical, cholangiographic and histologic criteria, as formulated by the EASL guidelines.[17] IBD diagnosis was scored based on accepted endoscopic, radiologic and histologic criteria.[18]

PSC related death was defined as death from liver failure, death from cholangiosepsis, death from cholangiocarcinoma or death from gallbladder carcinoma. The time-to-event phenotype liver transplant-free survival was defined as the time between PSC diagnosis and the composite endpoint of either liver transplantation or PSC related death.

We used genotype data of PSC patients as previously described.[13] Appendix A describes the quality control applied to this dataset. A total of 130,422 single nucleotide polymorphisms (SNPs) for 3,402 PSC samples remained after quality control and were used in the analysis.

Statistical analysis

The age at PSC diagnosis was expressed as median value and interquartile range. Categorical variables were expressed as numbers and percentages based on non-missing values.

Binary associations

Binary associations were calculated using multiple logistic regression. We corrected for clinical covariates by adding them to the regression model. To determine which clinical covariates to correct for, we performed a backwards elimination procedure per binary phenotype. We started with the full model including sex, country, date of PSC diagnosis, established IBD diagnosis, and smoking status, and removed covariates from the model until the AIC (Akaike information criterion) stabilized.

Time-to-event associations

Cox proportional hazards regression was used to estimate the effect of genetic variants on time-to-event PSC subphenotypes. Clinical variables that were significantly associated with the time-to-event phenotype in univariable Cox regression analyses ($P < 0.05$) were entered into a multivariable Cox model alongside the genotype. To visualize the effect of genotype on time-to-event phenotypes, Kaplan-Meier survival estimates were plotted. (Methods described in supplementary appendix A).

We used the SNP2HLA software[19] to impute classical HLA alleles from genotype data for *HLA-A, HLA-B, HLA-C, HLA-DRB1, HLA-DQA1, HLA-DQB1, HLA-DPA1* and *HLA-DPB1* and their corresponding amino acid polymorphisms from the genotype data (Methods described in supplementary appendix A).

Mouse experiments and in vitro experiments on the role of *RSPO3* in PSC

Mouse experiments

C57BL/6 (B6) mice were purchased from Charles River (Milan, Italy). Normal C57BL/6 mice were sacrificed at the age of 8-10 weeks. Organs were harvested and washed by cold phosphate buffered saline (PBS). Cholangiocytes were isolated both from normal mice (n=3) and from mice (n=3) fed 0.1%3,5-diethoxycarbonyl-1,4-dihydrocollidine (DDC) for 4 weeks, as a model of sclerosing cholangitis.[20] Total RNA was extracted and sequenced on an Illumina HiSeq 2000 machine. See Appendix A for more details.

In vitro experiments on human primary biliary tissue, cholangiocyte-like-cells and hepatic stellate cells

Human primary biliary tissue was obtained from a liver and pancreas organ donor after obtaining informed consent from the donor's family. A section of the bile duct was excised and homogenized and RNA was extracted. Cholangiocyte-like-cells were generated from human induced pluripotent stem cells and cultured. *RSPO3* expression was determined using qPCR and compared with the housekeeping gene using the 2−ΔCt method.[21] Next to that, we used previously published microarray data to verify *RSPO3* expression.[22] The R/Bioconductor package limma [23] was used to evaluate differential expression between pairs of conditions (hIPSCs and CLCs and hIPSCs and PBD). A linear model fit was applied and p-values were corrected using the method of Benjamini and Hochberg.[24] Methods are further described in Appendix A.

Primary human hepatic stellate cells were isolated and cultured from wedge sections of liver tissue, obtained from patients undergoing surgery at the Royal Free Hospital in London. Total RNA was extracted and retro-transcribed into cDNA, which was used for gene expression assessment with qPCR. Gene expression was compared with the housekeeping gene using the 2−ΔCt method.

RESULTS

Patient characteristics and natural history

Clinical characteristics of the PSC cohort are described in table 1. The cohort consisted of 2881 patients from Europe and 521 patients from the United States and Canada (Supplementary table 1). A total of 2,185 (65%) patients were male, and the median age at PSC diagnosis was 38.6 years (IQR 28.0 - 50.1). Concomitant IBD was diagnosed in 2,390 (75%) patients. The median follow-up was 8.7 years. In total 874 (26%) patients underwent liver transplantation and 181 (5%) patients died of PSC related causes. Over 11% of patients developed a malignancy, most often CCA (5.6%) or CRC (4.3%).

Genetic associations with binary subphenotypes

Genome-wide association analyses focusing on the occurrence of malignancy in PSC patients revealed several suggestive associations (Supplementary table 2). When comparing 107 PSC-AIH patients with 3,159 PSC patients without AIH overlap, a strong genetic association in the *HLA-DQB1* gene was identified (top SNP rs3891175, $P = 4.6 \times 10^{-11}$, OR = 2.41). After imputation of the classical HLA alleles, we found that the alleles DQA1*05:01 and DQB1*02:01 were most significantly associated with PSC/AIH overlap (P-values of 3.8×10^{-11} and 1.8×10^{-07}). For other binary subphenotypes - small duct PSC, the occurrence of HCC, gallbladder carcinoma, and proctocolectomy – no genetic associations were found.

Table 1 Clinical characteristics of the PSC cohort consisting of 3,402 patients

Variable	Groups	Number (%)
Age at PSC diagnosis*		38.6y (28.0y - 50.1y)
Sex	male	2185 (64.7%)
	female	1193 (35.3%)
	missing	24 (0.7%)
Main diagnosis	PSC	3159 (94.6%)
	small duct PSC	75 (2.2%)
	PSC with AIH overlap	107 (3.2%)
	missing	61 (1.8%)
Liver transplantation	Yes	874 (26.3%)
	No	2444 (73.7%)
	missing	84 (2.5%)
Colectomy	Yes	419 (12.6%)
	No	2897 (87.4%)
	missing	86 (2.5%)
IBD	No IBD	816 (25.5%)
	Ulcerative colitis	1940 (60.5%)
	Crohn's disease	357 (11.1%)
	IBD-U	93 (2.9%)
	missing	196 (5.8%)
Cholangiocarcinoma	Yes	188 (5.6%)
	No	3147 (94.4%)
	missing	67 (2.0%)
Colorectal carcinoma	Yes	127 (4.3%)
	No	2822 (95.7%)
	missing	453 (13.3%)
Gall bladder carcinoma	Yes	30 (1.0%)
	No	2977 (99.0%)
	missing	395 (11.6%)
Hepatocellular carcinoma	Yes	22 (0.7%)
	No	2984 (99.3%)
	missing	396 (11.6%)
Smoking status	Smoker	140 (6.0%)
	Ex-Smoker	529 (22.7%)
	Non-smoker	1657 (71.2%)
	missing	1076 (31.6%)
Death	Non PSC related	47 (1.5%)
	Liver failure	66 (2.1%)
	Cholangiosepsis	18 (0.6%)
	Gallbladder carcinoma	12 (0.4%)
	Cholangiocarcinoma	85 (2.6%)
	Hepatocellular carcinoma	6 (0.2%)
	Colorectal carcinoma in case of coexisting IBD	3 (0.1%)
	Alive	2977 (92.6%)
	missing	188 (5.5%)

*Values shown as median (IQR).
Quantitative data are expressed as counts and percentages excluding missing data.

Genetic associations with time-to-event subphenotypes

Next, we aimed to determine whether genetic variants are associated with important time-to-event variables reflecting the PSC disease course, e.g. time between PSC diagnosis and the development of a carcinoma. For this, we developed a framework to perform Immunochip-wide Cox proportional hazards analyses. We defined the liver transplant-free survival subphenotype as time from PSC diagnosis until liver transplantation or PSC related death. Univariable Cox regression analyses including clinical parameters showed statistically significant associations with the time to event endpoint transplant-free survival for sex, country, date of PSC diagnosis, established IBD diagnosis, and smoking status. Next, we tested 130,422 SNPs for association with liver transplant-free survival using multivariable Cox proportional hazards regression, including the genotype effect alongside the significant clinical co-variables. We found SNP rs853974 to be associated with liver transplant-free survival of PSC patients at genome wide significance ($P = 6.07 \times 10^{-9}$). Kaplan-Meier survival analysis showed a 50.9% (95% CI 41.5-59.5%) transplant-free survival for homozygous AA allele carriers of rs853974 compared with 72.8% (95% CI 69.6-75.7%) for GG carriers at ten years after PSC diagnosis (figure 1A). AA homozygotes had a 2.14 (95% CI = 1.66-2.76) increased hazard, indicating a 2.14 larger relative risk for need for liver transplantation or for PSC related death compared to GG homozygotes. Figure 1B shows a regional plot of this observed association.

SNP rs853974 is located on chromosome 6. We did not identify a direct functional effect of this SNP on gene expression or regulatory features (Appendix A). We hypothesized that neighbouring gene *R-spondin 3 (RSPO3)* would be the most likely positional candidate gene. The other neighboring gene, *CENPW*, has a fundamental role in kinetochore assembly and is required for normal chromosome organization and progress through mitosis and therefore not a good candidate. In addition to SNP rs853974, additional suggestive genetic associations with time-to-event phenotypes liver transplant-free survival and time to CCA were found (Supplementary table 3).

Expression of *RSPO3* in key liver-resident effector cells

To assess whether *RSPO3* is expressed in disease relevant cells (cholangiocytes and hepatic stellate cells), we performed RNA-sequencing on healthy and cholestatic cholangiocytes and multiple organs derived from normal C57BL/6 mice. *RSPO3* expression was 7 to 20 folds higher in cholangiocytes as compared with any of the organs. Furthermore, *RSPO3* expression was higher in healthy cholangiocytes than in cholestatic cholangiocytes (figure 2A).

Next, using microarrays, we assessed expression of *RSPO3* in human induced pluripotent stem cells, human induced pluripotent stem cell-derived cholangiocyte-like cells, and human primary bile duct samples. *RSPO3* expression was significantly higher in cholangiocyte-like cells and primary bile duct cells compared with human induced pluripotent stem cells (figure 2B). This finding was confirmed by qPCR (figure 2C).

Since activated hepatic stellate cells are the main cells involved in liver fibrosis,[25] we also investigated expression of *RSPO3* in human culture-activated hepatic stellate cells. We isolated, cultured and activated human hepatic stellate cells of three patients without PSC. Using qPCR we observed expression of the hepatic stellate cell marker gene Cytoglobin B, as well as expression of *RSPO3* in all three subjects (figure 2D). We did not observe *RSPO3* expression in human CD4 and CD8 T lymphocytes (data not shown).

DISCUSSION

To date, very few disease-modifying genes have been identified in rare complex diseases. Collaboration within the international PSC Study group enabled the establishment of a cohort of unprecedented size, for an orphan disease such as PSC, enabling the investigation of genetic variation underlying the progression of PSC through time. Overall, it is a major challenge to determine genetic variants associated with survival, and only few genetic studies investigating this have been published.[26] We present a conceptually new method to determine associations between genetic variants and disease course, using genome wide multivariable Cox proportional hazards regression analyses. Here we identify a genome wide significant association between SNP rs853974 – located close to the *RSPO3* gene – and liver transplant-free survival in PSC. Interestingly, this locus is not associated with PSC susceptibility and thus exemplifies different genetic regulation of disease susceptibility and disease progression.

This study is based on genotype data obtained using Illumina immunochip, a genotyping platform that densely covers genetic regions associated to immune mediated diseases. Use of GWAS arrays, that more uniformly cover genetic variants all over the genome, would have been ideal. However, a complete GWAS dataset for the entire international cohort was not available at the time of study. For that reason we started with the available immunochip data. Given the positive findings in this study, a similar study on GWAS arrays data could very well be of additional value.

For the binary phenotype of developing cholangiocarcinoma or not, we found an association at chromosome 5 at 150 Mb (Supplementary table 2). Of interest, this locus contains an established genetic association with Crohn's disease susceptibility in the autophagy gene *IRGM*. [27] For the phenotype of developing colorectal carcinoma or not we found an association at chromosome 14 at 35 Mb. This locus appeared not to be associated with sporadic colorectal cancer.

Since transplant-free survival is a combined and heterogeneous phenotype we

assessed to which extent the following three subgroups contribute to the association: 1) transplanted patients with indication for transplantation "end stage liver disease" and patients died because of "liver failure"; 2) transplanted patients with indication for transplantation "cca/high grade dysplasia" and patients died because of "cca or gallbladder carcinoma"; and 3) transplanted patients with indication for transplantation "intolerable complaints/pruritus/recurrent cholangitis" and patients died of "cholangiosepsis". We observed a stronger contribution of subgroups 1 and 3 to the association, indicating that the underlying biological mechanism is more likely one involved in causing progression of liver disease and/or cholangitis or cholangiosepsis rather than a mechanism involved in cancer development.

R-spondin 3 is a member of the R-spondin protein family (R-spondin 1-4).[28] These proteins are secreted agonists of the canonical Wnt/β-catenin signaling pathway.[28] They activate the pathway leading to induced transcription of Wnt target genes. Wnt/β-catenin signaling plays a central role in embryogenesis, organogenesis and adult homeostasis, and is a critical regulator of stem cell maintenance.[29,30] RSPO3 is a ligand of the Frizzled 8 and LRP5/6 receptors.[28] In the canonical form of the Wnt pathway, binding of ligands to the Frizzled (Fzd) receptor and LRP5 or 6 co-receptors causes β-catenin to dephosphorylate in the cytoplasm. Accumulated β-catenin translocates to the nucleus where it binds to T cell factor (TCF)/Lymphoid enhancer-binding factor (LEF), causing transcription of Wnt target genes – such as *Fibronectin*, *MMP-7, Twist,* and *Snail*. These factors activate hepatic stellate cells and induce liver fibrosis. Blocking the Wnt signaling pathway using Dickkopf-1 (Dkk-1), a Wnt co-receptor antagonist, restores hepatic stellate cells quiescence in culture.[31] Hence, Wnt signaling is involved in both progression and regression of liver fibrosis, either by inhibiting or promoting activation and survival of hepatic stellate cells.[31,32] Also, RSPOs have been shown to facilitate hepatic stellate cell activation and promote hepatic fibrogenesis.[33] Here, we demonstrate that *RSPO3* is expressed in key effector cells involved in the pathogenesis of PSC. Since we have shown that PSC patients that are homozygote AA carriers of rs853974 progress more rapidly towards PSC related death or

liver transplantation, *RSPO3* can be regarded a plausible candidate gene to be involved in PSC disease progression. Hypothetically, PSC patients might benefit from reduction of *RSPO3* or generally canonical Wnt signaling.

In an Immunochip analysis of the International IBD Genetics Consortium including over 75,000 individuals[34], an intronic SNP rs9491697 in *RSPO3* (which is not in linkage disequilibrium with rs853974, $r^2 = 0.014$) was identified to be associated with Crohn's disease ($P = 3.79 \times 10^{-10}$, OR = 1.077) but not with ulcerative colitis. Given the small number of Crohn's disease patients (n = 357) within the present study, the lack of linkage disequilibrium between the two "hit SNPs", and the fact that our multivariate Cox model corrected for IBD-status, the identified association signal does not seem to be driven by the co-occurrence of Crohn's disease in our cohort.

For several binary and time-to-event subphenotypes we found suggestive genetic associations. Two additional SNPs, rs1532244 on chromosome 3 and rs17649817 on chromosome 5, were suggestively associated with transplant-free survival. Furthermore, one SNP, rs7731017, was suggestively associated with the presence of CCA. We investigated whether any of the candidate genes in the locus overlapped with genes identified in tumour sequencing studies of cholangiocarcinoma. We did not find an overlap with the 32 genes reported to be significantly altered in intrahepatic, extrahepatic, and gallbladder cancer by Nakamura et al.[35] When comparing the genes in our CCA locus with 1146 genes containing non-synonymous somatic mutations in intrahepatic cholanciocarcinoma[36], we found that the *SYNPO* gene was both in the list of 1146 genes of the sequencing study as well as in the locus that we identified to be associated with the presence of CCA. There is little known about this gene and there is no connection with oncogenesis. Another gene, *SP100,* was found in this study to be in the locus associated with time to CCA and is also in the list of 1146 genes. *SP100* is associated with autoimmune disease of the urogenital tract and also with PBC. Interestingly, anti-sp100 autoantibodies have been described for PBC[37]. The genetic association of *SP100* with both PBC and the time to CCA within PSC patients as well as the existence of anti-sp100 autoantibodies makes this an interesting gene

for future follow-up studies.

When comparing PSC-AIH patients with PSC patients without AIH, we found a strong genetic association with PSC-AIH in the HLA-DQB1 gene. The identified variant was tagging the classical HLA haplotypes DQA1*05:01 and DQB1*02:01. These associations overlap the associations found by a previous genome-wide association study of AIH type 1 in The Netherlands,[38] suggesting that the genetic basis for AIH type 1 pathogenesis is similar for isolated AIH type 1 patients, compared with PSC-AIH patients.

This study is limited by the relatively small cohort size, when compared with other GWAS studies that incorporate tens of thousands of samples. The resulting lack of statistical power may have played a role in the binary analyses, in which suggestive hits were found for CCA and CRC but genome wide significance was not reached. However, PSC is a rare disease, and the present study has included patients recruited throughout the world in a joined effort. It is therefore not expected that a larger cohort of PSC cases will become available soon.

In conclusion, we present the largest association study of PSC genotypes with disease phenotypes to date. We identified several genetic variants associated with PSC disease course. Specifically, we report rs853974 to be genome-wide significantly associated with liver transplant-free survival in PSC. Findings of candidate gene *RSPO3* being expressed in both mouse and human cholangiocytes and human activated hepatic stellate cells warrant further assessments of the role of this potential key PSC modifier gene.

Acknowledgments

We thank all PSC patients for participation in this study. We acknowledge the members of the International PSC Study Group and the UK-PSC Consortium for their participation. We acknowledge Lukas Tittman for providing samples.

RA is supported by a PSC Partners Seeking a Cure grant "Unraveling genetics driving PSC subphenotypes; An IPSCSG study". LV and FS are supported by the ERC grant Relieve IMDs and the Cambridge Hospitals National Institute for Health Research Biomedical Research Center. SH is supported by a grant from the German Research Community (DFG) grant HO 4460/2-1. TM is supported by the German Research Community (DFG), grants MU 2864/1-1 and MU 2864/1-3. KNL is supported by the NIH RO1 DK 084960 and Sigismunda Palumbo Charitable Trust. EAMF is supported by a Career Development Grant from Dutch Digestive Foundation (Maag Lever Darm Stichting, MLDS). RKW is supported by a VIDI grant (016.136.308) from the Netherlands Organization for Scientific Research (NWO) and a PSC Partners Seeking a Cure grant "The Exome in PSC".

Contributors

RA, EMGdV, XJ, FS, KR, KS, ALM, and WW: statistical analysis and interpretation of data. CYP and RKW: study supervision. KR, KS, XJ, FS, and MP: performed experiments. RA, EMGdV, THK, SH, CS, TF, JRH, EM, FS, CYP, and RKW wrote the manuscript. JZL, AFranke, DE, and CAA performed genotyping, calling and QC. RA, EMGdV, ECG, XJ, FS, KR, KS, TF, TJW, ALM, WW, GA, DA, AB, NKB, UB, EB, KMB, CLB, MCB, MC, OC, AC, GD, JE, BE, DE, MF, EAMF, AFloreani, IF, DNG, GMH, BvH, KH, SH, JRH, FI, PI,BDJ, HL,WL, JZL, H-UM, MM, EM, PM, TM, AP, CRupp, CRust, RNS, CS, SS, ES, MSilverberg, BS, MSterneck, AT, LV, JV, AVV, BdV, KZ, RWC, MPM, MP, SMR, KNL, AFranke, CAA, THK, CYP, RKW, The UK-PSC Consortium, and The International PSC Study Group contributed to sample and clinical data collection. All authors revised the manuscript for critical content and approved the final version.

Competing Interests

None declared.

Transcript profiling

Previously published microarray data characterizing the transcriptomic profile of PBD and CLCs were used. This data is available on ArrayExpress, accession number: E-MTAB-2965.

References

1 Hirschfield GM, Karlsen TH, Lindor KD, *et al.* Primary sclerosing cholangitis. *Lancet* 2013;**382**:1587–99.

2 Boonstra K, Beuers U, Ponsioen CY. Epidemiology of primary sclerosing cholangitis and primary biliary cirrhosis: A systematic review. *J Hepatol* 2012;**56**:1181–8.

3 Boonstra K, Weersma RK, van Erpecum KJ, *et al.* Population-based epidemiology, malignancy risk, and outcome of primary sclerosing cholangitis. *Hepatology* 2013;**58**:2045–55.

4 Boberg KM, Lind GE. Primary sclerosing cholangitis and malignancy. Best Pract. Res. Clin. Gastroenterol. 2011;**25**:753–64.

5 Eaton JE, Talwalkar J a, Lazaridis KN, *et al.* Pathogenesis of primary sclerosing cholangitis and advances in diagnosis and management. *Gastroenterology* 2013;**145**:521–36.

6 Mitchell S, Thyssen M, Orchard T, *et al.* Cigarette smoking, appendectomy, and tonsillectomy as risk factors for the development of primary sclerosing cholangitis: a case control study. *Gut* 2002;**51**:567–73.

7 Chapman R, Varghese Z, Gaul R, *et al.* Association of primary sclerosing cholangitis with HLA-B8. *Gut* 1983;**24**:38–41.

8 Bergquist A, Montgomery SM, Bahmanyar S, *et al.* Increased risk of primary sclerosing cholangitis and ulcerative colitis in first-degree relatives of patients with

primary sclerosing cholangitis. *Clin Gastroenterol Hepatol* 2008;**6**:939–43.

9 Lamberts LE, Janse M, Haagsma EB, *et al.* Immune-mediated diseases in primary sclerosing cholangitis. *Dig liver Dis* 2011;**43**:802–6.

10 Saarinen S, Olerup O, Broomé U. Increased frequency of autoimmune diseases in patients with primary sclerosing cholangitis. *Am J Gastroenterol* 2000;**95**:3195–9.

11 Folseraas T, Melum E, Rausch P, *et al.* Extended analysis of a genome-wide association study in primary sclerosing cholangitis detects multiple novel risk loci. *J Hepatol* 2012;**57**:366–75.

12 Ellinghaus D, Folseraas T, Holm K, *et al.* Genome-wide association analysis in primary sclerosing cholangitis and ulcerative colitis identifies risk loci at GPR35 and TCF4. *Hepatology* 2013;**58**:1074–83.

13 Liu JZ, Hov JR, Folseraas T, *et al.* Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis. *Nat Genet* 2013;**45**:670–5.

14 Ellinghaus D, Jostins L, Spain SL, *et al.* Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat Genet* 2016;**48**:510–8.

15 Bartlett JR, Friedman KJ, Ling SC, *et al.* Genetic modifiers of liver disease in cystic fibrosis. *JAMA* 2009;**302**:1076–83.

16 Stickel F, Buch S, Zoller H, *et al.* Evaluation of genome-wide loci of iron metabolism in hereditary hemochromatosis identifies PCSK7 as a host risk factor of liver cirrhosis. *Hum Mol Genet* 2014;**23**:3883–90.

17 Lennard-Jones JE. Classification of inflammatory bowel disease. *Scand J Gastroenterol Suppl* 1989;**170**:2-6-19.

18 'European Association for the Study of the Liver'. EASL Clinical Practice Guidelines: management of cholestatic liver diseases. *J Hepatol* 2009;**51**:237–67.

19 Jia X, Han B, Onengut-Gumuscu S, *et al.* Imputing Amino Acid Polymorphisms in Human Leukocyte Antigens. *PLoS One* 2013;**8**.

20 Fickert P, Stöger U, Fuchsbichler A, *et al.* A new xenobiotic-induced mouse model of sclerosing cholangitis and biliary fibrosis. *Am J Pathol* 2007;**171**:525–36.

21 Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative CT method. *Nat Protoc* 2008;**3**:1101–8.

22 Sampaziotis F, Cardoso de Brito M, Madrigal P, *et al.* Cholangiocytes derived from human induced pluripotent stem cells for disease modeling and drug validation. *Nat Biotechnol* 2015;**33**:845–52.

23 Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 2004;**3**:Article3.

24 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B* 1995;**57**:289–300.

25 Mederacke I, Hsu CC, Troeger JS, *et al.* Fate tracing reveals hepatic stellate cells as dominant contributors to liver fibrosis independent of its aetiology. *Nat Commun* 2013;**4**:2823.

26 Wu C, Li D, Jia W, *et al.* Genome-wide association study identifies common variants in SLC39A6 associated with length of survival in esophageal squamous-cell carcinoma. *Nat Genet* 2013;**45**:632–8.

27 Parkes M, Barrett JC, Prescott NJ, *et al.* Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nat Genet* 2007;**39**:830–2.

28 Nam JS, Turcotte TJ, Smith PF, *et al.* Mouse cristin/R-spondin family proteins are novel ligands for the frizzled 8 and LRP6 receptors and activate ??-catenin-dependent gene expression. *J Biol Chem* 2006;**281**:13247–57.

29 van Amerongen R, Nusse R. Towards an integrated view of Wnt signaling in development. *Development* 2009;**136**:3205–14.

30 de Lau WB, Snel B, Clevers HC. The R-spondin protein family. *Genome Biol* 2012;**13**:242.

31 Cheng JH, She H, Han Y-P, *et al.* Wnt antagonism inhibits hepatic stellate cell

activation and liver fibrosis. *Am J Physiol Gastrointest Liver Physiol* 2008;**294**:G39–49.

32 Myung SJ, Yoon J-H, Gwak G-Y, *et al.* Wnt signaling enhances the activation and survival of human hepatic stellate cells. *FEBS Lett* 2007;**581**:2954–8.

33 Xinguang Y, Huixing Y, Linlin W, *et al.* RSPOs facilitated HSC activation and promoted hepatic fibrogenesis. *Oncotarget* 2016;**5**.

34 Jostins L, Ripke S, Weersma RK, *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 2012;**491**:119–24.

35 Nakamura H, Arai Y, Totoki Y, *et al.* Genomic spectra of biliary tract cancer. *Nat Genet* 2015;**47**:1003–10.

36 Gao Q, Zhao YJ, Wang XY, *et al.* Activating mutations in PTPN3 promote cholangiocarcinoma cell proliferation and migration and are associated with tumor recurrence in patients. *Gastroenterology* 2014;**146**:1397–407.

37 Norman GL, Bialek A, Encabo S, *et al.* Is prevalence of PBC underestimated in patients with systemic sclerosis? *Dig Liver Dis* 2009;**41**:762–4.

38 De Boer YS, Van Gerven NMF, Zwiers A, *et al.* Genome-wide association study identifies variants associated with autoimmune hepatitis type 1. *Gastroenterology* 2014;**147**.

Figure legends

Figure 1: Association of genetic variants on chromosome 6 with transplant-free survival of PSC patients.

(A) Kaplan-Meier curves of transplant-free survival. Patients are stratified according to their genotype for SNP rs853974. The *P*-value for genotype effect in the Cox proportional hazards model is $P = 6.07 \times 10^{-09}$. (B) Regional association plot for transplant-free survival. The Y-axis shows the $-\log_{10}$(*P*-value) for genotype effect in the Cox proportional hazards model.

Figure 2: *RSPO3* expression in mouse cholangiocytes and in human cholangiocyte-like-cells, primary bile duct and hepatic stellate cells

(A) RNAseq analysis of *RSPO3* expression in DDC induced cholestatic cholangiocytes, healthy cholangiocytes and multiple organs of normal C57BL/6 mice. FPKM: Fragments Per Kilobase of exon per Million mapped reads. (B) Microarray *RSPO3* expression in human induced pluripotent stem cells (hiPSCs), cholangiocyte-like-cells (CLCs) and primary bile duct (PBD). *RSPO3* expression is significantly increased in CLCs and in PBD compared to hiPSCs. n=3; error bars, standard deviation. Asterisks represent statistical significance (****adjusted P<0.0001, ***adjusted P<0.001, Benjamini and Hochberg corrected P-values). (C) Quantitative real time PCR (qPCR) analysis demonstrating the expression of *RSPO3* in hiPSC-derived CLCs and PBD samples compared to expression in hiPSCs. Expression levels are fold changes compared to housekeeping gene HMDS calculated using the $2^{-\Delta Ct}$ method. (D) Quantitative real time PCR (qPCR) analysis showing expression of *RSPO3* and Cytoglobin B in three patients without PSC. Cytoglobin B mRNA expression was evaluated as specific HSC marker. Target genes were normalized using *GAPDH* as endogenous control and their relative expression was calculated with the $2^{-\Delta Ct}$ method.