STUDY OF THE PHENOTYPIC CHARACTERISTICS AND GENETIC RISK FACTORS OF PRIMARY SCLEROSING CHOLANGITIS

Dr Brijesh Srivastava

Homerton College, University of Cambridge

September 2018

"This dissertation is submitted for the degree of Doctor of Medicine"

Declaration

I, Brijesh Srivastava, confirm that:

- This dissertation is the result of my own work and includes nothing, which is the outcome of work done in collaboration, except as specified in the text.

- It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University of similar institution.

- It does not exceed the prescribed word limit of 60,000 words.

Abstract

Primary sclerosing cholangitis (PSC) is a progressive chronic cholestatic liver disease, affecting young males predominantly. It causes stricturing of the intrahepatic and extrahepatic bile ducts, often leading to biliary cirrhosis, with an increased risk of colorectal and hepato-biliary malignancy. It is commonly associated with inflammatory bowel disease (IBD). Specific medical therapy is ineffective and for patients who develop end-stage liver disease, liver transplantation remains the only definitive treatment option.

The UK PSC study was set up in 2008 to recruit patients with PSC, to collect clinical data and DNA samples, to study phenotypic characteristics and the role of genetic risk factors in PSC. To date, more than 2,000 patients have been recruited from UK hospitals, making it the largest independent PSC cohort worldwide.

I studied the phenotypic characteristics of the cohort and identified an increased risk of requiring liver transplantation in patients with both intrahepatic and extrahepatic bile duct disease, in comparison to those with disease limited to intrahepatic ducts. Further study of the clinical and laboratory parameters, helped develop a UK PSC risk score, which allowed risk stratification of patients with PSC at the time of diagnosis.

I performed a replication/candidate gene study followed by a genome wide association study in a cohort of 1,030 and 1,020 patients respectively with 5,162 controls. Genome-wide signficant association was identified at two novel loci: 10p15 and 12q24, containing candidate genes *Il2RA* and *SH2B3/ATXN2* respectively. These genes have important roles in adaptive immune pathways, implicating an immue mediated disease process in the pathogenesis of PSC. As part of UK PSC, I also contributed to two large scale genome wide association studies in an international collaboration, which identified 13 novel loci associated with PSC.

The UK PSC cohort is a unique national resource of patients with PSC and further deep phenotyping and quality of life studies are planned, in addition to undertaking clinical trials for novel therapeutic agents. The study of phenotypic characteristics and genetic association, undertaken in patients with PSC are described in this thesis.

ACKNOWLEDGEMENTS

I am immensely grateful to the following individuals and organisations:

- My main supervisor, Dr Graeme Alexander, for giving me the opportunity to be part of the UK PSC study and for his constant guidance, encouragement and excellent support throughout my research
- My second supervisor: Dr Richard Sandford, in the Academic Department of Medical Genetics for providing excellent mentorship and support
- Dr George Mells for his immense support and guidance for analysis of genetic association study
- UK PSC investigators: Dr Simon Rushbrook (Norfolk and Norwich University Hospital NHS Trust), Dr Roger Chapman (Oxford Radcliffe Hospital NHS Trust) and Dr Carl Anderson (The Wellcome Trust Sanger Institute, WTSI)
- Dr Espen Melum and Dr Tom H Karlsen for sharing the summary data for the replication study
- Previous and present project managers for the UK PSC Genetics Study: Ms Agnes Muriithi, Mr Matthew Brown and Ms Kelly Spiess
- Department of Molecular Genetics, ATC, Addenbrooke's Hospital for extracting, normalizing and plating PSC DNA samples
- Institute of Clinical Molecular Biology (ICMB), Kiel, Germany for genotyping the UK PSC cases for the replication and candidate gene study
- Wellcome Trust Case Control Consortium (WTCCC) for providing the healthy control genotype data
- All participants of the UK PSC Study
- All principal investigators of the UK PSC consortium and the National Institute for Health Research (NIHR) Comprehensive Clinical Research Network (CCRN) led research nurse support received for recruiting patients enrolled in this study
- UK-based PSC support charity (Registered charity number 1115615) and the Norwegian PSC research center for funding this study
- **Most of all I would like to thank my wife Ankita Sahay for her unconditional support, patience and encouragement without which this work would not have been possible; and my daughter Anoushka for keeping me sane!**

TABLE OF CONTENTS

Abbreviations

AASLD – American association for study of Liver disease
AIH - autoimmune hepatitis
AIP - autoimmune pancreatitis
ALP - alkaline phosphatase
ALT - alanine aminotransferase
AST - aspartate aminotransferase
ANCA - anti-neutrophil cytoplasmic antibodies
CCA - cholangiocarcinoma
CRC - colorectal cancer
CBD - common bile duct
CD - Crohn's disease
EASL - European association for study of the liver
ERCP - endoscopic retrograde cholangio-pancreatography
FXR - Farsenoid X receptor
GB - gallbladder
GWAS - genome wide association study
IgG-SC - IgG4-related sclerosing cholangitis
IBD - inflammatory bowel disease
IPAA - ileal pouch anal anastomosis
KM - kaplan-meier
LFT - liver function tests
LT - liver transplantation
MELD - model for end-stage liver disease
MRCP - magnetic resonance cholangio-pancreatography
MAF - minor allele frequency
OR - odds ratio
PBC - primary biliary cholangitis
PI - Principal investigator
PREsTo - PSC Risk Estimate Tool
PSC - Primary sclerosing cholangitis
PTC - percutaneous transhepatic cholangiography
SLE - systemic lupus erythematosus
TBB5 - Beta-tubulin isotype 5
UDCA - Ursodeoxycholic acid
UC - ulcerative colitis

List of Tables

List of Figures

Chapter 7.

Chapter 1: Introduction.

Epidemiology of PSC

Primary sclerosing cholangitis (PSC) is a chronic cholestatic liver disease, characterized by progressive inflammation and fibrosis of the intrahepatic and extrahepatic bile ducts (1). PSC was first described by Delbet in 1924 (1) and then by Schwartz and Dale in a review of 6 patients with ulcerative colitis who developed inflammation of the bile ducts within as well as without the liver (2).

The first population based study in PSC was undertaken in a well defined Norwegian population in 1998, with an estimated mean annual incidence and point prevalence of 1.3 per 100,000 person-years and 8.5 per 100,000 persons respectively (3). Subsequent studies in the USA and UK have shown estimated incidence and prevalence rates of 0.41 – 0.9 per 100,000 person-years and 3.85 – 13.6 per 100,000 persons respectively (4, 5). PSC appears to be more prevalent in northern European countries, particularly Nordic countries (6). The most recent population based study performed in the Netherlands identified a mean incidence of 0.5 per 100,000 inhabitants (between 2000 and 2007) and a point prevalence of 6 per 100,000 inhabitants in a cohort of 590 well-characterised patients with PSC (7). PSC typically affects young males between the ages of 25 and 55 years; up to one third of affected patients are female (8). Male predominance in PSC was reported as early as 1966 (9) and has been confirmed in several follow-up studies (7). In contrast, a study by Takikawa et al, in Japanese patients concluded two distinct peaks in the age distribution at diagnosis: 20 – 30 years and 50 – 70 years when compared with the west (10). The prevalence rates of PSC in Japan are much lower when compared with northern Europe or U.S.A, at around 0.095 per 10,000 (11). The reason for the geographical difference is not entirely clear, but could reflect considerable variability in describing the PSC phenotype accurately as well as difficulty in diagnosis. It is apparent that

PSC differs considerably between various population groups and it is difficult to generalize about the epidemiological features based on any single population study alone. The main emphasis throughout this thesis however, will be based on studies done in the context of "Caucasian" (European & American) PSC cohorts, to allow a degree of homogeneity in the interpretation and conclusions of study findings across these sub-populations.

Clinical features of PSC

PSC should be considered in patients diagnosed with inflammatory bowel disease (IBD) who develop deranged (cholestatic) liver function tests (LFTs). Cholestasis is defined as decrease in bile flow, either due to impaired secretion from the hepatocytes or obstruction to bile flow within intrahepatic or extrahepatic bile ducts. However, the clinical presentation can vary considerably between different population sub-groups. Chapman *et al.* and Wiesner *et al* were the first groups to describe the clinical characteristics in large cohorts of patients with PSC in 1980 (1, 12). One key feature that emerged from this study was that approximately 7 - 10% of patients were asymptomatic at diagnosis with a raised alkaline phosphatase (ALP) the only clue to the diagnosis.

The largest single centre study by Boonstra *et al.* described the characteristics of 590 PSC patients (7): 64% of patients were male (male:female ratio 1.7: 1), with the mean age at diagnosis 38.9 years (1 standard deviation = 15.2). It would be reasonable to conclude from this study and other reported studies that PSC typically affects young males, usually in their 3rd or 4th decade (13-16). Up to half of the patients are asymptomatic at diagnosis (13, 16). The most prevalent symptom recorded is abdominal pain (35%), usually in the right upper quadrant, followed by jaundice (27 - 30%), pruritus (20 - 40%), fever (11 - 35%) and fatigue (17). 2 – 7% of patients may have stigmata of chronic liver disease at presentation such as ascites, oesophageal varices or variceal bleeding (13, 16). Findings at clinical examination are usually non-specific and dependent upon the stage of disease, and include hepatomegaly (44 - 55%) and splenomegaly (29 - 30%) (17).

The serum alkaline phosphatase (ALP) is usually elevated, but is normal in up to 8.5% of patients (13). Transient elevation of bilirubin level is common (≈ 40%) (13, 16) and usually reflects a degree of cholangitis or dominant bile duct stenosis. However, persistent conjugated hyperbilirubinaemia (for over 3 months), is associated with a poor prognosis, but was seen in only 15% of patients in one study (16).

Natural history of PSC

The pathogenesis of PSC is poorly understood and the disease course notably unpredictable; this has led to an ever-increasing research interest in this condition. The onset of PSC is usually insidious and patients tend to run a relatively stable course, at least in the early stages. Deterioration in the clinical picture is often due to an episode of ascending cholangitis secondary to biliary stasis. If treated promptly, patients usually make a full recovery to the pre-existing clinical state. Repeated attacks of cholangitis might lead to the development of progressive bile duct stenosis or strictures. Cholangitis is however, not a pre-requisite for the development of dominant strictures (discussed later in this chapter; page 45) and nearly half of patients with PSC have dominant strictures at follow-up (18).

As with other cholestatic liver disorders, PSC may lead eventually to biliary cirrhosis. However, only a small proportion (2 - 7%) of patients present with features of advanced liver disease at diagnosis (13, 16). Until recently, the estimated median survival from diagnosis to either death or OLT was estimated to be between 9.6 and 18 years (13, 15, 16). In the recent study by Boonstra *et al.* survival estimates were calculated for two separate PSC cohorts: a population based cohort (n = 140) and a tertiary referral centre cohort (n = 450) with the combined end-points of liver transplantation or PSC-related death (7). The population-based cohort had significantly extended survival when compared to the tertiary centre cohort (21.3 years versus 13.2 years), confirming the referral bias noted in previous reports (7).

Patients with a persistent elevation of the serum conjugated bilirubin from presentation have a reduced median survival of just 30 months (16). It is beyond doubt that the presence of sustained conjugated hyperbilirubinaemia marks the beginning of decompensated chronic liver disease. In this context it is imperative to search for the presence of a dominant stricture or more sinister pathology, such as a superimposed cholangiocarcinoma (CCA), which may present with cholestatic jaundice.

No therapy targeted at PSC has so far been shown to improve patient outcomes, other than liver transplantation.

Diagnosis of PSC

Defining PSC

This section describes the criteria used to make a diagnosis of PSC in this thesis and is in accordance with national and international guidelines.

PSC is diagnosed based on the following standard diagnostic criteria (19):

A) The presence of cholestatic LFTs.

B) The presence of intrahepatic and/or extrahepatic bile duct changes (including irregularity, narrowing, beading/segmental dilatation and/or stricturing) with endoscopic retrograde cholangio-pancreatography (ERCP), MR cholangiopancreatography (MRCP) or percutaneous transhepatic cholangiography (PTC).

C) Histological changes consistent with PSC.

Patients who have histological changes consistent with PSC but normal cholangiography are considered to have "small-duct PSC" (19).

It is important to exclude all possible causes of secondary sclerosing cholangitis before a diagnosis of PSC can be considered secure. Table 1.1 lists the recognised causes of secondary sclerosing cholangitis, which can mimic both the clinical and the radiological features of PSC.

Table 1.1 Causes of secondary sclerosing cholangitis

Choledocholithiasis
Ischemic cholangiopathy
Autoimmune pancreatitis/IgG4-associated cholangitis
Intra-arterial chemotherapy
Surgical biliary trauma
HIV associated cholangiopathy
Portal hypertensive biliopathy
Cholangiocarcinoma
Recurrent pyogenic cholangitis
Eosinophilic cholangitis
Histiocytosis X
Recurrent pancreatitis
Chronic ketamine use
Ductal plate malformation

Role of ERCP

Cholangiographic assessment of the biliary tract is essential for a definite diagnosis of PSC. Until the mid 1970s a definite diagnosis of PSC could only be made using operative cholangiography at laparotomy. The advent of endoscopic retrograde cholangio-pancreatography (ERCP) in the late 1970s allowed the diagnosis to be made in a less invasive manner, but with similar sensitivity and specificity (20). The characteristic cholangiographic findings include: multifocal, short, annular strictures alternating with normal or slightly dilated segments producing the classical beaded pattern of bile ducts (19, 21). A cholangiography based classification system - the "Amsterdam classification" (table 1.2) was first derived by Majoie *et al.* and subsequently modified by Rajaram *et al.* (22, 23). Patients usually have both intrahepatic and extrahepatic changes: however up to 25% of patients have disease restricted to the intrahepatic ducts (16, 19). Isolated extrahepatic disease is uncommon and seen in less than 5% of all patients (19). However, ERCP is

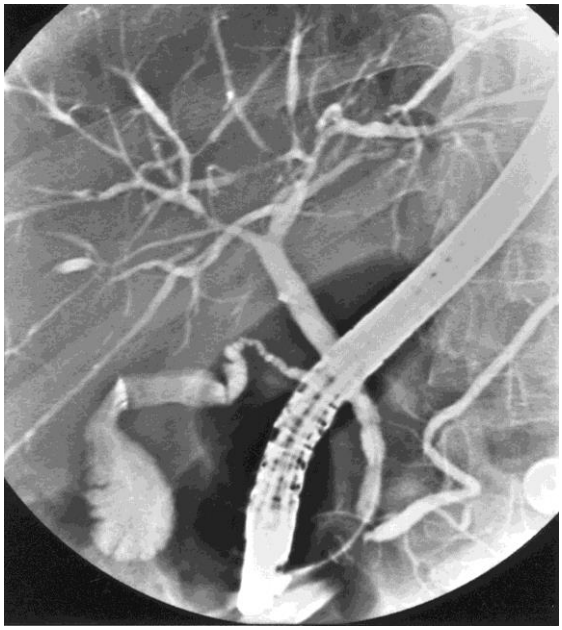
an invasive procedure and is associated with complications such as pancreatitis, biliary sepsis, bleeding, perforation and aspiration (24).

In the recent years, there has been a shift from using ERCP as a diagnostic tool to restricting its use for selected groups of patients who require therapeutic intervention to maximize the benefit to risk ratio. Figures 1.1(a) and 1.1(b) show intrahepatic and both intra- and extrahepatic duct involvement in PSC respectively at ERCP.

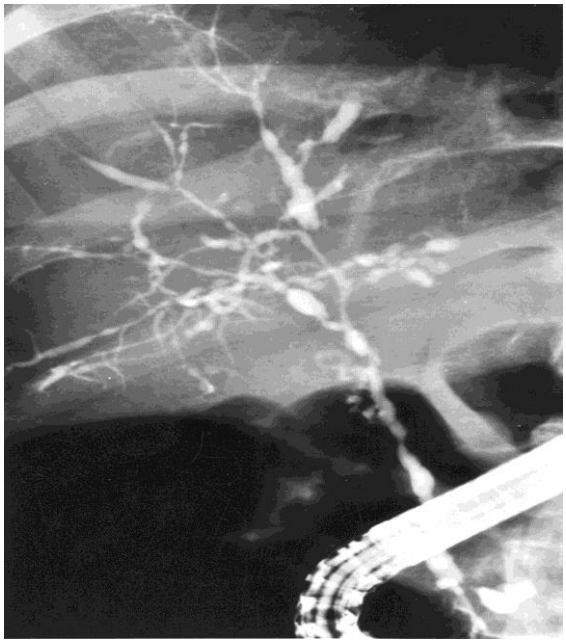
Table 1.2. Amsterdam classification system for PSC based on cholangiography

Morphologic type	Cholangiography findings
Intrahepatic	
Type 0	No abnormalities.
Type I	Multiple strictures, normal calibre of the bile ducts or minimal dilatation.
Type II	Multiple short, band-like strictures, saccular dilatations, decreased arborisation.
Type III	Only central branches filled despite adequate filling pressure; severe pruning, one or more outpouchings.
Extrahepatic*	
Type 0	No abnormalities.
Type I	Irregular contour of extrahepatic ducts, without distant narrowing.
Type II	Segmental narrowing of extrahepatic ducts, with smooth or irregular margin.
Type III	Irregular narrowing and beading of the entire length of the common duct.
Type IV	The margin of the extrahepatic ducts is extremely irregular with diverticulum like out-pouching.

*Extrahepatic ducts comprise the first order bile ducts (right or left main hepatic duct) and/or common bile duct at cholangiography allowing distinction between extrahepatic and intrahepatic disease.



(a)



(b)

Figure 1.1(a). ERCP showing intrahepatic bile duct beading and focal strictures with normal common bile duct; (b) intrahepatic and extrahepatic multifocal biliary strictures due to PSC.

Role of MRCP

With the introduction of MRCP in the early 1990s, its sensitivity and specificity were found to be comparable to ERCP in evaluating patients with suspected biliary tract diseases (25). Cholangiographic changes similar to that seen on ERCP, as described above, are considered diagnostic. A recent meta-analysis by Dave *et al*. concluded that the higher sensitivity (86%) and specificity (94%) obtained at MRCP for the diagnosing of PSC, allowed recommendation of MRCP as the first line investigation in patients with suspected PSC (26). However, where patients are unable to have MRCP, or require therapeutic intervention or when MRCP findings are inconclusive, ERCP is the preferred investigation.

Reporter variability is undoubtedly a concern but is likely to diminish with increased availability and experience of MRCP in the clinical setting.

The main advantages of MRCP over an ERCP are:

a) It is non-invasive and avoids the use of ionizing radiations.
b) Lower risk of significant complications.
c) Reduced cost.
d) It allows complete visualization of the biliary tree (intrahepatic and extrahepatic) irrespective of the presence of an obstructive large bile duct lesion.

Figures 1.2(a) and 1.2(b) depict the characteristic cholangiographic changes seen within the biliary tree at MRCP.

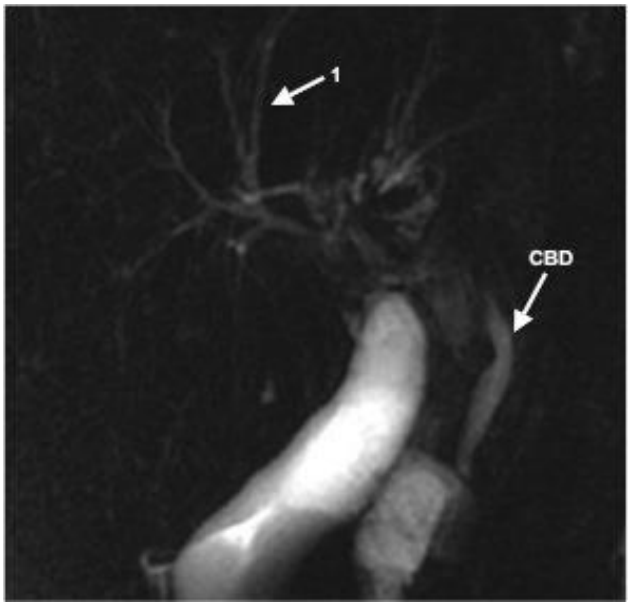


Figure 1.2(a). Cholangiography shows stricturing and beading of the intrahepatic ducts (arrow 1) with a normal appearing common bile duct (CBD).

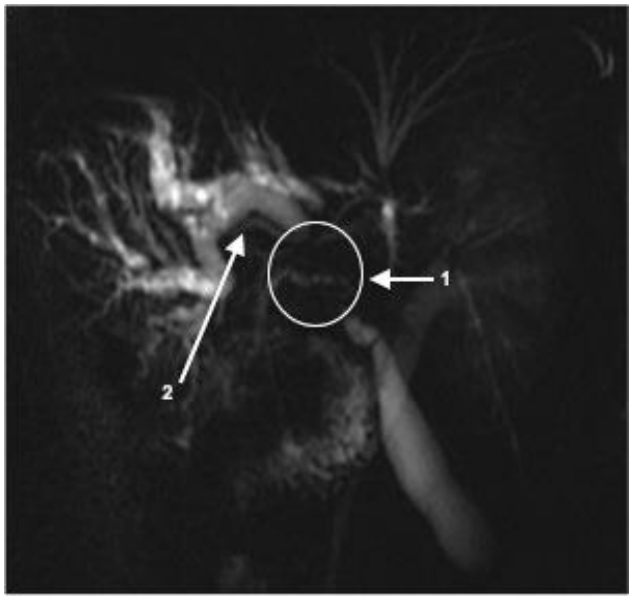


Figure 1.2(b). Cholangiography shows a severe hilar stricture involving the common hepatic duct (arrow 1) with marked dilatation of intrahepatic ducts (arrow 2).

Role of Liver biopsy and histological features of PSC

The classic histological hallmark of PSC is the presence of concentric onion skin-type periductal fibrosis. Other features that may be present include: portal fibrosis (60 - 80%), portal lymphocyte infiltration (69%), ductopenia with bile ductular proliferation (8 - 55%), cholestasis (7 - 50%) (17).

Figure 1.3 shows the characteristic histological changes pathognomonic of PSC.

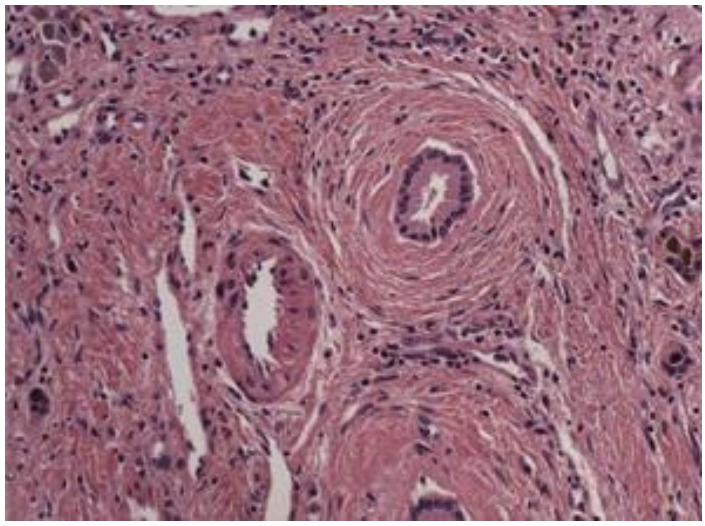


Figure 1.3. Liver histology showing lymphocytic infiltrate and concentric onion skin fibrosis.

Given the focal nature of PSC, liver biopsy specimens may not be a true representation of the pathological changes occurring in affected segments (sampling variation). In a retrospective study of patients with PSC confirmed by cholangiography, liver biopsy failed to add any useful diagnostic information (27). Moreover, a normal liver biopsy does not *per se* exclude PSC. As a consequence, routine liver biopsy is not recommended for patients with suspected PSC (19). It does however, have a role in patients with suspected PSC who might have early disease or those with normal cholangiography (to exclude small-duct PSC) and to help exclude co-existing liver disorders.

Serological markers in PSC

Serum autoantibodies form an important part of the routine diagnostic work-up of various immune mediated chronic liver disorders such as primary biliary cholangitis (PBC), autoimmune cholangitis, autoimmune hepatitis (AIH) as well as PSC. However, with few exceptions antibodies are not specific for any single disease.

Several autoantibodies have been studied in PSC but most of these have very low diagnostic specificity. The most relevant and prevalent of these is the perinuclear anti-neutrophil cytoplasmic antibody (p-ANCA), seen in 26 - 94% of patients with PSC (28) and this may even have a diagnostic role (29). The classical cytoplasmic ANCA (c-ANCA) and p-ANCA are directed against antigenic proteins found within the cytoplasm of human neutrophils. They are typically found in patients with systemic vasculitis, but have also been reported in AIH and ulcerative colitis (UC) (30). Another distinct type of ANCA known as "*atypical*" p-ANCA shows broad heterogeneous rim-like staining of the nuclear periphery associated with multiple intra-nuclear fluorescent foci (30). It is now clear that the *atypical* p-ANCA in fact represents anti-neutrophil nuclear antibodies and not the classical anti-neutrophil cytoplasmic antibodies (30). It is this *atypical* p-ANCA, which is associated most closely with PSC.

Unlike c-ANCA and p-ANCA, the auto-antigen for *atypical* p-ANCA was identified only recently. Terjung *et al.* identified Beta-tubulin isotype 5 (TBB5) as the ANCA autoantigen associated with both PSC and AIH (31). It is important to note that autoantibody titres (including that of *atypical* p-ANCA) do not often correlate with disease activity, the extent of organ involvement or with immunosuppressive therapy (29).

Based on the present evidence, a case cannot be made to use tests for autoantibodies in the diagnosis, stratification or monitoring of disease progression in patients with PSC. However, their presence can be used as circumstantial evidence of a co-existing immune-mediated process.

A list of various serum autoantibodies present in patients with PSC is shown in table 1.3.

Table 1.3. Serum autoantibodies present in patients with PSC

Antibody	Prevalence
Anti-neutrophil cytoplasmic antibody (ANCA) (c-ANCA; p-ANCA and atypical p-ANCA)	26 - 94%
Anti-nuclear antibody (ANA)	7 - 77%
Anti-smooth muscle antibody (SMA)	13 - 20%
Anti-endothelial cell antibody	35%
Anti-cardiolipin antibody	4 - 66%
Anti-thyroperoxidase antibody	7 - 16%
Thyroglobulin antibody	4%
Rheumatoid factor	15%

adapted from Chapman *et al.* (19)

Immune-mediated disease associations in PSC

There may be a role for an immune-mediated process in the pathogenesis of PSC (discussed later in this chapter). This association is further strengthened by the presence of various immune-mediated diseases in PSC patients and is discussed below.

Inflammatory bowel disease

Prevalence of IBD

PSC is often diagnosed in patients suffering with IBD (usually UC), who develop cholestatic LFT's. PSC is associated strongly with IBD, with a reported prevalence of IBD in the northern European PSC population of between 60 and 80% (1, 13, 32). The prevalence of IBD is significantly lower in Asia (20 - 50%) (33). The most common IBD phenotype observed in PSC is UC, present in 80 - 90% of patients with IBD (34). Crohn's disease (CD) is seen in up to 13% of cases and almost all of these have colonic involvement

(Crohn's colitis) (34, 35). Only 2 - 7.5% of patients with UC and 3.4% with CD appear to have co-existing PSC (34, 36).

It is important to note that not all patients with UC or CD are screened routinely for PSC and these figures are likely to underestimate the prevalence of PSC in the IBD population. This is elucidated very well in a recent study by Lunder *et al.* (37), in which 322 patients with UC were screened by MRCP for 20 years. A total of 7.4% patients were found to have features of PSC, of which only 2.2% had a prior diagnosis of PSC. It is possible that a proportion of patients with early PSC have normal liver biochemistry and so there is low clinical suspicion, until an ALP elevation becomes apparent.

Diagnostic timeline for IBD

The diagnosis of IBD precedes the diagnosis of PSC in most cases (19). However, there is no specific relationship between the onset of either of the two diseases and *de novo* cases of IBD are well recognized in patients after liver transplantation for PSC (38). Conversely, PSC can also develop several years after UC in patients who have undergone colectomy (1, 39).

PSC-IBD phenotype

The IBD associated with PSC (PSC-IBD) is considered indistinguishable from that of IBD in general, both in terms of endoscopic as well as histo-pathological findings. However, some phenotypic characteristics are seen more commonly in patients with PSC, suggesting that PSC-IBD may represent a distinct sub-type of IBD, distinct from UC and Crohn's disease (35).

Table 1.4 lists the key phenotypic features of UC seen in patients with PSC.

Table 1.4. Features of IBD seen in patients with PSC.

Pan-colitis with rectal sparing
Backwash ileitis
Quiescent disease course
Predominance of right sided inflammatory activity
Increased risk of colorectal cancer
Increased risk of pouchitis in patients with ileal pouch anal anastomosis following colectomy
Increased risk of peri-stomal varices in patients with ileostomy following colectomy

Non-IBD immune-mediated diseases associated with PSC

In a recent cohort study of 241 patients with PSC, the prevalence of non-IBD immune-mediated disease was 25%. 6% had features of autoimmune hepatitis (discussed on page 34) and the remaining (20%) had one or more extra-hepatic, non-IBD diseases with sarcoidosis, thyroid disease and type 1 diabetes the most prevalent (40). Other less common diseases reported in association with PSC patients and present in 0.4% - 2% of cases overall, include coeliac disease, rheumatoid arthritis, psoriasis, vitiligo and systemic lupus erythematosus (SLE) (40, 41).

Clinical variants in PSC

The diagnosis of PSC encompasses all cases of cholangiopathy in which secondary causes have been excluded. However, there are three distinct categories of patients, who despite having cholangiopathy differ from the usual clinical pattern of PSC either in terms of their imaging, biochemical parameters, treatment response or prognosis. Thus attempts should be made to confirm or exclude these clinical variants.

These include small-duct PSC; IgG4 associated cholangitis and PSC-AIH overlap and are discussed in more detail below. Whether these clinical phenotypes are part of the spectrum of PSC or distinct entities remains unclear.

Small-duct PSC

Cholangiography (ERCP, MRCP or PTC) can identify changes occurring only in the large bile ducts (> 100 μm in diameter), comprising ducts distal to the septal ducts. A cholangiopathy restricted to smaller bile ducts (interlobular and proximal septal bile ducts) can only be identified at liver biopsy and where MRCP or ERCP findings are considered normal, is a phenotypic variant of PSC and termed small-duct PSC (42).

True estimates of the incidence and prevalence of small-duct PSC remain largely unknown, but is estimated to be between 5 and 23% of all patients with PSC (43-45). In a recent large multi-centre study (published since the work described in this thesis was undertaken) with 7121 PSC cases, 3.4% of patients were diagnosed as having small-duct PSC (46).

The diagnostic criteria for small-duct PSC include:

A) Cholestatic liver biochemistry of otherwise unknown aetiology.

B) Normal cholangiography.

C) Liver histology consistent with PSC.

D) An absence of any risk factor for secondary sclerosing cholangitis (Table 1.1).

In general the clinical course of PSC is highly variable, but includes significant co-morbidity, with reduced survival consequent to progressive chronic liver disease and in some, a need for liver transplantation. In contrast, patients with small-duct PSC, tend towards a more benign clinical course with a significantly reduced risk of progression and superimposed malignancy.

In a large series of 83 patients with small-duct PSC patients followed for 7.4 years, liver transplant-free survival was significantly longer compared to well-matched patients with large-duct PSC (13 years versus 10 years) (44). Similar rates of IBD were seen in both the groups. None of the patients in the series developed cholangiocarcinoma, which is seen in almost a third of patients with large-duct PSC. Only 15.6% of patients with small-duct disease underwent

liver transplantation or died, demonstrating a much better prognosis in this group.

It is estimated that approximately 20% of such patients will progress to large duct involvement (44, 45), but it is unclear if small-duct PSC is part of the same clinical spectrum as large-duct disease. Since the pathogenesis for both disorders is unknown it is uncertain if these differ and this remains an area of active research.

IgG4-related disease

IgG4-related disease is a group of multisystem fibro-inflammatory disorders and has generated significant clinical interest in the last decade. This has led to the identification of robust biochemical and histo-pathological criteria to diagnose these conditions (47). Hepatic involvement is common and can be classified as:

- IgG4-related sclerosing cholangitis (IgG4-SC)
- IgG4-related hepatopathy
- IgG-related autoimmune hepatitis

IgG4-related sclerosing cholangitis (IgG4-SC), is the most common hepatic manifestation and has recently been recognized as a distinct clinical entity and can mimic PSC. Amongst various other causes of secondary sclerosing cholangitis, it is perhaps the most important to recognise because of its responsiveness to therapy. Serum IgG4 is raised (> 1.4g/l) in most but can be normal in up to 20% of patients and levels > 5.6g/l give a specificity and positive predictive value of 100% for differentiating IgG4-SC from PSC (48).

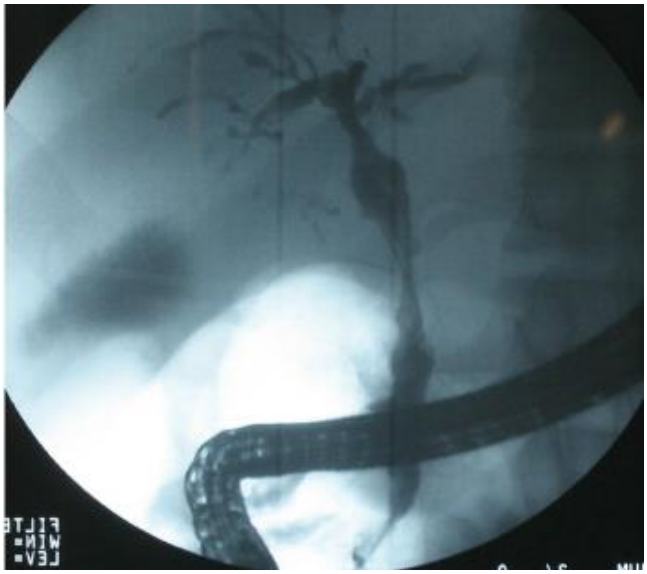
IgG4-SC is often associated with autoimmune pancreatitis (AIP) and is characterized by stricturing of the distal (intra-pancreatic) common bile duct (49). However, strictures involving other parts of biliary tree are often seen and can be indistinguishable radiologically from those in more classical PSC. IgG4-SC can be classified into sub-types based on cholangiogram findings (Table 1.5). Abnormalities of the pancreas include diffuse enlargement (atrophy in chronic cases), sausage shaped pancreas, a peri-pancreatic halo

and irregular pancreatic duct and are often present. By definition the serum IgG4 is elevated and is highly specific for the condition but is not diagnostic as it can be raised in other diseases. A cut-off greater than two times upper limit of normal is considered significant but up to 30% of AIP cases can have normal IgG4 levels (50). Histopathological examination remains the gold standard for diagnosis. Characteristic features include a lymphoplasmacytic infiltrate organized in a storiform pattern, obliterative phlebitis and a mild-to-moderate eosinophilic infiltrate. In addition, more than 30 IgG4 positive plasma cells per high-power field and a more than 50% ratio of IgG4 to IgG, provides compelling evidence of an IgG4-related disease (47). IgG4-related systemic diseases, including AIP and IgG4-SC, show dramatic responses to corticosteroid therapy with improved outcomes and corticosteroids are the treatment of choice in routine clinical practice. Figure 1.4 (a) and (b) shows cholangiographic response to steroid treatment in a patient with IgG4-SC (49).

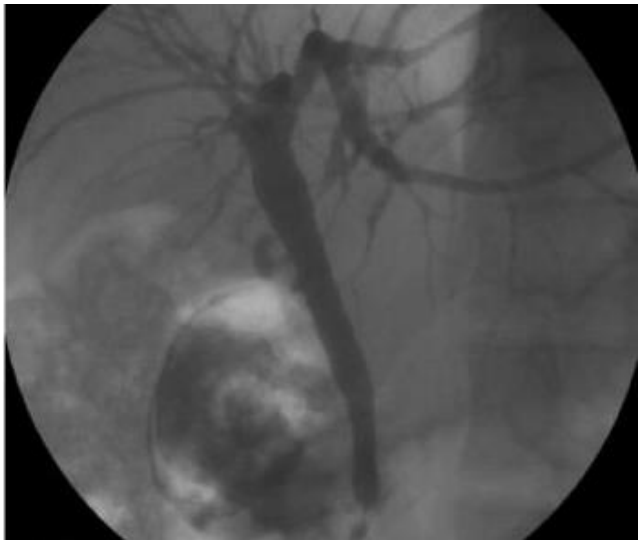
Table 1.5. IgG4-SC sub-classification based on cholangiogram.

Subtype	Biliary involvement
Type 1	Distal common bile duct (CBD) stricture
Type 2a	Diffuse intrahepatic cholangiopathy and a lower CBD stricture with prestenotic dilatation
Type 2b	Diffuse intrahepatic cholangiopathy and a lower CBD stricture without prestenotic dilatation
Type 3	Hilar and distal CBD stricture
Type 4	Hilar stricture

adapted from Culver *et al*. (51)



(a)



(b)

Figure 1.4(a). ERCP showing intrahepatic and hilar strictures in a patient with raised serum IgG4. (b) Repeat ERCP after 3 months of corticosteroid treatment showing marked improvement in biliary strictures.
Images taken from Webster et al. (49).

Raised serum IgG4 levels are seen in 9 – 36% of patients with PSC (52, 53) and an increased number of intra-hepatic IgG4 positive plasma cells were found in 23% of explanted livers from patients with PSC undergoing liver transplantation and not thought beforehand to have IgG4-SC (49). PSC patients with raised IgG4 can be distinguished from those with IgG4-SC using an IgG1:IgG4 ratio > 0.24 with a sensitivity of 86% and specificity of 95% (48).

The role of IgG4 in disease pathogenesis is unclear at present; raised levels could merely be secondary to the inflammatory response but patients with PSC and a raised IgG4 may run a more progressive course (54). It is evident that a mild elevation of the serum IgG4 is present in a small proportion of patients with PSC and does not indicate the presence of IgG4-related disease (IgG4-SC +/- AIP).

Conversely, some patients with IgG4-SC may be misdiagnosed as having PSC. The importance of distinguishing IgG4-SC from PSC cannot be overemphasized for the simple reason that the former condition is corticosteroid-responsive and the latter is not. If left untreated, IgG4-SC can progress quickly to hepatic failure in months (55) and biliary cirrhosis/chronic liver disease in longstanding cases. There are no clear-cut laboratory tests at present that would allow definitive distinction between these two conditions and emphasis should be laid on the presence of morphological features and other circumstantial evidence. If these are present, then a trial of corticosteroid therapy may be considered.

PSC with autoimmune features ("Overlap syndrome")

As highlighted previously in this chapter (on page 26), several autoantibodies are prevalent in patients with PSC but none has sufficient specificity for use as a diagnostic or prognostic marker for PSC. The lack of a specific auto-antigen with a corresponding antibody, male predominance and the lack of response to immunosuppressive treatment make it difficult to call PSC a classical autoimmune disease.

However, a small proportion of patients with PSC also have features of autoimmune hepatitis (AIH). These groups of patients are sometimes referred to having 'overlap syndrome'. The overlapping features include symptoms, clinical findings, biochemical tests, variety of immunological findings, as well as histology. A position paper from the International autoimmune hepatitis group (IAIHG) discourages the use of term 'overlap syndrome'; instead, each diagnosis should be considered separately. Diagnosis of PSC should be made on the standard criteria and if features of autoimmune hepatitis are present,

then patient has PSC with features of autoimmune hepatitis (56). The prevalence of PSC in patients with AIH has been reported to be between 1.7 and 12% in recent studies (57, 58).

It is possible that the two conditions occur sequentially (usually AIH precedes PSC) and this is commonly seen in adult transition of patients with paediatric AIH. However, in the prospective study by Gregorio *et al., which* evaluated 55 children with serological features of AIH, 50% of these cases already had features of cholangiopathy at MRCP and were labeled as having autoimmune sclerosing cholangitis (59). An alternative possibility is that PSC and AIH occur concurrently due to a predisposition to immune mediated diseases in PSC (e.g. thyroid disease and type 1 diabetes mellitus). 'Overlap' at presentation is often seen in younger patients (< 25 years) (54) and when present, the first therapeutic approach should be immunosuppression (19).

Malignancy in PSC

As with most other chronic inflammatory disease states, which are known to promote carcinogenesis, PSC is also associated with an increased risk of malignancy and this represents the major cause of mortality.

The most commonly reported malignancies are those of the biliary tract and colorectal mucosa. However, malignancy affecting other organs have been reported (60). In a recent study (published in 2017) by Weismuller *et al.* 21% of patients with PSC developed HPB malignancy over a 20-year period from diagnosis (46). The majority of these were cholangiocarcinoma, with a third diagnosed within the first year of PSC diagnosis.

Several risk factors have been postulated to increase the risk of neoplastic transformation including older age, alcohol consumption as well as the duration of co-existent IBD, but none have been validated and are of limited utility in clinical practice. The most commonly reported malignancies are discussed in this section.

Cholangiocarcinoma

Cholangiocarcinoma (CCA) is by far the most common malignancy reported in patients with PSC, with an estimated prevalence of between 5 and 20% (13, 60-62). CCA often develops considerably sooner (in the 5th decade) in patients with PSC when compared to those without PSC (usually in the 7th decade) and up to half of patients with CCA are diagnosed in the first year following the diagnosis of PSC (61, 63). This suggests that prior to the manifestation of CCA, PSC may have been present but sub-clinical until the evolution of cholestasis.

Beyond the first year of diagnosis of PSC, the annual incidence of CCA is between 0.5 and 1.5% per year (64). Diagnosing CCA is often challenging, as it grows longitudinally with early perineural and pervascular invasion (65) and typically presents as a bile duct stricture (rather than a mass) usually at the hilum.

As many as 50% of patients with PSC develop a dominant bile duct stricture (discussed on page 45). In most the cause is inflammation alone, but CCA is reported in this context in 5 to 20% of cases (66). Clinical difficulty arises as the images in the presence or absence of superimposed CCA may be identical. The diagnosis of CCA should always be considered in a PSC patient with recent clinical deterioration in the form of obstructive jaundice, worsening pruritus, cholangitis or weight loss.

No single test can reliably exclude CCA and results can be inconclusive despite repeated tests. The carbohydrate antigen 19-9 (CA 19-9) is often measured as part of routine clinical follow up in PSC. Values over 130 U/ml have a sensitivity of 79% and specificity of 98% in diagnosing CCA (66), but the results should be always be interpreted with caution as the Ca19-9 is often raised markedly in patients with bacterial cholangitis.

The EASL and AASLD guidelines do not recommend any biochemical or imaging modality to screen for CCA in patients with PSC due to lack of

evidence. Other tests that are routinely undertaken to investigate suspected CCA include MRI with MRCP, triple phase CT liver, ERCP +/- cholangioscopy, with biopsy or brushings and EUS. The prognosis is extremely poor with a median survival of only 5 months after diagnosis (67).

Gallbladder disease and malignancy

An increased risk of gallbladder (GB) disease including stones, cholecystitis, polyps and neoplasia is well established in patients with PSC (68-73). Cholecystectomy is undertaken more often in this population in as many as 20 to 25% of patients (74). In a recent study 25% of patients were found to have gallstones and cholecystitis (irrespective of the gender) and another 6% of cases had a mass lesion within the GB (73). More than half of the cases with a mass lesion were subsequently confirmed to have GB carcinoma. In contrast, only 0.2 to 20% of the general population with a mass lesion develop carcinoma (73).

Furthermore, GB polyps < 1.0 cm in size in the general population are unlikely to develop malignancy (75). This is in contrast to the risk in PSC, which appears to be present independent of polyp size. In a study in PSC by Said et al. 9/53 (13%) GB specimens were found to demonstrate dysplasia, despite the absence of a mass lesion on pre-operative imaging (73).

It is possible that GB carcinoma also follows the inflammation-dysplasia-carcinoma sequence and so it is recommended currently to refer patients with PSC patients and a GB mass lesion of any size for cholecystectomy (19, 73). It is assumed that chronic cholestasis is one risk factor for developing pigment gallstones, which are the variety seen most often in patients with PSC, but the precise mechanism remains elusive.

Colorectal malignancy

Studies have consistently shown an increased risk of colorectal cancer (CRC) in patients with PSC and co-existing IBD. This risk is increased for patients suffering with UC as well as CD and appears to be significantly higher than the risk of developing CRC in patients with IBD alone (76, 77). The absolute cumulative risk of CRC in a group of patients with PSC and UC has been

estimated to be 9%, 31% and 50% after suffering with UC for 10, 20 and 25 years respectively, in comparison to a risk of 2%, 5% and 10% for patients with UC alone at the same time points (78). A recent population based study has provided further evidence of an increased CRC risk with an estimated odds ratio (OR) of 6.9 (79).

One of the plausible explanations for the increased susceptibility may be the relatively quiescent nature of UC in PSC patients, allowing it to remain covert for a number of years, until PSC is diagnosed. However, the exact pathological basis of increased risk is unexplained.

Hepatocellular carcinoma

Patients with PSC who progress to cirrhosis are also at increased risk of developing hepatocellular carcinoma (HCC) in between 1.5 and 4% of cases (60, 80). Such patients should be managed according to the same guidelines for patients with HCC developing due to cirrhosis of any cause.

Pancreatic cancer

An increased risk of pancreatic carcinoma in PSC was first described in a case-control study by Bergquist *et al.* in 2002 (80). 5/604 (0.8%) patients were found to have pancreatic cancer giving a standard incidence ratio of 14.3. More recently, 5/200 (2.5%) patients were found to have pancreatic cancer, all of whom died within 5 years of diagnosis. No differences were seen with respect to the age at diagnosis of pancreatic cancer in comparison to the general population (60).

Pathogenesis of PSC – Current views and hypotheses

The pathogenic mechanisms leading to the development of PSC remain poorly understood, but are likely to be multifactorial. There is evidence to support an immune-mediated basis to the development of PSC, involving a complex interplay between innate and adaptive immune responses. According to this

hypothesis PSC results from exposure of a genetically susceptible individual to one or a combination of unidentified environmental factors.

There are several aspects of immune function that differ between patients with PSC and both healthy individuals and patients with other varieties of liver injury. Some of these differences may provide insight into the pathogenesis of the condition and in turn might guide future therapy.

Immunological changes in PSC

Immunological changes that occur in PSC at the cellular level have largely been studied on liver biopsy specimens from patients with established PSC. As such, it is difficult to study the cellular immune responses that take place in the earlier stages of the evolution of PSC; these early changes may be more relevant than the later changes in determining any role of the immune system in PSC. Moreover, given the patchy intra-hepatic involvement in PSC, the changes seen on a liver biopsy specimen may not reflect or represent the underlying disease process accurately. A role for both innate as well as adaptive immune response in PSC has been proposed.

An organ specific, T-cell infiltrate is characteristic in patients with autoimmune diseases. In PSC, a mononuclear cell infiltrate (predominantly both $CD4^+$ and $CD8^+$ T-cells) and an increased proportion of natural killer (NK) cells is typical, especially within and surrounding portal tracts (81, 82). However, the role of lymphocytes in mediating the pathological changes of PSC has not been well established. There is increased production of tumour necrosis factor-α (TNF-α) and IL-1β by liver derived lymphocytes; liver derived lymphocytes in PSC have been shown to produce higher levels of TNF-α at different stages of the disease in comparison to patients with PBC, suggesting the presence of a TNF-α rich environment, irrespective of disease stage (83, 84). The stimulus to drive lymphocytes to produce cytokines in PSC is not known. However, the proliferative and functional capacity of intrahepatic T-lymphocytes and NK cells is impaired in patients with PSC (83). Whether that is a direct consequence of exposure to high levels of TNF-α, or some other mechanism, is not clear.

It is noteworthy that genetic variation (polymorphisms) within the TNF-α gene has been associated with certain inflammatory and immune-mediated diseases, although this association has not yet been observed in PSC.

Leaky gut hypothesis

The atypical p-ANCA seen in patients with PSC cross-reacts with human beta-tubulin isotype 5 and the bacterial protein FtsZ expressed by intestinal flora (85). The strong association of PSC with IBD (mainly UC) has led to the hypothesis that the initial trigger in PSC could be exposure of liver cells to bacterial cell wall products and/or PAMPs, which enter the portal circulation through a permeable intestinal mucosa resulting in an aberrant immune response. Macrophages (Kupffer cells) and dendritic cells are activated through the pattern recognition receptors (PRRs) resulting in phagocytosis and subsequent production of inflammatory cytokines and chemokines (including IL-12, TNFα, IL-1β and CXCL8) (86). The resulting activation of natural killer (NK) cells and lymphocyte recruitment results in a pro-inflammatory and a pro-fibrotic environment. In addition, exposure to bacterial cell wall products may result in disruption of tight junctions within the biliary epithelial cells, thereby exposing them to toxic effects of bile acids and subsequently inflammation and fibrosis (87).

Aberrant lymphocyte homing

Although the leaky gut hypothesis is plausible, it does not fully explain the development of PSC in patients who have previously undergone total colectomy i.e. the absence of a diseased colon. It has been postulated that lymphocytes activated in the gut undergo entero-hepatic circulation and mediate hepatic inflammation leading to the development of PSC (88). Usually, intestinal mucosal lymphocytes express integrin α4/β7, which binds to its ligand mucosal addressin cellular adhesion molecule-1 (MAdCAM-1) expressed on the mesenteric endothelium, thereby ensuring intestinal homing of activated T-lymphocytes. In inflammatory liver diseases (including PSC) increased expression of MAdCAM-1 and vascular adhesion protein-1 (VAP-1) is seen in the hepatic portal endothelium, allowing liver recruitment of

lymphocytes activated in the intestine and expressing integrin α4/β7 (89). Memory cells activated in the intestine remain in circulation and may mediate hepatic inflammation, even in the absence of a diseased colon.

Bile acid toxicity

Some animal models of biliary transporter abnormalities have been shown to develop a phenotype resembling PSC. The most commonly studied amongst these is the phospholipid transporter - multidrug resistance protein 2 (mdr2) knock-out mouse, which develops inflammation and progressive fibrosis of the bile ducts following bile acid toxicity (90).

Abnormalities in the human bile acid transporter genes *ABCB4* (encodes MDR3) and *ABCB11* (encodes bile salt export pump) typically cause intra-hepatic cholestasis but variants at these loci have not yet been found to be associated with PSC. However, patients with PBC who happen to have variants at the *MDR3* locus tend to have a more severe disease course, which is most likely a consequence of hepatotoxicity secondary to persistent cholestasis (91). It is possible that patients with cholestatic liver disease such as PSC or PBC, who have mutations in the bile acid transporter genes, develop a more severe disease phenotype, but a causal relationship has not been established.

Gut microbiome

The gut microbiome in PSC has been studied using 16S rRNA sequencing and has shown an overall reduction in bacterial diversity and altered abundance of certain bacteria compared with healthy controls. No causal link has been established yet between altered gut microbiome and development of PSC, but immune dysregulation may play a part.

Treatment of PSC

Despite recent advances in understanding the pathogenesis of PSC, one might say without fear of contradiction that little progress has been made in developing effective medical therapy for PSC. Liver transplantation offers the only chance of cure. Several drugs have been studied in PSC patients including immunosuppressive drugs (including corticosteroids, calcineurin inhibitors, azathioprine and methotrexate); anti-fibrotic drugs (colchicine, pentoxyphylline, D-penicillamine and pirfenidone); and anti-TNF drugs (Infliximab and Etanercept). But none alter disease progression or improve liver related outcomes (92-95).

In contrast, systemic corticosteroid therapy has a major role in patients with proven AIH/PSC overlap syndrome or IgG4-SC as discussed above.

Role of UDCA in treatment of PSC

The most widely studied therapeutic approach in PSC is UDCA. It is a di-hydroxy bile acid, which constitutes ~ 3% of human bile acid pool and undergoes enterohepatic circulation from the distal small bowel. Enrichment of the bile acid pool with UDCA can be achieved with systemic therapy. UDCA protects hepatocytes and cholangiocytes from the toxic effects of hydrophobic bile acids present in cholestatic liver disease by detoxification, facilitation of bile acid secretion as well as inhibition of apoptosis (96). UDCA has been studied at varying doses in patients with PSC (7 to 28 mg/Kg/day) and although there was biochemical and histological improvement in some studies, almost all were underpowered to detect a survival benefit (97, 98).

A recent randomized controlled trial investigated high dose (28 to 30 mg/kg/day) UDCA in 150 patients with PSC, but the trial had to be terminated early due to a higher risk of death, need for liver transplantation and adverse effects in the UDCA treated arm (99).

A meta-analysis of 8 randomised controlled trials (5 using a standard dose ≤ 15mg/Kg/day and 3 using high dose > 15mg/Kg/day) involving 567 patients, revealed biochemical improvement in UDCA treated patients but no effect on survival. Limitations of all the studies were the small total number of patients and short follow-up periods (3 months to 6 years). There was insufficient evidence to support the use of UDCA in PSC (100).

Role of UDCA in chemoprevention of colorectal cancer

A chemopreventive role of UDCA has been suggested to reduce the risk of colorectal dysplasia and/or cancer in patients with PSC and IBD. However, these studies have been limited by sample size (50 to 100). In a recent study Lindstrom *et al*. looked at colonic epithelial dysplasia and cancer-free survival in 98 patients with PSC and IBD randomised to treatment with UDCA or placebo for 5 years and followed for a median of 12 years. No effect of UDCA was seen with either dysplasia or cancer prevention (log rank *p*-value = 0.73) (101).

Singh *et al.* (102) performed a meta-analysis of 8 studies involving 763 patients with both PSC and IBD but again, did not identify a protective effect of UDCA use on colorectal neoplasia (defined as colorectal cancer and/or any dysplasia) prevention. However, significant protective association was seen for prevention of advanced colorectal neoplasia (defined as high-grade dysplasia and/or colorectal cancer) in UDCA treated patients. The studies included in the meta-analysis had significant heterogeneity and the median duration of exposure to UDCA was only 3.5 years so the results need to be interpreted with caution and further long-term randomised trials are needed to determine a role for UDCA in prevention of colorectal neoplasia.

EASL and AASLD recommendations for UDCA use

Due to limited evidence at present, the EASL guidelines do not make a specific recommendation for the general use of UDCA in PSC. However, based on suggestive but limited evidence, EASL supports the use of UDCA for chemoprevention of colorectal cancer in high-risk groups, such as those with a strong family history of colorectal cancer, previous colorectal neoplasia

or longstanding ulcerative colitis (103). The AASLD guidelines do not recommend the use of UDCA in PSC (19).

Novel therapeutic agents

24-norursodeoxycholic acid

24-norUDCA was compared with placebo in 161 patients with PSC with raised ALP for 12 weeks in a randomised control phase II trial (104). It significantly reduced ALP level compared to placebo with a very favourable safety profile. Phase III trials are currently in progress.

It has been shown to reduce the production of toxic bile acids *in-vitro* and facilitate bicarbonate excretion into bile thereby rendering an alkaline environment, which may protect injured bile ducts from toxic bile acids. In addition, it has also been shown to have anti-fibrotic and anti-inflammatory effects in *ABCB4*-/- mice models (105).

Farnesoid X receptor (FXR) agonists

FXR agonists play a role in bile acid homeostasis by activating the farnesoid X receptor. They up-regulate bile salt export pump thereby facilitating excretion of bile acids and may have an additional role in controlling inflammation. Obeticholic acid (OCA) is an FXR agonist, which showed promising results in phase II and phase III trials in PBC (106) and is now licensed to be used as second line treatment for patients with PBC (UDCA non-responders) in the UK. OCA has also been studied in PSC in a phase II trial and showed dose-dependent reduction in ALP compared to placebo after 24 weeks of treatment (published as abstract), but longer-term results are awaited.

Another FXR agonist that has been studied recently in PSC is NGM282, which is a bioengineered analogue of FGF19 (fibroblast growth factor 19) and regulates bile acid biosynthesis. In a recent multicenter phase II trial in patients with PSC (NGM282 versus placebo) for 12 weeks, the primary end point of a change in ALP was not reached. However, there was a significant

reduction in bile acid levels and biomarkers of hepatic fibrosis (107). Further loner-term studies are needed to determine if NGM282 has a role in treatment of PSC or not.

Although theoretically, these are therapeutic options for patients with PSC, there have been concerns about the use of such prominent choleretic agents in patients with downstream biliary obstruction such as seen in PSC (108). Further studies are necessary to establish their efficacy and safety in PSC.

Microbiome modulators

Several antibiotics have been studied in PSC but despite improvements in liver biochemistry, evidence for long-term clinical benefit is lacking. Vancomycin has been consistently shown to improve ALP levels as well as the PSC Mayo risk score (discussed on page 46). A phase III trial is currently in progress to evaluate the role of oral vancomycin for PSC in IBD patients.

In addition to the novel therapeutic agents discussed above, various immunomodulatory (e.g. Vedolizumab and Timolumab) and anti-fibrotic (e.g. Simtuzumab) agents have also been evaluated in PSC, but have not shown any significant clinical benefit.

Role of endoscopy

Patients with PSC are at risk of both intra-hepatic and extra-hepatic biliary strictures. Endoscopic therapy is reserved for patients with extra-hepatic strictures (distal to the second order bile ducts), especially in the presence of persistent cholestasis/hyperbilirubinaemia. A dominant stricture, defined by presence of stenosis ≤ 1.5mm in diameter in the common bile duct (CBD) or ≤ 1mm in the intra-hepatic duct, develops in up to 50% of patients with PSC and should be treated with endoscopic therapy (109). Patients with a dominant stricture should be investigated for the presence of a superimposed cholangiocarcinoma (usually with a combination of the tumour marker CA19.9, ERCP with brushings +/- FISH).

Treatment of biliary strictures is performed endoscopically, while biliary reconstruction surgery (in the form of hepatico-jejunostomy) is reserved for patients with complex hilar/biliary strictures. Endoscopic treatment options include serial balloon dilatation with or without stent insertion (110). A multi-centre randomised controlled trial published in 2018, has shown similar clinical benefits with serial dilatation versus stenting, but the latter group had more side effects with increased risk of pancreatitis and bacterial cholangitis (111).

Liver transplantation in PSC

Liver transplantation offers the only chance of cure in PSC. PSC is the leading indication for liver transplantation in some Nordic countries and amongst the top five indications in the USA (33). Outcomes after liver transplantation in PSC are encouraging; a recent retrospective analyses of the United Network for Organ Sharing (UNOS) database showed 1-year and 5-year patient survival for cadaveric transplants of 93% and 87.5% respectively and even better survival at over 95% after 1-year and 5-years with living donor transplants (112).

Role of prognostic models

Patients with PSC usually run a variable, unpredictable disease course and the optimal timing for liver transplantation remains difficult to determine in many. Some studies have identified prognostic markers/models to predict outcomes in patients with PSC (113, 114). The most common model used is the revised Mayo score (115). It includes the following clinical parameters to calculate the risk score: age, bilirubin, albumin, AST and variceal haemorrhage. The major limitation of the revised Mayo score is that it only estimates a 4-year outcome risk and is a less useful discriminant for patients presenting early during the course of their disease.

Two new risk scores for PSC have recently been developed and are briefly discussed below:

Amsterdam-Oxford prognostic model

The Amsterdam-Oxford prognostic model for PSC is based on seven objectively measured and readily available variables: PSC subtype, age at PSC diagnosis, albumin, platelets, aspartate aminotransferase, alkaline phosphatase and bilirubin (116). The derivation cohort comprised of 692 patients with large duct PSC from Netherlands with a median follow up of 110 months. Model was validated in an external cohort of 264 patients with PSC from Oxford, UK. The test C-statistic for the model developed was 0.68 (0.51 - 0.85) and allowed accurate prediction of transplant-free survival using variables within the first few years of diagnosis.

PSC Risk Estimate Tool (PREsTo)

Eaton, JE *et al.*, derived this model in a cohort of 509 patients with PSC from North America and validated in an independent international cohort of 278 patients, using a novel machine based learning tool (gradient boosting) to predict the risk of hepatic decompensation (ascites, variceal haemorrhage or hepatic encephalopathy) (117). It consists of nine variables: bilirubin, albumin, serum ALP, platelets, aspartate aminotransferase (AST), haemoglobin, sodium, patient age, and number of years since PSC was diagnosed. PREsTo accurately predicted hepatic decompensation with a C-statistic 0.90 (0.84 – 0.95) and outperformed the Mayo and MELD scores.

However, the current clinical guidelines do not recommend the use of any prognostic models to predict clinical outcome, as there is insufficient evidence to support the use of such models (19). Further validation of the prognostic models described above in independent cohorts is needed before they can be used routinely in clinical practice.

Conversely, female sex, the presence of Crohn's disease (as opposed to UC) and small-duct PSC protect against a need for liver transplantation or death and may have a role in risk stratification (46).

Indications for liver transplantation in PSC

In the presence of cirrhosis, the indications for considering transplantation are similar to those with cirrhosis of other aetiology. However, additional, perhaps softer, indications in PSC include intractable pruritus, fatigue or recurrent cholangitis. For those with a dominant stricture and obstructive jaundice, which fails to respond to endoscopic intervention, the decision to proceed to liver transplantation can be very difficult.

It is also important to screen patient for colorectal dysplasia and cancer and attempt to exclude cholangiocarcinoma before transplantation given the poor post-transplant survival in these patients with such a high risk of tumour recurrence. Most centres (including those in the UK) exclude patients with confirmed CCA from liver transplantation although one group argues that outcomes can be satisfactory in a carefully selected minority patient group who receive additional neo-adjuvant chemo-radiotherapy (118).

Impact of liver transplantation on the course of IBD

Patients with PSC and IBD, who have undergone OLT but have an intact colon, may be at a higher risk of an IBD flare, despite immunosuppression, although this is contentious and some studies demonstrated no change in disease activity (119).

However, the cumulative 5-year and 10-year risk of developing colorectal dysplasia or cancer in patients with PSC and IBD after liver transplantation is significantly elevated at 14% and 17% respectively (120). All such patients should continue to have annual colonoscopy surveillance.

Recurrent PSC

The diagnosis of recurrent PSC (rPSC) after OLT is very difficult to make. There is no gold standard test and patients who develop ischaemic biliary strictures whatever the original aetiology have cholangiographic appearances identical to those seen in PSC.

In 1999, the Mayo clinic proposed criteria for this condition (121):

A) A diagnosis of PSC before liver transplantation.
B) Cholangiography demonstrating non-anastomotic intra-hepatic and/or extra-hepatic biliary strictures occurring 90 days after OLT and fibro-obliterative lesions with or without ductopenia on liver histology.
C) The absence of: hepatic artery thrombosis/stenosis, chronic rejection (causes pruning of biliary tree), anastomotic strictures alone, ABO incompatibility (ABO mismatch) and antibody mediated rejection.

Recurrent PSC develops in approximately 20 to 25% of liver transplant recipients at a median duration of 4 to 10 post operative years (119). Several risk factors have been associated with development of recurrent PSC including: male sex, active IBD (post-liver transplant), the presence of an intact colon, acute cellular rejection, HLA DRB1*08 type and extended donor criteria grafts (122-125).

No intervention has been found to reduce the risk of recurrent PSC and patients are followed carefully in any case for the presence of cholestasis. There is evidence suggesting a significantly reduced risk of recurrent PSC in patients (with IBD) who have undergone pan-procto-colectomy before transplantation, but a prophylactic colectomy cannot be recommended based on current evidence, although it is a frequent topic for discussion at transplant meetings.

In the absence of effective medical therapy, re-transplantation is the only treatment option available for patients with recurrent PSC who develop progressive cholestasis unresponsive to endoscopic or surgical intervention..

Role of genetic association studies in complex diseases

A brief explanation of the common terminologies used in genetic association studies is in chapter 5 (page 97) and glossary Box 1 (page 160).

Genetic association studies help identify disease-causing genes or determine genetic risk factors correlated with a disease specific phenotype. For most of the 20th century, linkage analysis and positional cloning remained the mainstay of identifying genetic risk factors in diseases and led to the association of several disease traits with specific genes by tracing transmission in affected families. However there were significant limitations to this approach, mostly due to small family size and a relative paucity of available genetic markers, making it difficult to trace inheritance accurately. When applied to complex disease traits such as type II diabetes mellitus, hypertension and heart disease, studies using linkage analysis failed to detect a single causal gene association, suggesting a more complex polygenic model of inheritance in such traits (126). Similarly in PSC, the lack of families with affected sibling pairs precluded linkage-based studies to determine genetic susceptibility loci.

To study the genetic risk profile of complex disease traits in the general population, genetic association studies were designed, which allowed comparison of frequencies of genetic variants occurring among affected and unaffected individuals in a simple case-control study design. This approach was initially limited to plausible candidate genes based on their prior probability of causing disease. However, the identification of genes, some of which were completely unsuspected on the basis of previous knowledge as causing disease in linkage studies, made candidate gene studies difficult to interpret reliably.

This led to the emergence of genome-wide association studies (GWAS), which analyses thousands of common genetic variants (usually single nucleotide polymorphisms (SNPs)) in the human genome, between cohorts of patients with the disease phenotype of interest and healthy controls using a simple case control association study. The first GWAS was published in 2005 on age-related macular degeneration (127).

At the time of undertaking genetic association studies in the UK PSC cohort, GWAS had become a well-established and effective method of identifying genetic loci associated with common diseases. As of 1st September 2016, more than 24,000 unique SNP-trait associations have been confirmed across hundreds of diseases (128). For example, in primary biliary cholangitis (PBC), 27 non-HLA genetic risk loci have been confirmed to date and have implicated key immunoregulatory pathways in PBC pathogenesis (129).

The key concepts, limitations and statistical methods used in GWAS are discussed in detail in chapter 5 (page 97).

Genetic susceptibility in PSC

PSC is considered to be a complex disease with evidence pointing towards a role for genetic and environmental risk factors in disease development (130).

Evidence for a genetic risk in PSC pathogenesis derives mainly from:

A) Increased relative risk in siblings:

Relative risk in siblings (λs) is a measure of the heritability of a disease and is calculated by dividing the disease prevalence in siblings by the disease prevalence in the general population. For monogenic disorders, λs usually ranges from 100 to 1000. The reported λs in PSC is 9 to 39 fold when compared to the general population (131).

B) Association with other immune-mediated diseases with strong genetic risk (e.g. type 1 diabetes, Coeliac disease and auto-immune thyroid disease).

Genetic susceptibility loci associated with PSC

Loci associated with risk of developing PSC can be separated in to those seen within the HLA complex and those identified by candidate gene studies or genome wide association studies (GWAS).

HLA association in PSC

The HLA complex spans across 7.6 million base pairs of DNA on the short arm of chromosome 6 (6p21) and contains 252 expressed protein-coding

genes; up to a quarter of these have possible immunological function (132). The HLA complex comprises three distinct, but closely linked regions denoted as Class I, Class, II and Class III HLA loci. Class I locus has an extended region referred to as Extended Class I, which includes the telomeric limit of the histone gene cluster and tRNA. The HLA region is highly conserved in the European population and is characterised by very strong linkage disequilibrium (LD) extending across HLA class I and II loci. A brief explanation about linkage disequilibrium is in chapter 5 (page 97).

HLA class I molecules (HLA-A, -B and -C) are expressed on all nucleated cells, present intracellular antigens to the $CD8^+$ T-lymphocytes and also serve as ligands for inhibitory killer immunoglobulin-like receptors (KIRs) on natural killer cells (NK) cells and γδ T-lymphocytes. HLA class II molecules on the other hand are expressed on antigen presenting cells (e.g. macrophages, dendritic cells and B-lymphocytes) and present extracellular/exogenous antigens to $CD4^+$ T-lymphocytes (130).

The first HLA association in PSC was reported in 1982 with HLA-B8 and subsequently HLA-DR3 (133, 134). Several studies have since confirmed an HLA association with various haplotypes but the most widely replicated risk haplotypes include the "8.1 ancestral haplotype" (*A1-B8-DR3-DQ2)*; *DRB1*1301-DQA1*0103-DQB1*0603*; *DRB1*0301-DQA1*0501-DQB1*0201* and *DRB1*1501-DQA1*0102-DQB1*0602* (135, 136). One of the protective haplotypes associated with PSC is the DR4 haplotype - *DRB1*0401-DQB1*0302*, which, when present, has been associated with a poor prognosis and possibly cholangiocarcinoma (136, 137). Other known protective HLA haplotypes include:
*DRB4*0103-DRB1*0701-DQA1*0201-DQB1*0303*;
*DRB4*0202-DRB1*1101-DQA1*0501-DQB1*0301*
and *MICA*002* (137, 138). Due to the strong linkage disequilibrium across this region, it has been very difficult to dissect the association with a specific allele/gene within the associated haplotype.

This is true for most immune mediated diseases associated with the HLA complex. One way to refine the HLA association is to study the association in ethnically different population cohorts such as Africans, which have less linkage disequilibrium. One such study was performed to define shared and distinct features with respect to distribution of HLA alleles in two different population groups: European Americans (n = 1834), Hispanics and African Americans (n = 116) who were listed for liver transplantation in the USA (139). The linkage disequilibrium between HLA-B8 and DR-3 in the African-American population was very low and the association was present for the HLA-B*08 but not HLA-DRB1*03, in contrast to the European-American population, where the association was present for both. This suggests that the association with the "8.1 ancestral haplotype" may arise primarily from the HLA-B*08 i.e. class I effect, but does not exclude associations in class II.

Ideally further studies should be attempted in such populations but given the very low prevalence of PSC in this population, such a study would be difficult to replicate. The detailed role and significance of the HLA association in PSC is beyond the scope of this thesis and is not discussed further.

Candidate gene studies in PSC (Pre-GWAS era)

Prior to the widespread availability of GWAS, genetic studies in PSC were confined to candidate gene studies. Several candidate genes (outside the HLA complex) have been studied in PSC and were selected on the basis of their priory probability as causing disease based on their association with inflammatory bowel disease, cholestatic liver disease or other immune mediated diseases, which are also prevalent in the PSC population (Table 1.6).

A true candidate gene (see glossary Box 1) shows consistent statistical association, has relevant tissue expression and is associated with a functional consequence of the identified mutation (140). Unfortunately, none of the studies performed in PSC has been able to identify any susceptibility gene for PSC. Table 1.6 lists the candidate gene studies (non-HLA loci) performed in PSC. There are two important points to note. First, the sample size of the PSC cohort studied was small and the studies may not have been powered

adequately to detect association of susceptibility loci with relatively modest effect size (OR < 2). Second, an important negative result from these studies was a lack of any association with established IBD risk loci in the PSC cohort, especially given the high prevalence of IBD (mainly UC) in all studies. This raises the possibility that the IBD phenotype observed associated with PSC might be distinct from that seen in UC and Crohn's disease. The advent of GWAS has allowed this aspect to be explored in more detail and is discussed later in the thesis.

Table 1.6. Candidate gene studies performed in PSC

Chr	Gene	Prior disease association	N (PSC)	Result	Ref
1q31	*IL-10*	IBD/RA	96	No association	(141)
2q13	*IL-1*	IBD/RA	96	No association	(141)
2q24	*BSEP*	PFIC	31	No association	(142)
2q33	*CTLA4*	T1D, Coeliac disease; Grave's disease	144	No association	(143)
3p21	*CCR5*	RA/MS	71	No association	(144)
7q	*CFTR*	CF	29	No association	(145)
7q21	*MDR3*	PFIC	30	No association	(142)
11q22	*MMP1/MMP3*	UC, Crohn's	165	No association	(146)
16q12	*CARD15*	IBD	365	No association	(147)
	TLR-4	IBD	365	No association	
	CARD-4	IBD	365	No association	
	SLC22A4	IBD	365	No association	
	SLCC22A5	IBD	365	No association	
	DLG5	IBD	365	No association	
	MDR1	IBD	365	No association	
19p13	*ICAM-1/MAdCAM-1*		104	No association	(148)
21q22	*AIRE*	APS-1	60	No association	(149)

APS-1 - autoimmune polyglandular syndrome type-1; CTLA4 - cytotoxic T-lymphocyte-associated protein 4; CARD15 – capsase activating recruitment domain 15; TLR-4 – toll-like receptor 4; CARD-4 - capsase activating recruitment domain 4; SLC22A4 – solute carrier family 22, member 4; SLC22A5 - solute carrier family 22, member 5; DLG5 – Drosophila discs large homolog 5; MDR1 – multidrug resistance gene 1.

Genome-wide association studies in PSC

Prior to undertaking the genetic association studies in the UK PSC cohort (discussed in chapters 6 and 7), two GWAS and an extended analysis of GWAS dataset by the same group had been reported in PSC.

Karlsen T *et al.* published the first GWAS in PSC in 2010 (150). The discovery panel comprised 285 Norwegian patients with PSC and 298 healthy controls, genotyped for 443,816 SNPs on the Affymetrix SNP array 5.0.

Three replication panels were used to verify the findings from discovery panel and comprised:

a) 137 patients with PSC and 368 healthy controls from Norway and Sweden (panel 1).
b) 229 PSC cases and 735 healthy controls from Belgium and Netherlands (panel 2).
c) 400 PSC cases and 1832 healthy controls from Germany (panel 3).

The strongest associations were detected near HLA-B at chromosome 6p21 [rs3099844: odds ratio (OR) = 4.8 (3.6 – 6.5); p-value = 2.6×10^{-26} and rs2844559: OR = 4.7 (3.5 – 6.4); p-value = 4.2×10^{-26}] in the discovery panel.

Outside the HLA complex, none of the SNPs reached genome-wide significance ($p < 5 \times 10^{-8}$). However, rs9524260 at chromosome 13q31 showed a suggestive association in the discovery panel [OR = 0.67 (0.53 – 0.85); p-value = 8.1×10^{-4}], but did not reach the agreed conventional threshold of genome-wide significance (discussed in chapter 5; page 102).

A similar trend towards an association was seen in two of the three verification panels (panels 1 and 2) giving a combined association for the rs9524260 (A allele) in all 1051 PSC patients and 3233 healthy controls with an OR = 0.77 (0.61 – 0.98); p-value = 2.7×10^{-3}). It was suggested that the most likely candidate gene at this locus was *GPC6* (*glypican 6*). Lentiviral

silencing of glypican 6, encoded at this locus, led to the up-regulation of pro-inflammatory markers in a cholangiocyte cell line. However, as the association at this locus did not reach genome-wide significance, its pathophysiological role in PSC remains speculative.

As part of the study, 15 established UC susceptibility loci were also replicated in the combined discovery and replication panel 1 (i.e. the Scandinavian cohort comprising 422 PSC cases and 666 healthy controls). A suggestive association was obtained at chromosomes 2q35 [rs12612347: OR = 1.26 (1.06 – 1.50); p-value = 8.8×10^{-3}] and 3p21 [rs3197999: OR = 1.22 (1.02 – 1.47); p-value = 3.3×10^{-2}], but this association did not meet threshold for statistical significance for multiple testing (chapter 5; page 102). Circumstantial evidence points towards the G-protein-coupled bile acid receptor-1 (GPBAR-1) and macrophage-stimulating 1 (MST1), respectively, as plausible candidate genes at these loci.

Melum *et al.* reported the second GWAS in PSC in 2011 (151). The discovery panel included 332 Scandinavian and 383 German PSC cases, along with 262 Scandinavian and 2700 German controls. Affymetrix genome-wide Human SNP array 6.0 was used as the genotyping platform and a total of 2,466,182 SNPs were analysed. The replication panel comprised three different cohorts of cases/controls as described below:

a) 259 PSC cases and 729 healthy controls from Scandinavia
b) 498 PSC cases and 891 healthy controls from Central Europe
c) 268 PSC cases and 554 healthy controls from United States

The strongest association was detected at 6p21 for SNPs within the HLA complex (rs3134792: p-value = 6.8×10^{-49}). To study association outside HLA loci, 23 SNPs (pruned from 379 SNPs on the basis of LD) with p-value $< 10^{-4}$ in the discovery panel were taken forward for replication.

Genome-wide significant association was detected for two loci at 2q13 and 3p21. The strongest non-HLA association was detected in the discovery panel

at 3p21 (p-value = 1.4 x 10^{-9}) and confirmed in a replication panel with a combined p-value = 1.1 x 10^{-16}; OR = 1.39 (1.24 - 1.56) for the non-synonymous (p.Arg689Cys) SNP rs3197999 located in the MST1 gene. The most plausible candidate gene at this locus is macrophage stimulating 1 (*MST1*) and has been associated with both UC and Crohn's disease. An association at 2q13 was also replicated for rs6720394 with genome-wide significance and a combined p-value = 4.1 x 10^{-8}; OR = 1.29 (1.10 – 1.51), with the most likely candidate gene *BCL2L11*.

Following on from this study Folseraas *et al*. performed an extended analysis of the results of the GWAS (152). A total of 59 non-HLA SNPs with nominal significance ($p < 0.05$) in the discovery panel, were selected for replication analysis based on their prior association with immune mediated or chronic inflammatory diseases.

The strongest association in the replication panel was detected at 1p36 for the coding SNP rs3748816 located in the Membrane metallo-endopeptidase-like 1 (*MMEL1*) locus and this was the only SNP to achieve genome-wide significance with combined p-value = 2.1 x 10^{-8}. The association signal in this region extends into the tumour necrosis factor receptor superfamily member 14 (*TNFRSF14*) making it a plausible candidate gene.

In summary:
four genetic loci (with most plausible candidate genes) were found to be associated at genome-wide significance with PSC, prior to undertaking the genetic association studies in the UK-PSC cohort:

a) 6p21 (*HLA complex*)
b) 2q13 (*BCL2L11*)
c) 3p21 (*MST1*)
d) 1p36 (*MMEL1/TNFRSF14*)

Chapter 2: UK PSC cohort selection and recruitment.

I identified the principal investigators (PI), obtained study approval at each site and led patient recruitment for the UK PSC study. The respective PI identified patients with PSC at each hospital.

The UK PSC project

The UK PSC project was established in 2009 as part of a multicentre collaboration to collect a large cohort of 2000 or more patients with PSC in the UK, to characterise the genetic risk factors that may be associated with this condition. Cambridgeshire 4 National Ethics committee (MREC Number 08/H0305/45) provided ethical approval for the study (*Appendix 1; page 216*). The consortium comprises principal investigators (PI) at each of the recruiting NHS hospital site and is led by a steering committee; the members are listed in *Appendix 2* (page 219).

I joined the study at the beginning of patient recruitment in January 2010 and worked alongside a project administrator from the initial stages to recruit patients and collect detailed phenotype data. The study was adopted as one of the UK Comprehensive Local Research Network (CLRN) portfolio studies, which enabled recruitment by local research nurses at most of the recruiting hospitals to facilitate study approval and patient recruitment. The study site was based at the Academic Department of Medical Genetics, Addenbrooke's Hospital, Cambridge.

Funding for the study was secured from UK and Norwegian PSC Charities and included costs for phenotyping and genotyping (genome-wide) of recruited patients. Funding had also been provided to genotype the UK cohort using the Immunochip platform as part of the Immunochip Consortium (details of the Immunochip study are discussed in chapter 8). These two strategies aimed to provide detailed phenotype and genome-wide association

data, including high density SNP genotype data for several autoimmune loci (as part of the Immunochip project).

Identification of recruitment sites

All NHS hospital trusts within the UK were eligible for inclusion in the study. The research team (Dr Simon Rushbrook [who chaired UK PSC at the time], an administrator and myself) identified a principal investigator at each potential site to recruit suitable patients for the study. Study approval was obtained at each site by submitting a formal application via the Integrated Research Application System (IRAS) to the respective Research and Development department.

Inclusion criteria

Inclusion criteria were based on standard diagnostic criteria (1) as detailed in chapter 1 (page 19) and included the presence of cholestatic liver function tests with characteristic bile duct changes on either endoscopic retrograde cholangio-pancreatography (ERCP), magnetic resonance cholangio-pancreatography (MRCP) and/or liver histology.

Exclusion criteria

Exclusion criteria were presence of any secondary causes of sclerosing cholangitis including congenital abnormality of the biliary tree; previous biliary tract surgery (excluding cholecystectomy); primary bile duct carcinoma; HIV cholangiopathy; primary biliary cholangitis; hepatic sarcoidosis and drug induced liver injury.

Recruitment strategy

The PI or the research nurses identified potential patients at each recruiting site. This was done by:

a. searching clinic letters.

b. searching existing databases.
c. searching histology and/or radiology databases.
d. searching databases in the clinical coding department for patients with either the local or ICD-9 code 576.1 (for PSC).

Patients were also eligible to participate in the study by self-presenting through the PSC patient support group website.

Patients willing to participate sent a reply slip to the study centre confirming interest in the study. A recruitment pack was then sent to each patient with a participant questionnaire, consent form and blood test kit. Patients were requested to provide a blood sample for DNA extraction. EDTA vacutainer tubes were sent with the recruitment pack (including a return envelope suitable for biological specimens) and patients were advised to have the blood test done at their local hospital. Blood samples were sent back to the research site by the patients in pre-paid, self-addressed envelope provided in the recruitment pack.

Please note that patients self-consented at home and returned the questionnaire and consent form to the research team at Cambridge. Upon receipt of the consent form, a clinician questionnaire was sent to the respective PI or research nurse to verify the diagnosis and obtain data on cholangiography, the presence of IBD and transplant status.

Versions of the participant and clinician questionnaire are attached in *Appendix 3* (page 222).

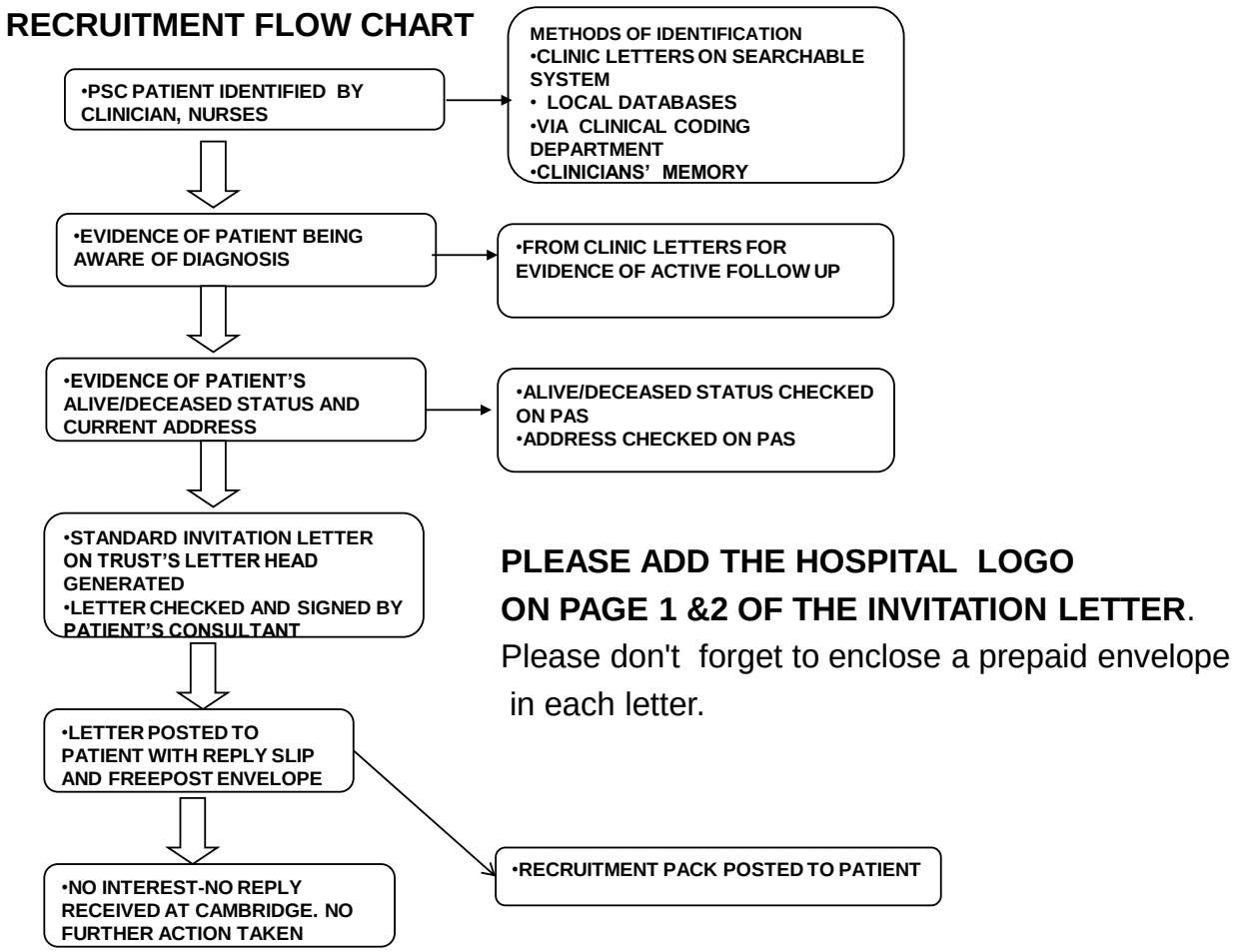


Figure 2.1. Patient recruitment flow chart.

Blood DNA extraction and preparation

Blood samples received (EDTA tubes) were forwarded immediately upon receipt for genomic DNA isolation in the molecular genetic laboratory at Addenbrooke's Hospital, Cambridge. Whole blood samples in EDTA vacutainer tubes were stored at -80˚C and thawed overnight or for 3 hours at room temperature. DNA was isolated using the Autopure LS (Large Sample Nucleic Acid Purification) procedure on the Gentra Automated System. This system is summarised briefly in the following steps:

a) *Red blood cell (RBC) lysis*: Red blood cells are lysed and separated from the white blood cells, which form a pellet on centrifuging. The supernatant (of lysed red blood cells) is discarded.
b) *Protein Precipitation:* Protein precipitation solution is mixed thoroughly with this cell lysate and centrifuged to precipitate proteins and leave DNA in solution in the supernatant.
c) *DNA Precipitation and hydration*: 100% isopropanol added to the supernatant precipitates the DNA, which is pelleted by centrifuging. The DNA pellet is re-suspended in 100 μ l of DNA Hydration solution (or 1xTE buffer) for storage.
d) *DNA normalization*: DNA samples were diluted by adding variable quantity of 1xTE buffer (based on the initial DNA concentration) at room temperature to achieve a final concentration of between 50 -300ng/µl for downstream genetic experiments. All DNA samples were stored at -40˚C in the Academic Department of Medical Genetics, Addenbrooke's Treatment Centre, Addenbrooke's Hospital.

Chapter 3: Phenotypic description of the UK PSC cohort and factors determining liver transplant-free survival.

I collected and analysed data from the participant and clinician questionnaire to obtain phenotypic data for all the recruited patients with PSC for this chapter.

Study design

All patients recruited and consented up to 1st December 2013 were included in the study. Each participant completed a participant questionnaire at recruitment (*Appendix 3;* page 222). Information collated and stored included age, gender, medical and surgical co-morbidity including IBD status, a history of malignancy, symptoms at presentation, family history of PSC and IBD and medical therapy (including the use of and dose of UDCA). Demographic and clinical data for each patient were extracted from the questionnaire and analysed.

A clinician questionnaire (*Appendix 3*) sent to the clinician or research nurse at each recruiting site who collated data on findings from the most recent cholangiogram and/or liver biopsy, as well as IBD status. Involvement of first order bile ducts (right or left main hepatic duct) and/or common bile duct at cholangiography allowed distinction between extrahepatic and intrahepatic disease. The data extracted from the clinician questionnaire were also used as a quality control measure to verify the diagnosis and determine the validity of patient-reported IBD status.

Statistical analysis

Descriptive statistics were calculated and reported as the number or proportion of patients who had completed the questionnaire for the relevant dataset. Categorical variables were analysed using Fisher's exact test. Continuous variables are reported as median and range (minimum, maximum).

Site-specific recruitment

At 1st December 2013, 1441 patients had been consented and had completed the patient questionnaire. Of these 1441 patients, 583 patients were recruited from transplant centres and 858 patients from non-transplant centres.

Clinician questionnaire

The clinician questionnaire was received for 1230 of those 1441 patients (85%). Cholangiography and/or liver histology compatible with PSC was present in 1147 (93%) patients [6/1230 (0.5%) patients had primary biliary cholangitis and were excluded from further analysis]. For the remaining 288 patients, the diagnosis of PSC was made by the recruiting clinician, but the clinical questionnaire was either incomplete or not available, so that the research team could not verify the diagnosis of PSC. Thus, 1435 patients were included in the final analysis. Table 3.1 shows the baseline characteristics of the cohort.

Table 3.1. Demographics of the study cohort (n = 1435)

Parameter	Number (range or %)
Median age at recruitment (n = 1435)	60 years (16 - 89 years)
Median age at diagnosis (n = 1198)	45 years (8 - 85 years)
Gender	Male: 903 (63%)
	Female: 532 (37%)
IBD	954 (66.5%)
	UC = 821
	Crohn's disease = 124
	Indeterminate colitis = 9
Active smokers	64 (4.5%)
Non-smokers	884 (62.5%)
Auto-immune disease	207 (14.5%)
Small-duct PSC	56 (4%)
Liver transplant recipients	337 (23.5%)

Results

Demographics

The median age at recruitment was 60 years (range 16 - 89 years). 903/1435 (63%) were male. 83% (1198/1435) were able to provide the date of diagnosis. For those, the median age at diagnosis was 45 years (range 8 - 85 years) with a median follow-up of 85 months (interquartile range 48 - 133 months) (*figure 3.1*). 64 (4.5%) patients were smokers at recruitment and 464/1435 (33%) patients were ex-smokers. Thus, the majority (62.5%) had never smoked. Smoking status was unknown for 23 patients.

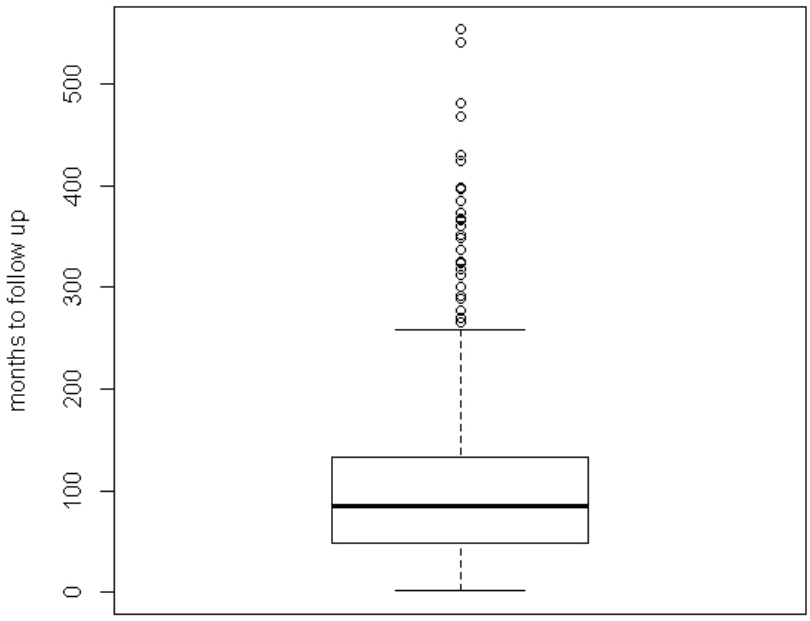


Figure 3.1. Box plot representing median follow up duration of the cohort with interquartile range.

Inflammatory bowel disease

954 of 1435 (66.5%) suffered with IBD; 356 patients did not have any bowel disease and IBD status was unknown in 125 (8.5%). Ulcerative colitis was the most common type of IBD, present in 821 (86%); Crohn's disease in 124 (13%); and indeterminate IBD in 9 (1%). 741/954 patients were able to

provide the date that IBD was first diagnosed and for those, the median age at diagnosis of IBD was 27 years (range 6 - 76 years). 638 were able to provide the date of onset for both PSC and IBD. For those, the median age at diagnosis of IBD was 27 years (range 5 - 76 years) and of PSC was 42 years (range 6 - 79 years). Clinical manifestations of PSC preceded IBD in only 8% (50/638).

Inflammatory bowel disease and smoking

Smoking status was known for 1286 (98%) patients. Only 3.6% of patients with concomitant IBD reported smoking at recruitment and an additional 29% were ex-smokers. However, the majority (67%) with IBD had never smoked. A history of smoking was associated with a reduced risk of IBD (p = 0.0005; odds ratio (OR) = 0.63, 95% confidence interval (CI), 0.49 - 0.81) (Figure 3.2(a)); this effect was more pronounced in patients with ulcerative colitis (p = 0.0004; OR = 0.57, 95% CI, 0.44 - 0.75) (Figure 3.2(b)). 52% (64/124) of patients with Crohn's disease had never smoked and only 8% (10/124) reported smoking at recruitment.

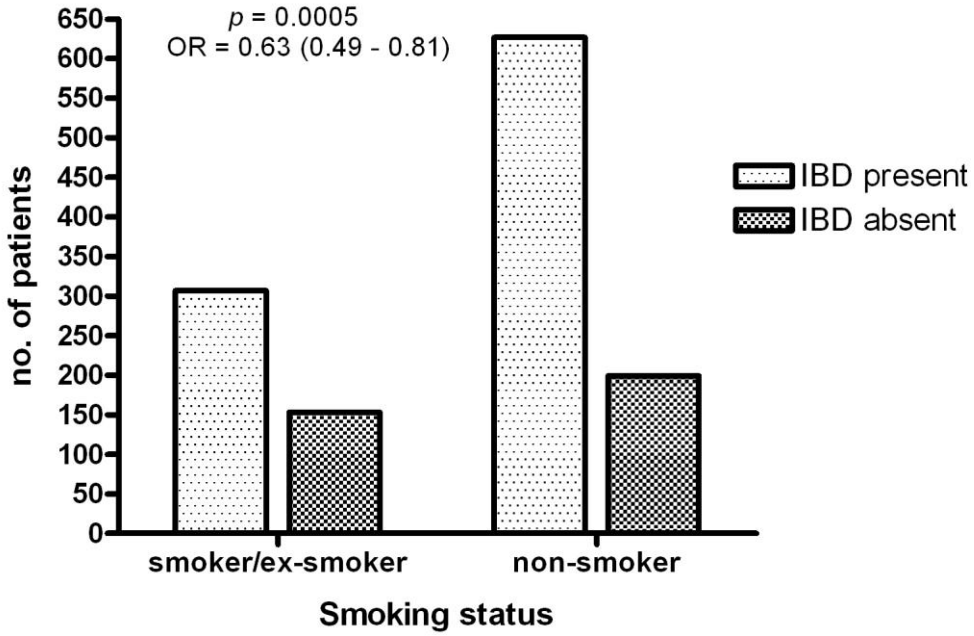


Figure 3.2 (a). The relation between smoking and inflammatory bowel disease.

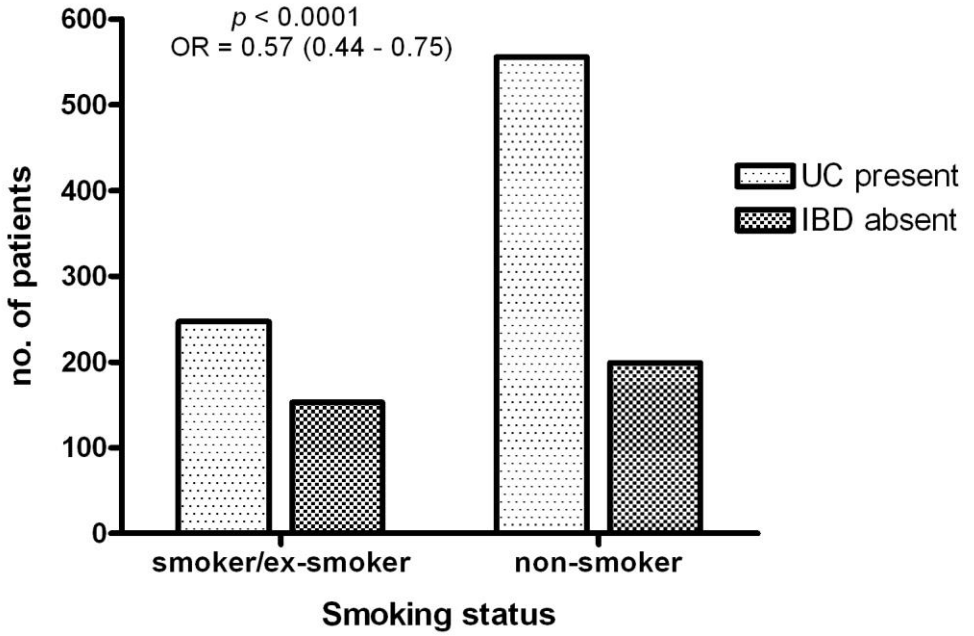


Figure 3.2 (b). The relationship between smoking and ulcerative colitis.

Autoimmune diseases

An increased prevalence of co-existing autoimmune disease was observed. 207 (14.5%) had one or more autoimmune disease (*table 3.2*). Thyroid disease was most common and present in 136 (9.5%), coeliac disease in 35 (2.5 %) and type 1 diabetes mellitus in 26 (1.8%). These autoimmune diseases appear to be more prevalent within the PSC population in comparison to the reported prevalence of these conditions in the UK population (*table 3.2*). However, a statistical analysis comparing the prevalence rates in age matched population cohort and patients with PSC has not been performed.

Table 3.2. Autoimmune diseases identified in the cohort.

Auto-immune disease	Number (%)	Prevalence in the UK
Thyroid disease	136 (9.5%)	3% (153)
Coeliac disease	35 (2.5%)	1% (154)
Type 1 Diabetes	26 (1.8%)	0.4%
Systemic lupus erythematosus	7 (0.5%)	0.1%
Rheumatoid arthritis	5	-
Sjogrens syndrome	4	-
Fibrosing alveolitis	1	-
Raynaud's	1	-
Auto-immune pancreatitis	1	-

Autoimmune diseases and the presence of IBD

Patients with IBD have increased predisposition to autoimmune diseases. Given the high prevalence of IBD in PSC population, I evaluated the effect of IBD status on the presence of autoimmune diseases. This was to determine whether the increased prevalence of autoimmune disease in the cohort was due to co-existing IBD or associated with PSC independent of IBD. The presence or absence of IBD was unrelated to an association between PSC and other autoimmune conditions (p-value = 0.13) (*Figure 3.3*).

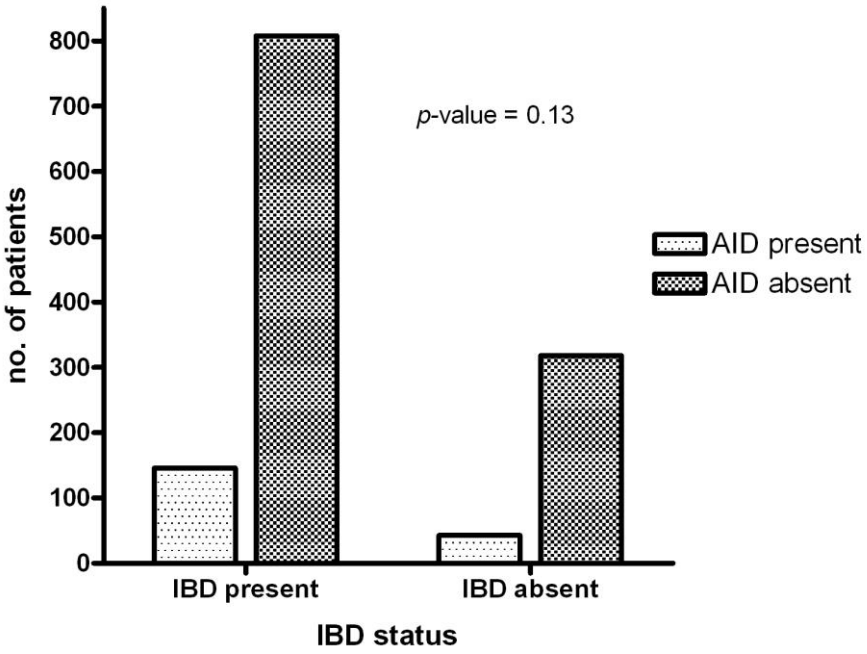


Figure 3.3. The relation between IBD status and autoimmune disease.

Symptoms at presentation

All symptoms reported by patients are listed in *Table 3.3*. At diagnosis 26% patients were asymptomatic, 51% had fatigue and/or pruritus at presentation, 35% were jaundiced and 6% reported ascites and/or gastro-intestinal bleeding.

Table 3.3. Symptoms reported by patients at presentation.

	Number (%)
Asymptomatic	376 (26)
Fatigue	733 (51)
Pruritus	700 (50)
Abdominal pain	526 (36.5)
Jaundice	510 (35.5)
Ascites or gastro-intestinal bleeding	86 (6)
Weight loss	15 (1)

Malignancy

A history of hepatobiliary malignancy was reported in 24 (1.5%), with cholangiocarcinoma most common (n = 13). Other hepatobiliary malignancy included gall bladder carcinoma (n = 6), hepatocellular carcinoma (n = 4) and one case of hepatic leiomyosarcoma.

61 (4%) had a history of colorectal cancer and an additional 13 (1%) had high-grade colonic dysplasia. Of 74 patients with colorectal cancer or dysplasia, 90% had underlying IBD (UC = 64; CD = 3).

A history of malignancy was present in an additional 96 (6.6%) patients divided into: skin (n = 39); urological (n = 23); breast (n = 22); haematological (n = 8) and gynaecological (n = 4). 16/39 patients with skin cancer were liver transplant recipients. The distribution of the various malignancies is shown in *Table 3.4*.

The rates of malignancy described above are those observed in patients who were recruited in this study and were under regular hospital follow-up. A proportion of patients with malignancy may not have been recruited introducing a selection bias (discussed on page 78).

Surgical co-morbidity

A history of total or sub-total colectomy was reported in 256 (18%). 161/256 patients were able to specify the type of bowel anastomosis following colectomy (permanent ileostomy = 107; ileal-anal pouch = 54). 240 (94%) of the colectomies were performed in patients with underlying IBD (214 with ulcerative colitis and 24 with Crohn's disease and 2 with mixed features). Of the 16 patients without IBD, 4 underwent colectomy for colorectal cancer, but the indication was unknown in the remainder. Cholecystectomy was reported in 182 patients (12.5%). A history of biliary reconstructive surgery was reported by 27 (1.8%) patients, all of whom had evidence of either extra-hepatic or both intra and extra-hepatic disease at cholangiography. An additional 12 patients had undergone liver resection (indication not known) and 8 had a history of undergoing a Whipple’s procedure for suspected carcinoma of the pancreas.

Table 3.4. Distribution of malignancy in the cohort.

Malignancy	Type of malignancy	Number
Gastro-intestinal*	Colorectal cancer	61
	High grade dysplasia	13
Skin¶	Malignant melanoma	12
	Basal cell carcinoma	6
	Unknown	21
Hepatobiliary	Cholangiocarcinoma	13
	Gall bladder cancer	6
	Hepatocellular carcinoma	4
	Leiomyosarcoma	1
Urological	Prostate cancer	8
	Renal cell cancer	7
	Bladder cancer	5
	Testicular tumour	3
Breast	-	22
Haematological	Non-Hodgkin's lymphoma	5
	Chronic lymphocytic leukaemia	2
	Post-transplant lymphoproliferative disorder	1
Gynaecological	Cervical cancer	2
	Endometrial cancer	2

*IBD present = 67. ¶ Liver transplant recipients = 16

Cholangiography

Cholangiography consistent with PSC was reported in 866 cases. Details regarding cholangiography findings from clinicians were available for 791 patients (ERCP = 274; MRCP = 517). Intra-hepatic disease was present in 405/791 (51%), extra-hepatic in 28 (3.5%) and both intra-hepatic and extra-hepatic in 281 (35.5%). Cholangiography was normal in 77 (10%), 56 of whom had liver histology compatible with PSC (i.e. small-duct PSC). Liver histology was not available for the remaining 21 patients.

Liver histology consistent with PSC was present in additional 225 patients, in whom reliable cholangiography data were not available (comprising 117 diagnostic liver biopsies and 108 liver explants).

UDCA therapy

Data on the use of UDCA were available in 1294 (90%), while 895 patients were also able to provide information on dose. 1022 (79%) patients were receiving or had been on UDCA treatment (until liver transplantation) and 272 (21%) patients were not on medication. The median dose of UDCA was 1000 mg daily (range 150 - 4000 mg) and the median dose corrected per kg body weight was 12 mg/kg/day (range 2 - 46 mg/kg).

For those with underlying IBD (n = 954), 682 of 866 for whom data were available (79%) were treated with UDCA. After excluding patients with IBD who had undergone colectomy, only 530 (75%) of those were treated with UDCA; 121 (17%) were not on UDCA and data were unavailable for 63. The proportion of patients who could not take UDCA because of side effects is not known.

Liver transplantation

337 (23.5%) patients were liver transplant recipients and 28 (2%) underwent a second transplant.

Liver transplantation and time to event analysis

Brief Introduction

A proportion of patients with PSC, develop advanced chronic liver disease and ultimately require liver transplantation 12 to 17 years following diagnosis. Liver transplantation remains the only definitive treatment for advanced PSC (33). However, it is not possible currently to identify patients at risk of developing more advanced disease at an early stage and there have been few large prospective studies examining prognosis or a susceptible clinical phenotype (113, 155, 156). Furthermore, most descriptive studies derive from single centres or tertiary referral institutions with small sample size and so are subject to referral bias. A recent retrospective study identified an association

between the reduction in serum alkaline phosphatase to < 1.5 upper limit of normal with improved liver-related outcome (based on liver decompensation, transplantation and liver-related deaths) in 139 PSC patients (157). Our aim was to identify those risk factors that might be associated with an increased risk of developing end stage liver disease necessitating liver transplantation.

Methods and statistical analysis

We performed time-to-event analysis using the Kaplan-Meier method and Cox proportional hazards regression analysis. The following variables were analysed: age at diagnosis; gender; symptoms at diagnosis; smoking; presence of IBD; ulcerative colitis; coeliac disease; autoimmune disease; UDCA dose and cholangiographic changes. Involvement of first order bile ducts (right or left main hepatic duct) and/or common bile duct at cholangiography allowed distinction between extra-hepatic and intra-hepatic disease. In time-to-event analysis, the start-point was defined as the date of diagnosis of PSC; the end-point was liver transplantation and the censor point was the end of the study period (01 February 2012). A complete dataset for all the variables was available for 362 patients and the time to event analysis was undertaken for this sub-cohort. Univariate analysis using the Cox proportional hazards regression was performed to identify variables associated with transplant-free survival. All variables with a *p*-value < 0.10 were taken forward for multi-variable analysis using a Cox-proportional hazard regression model, to determine the independent prognostic value of each variable. Cox proportional regression analysis was performed using the package 'Survival' in R 2.15.2. Kaplan-Meier survival plots were generated for variables associated (*p*-value < 0.05) with transplant-free survival using Graphpad Prism version 4.0. For survival analysis by UDCA dose, patients were divided into two sub-groups based on a dose above or below 15mg/kg/day.

Results

20% of patients (72/362) were liver transplant recipients. Characteristics of the cohort are shown in *Table 3.5*. Variables associated with a shorter transplant-free survival in univariate analysis included: cholangiography (p = 0.0008; hazard ratio (HR) = 2.28, 95% CI, 1.40 - 3.69) and UDCA dose (mg) (p = 0.0007; HR = 0.92, 95% CI, 0.88 - 0.96) (*Table 11*). Multivariate analysis using a Cox-proportional hazard regression model identified clarify intra-hepatic and extra-hepatic disease on cholangiography (p = 0.0005) and clarify UDCA dose (p = 0.001) as independent risk factors associated with shorter transplant-free survival (*Table 3.6*).

Table 3.5. Characteristics of the sub-cohort and time to event analysis (n = 362).

Variable	Number (%)
Gender	Male = 231 (64%). Female = 131 (36%).
Median age at diagnosis	46 years (10 - 80)
Liver transplant recipient	72 (20%)
Cholangiography changes	Intra-hepatic – 205 (56.5%) Intra- and extra-hepatic – 157 (43.5%)
Inflammatory bowel disease	n = 264 (73%) Ulcerative colitis – 226 (86%) Crohn's disease – 34 (13%) Indeterminate – 4 (1%)
Non-smoker	240 (66%)
Active or ex-smoker	122 (34%)
Autoimmune disease	47 (13%)

Table 3.6. Risk factors associated with transplant-free survival (n = 362).

Variable	Univariate		Multivariate	
	p-value	HR (95% CI)	*p*-value	HR (95% CI)
Age (years)*	0.28	0.99 (0.97 – 1.00)		
Gender (M/F)	0.25	0.73 (0.43 – 1.24)		
Symptoms (Yes/No)	0.12	1.62 (0.87 – 3.02)		
Non-smoker (Yes/No)	0.07	1.68 (0.95 – 2.97)	0.26	1.38 (0.77 – 2.48)
Autoimmune disease (present/absent)	0.35	1.39 (0.69 – 2.82)		
IBD (present/absent)	0.61	1.15 (0.66 – 1.99)		
Ulcerative colitis (present/absent)	0.34	1.28 (0.76 – 2.16)		
Crohn's disease (present/absent)	0.08	0.35 (0.11 – 1.13)	0.16	0.43 (0.13 – 1.39)
UDCA dose (mg)¶	**0.0007**	**0.92 (0.88 – 0.96)**	**0.001**	**0.92 (0.88 – 0.97)**
Cholangiography (IH+EH/IH)	**0.0008**	**2.28 (1.40 – 3.69)**	**0.0005**	**2.34 (1.44 – 3.8)**

IH+EH – intra-hepatic and extra-hepatic disease; IH – intra-hepatic disease only; HR = hazard ratio; CI = confidence interval). (*Median age for this cohort = 46 years (10 – 80 years); ¶ Median dose = 1000 (150 – 4000 mg).

The significance of cholangiography changes and UDCA dose were then evaluated in the entire cohort where data were complete, using Kaplan-Meier survival analysis. Data on cholangiography and time to liver transplantation or the censor point were available in 587 patients. Of 250 (42.5%) with both intra-hepatic and extra-hepatic disease, 74 (29.5%) patients were transplant recipients, in comparison to 63 of 337 (18.5%) with disease confined to intra-hepatic ducts. The median survival in the two groups was 17 years and 29 years respectively (log-rank p = 0.002; HR = 2.1, 95% CI, 1.3 – 3.3) (*Figure*

3.4). Data on the dose of UDCA and time to transplant or to the censor point were available in 775 patients. Patients were divided into two sub-groups based on the prescribed dose of UDCA above or below 15mg/kg/day. 69% (n = 533) patients were on < 15mg/kg/day. 20% of patients taking < 15mg/kg/day of UDCA needed a transplant in comparison to 11% taking ≥ 15mg/kg/day of UDCA (log-rank *p*-value = 0.004; HR = 2.16, 95% CI, 1.23 - 3.25) (*Figure 3.5*).

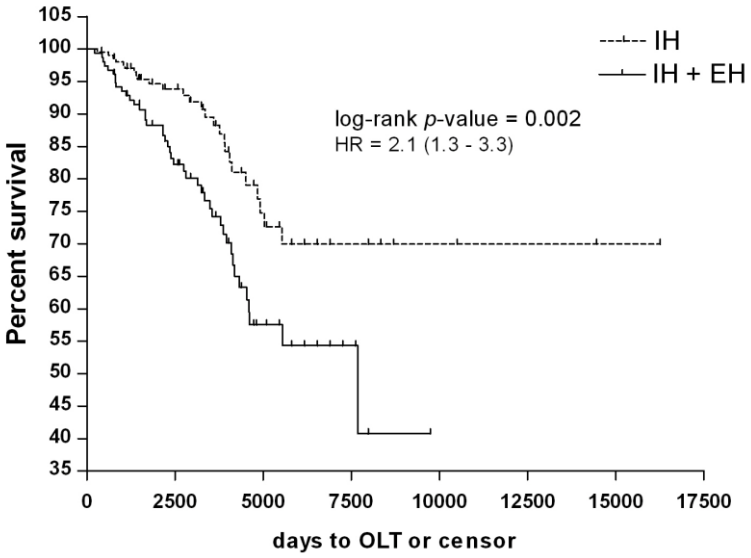


Figure 3.4. Kaplan-Meier transplant-free survival curves for patients with both intrahepatic (IH) and extrahepatic disease (EH) and those patients with isolated intrahepatic disease. (HR: hazard ratio (95% confidence interval)).

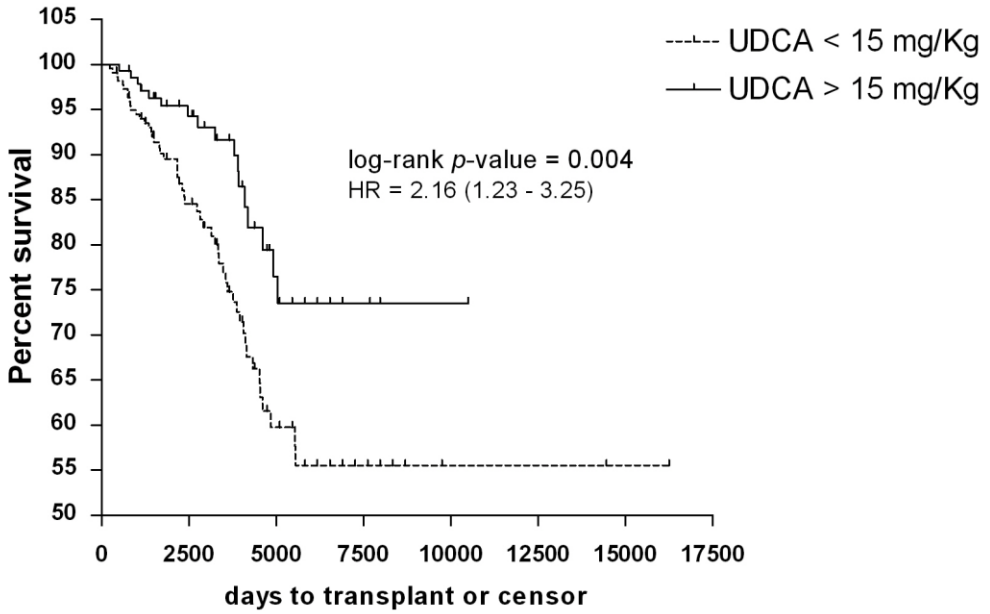


Figure 3.5. Kaplan-Meier transplant-free survival curves by UDCA dose (< 15 mg/kg/day or ≥ 15 mg/kg/day).

(UDCA: ursodeoxycholic acid; HR: hazard ratio (95% confidence interval)).

Discussion

All patients included in the final analyses were recruited according to rigorous standard criteria; all were under regular hospital follow-up until the censor point or liver transplantation and diagnostic accuracy was confirmed. This represents the largest reported series of patients with PSC that includes demographic and phenotypic characteristics. In agreement with previous series, PSC was more common in young, non-smoking males and 66.5% of the cohort had IBD, which was most often ulcerative colitis. The diagnosis of IBD preceded the onset of PSC by a median 15 years. The known protective effect of smoking against ulcerative colitis was maintained in the PSC population. In contrast to ulcerative colitis, smoking is a recognised risk factor for Crohn's disease (158), but in this cohort defined first by the diagnosis of PSC, more than half of the patients with Crohn's disease were non-smokers; a protective effect of smoking in Crohn's disease could not be established, but the number of such patients in the cohort was small. Of note,

26% patients were asymptomatic with the remainder having mainly non-specific symptoms or fatigue with or without pruritus at presentation. The non-specific nature of the presenting symptoms makes it difficult to make an early diagnosis of PSC especially in general practice. As a consequence, a significant proportion of patients with PSC are identified later in the course of their disease, sometimes with features of established chronic liver disease.

The increased prevalence of autoimmune diseases in patients with PSC was confirmed. IBD is also associated with an increased prevalence of other immune-mediated diseases. However, patients with ulcerative colitis and PSC in this series did not have an increased prevalence of autoimmune disorder compared to PSC without ulcerative colitis and the spectrum of autoimmune disease identified was more characteristic of the associations described with PSC rather than those described in association with ulcerative colitis, suggesting that the observed increased prevalence of autoimmune disorder in the current series was related to PSC. It is plausible that PSC shares key immunological risk factors and disease pathways with other immune-mediated diseases, which are also prevalent in the PSC population.

Colorectal cancer was the malignancy reported most commonly found in association with PSC, but was mainly confined to patients with underlying IBD, consistent with the increased prevalence of colorectal cancer observed in patients with both PSC and ulcerative colitis in this series. Only 13 patients in this series reported a history of cholangiocarcinoma at recruitment. Cholangiocarcinoma has an extremely poor prognosis and in a recent study the median time from diagnosis of cholangiocarcinoma to death was 6 months (159). All the patients recruited were under active hospital follow-up for PSC. It is possible that a sub-set of PSC patients with cholangiocarcinoma that presented soon after a diagnosis of PSC could not be captured during recruitment for this study because of the poor prognosis and the likelihood that such patients would be less likely to be recruited to such studies. Thus, the observed prevalence rate of cholangiocarcinoma in our cohort is unlikely to be a true representation of its prevalence in PSC. Likewise, a direct comparison of rates of malignancy with a control population was not possible in this study.

Colectomy, followed by cholecystectomy, was the surgical procedure undertaken most often and confined largely to those with PSC and IBD. The age-standardised prevalence of cholecystectomy in the general UK population is 2.6% (females = 3.9%; males = 1.2%) (160), suggesting an increased cholecystectomy rate in this male predominant series. The indications for cholecystectomy in this cohort are unknown, but given the increased prevalence of gallstones in PSC (73), it is likely to have been the most common indication.

Cholangiography (ERCP or MRCP) remains the gold standard for the diagnosis of PSC and allows differentiation between intrahepatic and extrahepatic disease. A proportion of patients with PSC develop chronic liver disease necessitating liver transplantation eventually. Few risk factors for severe or progressive disease in PSC are known and there are no validated prognostic models for PSC. Patients with small duct PSC have a better prognosis than those with larger duct disease and those with an elevated bilirubin are more likely to develop advanced liver disease (45, 113).

In this series patients with both intrahepatic and extrahepatic disease had a median transplant-free survival of 17 years in comparison to 29 years for those in whom disease was restricted to the intrahepatic biliary tree. The transplant-free survival in this series was longer than that reported previously and was even more marked in those with disease limited to intra-hepatic ducts, consistent with recent studies showing a prolonged liver related end-point free survival in PSC (7, 15, 157). Al Mamari *et al*. also identified an association between disease confined to the intra-hepatic ducts and sustained improvement in the serum alkaline phosphatase, which in turn was associated with an improved liver related outcome (157). One possible explanation is that patients with extensive bile duct strictures have more marked intra-hepatic cholestasis, which in itself may exacerbate inflammation and fibrosis. An increased clinical focus on this high-risk group might be beneficial.

UDCA is the only available treatment for patients with PSC and has been shown to improve both symptoms and liver function tests (100, 161). One pilot study showed improved biochemistry and improved predicted survival at 4 years (161), but in contrast, several studies have failed to demonstrate efficacy of UDCA at either the standard dose or a higher dose (> 15mg/kg/day) (100). Current EASL guidelines recommend UDCA in selected patients while AASLD guidelines do not recommend UDCA (19, 162). It has been shown that cholestasis reduces UDCA mediated enrichment of bile acid pool and higher doses may be required to maintain the same level of enrichment (163, 164). Thus, we divided our patients in to those receiving the standard dose (< 15mg/kg/day) or a higher dose (≥ 15mg/kg/day) of UDCA. In this large series involving most UK centres and analysed retrospectively, there was a dose-dependent benefit of UDCA therapy on transplant-free survival at ≥ 15 mg/kg/day. It must be noted that higher doses of UDCA (28 - 30 mg/Kg/day) are reported elsewhere to be associated with increased mortality, although cause and effect are unproven (99).

This study has several limitations. It was an observational study and data were collected retrospectively, while the demographic and phenotypic characterisation was extracted from well-designed participant and clinician questionnaires. Further, clinician questionnaires were incomplete in a small proportion of patients and data were not available regarding the duration of either UDCA therapy or the biochemical response to treatment. Finally, liver biochemistry as a prognostic factor was not assessed and IgG4 levels were available in too few patients to allow meaningful analysis. Thus, the study team felt there was a need to collect detailed clinical data, which was subsequently undertaken (discussed in chapter 4; page 81).

In conclusion, the phenotypic characteristics of a large, comprehensive cohort of UK patients with PSC were established and we identified two clinical factors associated with liver transplant-free survival, which may have a role for risk stratification in patients suffering with PSC.

Chapter 4: Factors associated with clinical outcome and the development of a risk scoring system for patients with PSC.

For this chapter, I designed the clinical questionnaire, recruited patients and managed the clinical database to capture the additional clinical data.
The data included in this chapter was analysed by Dr Elizabeth Goode (Clinical research fellow, UK PSC study) and Dr AB Clark (Senior lecturer, University of East Anglia).

Introduction

Patients with PSC run an unpredictable disease course, with liver transplantation as the only definitive treatment option. Individualised clinical risk prediction is lacking, especially during the early course of the disease due to various limitations of the existing risk scores as outlined in chapter 1 (page 46).

The phenotypic characteristics discussed in the previous chapter were obtained from participant and clinician questionnaire, but did not include detailed biochemical and clinical outcome data. This study was an extension of the phenotypic study of PSC patients in the UK and sought to collect more complete biochemical and other relevant clinical data from patients recruited in the UK PSC study, to help identify clinical risk factors predicting outcome and to develop and then validate a clinical risk scoring system.

Study Design

Patients recruited to the UK PSC study, ≥ 18 years of age, with PSC incident or prevalent between1st August 2008 and 31st March 2015, including those who had undergone liver transplantation prior to 31st March 2015, were included in the study.

Inclusion and Exclusion criteria were as described previously in Chapter 2 (page 59). Patients were classified as having extrahepatic disease based on the criteria described in chapter 3 (page 63).

Data Capture

Data were collected onto pre-specified questionnaires (*Appendix 3*), through review of case notes and included patient demographics, diagnostic cholangiography, histology reports, haematology and biochemical blood test results including serial liver biochemistry at diagnosis (t_o), 1-year following diagnosis (t_1) and 2-years following diagnosis (t_2), IBD status, use of UDCA, development of liver decompensation and progression to liver transplantation or death. Incomplete questionnaires or missing data received, were systematically queried with the respective research nurse or clinician at the participating site, to ensure accurate and complete capture of data.

Study endpoints

Date of diagnosis (t_0) was defined as the date of first cholangiography or liver histology confirming a diagnosis of PSC. The main primary endpoint of the study was liver transplantation. The second primary endpoint was all-cause mortality (to encompass all liver related deaths). Time from diagnosis of PSC to study endpoint was calculated. Patients were censored at their most recent follow-up or blood test if they did not reach an endpoint.

Variables evaluated for risk score

The variables were included based upon clinical relevance or prior reported association. These included alkaline phosphatase (ALP), haemoglobin (Hb) g/l, platelet count (Plts) x 10^9/l, serum albumin (Alb) g/l and bilirubin (Bili) μ mol/l.

Data analysis

Descriptive statistics were calculated and reported as numbers or percentages. Variables with > 40% missing data were excluded from further analysis. This cut-off was chosen arbitrarily to allow inclusion of certain clinically relevant parameters (e.g. ALT). Time-to-event analysis was conducted using Cox's proportional hazards model, ensuring at least 10 events per risk factor included in the model. Events were truncated at 10 years follow-up to facilitate accurate prediction of risk. Categorical variables were only considered if the categories had > 5% of the cohort in each category.

Univariate analysis was performed in the raw dataset to show association between risk factors and outcome. Where data were missing, multivariable imputation was performed using iteratively chained equations. Results of ten imputed data sets were combined using Rubin's equation and used to estimate the multivariate model. Method of backward elimination was used to select variables for the final risk score, with removal of risk factors not significant at the 10% level. Continuous variable were assessed for non-linear association using cubic splines. Variables demonstrating a linear association were included in a standard continuous fashion and those showing a non-linear association were categorised using cubic splines and clinical judgment to allow for ease of interpretation.

Association with alkaline phosphatase (ALP)

Association between ALP at t_1 and t_2 with outcome was analysed to determine the optimal threshold for predicting 10-year hazard of outcome. ALP was divided into categorical variables from </≥0.5 to 4xULN, with increments of 0.1. Each ALP cut-off was plotted against the hazard of reaching an endpoint. The optimal threshold for ALP was determined using Harrell's C statistic.

Derivation of UK PSC risk scores

We derived three separate risk scores, to determine the model with the best discrimination. The first was a score using t_0 data to predict 10-year risk of outcome, the second a short-term risk score using t_0 data to predict two-year risk of outcome (RS_{ST}) and the third, a long-term risk score using t_0 and t_2 data to predict 10-year risk of outcome (RS_{LT}). The RS_{LT} included only those patients who did not reach a primary endpoint within 2 years of diagnosis. Harrell's C statistic was used to compare discrimination of each risk score. Calibration curves for RS_{ST} and RS_{LT} were generated by creating deciles of data and comparing the model's predicted survival rates with the observed rates in the cohort, estimated by Kaplan-Meier (KM) curve.

Independent validation of UK PSC risk scores

An independent data set from two external PSC patient cohorts, not included in the original analysis was used to validate the UK PSC risk scores. The first was a national validation cohort, n = 352, from two UK hospitals (John Radcliffe Hospital, Oxford and Queen Elizabeth Hospital, Birmingham). The second validation cohort was from Norway (n = 99). Validation of the scoring system was performed by fitting a Cox-model to the validation cohort using the scoring system derived from the derivation cohort. Further visual validation was performed by displaying Kaplan-Meier (KM) survival curves for four separate risk groups in both cohorts. Risk groups were defined by dividing the derivation cohort into four equal sized groups with increasing RS_{LT} and the validation cohort divided into four groups according to the same RS_{LT} categories.

Comparison of the UK PSC score with existing scores

Two existing risk scores were compared with the UK PSC risk score in both derivation and validation cohorts: revised Mayo risk score and AST:platelet ratio index (APRI) score using Harrell's C statistic.

Statistical analysis

All analyses were performed using Stata software (version 14.0/SE; StataCorp LP, College Station, Tx).

Results

Cohort characteristics

1749 patients were recruited to the UK PSC cohort; 1252 questionnaires distributed and 1131 returned. 130 were excluded following quality control, leaving 1001 patients for analysis. 57% of the patients were recruited from non-transplant centres. Table 4.1 shows the cohort characteristics.

Table 4.1. Demographics of the UK PSC (derivation) cohort and national and international validation cohorts.

Parameter	Variable	UK cohort (n = 1001) %	Validation cohorts	
			National (n = 352) %	International (n = 99) %
Demographics	Male	63.8	62.4	75.7
	Median age at diagnosis (yrs)	46.8	45	35
	Median age at transplant (yrs)	47	47	39
	Median follow-up (yrs)	14.8	6	8
Disease distribution	Extra-hepatic biliary disease present (%)	44.1	47.8	33.3
IBD	IBD	72.5	71.0	86.0
	UC	80.4	73.6	77.6
	CD	14.2	10.7	15.3
	IC	5.4	3.2	7.1
Events	Total events	35.7	39.2	32.3
	Transplants	27.8	13.9	11.1
	Deaths (all-cause)	7.9	25.3	21.2

IBD – Inflammatory bowel disease; UC – ulcerative colitis; CD – Crohn's disease; IC – indeterminate colitis

Study endpoints

35.7% of patients reached a primary endpoint over a cumulative follow-up period of 7,904 years. 27.8% of the cohort underwent OLT at a median age of 47 years. 7.9% of the cohort died without a transplant; 47.8% of all deaths were PSC-related. The overall proportion of the cohort that was event-free at 2, 5 and 10 years was 92%, 82% and 64% respectively. 39% of men reached an outcome, compared to 29% of females (p-value = 0.002). 43% of those with extra-hepatic biliary disease reached an outcome compared to 23% of those without (p-value = 0.001). Patients with extra-hepatic biliary disease had a reduced median transplant-free survival compared to those without extra-hepatic disease (11.7 versus 23 years). UDCA use in the first 2 years following diagnosis was not associated with outcome.

Serum alkaline phosphatase is associated with PSC outcome

ALP data at t_1 and t_2 was available for 72% and 70% of the cohort respectively. At both t_1 and t_2, raised ALP was associated with an increased 10-year hazard of reaching an outcome (p-value < 0.001) (figure 4.1a and 4.1b). There was a log-linear association between serum ALP and outcome. At t_1 the optimal threshold for predicting 10-year outcome was ALP $\geq$ 2.4 x ULN (HR = 3.05, C = 0.63) (supplementary figure 4.1a). Median transplant-free survival at t_1 for those with ALP $\geq$ 2.4 x ULN was 63 versus 108 months for those with ALP < 2.4 x ULN (p-value < 0.0001 (log-rank test)) (figure 4.1c). At t_2, the optimal threshold for predicting 10-year outcome was ALP $\geq$ 2.2 x ULN (HR = 3.05, C = 0.66) (supplementary figure 4.1b). Median survival for those with a t_2 ALP $\geq$ 2.2 x ULN was 44 versus > 96 months for those with a t_2 ALP < 2.2 x ULN (p-value < 0.0001 (log-rank test)) (figure 4.1d).

Disease distribution is associated with outcome

Cholangiographic data at t_0 were available in 87.2% of the cohort. The presence of extrahepatic biliary disease was associated with an adverse

outcome (HR = 1.45 (95% CI = 1.09 – 1.92); p-value = 0.010). However, > 50% of patients in both groups (intrahepatic and extrahepatic) were event free at 10 years and thus median survival was not reached.

Derivation of a UK PSC risk score

UK PSC risk score was designed using clinical variables available at diagnosis (t_0) to predict 10-year risk of outcome. Following multivariate analysis, seven variables were included in the score: age at t_0, serum bilirubin, ALP, albumin, haemoglobin, platelet count and presence of extra-hepatic biliary disease at t_0 (supplementary table 4.1). The C statistic was 0.78 with shrinkage of 0.94. The event rate within the cohort in the first two years of PSC diagnosis was high (8%). Therefore, to determine if variables predicting short-term and long-term risk were different, two different risk scores were developed as follows:

a) **Short-term risk score (RS_{ST})**: clinical variables at t_0 were analysed to predict risk of outcome within 2 years following diagnosis. The best fitting Cox model included four variables – bilirubin, albumin, haemoglobin and platelet count (table 4.2). Based on these coefficients, a prognostic model was developed to predict risk of death or liver transplantation by year 2 (C = 0.81, shrinkage = 0.92) (supplementary Box 4.1).

b) **Long-term risk score (RS_{LT})**: clinical variables at t_2 were analysed to predict 10-year risk of outcome. The best fitting Cox model included seven variables: age at t_0, presence of extra-hepatic biliary disease at t_0, serum bilirubin at t_2, ALP at t_2, albumin at t_2, platelet count at t_2 and variceal haemorrhage at t_2 (C = 0.80, shrinkage = 0.96) (table 4.3 and supplementary Box 4.1).

Both RS_{ST} and RS_{LT} demonstrated improved predictive ability over the original risk score model. Calibration of RS_{ST} and RS_{LT} using predicted versus

observed survival rates estimated by the Kaplan-Meier curve demonstrated good correlation.

Table 4.2. Univariate analysis of factors at diagnosis associated with 2-year risk of transplant or death (using un-imputed data and multivariate analysis using 10 imputed data sets).

	Univariate Analysis		Multivariate Analysis	
Factor	**HR (95% CI)**	**p-value**	**HR (95% CI)**	**p-value**
Female	0.88 (0.54,1.42)	0.596		
Age at diagnosis	1.01 (1.00,1.03)	0.126		
Extra-hepatic disease	1.30 (0.77,2.21)	0.332		
IBD presence	1.09 (0.49,2.44)	0.832		
UC				
CD				
IC				
Autoimmune disease	0.90 (0.46,1.75)	0.757		
Smoker	1.22 (0.74,2.02)	0.426		
Bilirubin (μ mol/l)				
35 -49	4.03 (1.36,11.98)	0.012	2.11 (0.74,5.96)	0.159
50+	14.12 (7.89,25.3)	<0.001	5.02 (2.76,9.13)	0.000
ALP (ratio of ULN)				
1.5 - <2.5	1.25 (0.49,3.17)	0.634		
2.5+	2.64 (1.35,5.17)	0.005		
ALT (IU/l)	1.02 (0.98,1.05)	0.331		
Albumin (g/l)	0.87 (0.84,0.90)	<0.001	0.94 (0.90,0.99)	0.011
Haemoglobin (g/l)	0.79 (0.71,0.89)	<0.001	0.99 (0.97,1.00)	0.095
Platelets group ($\times 10^9$/l)				
150 - 199	0.23 (0.08,0.72)	0.011	0.62 (0.26,1.48)	0.283
200 - 399	0.22 (0.11,0.45)	<0.001	0.50 (0.25,0.98)	0.045
400+	0.32 (0.13,0.78)	0.012	0.38 (0.15,0.98)	0.046
Eosinophils ($\times 10^9$/l)	1.10 (0.89,1.36)	0.368		
Sodium (mmol/l)	0.89 (0.82,0.98)	0.015		
Creatinine >120 (µmol/l)	4.21 (1.66,10.68)	0.002		
IgG (g/l)	1.08 (0.93,1.25)	0.313		

Table 4.3. Univariate analysis of factors at 2 years associated with 10-year risk of transplant or death (using un-imputed data and multivariate analysis using 10 imputed data sets).

	Univariate Analysis		Multivariate Analysis	
Factor	**HR (95% CI)**	**p-value**	**HR (95% CI)**	**p-value**
Female	0.81 (0.60,1.10)	0.181		
Age at diagnosis	1.01 (1.00,1.03)	0.005	1.03 (1.01,1.04)	<0.001
Extra-hepatic disease	1.95 (1.42,2.69)	<0.001	1.70 (1.15,2.48)	0.008
IBD	0.91 (0.59,1.38)	0.646		
Autoimmune disease	1.27 (0.88,1.83)	0.200		
Smoker	0.96 (0.70,1.32)	0.790		
Bilirubin (μ mol/l)				
35 -49	6.77 (3.87,11.85)	<0.001	3.31 (1.65,6.62)	0.001
50+	7.92 (5.62,11.18)	<0.001	3.96 (2.37,6.62)	<0.001
ALP (ratio of ULN)				
1.5 - 2.4	1.75 (0.98,3.15)	0.061	1.50 (1.09,2.30)	0.015
2.5+	1.40 (1.04,1.88)	0.025	1.57 (1.12,2.52)	0.011
ALT (IU/l)	1.05 (1.03,1.08)	<0.001		
Albumin (g/l)	0.88 (0.85,0.90)	<0.001	0.93 (0.90,0.96)	<0.001
Haemoglobin (g/l)	0.75 (0.69,0.81)	<0.001		
Platelets group ($\times 10^9$/l)				
150 - 199	0.35 (0.20,0.60)	<0.001	0.58 (0.31,1.10)	0.092
200 - 399	0.29 (0.20,0.43)	<0.001	0.60 (0.40,0.91)	0.016
400+	0.32 (0.17,0.60)	<0.001	0.46 (0.23,0.92)	0.028
Eosinophils ($\times 10^9$/l)	0.81 (0.52,1.29)	0.380		
Sodium (mmol/l)	0.90 (0.96,0.93)	<0.001		
Creatinine >120 (µmol/l)	0.66 (0.21,2.07)	0.474		
IgG (g/l)	1.01 (0.92,1.12)	0.774		
UDCA use	0.96 (0.72,1.28)	0.795		
Variceal bleed by yr 2	5.97 (2.93,12.16)	<0.001	2.76 (1.14,6.66)	0.024

Defining risk groups based on long-term risk score (RS_{LT})

In order to define low and high-risk disease groups according to RS_{LT}, we divided the cohort into four equal quartiles. Event-free survival, plotted on a Kaplan-Meier survival curve (figure 4.2a) demonstrated an observed event rate of 6.0%, 8.4%, 19.1% and 55.8% in the four respective risk groups. Curves for the four risk groups were generally well separated, although the

model was less able to distinguish between the two lowest risk groups. RS_{LT} defining the four risk groups is shown in table 4.4.

Table 4.4. Event rates in four groups defined according to long-term risk score in the derivation and validation cohorts.

		Derivation cohort		Validation cohort	
Risk group	Score (RS_{LT})	No. of individuals	Observed event rate (%)	No. of individuals	Observed event rate (%)
1	$RS_{LT} < -2.019$	216	13 (6.0%)	105	3 (2.9%)
2	$-2.019 < RS_{LT} < -1.463$	215	18 (8.4%)	77	8 (10.4%)
3	$-1.463 < RS_{LT} < -0.814$	215	41 (19.1%)	60	12 (20.0%)
4	$-0.814 < RS_{LT} < 2.737$	215	120 (55.8%)	94	45 (47.9%)

Validation of the UK PSC risk scores

The predictive ability of both risk scores was analysed in a separate UK international cohort. In the respective national and international validation cohorts (table 1), 62.4% and 75.7% of the cohort were male, diagnosed at a median age of 47 and 39 years, with 71% and 86% diagnosed with concomitant IBD. The most notable differences between the two cohorts were the shorter median follow-up (6 and 8 years in the national and international cohorts respectively versus 14.8 years), higher death rate (25.3% and 21.2% versus 7.9%) and lower transplant rate (13.9% and 11.1% versus 27.8%).

Both the RS_{ST} and RS_{LT} were associated with outcome in the national validation cohort (p-value < 0.001). In the international validation cohort, the lack of events within the first two years meant only RS_{LT} could be validated, which was associated with long-term outcome (p-value < 0.001).

Further visual validation of the RS_{LT} was performed by comparing KM survival curves for the validation cohort according to the same four previously defined risk groups as the derivation cohort (figure 4.2b). Event rates were similar to the derivation cohort at 2.9%, 10.4%, 20.0% and 47.9% (table 4.4). Both set of four curves were quite well separated, confirming that the model

had discrimination in both cohorts, however the model was less able to distinguish between the two intermediate risk groups in the validation cohort.

Comparison of UK PSC risk score with existing scores

We compared the predictive accuracy of the revised Mayo and APRI scores to the RS_{ST} and RS_{LT} in the derivation dataset, both of which use AST levels to calculate the risk scores. ALT rather than AST is measured in most biochemistry laboratories in the UK.

Based upon a subset of 170 patients from the validation cohort, for which both AST and ALT measurement were available for t_0 and t_2, there was strong correlation ($r = 0.94$, p-value < 0.0001) and strong concordance ($c = 0.92$, p-value < 0.0001) between the two variables. Therefore, ALT was used instead of AST for calculation of the Mayo and APRI scores in the derivation cohort.

In predicting 2-year outcome, the RS_{ST} out-performed the APRI and revised Mayo scores with C statistics of 0.81, 0.63 and 0.75 respectively. In predicting 10-year outcome the RS_{LT} demonstrated an incremental improvement over the APRI and Mayo scores with statistics of 0.80, 0.59 and 0.79 respectively. Similar observations were noted in the validation cohort as well.

Discussion

The UK PSC cohort is a large ethnically homogeneous cohort of patients recruited from both transplant and non-transplant centres across the UK, thus representing the full spectrum of PSC disease severity. This study has identified clinically relevant variables associated with disease outcome.

Akin to PBC, we have confirmed the importance of ALP as a prognostic indicator in PSC. ALP level < 2.4 x ULN and < 2.2 x ULN at 1 and 2 years following PSC diagnosis, is associated with improved transplant-free survival.

Significance of ALP level at the time of PSC diagnosis remains unclear, as it was not associated with short-term outcome. This may partly be due to the fluctuations in ALP seen at time of diagnosis (e.g. due to untreated cholangitis).

In addition, I have provided further confirmation of an association between extrahepatic biliary disease and reduced transplant-free survival, which we found on the initial analysis (as discussed in chapter 3). This highlights the importance of accurate assessment of disease extent on cholangiography (by radiologists) at the time of PSC diagnosis to allow early disease risk stratification. There is no clear evidence based guidance on appropriate surveillance strategy in PSC. Based on the findings from this study, it might be possible to identify patients, who present with or develop poor prognostic indicators and thus be monitored closely with regular clinic review and interval follow-up scans.

A high event rate (8%) was observed within the first two years following diagnosis, suggesting there is a subset of patients who present late in their disease course, or who have a rapidly progressive disease. Hence, separate risk scores were developed for short-term (RS_{ST}) and long-term (RS_{LT}) risk prediction. The key differences between the two are the laboratory parameters included for calculating the score. RS_{ST} includes parameters of intrinsic liver function (bilirubin, albumin, haemoglobin and platelet count), which are probably more relevant in predicting immediate outcome.

Conversely, RS_{LT} includes laboratory parameters (bilirubin, albumin, platelet count, ALP) in addition to variceal haemorrhage and cholangiographic disease distribution. This would allow clinicians to recalculate risk at 2 years following diagnosis for greater prognostic accuracy.

In addition, the new PSC risk score out-performed the existing revised Mayo score, with the latter only allowing a 4-year risk prediction of all-cause mortality but not liver transplantation. In comparison, the UK PSC risk score predicts short-term (2 year) and long-term (10 year) risk of liver transplantation in addition to all-cause mortality.

Based on this study and using the RS_{LT} risk score, it is possible to define a low-risk disease group (patients with RS_{LT} score < -2.02 had < 10% chance of an event by 10 years follow-up) and a high-risk group (patients with RS_{LT} score -0.81 < RS_{LT} < 2.74 had ~ 50% chance of an event by 10 years) (table 4.4).

The main limitation of this study is the retrospective nature of case ascertainment and data capture. It is thus inherently biased towards those patients who were transplanted and recruited in to the study in comparison to patients who died or developed cholangiocarcinoma (median survival 4 – 6 months).

In addition, there was a proportion of patients with missing data. Rates of missing data were higher for patients diagnosed many years previously and as it was predominantly related to the year of diagnosis, the data were considered to be 'missing at random', which is defined as any systematic difference between the missing values and the observed values that can be explained by differences in the observed data (165). Thus, imputation was used to improve the validity of the results.

In conclusion, this study was based on a detailed clinical evaluation of a large representative cohort of patients with PSC and has identified clinical parameters that predict their outcome. Further studies are needed in even larger cohorts to validate the findings from this study with greater confidence, in particular the predictive accuracy of the UK PSC risk score.

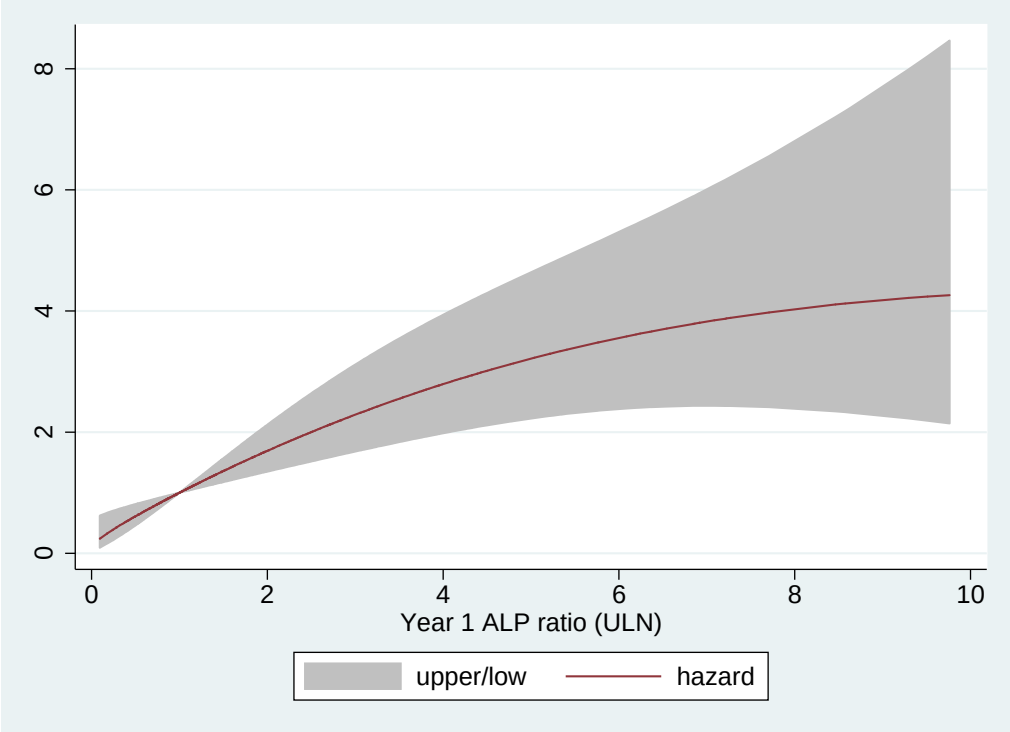


Figure 4.1 (a). Association between alkaline phosphatase (as ratio of ULN) at year-1 and hazard of reaching a clinical endpoint within 10 years, with 95% CI.

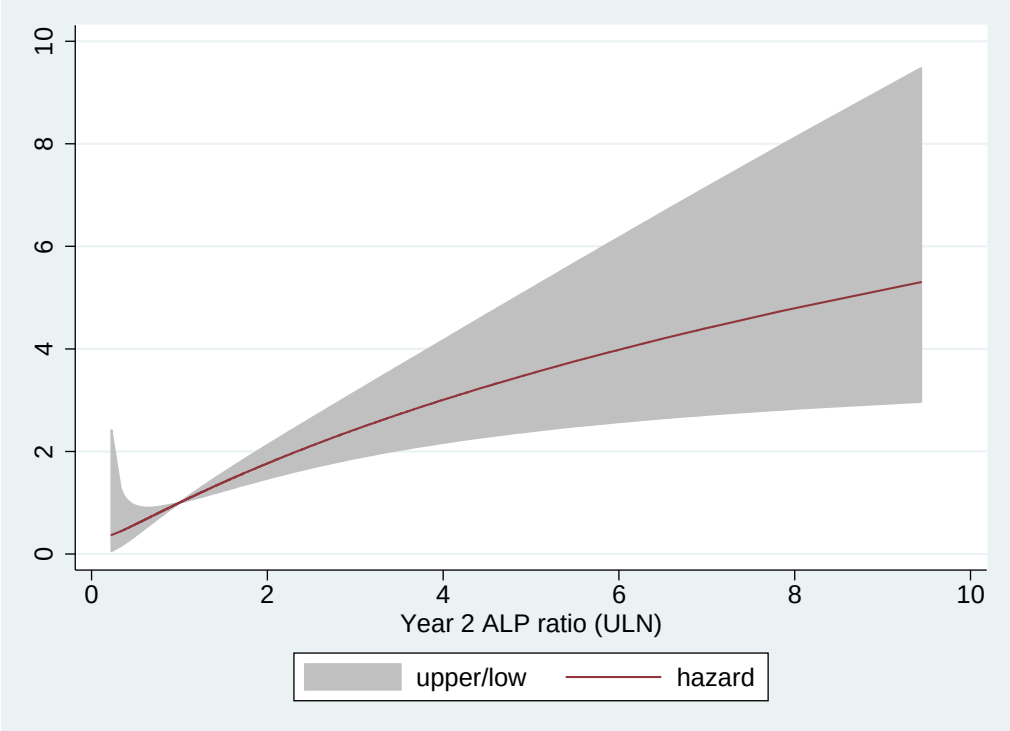


Figure 4.1 (b). Association between alkaline phosphatase (as ratio of ULN) at year-2 and hazard of reaching a clinical endpoint within 10 years, with 95% CI.

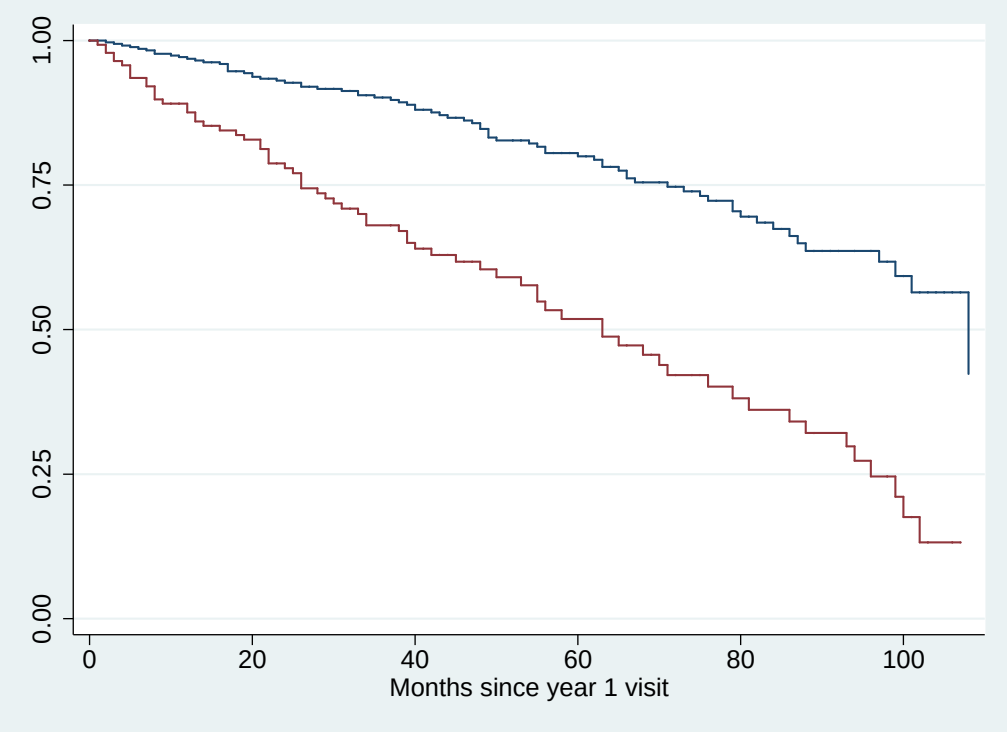


Figure 4.1 (c). Kaplan Meier survival curve for transplant-free survival in patients with ALP ≤ 2.4 x ULN (blue line) versus ALP > 2.4 x ULN (red line) at 1-year following diagnosis (0 = 12 months post diagnosis).

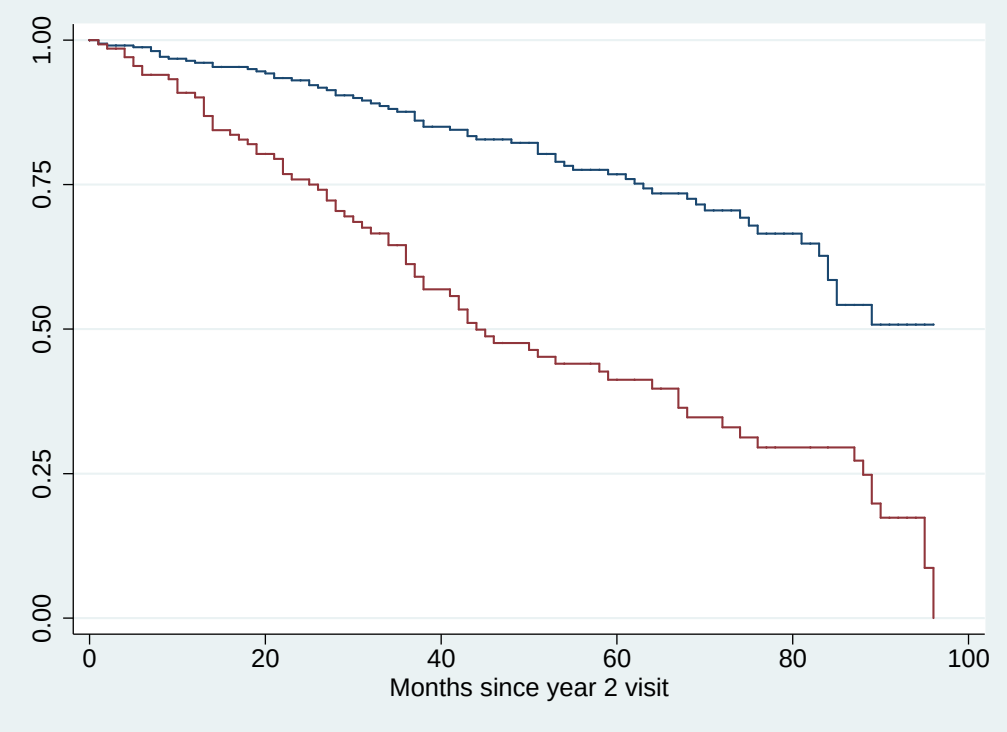


Figure 4.1 (d). Kaplan Meier survival curve for transplant-free survival in patients with ALP ≤ 2.2 x ULN (blue line) versus ALP > 2.2 x ULN (red line) at 2-years following diagnosis (0 = 24 months post diagnosis).

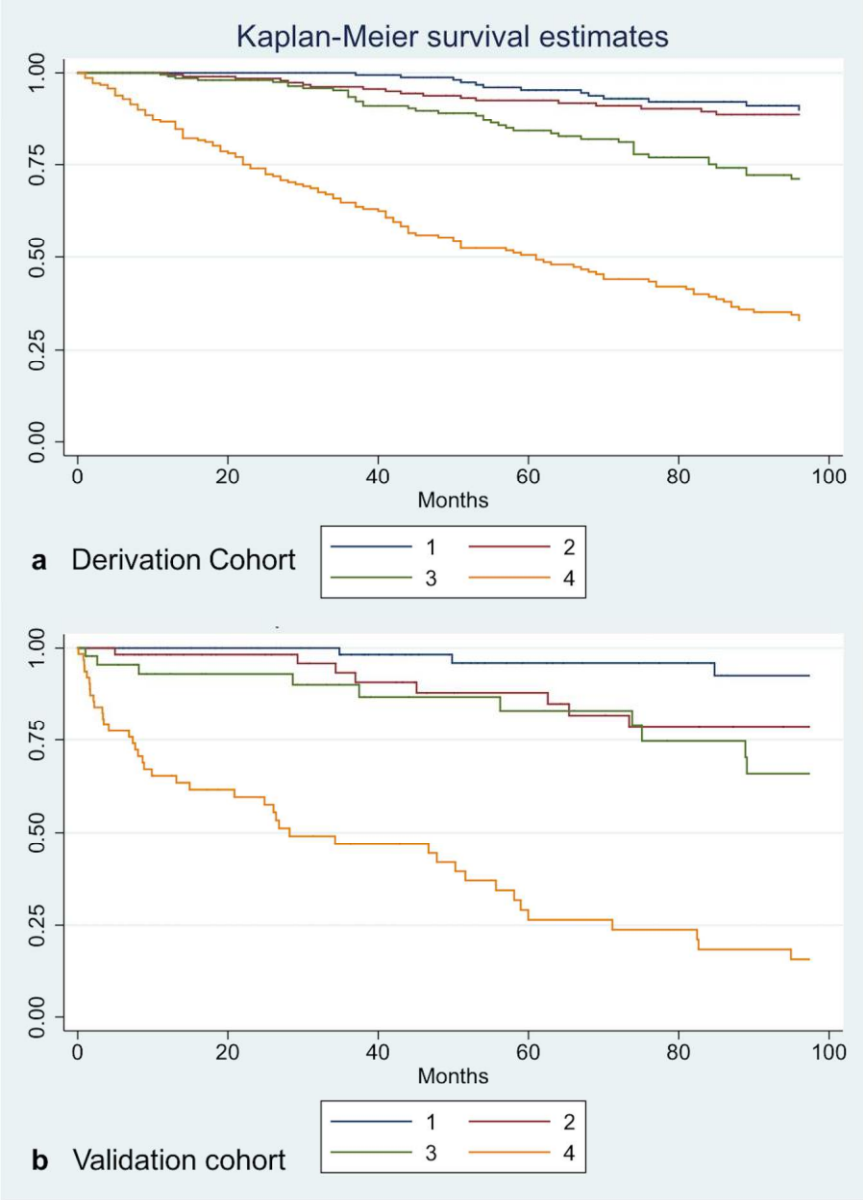


Figure 4.2 (a) Derivation cohort Kaplan Meier survival curves for 4 risk groups.
(b) Validation cohort Kaplan Meier survival curves for 4 risk groups.

Chapter 5: Genetic association studies – key concepts.

This chapter gives a brief introduction to the key terminologies used in subsequent chapters, the concepts underlying study design and statistical analysis, in population based case-control genetic association studies, including its limitations.

Introduction

The main aim of case-control genetic association studies is to identify genetic risk factors associated with various diseases, especially complex disease traits where genetic factors may contribute towards disease development or pathogenesis. The field of bioinformatics is rapidly evolving and offers a variety of technology, study designs and analytical tools to aid genetic studies.

The two most popular study designs for case-control genetic association studies are:

a) **Candidate gene study:** Candidate gene studies are hypothesis-driven, and focus on a particular gene or genetic region, due to compelling evidence or a priory probability of association with the disease.
b) Genome wide association study (GWAS): GWAS are hypotheses free and seek to identify novel genes or genetic regions, associated with the disease.

The most common genetic variation studied is the single nucleotide polymorphism (SNP).

Single Nucleotide Polymorphism

A SNP is a variation at single base-pair position in a DNA sequence occurring at an appreciable frequency (typically greater than 1%) in the population. A SNP has two alleles (genotype) and its frequency is denoted by the frequency

of the less common allele in the general population i.e. minor allele frequency (MAF).

In genetic association studies, a large number of SNPs are genotyped in individuals by using commercially available chip based microarray platforms (e.g. Illumina or Affymetrix). Comparing the genotype frequency at each SNP between cases and healthy controls, allows identification of genetic regions that differ between the two cohorts, thereby implicating genes in these regions in disease pathogenesis. Hence, SNPs are considered to be markers of the genetic region in the genome.

A SNP is usually present in the non-coding region of the genome but can be present in the coding region as well. Commonly occurring SNPs have the ability to highlight potential candidate genes within the associated genetic region, which may have role in disease pathogenesis.

Common disease Common variant hypothesis (CDCV)

Monogenic disorders (e.g. cystic fibrosis) are usually caused by genetic variants occurring within a single gene. By studying the genetic make-up and segregation in families affected by specific diseases using linkage analysis, it was possible to map heritable trait genes to their chromosome locations. Linkage typically tested highly polymorphic markers or SNPs in pedigrees segregating a trait to identify chromosome regions likely to harbor genes for the trait. Sequencing was then used to find the causal variant that was typically rare but had large effect size and high penetrance (166) (167). However, when linkage was applied to more commonly occurring diseases, it failed to identify chromosomal regions associated with such diseases.

The CDCV hypothesis was proposed following the identification of relatively common genetic variants (defined as MAF > 0.01) associated with common diseases; for example, the *apolipoprotein E* (*APOE*) gene for Alzheimer's disease (168).

The hypothesis states that, commonly occurring disorders are likely to be influenced by genetic variants that occur commonly within the same population. An important implication of this hypothesis is in relation to the penetrance (glossary Box 1; page 160) or the effect size of the common variant identified. It is believed that, a common genetic variant will invariably have a small effect size. If a common variant exerts a very strong effect, then the phenotype (disease) will also be equally common in the general population (169). Hence, for commonly occurring diseases in the population, it is likely that many common variants, each with small effect size, exert a combined effect, thereby contributing to a proportion of the disease risk in the population (i.e. heritability – glossary Box 1; page 160).

It is likely that there is a combination of genotypic, environmental and epigenetic interactions in disease causation and such diseases are referred to as complex diseases (e.g. diabetes mellitus and cardiovascular diseases). This led to the emergence of population based, case-control genetic association studies especially for complex disease traits.

The International Human Haplotype Map project

The International HapMap project was established in 2002 to identify commonly occurring genetic variations (MAF ≥ 0.05) across the genome with a view to developing a haplotype map of the human genome, HapMap, to describe the common patterns of human DNA sequence variation (i.e. haplotypes) (170). Four populations were selected for inclusion in the HapMap: 30 adult-and-both-parents Yoruba trios from Ibadan, Nigeria (YRI), 30 trios of Utah residents of northern and western European ancestry (CEU), 44 unrelated Japanese individuals from Tokyo, Japan (JPT) and 45 unrelated Han Chinese individuals from Beijing, China (CHB). Approximately, 10 million SNPs were genotyped as part of this project in 3 different phases across different populations, which act as the reference point. The complete dataset obtained in Phase I was published on 27 October 2005. The analysis of the Phase II dataset was published in October 2007. In Phase III of the

project, a total of 11 global ancestry groups were assembled and their dataset was released in 2009.

Haplotype and concept of Tag SNPs

Sets of SNPs in close proximity on the same chromosome segregate together in blocks. The pattern of SNPs within a block is referred to as the Haplotype, which represents that genetic region. A few SNPs within each block are sufficient to identify the haplotype blocks by virtue of linkage disequilibrium between different markers. These SNPs are referred to as Tag SNPs. Information on common haplotypes allowed identification of a minimum set of Tag SNPs that would allow capture of majority of the genetic variation across the genome. This concept allows reduction in the total number of SNPs needed to be genotyped from 10 million to approximately 500,000 SNPs. Only the tag SNPs are genotyped to cover the entire genome. This was a crucial development, which facilitated the planning of genome-wide association studies in a cost effective manner.

All the resources were made freely available to the public for use in genetic association studies. International HapMap data were used in the UK PSC genetic association studies discussed in Chapters 6 and 7.

As of June 2016, the HapMap resource site has been decommissioned. It has been surpassed by the 1,000 genomes project, which was set up in 2008, to develop a comprehensive catalogue of genetic variants (SNPs) with a frequency of ≥ 0.01. That is now established as the research standard for population genetics and genomics for all future genetic studies.

Linkage Disequilibrium

In a given population, which is undergoing random mating, repeated random recombination events over several generations will break apart contiguous segments of a chromosome. Eventually, over a period of time, it will lead to

random distribution of alleles (i.e. independent of each other) throughout the genetic region – this is referred to as 'linkage equilibrium'.

Linkage disequilibrium refers to the combination of an allele with an allele of another SNP, more often than would be expected by chance (i.e. non-random association). Two loci are in linkage disequilibrium, if the frequency of association of their respective alleles is higher or lower than what would be expected if they were associated randomly.

Linkage disequilibrium is measured statistically by calculating the difference between the observed frequency of the combination of two alleles and the frequency expected if the two markers were independent. The measure of linkage disequilibrium is often denoted by the variance D`, or by the squared correlation between the presence and absence of alleles at different loci, denoted by '$\mathbf{r^2}$'. D` is a measure of population genetics, related to the recombination events between markers. A D` of 1 between two markers indicates complete LD.

The values of r^2 range from 0 to 1, with 0 suggesting no linkage disequilibrium and 1 complete linkage disequilibrium (169). In genetic studies, $r^2 > 0.80$ suggests strong genetic correlation between two loci and is often used as the cut-off to determine marker/SNP correlation.

The concept of linkage disequilibrium is vital in planning genetic association studies. In the presence of high linkage disequilibrium between two loci, genotyping one loci will ascertain the alleles (genotype) at both loci.

Power calculation in genetic association studies

The success of a genetic association study is dependent upon the strength of statistical power of the study. It measures the probability of rejecting a false null hypothesis. Various factors influence the statistical power of the study to detect associations and it is crucial that these factors are taken into

consideration while planning genetic association studies. The following factors play an important role in determining the power of a GWAS:

a) Disease phenotype and prevalence
b) Sample size (cases and controls)
c) Number of SNPs genotyped and their frequency
d) Linkage disequilibrium
e) Effect size of the associated variants
f) Inheritance models used for analysis (e.g. additive, dominant or multiplicative)

Estimating the statistical power of a study is a useful tool, which allows selection of the appropriate sample size. Adequate statistical power is generally estimated to have 80% probability of detecting an association. It is inversely correlated with Beta (type II error): the higher the power of a study, the lower the risk of a false negative result.

Due to the low prevalence of PSC, a small sample size is a major limiting factor for adequate statistical power in genetic association studies. The concept of varying degrees of statistical power of an allelic test in a case control association study according to: (A) minor allele frequency (MAF), (B) disease prevalence, (C) linkage disequilibrium (LD), and (D) case-to-control ratio, is illustrated very well by Hong, E.P. *et al.* using genetic power calculator (171) (supplementary figure 5.1).

As a general guide, for low prevalent complex disease traits, a sample size of 2000 cases and 2000 controls will give a power of 80% to detect a common genetic variant (MAF > 0.05) with a modest effect size (OR > 1.3).

Significance threshold in GWAS

Conventionally, a test is statistically significant if the p-value is below a preset threshold (α) < 0.05. This is applicable to a single statistical test. However, in GWAS, several independent tests are conducted depending upon the number of SNPs genotyped. When several independent statistical tests are performed, such as in GWAS (i.e. 500,000 to 1 million), the probability of false positive test results (type 1 error; glossary Box 1; page 160) is much higher unless correction is made for multiple testing.

The most common method used to correct for multiple testing is to apply Bonferroni's correction threshold, which is calculated by dividing the significance threshold (0.05) by the total number of independent tests performed (500,000 – 1 million SNPs in GWAS). Thus, the accepted threshold for genome-wide significance is p-value < 5×10^{-8}.

This stringent threshold should not be applied to candidate gene studies and replication studies, where there is already *a priori* a probability of an association. Instead, for validation, the conventional p-value of < 0.05 should be used to confirm statistical significance.

Genotyping

Development of chip based microarray technology with capacity to genotype several hundred thousand SNPs simultaneously, has made GWAS possible. At the time of undertaking genetic association studies in the UK-PSC cohort, the two main commercially available genotyping platforms were: Illumina (San Diego, CA) and Affymetrix (Santa Clara, CA). The genotyping chips used in GWAS were designed to capture common genetic variants (SNPs with MAF > 0.05) and the genotyping platforms preferentially included SNPs in high LD (r^2 > 0.8) with other variants in the genetic region to maximize the coverage and reduce the cost by genotyping fewer SNPs.

Quality control in genetic association studies

In this section, I have briefly outlined the key data quality control (QC) steps carried out in genetic association studies, focusing on GWAS. QC steps performed in the UK PSC genetic association studies are discussed individually in chapters 6 and 7.

The main purpose of stringent quality control steps is to identify and remove DNA samples and SNPs/markers that may introduce bias and/or errors, resulting in a high number of false positive and/or false negative results.

It is usual practice to implement the QC steps in individuals (cases and controls) first, prior to 'marker/SNP' QC. This minimises the risk of markers being removed (in error), due to a set of poorly genotyped individuals.

The 'individual' and 'marker' QC steps usually undertaken in genetic case-control association studies are outlined below.

'Individual' quality control

There are four key steps involved in 'individual' QC which are performed sequentially as outlined below:

a) *Identification of discordant sex information*

The genotype data from the X-chromosome is used to check for discordance with the ascertained sex information. It is particularly relevant if sex-stratified analysis is planned, but may also help detect DNA plating errors. As males only have one copy of the X-chromosome, they can only be homozygote for SNPs genotyped on this chromosome (homozygosity rate ~ 1). On the other hand, females have a much lower homozygosity rate (< 0.2). The genotype-calling algorithm (glossary Box 1; page 160) calculates the homozygosity rate across all X-chromosome SNPs in all individuals and compares it with the expected rate. Individuals with discordant sex information should be excluded from further analysis.

b) *Calculating the missing genotype and heterozygosity rate*

The missing genotype rate and sample heterozygosity rate are both measures of the DNA quality of the individuals. It measures the proportion of SNPs with missing genotype data for that individual. Individuals who are missing genotypes above a pre-determined threshold (usually 3 to 5%) are removed from further analysis.

Likewise, the proportion of individuals with an excessive or reduced number of heterozygote genotypes (heterozygosity rate) could suggest sample contamination. Mean heterozygosity rate (*H*) is calculated by using the formula:

$$H = (N - O)/N$$

(N is the total number of non-missing genotypes, and O is the number of homozygous genotypes for a given individual).

Distribution of the mean heterozygosity rate is inspected to identify individuals with extreme heterozygosity rates (more than ± 3 standard deviations from the mean) and should be excluded from further analysis.

c) *Identification of duplicate or related individuals*

In population based studies, it is important to identify duplicate samples and individuals who are closely related, as it can result in false overrepresentation of genotypes and inflation of allele frequencies in the studied population. Information about duplicate samples is obtained by calculating the proportion of alleles shared at genotyped SNPs (excluding sex chromosome) between individuals and referred to as Identity by state (IBS).

The degree of relatedness can be measured by calculating identity by descent (IBD), using genome wide IBS data. The IBS approaches 1 for duplicate samples or monozygotic twins; 0.5 for 1st degree relatives and

0.25 for 2nd degree relatives. Individuals who are closely related (i.e. 1st or 2nd degree) are typically removed from further analysis.

d) Population stratification

Prior to undertaking a genetic association study, it is vital that cases and controls arise from the same population, to minimise the risk of confounding due to population stratification (glossary Box 1; page 160).

Population structure variations in the study subjects can lead to confounding in genetic association studies. Despite strict inclusion criteria, individuals from a different ethnic and population background may occasionally be inadvertently included in the study. This may result in differences between allele frequencies in the population studied, not because of association with disease risk, but because certain alleles may be over or under represented in a different population. For example, if a disease is studied in a European population to identify genetic risk factors, then it would not be ideal to have individuals from Asia or Africa in the same group, as their genetic make-up would be different to that of Europeans. Population stratification can introduce significant bias in GWAS and must be addressed prior to association analysis.

One of the statistical methods to identify individuals of divergent ancestry is principal component analysis, which is discussed below.

Principal component analysis (PCA)

In simple terms, PCA measures the genotype correlation between individuals and SNPs in reference to populations of known ancestry (using the HapMap reference data).

Principal component modeling is done first, using the available HapMap genotype data and allows clustering of the population into different ancestral groups based on their genetic make-up. The model is then applied to individuals in the study to calculate their respective principal

component scores based on genotype data and group them closest to their reference ancestry groups from HapMap samples. Usually, the first principal component score would allow detection of most of the variation in the study group, followed by second, third and fourth component scores. In well-defined population study groups the first two principal components are usually sufficient to identify ancestral outliers, who should then be removed from further analysis.

'Marker/SNP' quality control

The following steps are typically performed to identify SNP genotyping error:

a) *Calculating the 'genotype call rate'*
This is the proportion of individuals per SNP with non-missing data. Poorly genotyped SNPs could introduce bias and should be removed from the study. SNPs with a call rate less than 95 to 97% are typically removed.

b) *Hardy-Weinberg equilibrium (HWE)*
HWE assumes that the allele and genotype frequencies in successive generations remain constant in a large population in the absence of mutation, natural selection or migration. Significant deviation from the HWE could indicate genotyping or genotype calling error. It is important to note that, SNPs associated with a disease may deviate significantly from HWE, due to their underlying disease association and removing such SNPs would fail to identify true disease association.

Thus, HWE thresholds should only be applied to control samples to detect genotyping or calling error. The significance threshold applied for HWE is variable and in the UK PSC GWAS, we used a stringent threshold of p-value $< 1 \times 10^{-6}$.

c) *Minor allele frequency (MAF)*
SNPs with very low MAF should be removed to reduce false positive results. The threshold applied to remove SNPs varies, depending upon the size of the

study. In the UK-PSC genetic association study, SNPs with MAF < 0.05 (5%) were removed from further analysis.

Despite the above steps, genotyping errors may still be present. The genotype cluster plots for all SNPs showing an association, should be visually inspected to ensure that genotypes have been called correctly.

Imputation

Imputation is a statistical tool, which allows prediction of the genotype of a SNP that was not directly genotyped in the study. It uses information from the genotyped SNPs and haplotype patterns from a reference panel (e.g. HapMap or 1000 genomes) to predict the unobserved genotypes in a dataset (172). Imputation increases the power of GWAS to detect associations and allows meta-analysis of genome wide association scans in samples that were originally genotyped using a different platform.

Chapter 6: Replication and fine mapping of genetic risk loci in PSC.

This chapter focuses on a replication and a candidate gene study that was undertaken in 2012 in the UK PSC cohort. The aim of this study was to validate the loci identified in previous GWAS using the UK PSC cohort.

I designed this study and identified the SNPs for replication. After obtaining extracted DNA samples from the laboratory, I diluted and normalised the DNA samples to the appropriate concentration before sending it for genotyping, which was performed at the Institute of Clinical Molecular Biology (ICMB), Kiel, Germany.
I performed the statistical analysis for this study under the supervision of Dr George Mells, MRC research fellow, Addenbrooke's Hospital, Cambridge.

Introduction

At the time of undertaking this study, two genome-wide association studies (GWAS) had identified susceptibility loci at 2q13, 2q35, 3p21, and 13q31 with the most likely candidate genes proposed as BCL2-like 11 (*BCL2L11*), G-protein- coupled bile acid receptor 1 (*GPBAR1*), macrophage stimulating 1 (*MST1*) and glypican 6 (*GPC6*) respectively (150, 151). Details of these two GWAS were covered in chapter 1 (page 54).

In addition to the genome-wide significant associations, several SNPs at the interleukin-2 receptor-α (*IL2RA*) locus showed a suggestive association in the discovery panel of the study by Melum *et al*., but were not replicated, probably due to effect size heterogeneity in the replication panels (151). Intriguingly, the 4q27 locus harboring the ligand interleukin-2 (IL-2) of IL2RA has also been implicated in a candidate gene study in PSC (173).

Genetic risk variants at 4q27 and 10p15 are shared across multiple immune-mediated diseases and harbor plausible candidate genes *IL2/IL21* and *IL2RA* respectively (174). They are of interest because of the role of IL-2 signaling in immune tolerance. Furthermore, Il-2ra-/- mice spontaneously develop

intestinal and biliary inflammation (175), making IL-2/IL2RA a highly plausible signaling pathway in PSC pathogenesis.

The aim of this study was to:

a) Replicate and validate the association at four non-HLA loci previously reported to be associated with PSC at 2q13, 2q35, 3p21, and 13q31.

b) Further characterize and refine the associations at 4q27 (*IL-2/IL-21*) and 10p15 (*IL2RA*) in a large UK cohort of PSC cases and healthy UK controls.

Methods

Cohort ascertainment

At the time of the study, a total of 1030 PSC cases (with DNA samples) were recruited, including all transplant centres in the UK. All patients with PSC were diagnosed using standard diagnostic criteria as described in chapter 2 (page 58), and were mostly Caucasian. The characteristics of the cohort are shown in Table 6.1.

The control population comprised 5162 individuals from the 1958 British Birth Cohort and National Blood Service samples genotyped as part of the Wellcome Trust Case Control Consortium 2 (WTCCC2) project.

Table 6.1. Characteristics of patients included in the replication study (n = 130).

Ethnic origin	100% Caucasian (97.5% British)
Median age at recruitment	58 (range, 16 - 86) years
Gender	Males = 650, Females = 380 (M:F = 1.7 : 1)
Inflammatory bowel disease	Total = 658 (64%); UC = 570 (86%), CD = 83 (13%), IC = 5 (1%)
Liver transplant recipients	252 (24%)

UC – Ulcerative colitis; CD – Crohn's disease; IC – Indeterminate colitis

Selection of markers and SNPs

a) Four SNPs previously associated with PSC identified from two GWAS were included. These were rs6720394 (2q13; *BCL2L11*), rs12612347 (2q35; *GPBAR1*), rs3197999 (3p21; *MST1*) and rs9524260 (13q31; *GPC6*).

b) To study the interleukin 2 receptor alpha (*IL2RA*) locus, an 80Kbp region was selected on chromosome 10p15 between 6080Kbp - 6160 Kbp spanning *IL2RA*. SNP data were downloaded for the CEU population from HapMap (170) Data Rel 24/phase II Nov 08, on NCBI B36 assembly, dbSNP b126 and Haploview v4.2 (176) was used for selecting tagging SNPs. The HapMap data was filtered based on the following quality control (QC) thresholds: minimum genotype rate = 80%; maximum Mendelian errors = 1; minimum minor allele frequency (MAF) $\geq$ 0.05 and Hardy-Weinberg Equilibrium (HWE) test *P*-value > 0.001. 77 markers met QC criteria. The SNP list was refined to 48 tag SNPs using a pairwise tagging approach (r^2 threshold $\geq$ 0.8), force including eight key SNPs because of their reported association with immune-mediated diseases.

c) To study the *IL-2/IL-21* locus, a 564Kbp region was selected on chromosome 4q27 between 123236Kbp - 123800 Kbp that included the *KIAA1109*, *ADAD1*, *IL-2* and *IL-21* genes. A total of 252 markers were selected for generating the Tag SNP list with a pairwise tagging approach (r^2 threshold $\geq$ 0.8). Five SNPs were force included based on their reported association. As this region shows very high linkage disequilibrium (LD), 39 tag SNPs adequately covered this region.

A list of *IL-2* (five SNPs) and *IL2RA* (eight SNPs) SNPs, which were force included and their respective prior disease associations, is shown in Table 6.2. Please note that the summary discovery panel statistics from a recent GWAS (151) were also used as a reference to select the SNPs to be force included for genotyping. Altogether, 91 SNPs were taken forward for design of genotyping assays.

Table 6.2 *IL-2/IL-21* and *IL2RA* SNPs force included because of their published association with immune-mediated diseases.

SNP	Locus	Previous association (Reference)
rs2104286	*IL2RA*	MS (177), T1D (178)
rs12722489	*IL2RA*	MS (179)
rs4147359	*IL2RA*	T1D (180), PSC (151)
rs10905718	*IL2RA*	PSC (151)
rs4749955	*IL2RA*	T1D (178)
rs12251307	*IL2RA*	T1D (178, 180)
rs706778	*IL2RA*	PSC (151), RA (181)
rs11594656	*IL2RA*	MS (177), T1D (180), RA (181)
rs6822844	*IL2/IL21*	CoD (182), MS (177), CD (183), RA (181), T1D (184), UC (185), PS (186)
rs11938795	*IL2/IL21*	CD (183)
rs13151961	*IL2/IL21*	UC (185), CoD (182), PS (186)
rs13119723	*IL2/IL21*	RA (181), UC (185)
rs6840978	*IL2/IL21*	CoD (182), UC (185), PS (186)

MS – multiple sclerosis; T1D – type 1 diabetes mellitus; PSC – primary sclerosing cholangitis; RA – rheumatoid arthritis; CoD – coeliac disease; CD – Crohn's disease; UC – ulcerative colitis; PS – psoriasis.

DNA and plate preparation

DNA extraction and plating was undertaken as described in chapter 2 (page 62).

Genotyping

Genotyping in cases

The Sequenom iPlex Gold™ genotyping platform was used for genotyping. SNP assay design failed for 25/91 SNPs and these were discarded. However, all thirteen key SNPs listed in Table 13 passed assay design. Thus, 66 SNPs were taken forward for genotyping in 1030 cases. Genotyping was performed at the Institute of Clinical Molecular Biology (ICMB), Kiel, Germany with the Sequenom platform, using an automated laboratory setup and all process data

were written to and administered by a database-driven laboratory information management system (LIMS).

One SNP (rs3197999) failed genotyping with very low call-rate on Sequenom. This SNP alone was subsequently re-genotyped using TaqMan technology at ICMB, Kiel according to the manufacturer's recommendations. Results from the genotyping assay were analysed using the SDS 2.3 software.

Genotyping in controls

We did not genotype the control population. Instead, control genotype data (fully quality controlled for heterozygosity, gender, ethnicity and relatedness) were obtained for 5162 individuals from the WTCCC2, that had previously been genotyped using the Illumina Human1M-Duo array (Illumina, San Diego, California). Control genotype data were available for 47/66 SNPs. Imputed data were generated for the remaining 19 SNP controls.

Imputation of SNPs in the control population

Imputation for 19 SNPs in the control population was carried out by Prof Heather Cordell, Professor of Statistical Genetics, Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, UK, where genotype data were not available. The steps involved in generating the imputed data are summarised below:

Imputed data were generated for controls (on the basis of genotyped SNPs from the Illumina Human1M-Duo array) using the MACH software (187). Phased CEU haplotypes from HapMap Phase II were used as the reference sample (170). Prior to imputation, SNPs with genotype call rate < 95%, minor allele frequency < 1% or HWE (*P*–value < 10^{-4}) were removed. Any A/T or C/G SNPs (for which strand flips would not necessarily be automatically detected) were also removed. Following imputation, only those genotypes that had been imputed with a posterior probability of > 99%, according to the MACH output were retained. Two SNPs, rs4833248 (4q27) and rs1996077 (4q27), showed unacceptably low overall rates of genotype assignment when using this 99% threshold and were excluded from the study.

Quality control

Stringent individual and SNP QC thresholds were applied to both cases and controls as describe below.

QC in Cases

38 cases were excluded due to low genotyping rate (< 90%); the genotyping rate in the remaining cases was > 99%. None of the SNPs showed significant deviation from HWE (*P*-values > 10^{-4}). One SNP rs3197999 (3p21; MST1) had a genotype rate of < 95%. This key SNP was found to be associated with PSC in a previous GWAS (151). For this SNP genotyping was repeated using TaqMan technology with a genotyping success rate > 98%.

QC in Controls

QC thresholds were applied to 47 genotyped SNPs in 5162 controls before combining the dataset with imputed genotype data for the remaining 17 SNPs. The total genotyping rate in controls was > 99% and all the markers had acceptable genotyping rate (> 95%). None of the markers deviated significantly from the Hardy-Weinberg equilibrium (*P*-value > 10^{-4}). One SNP rs2390352 (4q27) had MAF < 0.05 and was removed from further analysis.

After QC, 63 SNPs were analysed in 992 cases and 5162 controls. Details of all the SNPs, which passed QC and were analysed and the method of genotyping (genotyped or imputed) in control population are given in Table 6.3.

Table 6.3 Genotyping method used in controls for SNPs included in the analysis.

Chr	Locus	SNP	Position	Alleles	Genotyping method in controls
2	*BCL2L11*	rs6720394	111,705,843	G/T	Genotyped
2	*GPBAR1*	rs12612347	218,765,583	G/A	Imputed
3	*MST1*	rs3197999	49,696,536	A/G	Genotyped
13	*GPC6*	rs9524260	93,311,791	A/G	Genotyped
4	*IL-2/IL-21*	rs12511287	123,797,981	A/T	Imputed
4	*IL-2/IL-21*	rs6822844	123,728,871	T/G	Genotyped
4	*IL-2/IL-21*	rs13151961	123,334,952	G/A	Genotyped
4	*IL-2/IL-21*	rs13125087	123,238,043	T/A	Imputed
4	*IL-2/IL-21*	rs11938795	123,292,459	C/T	Genotyped
4	*IL-2/IL-21*	rs6840978	123,774,157	T/C	Genotyped
4	*IL-2/IL-21*	rs6827444	123,352,628	G/C	Imputed
4	*IL-2/IL-21*	rs13143866	123,760,208	A/G	Genotyped
4	*IL-2/IL-21*	rs7693745	123,455,033	G/A	Imputed
4	*IL-2/IL-21*	rs2137497	123,777,704	T/G	Genotyped
4	*IL-2/IL-21*	rs12642902	123,727,951	A/G	Genotyped
4	*IL-2/IL-21*	rs11722421	123,491,118	C/G	Imputed
4	*IL-2/IL-21*	rs6849146	123,545,541	C/T	Genotyped
4	*IL-2/IL-21*	rs975404	123,740,742	C/T	Imputed
4	*IL-2/IL-21*	rs17454584	123,572,882	G/A	Genotyped
4	*IL-2/IL-21*	rs1512971	123,744,785	T/C	Imputed
4	*IL-2/IL-21*	rs6534355	123,781,100	T/C	Imputed
4	*IL-2/IL-21*	rs4295278	123,766,991	C/T	Genotyped
4	*IL-2/IL-21*	rs4833810	123,237,840	T/C	Imputed
4	*IL-2/IL-21*	rs4833834	123,685,801	G/A	Genotyped
4	*IL-2/IL-21*	rs6534338	123,246,319	T/C	Genotyped
4	*IL-2/IL-21*	rs13110000	123,797,510	C/T	Genotyped
4	*IL-2/IL-21*	rs7678445	123,502,222	T/G	Genotyped
4	*IL-2/IL-21*	rs13119723	123,437,763	G/A	Genotyped
4	*IL-2/IL-21*	rs11931332	123,236,177	C/T	Imputed
4	*IL-2/IL-21*	rs6848868	123,369,736	C/T	Genotyped
4	*IL-2/IL-21*	rs6419221	123,783,569	T/C	Genotyped
4	*IL-2/IL-21*	rs1022234	123,782,528	G/A	Genotyped
10	*IL2RA*	rs4147359	6,148,445	A/G	Genotyped
10	*IL2RA*	rs706778	6,138,955	T/C	Imputed
10	*IL2RA*	rs7090530	6,150,881	C/A	Genotyped
10	*IL2RA*	rs10905718	6,154,862	G/A	Imputed
10	*IL2RA*	rs10905669	6,132,099	T/C	Genotyped
10	*IL2RA*	rs11594656	6,162,015	A/T	Genotyped
10	*IL2RA*	rs10905716	6,154,016	T/C	Imputed
10	*IL2RA*	rs706779	6,138,830	T/C	Genotyped
10	*IL2RA*	rs7090512	6,150,835	C/T	Genotyped
10	*IL2RA*	rs706781	6,126,391	C/T	Imputed
10	*IL2RA*	rs4749924	6,122,402	C/A	Genotyped
10	*IL2RA*	rs11256456	6,120,718	C/T	Genotyped
10	*IL2RA*	rs2104286	6,139,051	C/T	Genotyped
10	*IL2RA*	rs2476491	6,135,416	T/A	Genotyped

10	*IL2RA*	rs11256497	6,127,800	A/G	Genotyped
10	*IL2RA*	rs12722563	6,109,567	A/G	Genotyped
10	*IL2RA*	rs2256774	6,137,171	C/T	Genotyped
10	*IL2RA*	rs12251307	6,163,501	T/C	Genotyped
10	*IL2RA*	rs12359875	6,091,113	T/C	Imputed
10	*IL2RA*	rs6602392	6,118,085	A/C	Genotyped
10	*IL2RA*	rs791587	6,128,705	A/G	Genotyped
10	*IL2RA*	rs9663421	6,095,610	T/C	Genotyped
10	*IL2RA*	rs11598648	6,124,031	A/G	Imputed
10	*IL2RA*	rs11256457	6,120,800	G/C	Genotyped
10	*IL2RA*	rs2076846	6,103,259	G/A	Genotyped
10	*IL2RA*	rs4749955	6,158,972	C/T	Genotyped
10	*IL2RA*	rs12722596	6,096,300	C/T	Genotyped
10	*IL2RA*	rs12722588	6,100,439	T/C	Genotyped
10	*IL2RA*	rs7093069	6,103,325	T/C	Genotyped
10	*IL2RA*	rs12722489	6,142,018	T/C	Genotyped
10	*IL2RA*	rs12244380	6,093,380	G/A	Genotyped

Position refers to NCBI's build 36. Chr – chromosome.

Statistical analysis

Case control association analysis

Allelic association analysis for all the SNPs was performed using logistic regression in PLINK v. 1.07 (188). I only included control genotypes that had been imputed with > 99% probability to reduce the risk of false positive results. To correct for multiple testing for the 63 SNPs analysed, I determined a Bonferroni corrected threshold for a significant *P*-value at $< 7.9 \times 10^{-4}$ (0.05/63).

Meta-analysis

For selected SNPs, previously published summary statistics were used to perform meta-analysis using the Metagen (inverse variance method) package in the R statistical software package (R v 2.13.1). Most significant SNPs reported in the two previous GWAS were included: rs6720394 (*BCL2L11*); rs12612347 (*GPBAR1*); rs3197999 (*MST1*) and rs9524260 (*GPC6*). In addition, most significantly associated SNPs at 4q27 (*IL2/IL21*) and 10p15 (*IL2RA*) loci (whose summary statistics from previous studies were available)

were included in the meta-analysis: rs6822844 (*IL2/IL21*), rs12511287 (*IL2/IL21*), rs4147359 (*IL2RA*), rs706778 (*IL2RA*) and rs7090530 (*IL2RA*).

To account for differences between the population groups, I used a random effects model even though a test for heterogeneity was not statistically significant for any of the associated SNPs.

Conditional analysis

Conditional analysis is usually performed at an associated locus, to identify if there is effect of associated SNPs is independent of each other or not. I performed a logistic regression analysis conditioned on the most associated SNPs at 4q27 and 10p15 loci in PLINK v. 1.07, to determine the strength and validity of association of the SNPs in relation to each other.

Phenotype based sub-group analyses

I performed a sub-group analysis based on:

a) Auto-immune disease status

I hypothesized that the presence of co-existing immune-mediated diseases in the cohort could influence the association analysis results in particular because of the reported association of SNPs with immune loci. Thus, I identified and excluded 122 cases suffering with one or more autoimmune diseases from the original cohort: n = 908 (1030 - 122) and then applied quality control (QC) thresholds to these 908 cases. 33 cases were further removed for low genotyping rate (< 90%), leaving 875 cases. None of the 63 markers studied had low genotyping rate (< 95%) or deviated significantly from HWE ($P > 10^{-4}$).

Thus, 63 SNPs were analysed in 875 cases and 5162 controls using logistic regression in PLINK v 1.07.

b) IBD status

For the IBD sub-group analysis I used fully quality-controlled dataset (n = 992 cases) and identified the cases based on their IBD status and then performed a logistic regression analysis in PLINK comparing:

- PSC cases with IBD (PSC + IBD = 625) with 5162 controls
- PSC cases without IBD (PSC - IBD = 367) with 5162 controls.

Results

Allelic association analysis

After quality control, 992 PSC cases were compared with 5162 controls at 63 SNPs. Association results for all the studied SNPs are shown in supplementary table 6.1.

a) Results of SNPs associated in the previous GWAS

Only one of the four previously reported non-HLA loci SNPs was replicated. This was at 3p21 (rs3197999); *p*-value = 1.9 x 10^{-6}; $OR_{A\ vs\ G}$ = 1.28, 95% confidence interval [CI] (1.16 - 1.42).

SNPs at 2q13 (*BCL2L11*); 2q35 (*GPBAR1*) and 13q31 (*GPC6*) did not reach threshold for statistical significance (i.e. *p*-value < 7.9 x 10^{-4}) (Table 6.4).

Table 6.4. Association results for SNPs previously associated in GWAS.

Chr	Locus	SNP	Position	Alleles	MAF (cases/controls)	*P*-value (UK cohort)	OR (95% CI) (UK cohort)
2	*BCL2L11*	rs6720394	111,705,843	G/T	0.13/0.12	0.40	1.07 (0.92 - 1.23)
2	*GPBAR1*	rs12612347	218,765,583	G/A	0.48/0.50	0.21	0.94 (0.85 - 1.04)
3	***MST1***	**rs3197999**	**49,696,536**	**A/G**	**0.33/0.28**	**1.9E-06**	**1.28 (1.16 - 1.42)**
13	*GPC6*	rs9524260	93,311,791	A/G	0.40/0.40	0.77	0.99 (0.89 - 1.09)

Association results for four SNPs in the UK cohort (992 PSC cases and 5162 controls) that have previously been associated in GWAS are shown. Logistic regression analysis was performed in PLINK v1.07. Position refers to NCBI's build 36. Bonferroni corrected *P*-value threshold for significance was determined to be < 7.9 x 10^{-4}. All the reported odds ratios are with reference to minor allele vs. major allele. Chr – chromosome, MAF – Minor allele frequency, OR – Odds ratio, CI – Confidence interval.

b) Association at 4q27

At 4q27 locus, a significant association was observed for rs12511287 (*P*-value = 3.0 x 10^{-4}, $OR_{A\ vs\ T}$ = 1.21, 95% CI (1.09 - 1.35)) (Table 6.4). The second most associated SNP at this locus (rs6822844) was associated at nominal significance but was insufficiently robust for multiple corrections (*P*-value = 1.9 x 10^{-3}, $OR_{T\ vs\ G}$ = 0.81, 95% CI (0.70 - 0.92)).

In addition, eight other SNPs at this locus were nominally associated (Table 6.5). A regional association plot for all the SNPs analysed at 4q27 locus is shown in Figure 6.1 (page 128).

Table 6.5. Association results for SNPs analysed at 4q27 locus.

Chr	Locus	SNP	Position	Allele	MAF (cases/controls)	*P*-value (UK cohort)	OR (95% CI) (UK cohort)
4	***IL-2/IL-21***	**rs12511287**	**123,797,981**	**A/T**	**0.31/0.27**	**3.0E-04**	**1.21 (1.09 - 1.35)**
4	***IL-2/IL-21***	**rs6822844**	**123,728,871**	**T/G**	**0.15/0.17**	**1.9E-03**	**0.81 (0.70 - 0.92)**
4	***IL-2/IL-21***	**rs13151961**	**123,334,952**	**G/A**	**0.15/0.17**	**3.2E-03**	**0.82 (0.71 - 0.93)**
4	***IL-2/IL-21***	**rs13125087**	**123,238,043**	**T/A**	**0.16/0.19**	**3.9E-03**	**0.82 (0.72 - 0.94)**
4	***IL-2/IL-21***	**rs11938795**	**123,292,459**	**C/T**	**0.23/0.26**	**5.9E-03**	**0.85 (0.76 - 0.95)**
4	***IL-2/IL-21***	**rs6840978**	**123,774,157**	**T/C**	**0.18/0.21**	**8.5E-03**	**0.85 (0.75 - 0.96)**
4	***IL-2/IL-21***	**rs6827444**	**123,352,628**	**G/C**	**0.22/0.19**	**0.02**	**1.15 (1.02 - 1.30)**
4	***IL-2/IL-21***	**rs13143866**	**123,760,208**	**A/G**	**0.26/0.28**	**0.028**	**0.89 (0.79 - 0.99)**
4	***IL-2/IL-21***	**rs7693745**	**123,455,033**	**G/A**	**0.29/0.26**	**0.034**	**1.12 (1.01 - 1.25)**
4	***IL-2/IL-21***	**rs2137497**	**123,777,704**	**T/G**	**0.43/0.40**	**0.040**	**1.11 (1.01 - 1.22)**
4	*IL-2/IL-21*	rs12642902	123,727,951	A/G	0.32/0.34	0.066	0.91 (0.82 - 1.01)
4	*IL-2/IL-21*	rs11722421	123,491,118	C/G	0.35/0.33	0.20	1.07 (0.97 - 1.18)
4	*IL-2/IL-21*	rs6849146	123,545,541	C/T	0.39/0.38	0.21	1.06 (0.97 - 1.17)
4	*IL-2/IL-21*	rs975404	123,740,742	C/T	0.35/0.36	0.28	0.94 (0.85 - 1.05)
4	*IL-2/IL-21*	rs17454584	123,572,882	G/A	0.23/0.22	0.34	1.06 (0.94 - 1.19)
4	*IL-2/IL-21*	rs1512971	123,744,785	T/C	0.28/0.27	0.36	1.05 (0.94 - 1.17)
4	*IL-2/IL-21*	rs6534355	123,781,100	T/C	0.24/0.24	0.37	1.05 (0.94 - 1.18)
4	*IL-2/IL-21*	rs4295278	123,766,991	C/T	0.06/0.05	0.38	1.10 (0.89 - 1.34)
4	*IL-2/IL-21*	rs4833810	123,237,840	T/C	0.39/0.38	0.45	1.04 (0.94 - 1.15)
4	*IL-2/IL-21*	rs4833834	123,685,801	G/A	0.09/0.10	0.45	0.94 (0.79 - 1.11)
4	*IL-2/IL-21*	rs6534338	123,246,319	T/C	0.30/0.29	0.52	1.04 (0.93 - 1.15)
4	*IL-2/IL-21*	rs13110000	123,797,510	C/T	0.44/0.43	0.52	1.03 (0.94 - 1.14)
4	*IL-2/IL-21*	rs7678445	123,502,222	T/G	0.08/0.07	0.53	1.06 (0.89 - 1.27)
4	*IL-2/IL-21*	rs13119723	123,437,763	G/A	0.15/0.15	0.65	0.97 (0.85 - 1.11)
4	*IL-2/IL-21*	rs11931332	123,236,177	C/T	0.34/0.34	0.78	1.01 (0.92 - 1.12)
4	*IL-2/IL-21*	rs6848868	123,369,736	C/T	0.08/0.08	0.81	0.98 (0.82 - 1.16)
4	*IL-2/IL-21*	rs6419221	123,783,569	T/C	0.37/0.37	0.87	0.99 (0.90 - 1.10)
4	*IL-2/IL-21*	rs1022234	123,782,528	G/A	0.31/0.32	0.91	0.99 (0.90 - 1.10)

Association results for all the SNPs analysed in the UK cohort (992 PSC cases and 5162 controls) at 4q27 locus are shown. Logistic regression analysis was performed in PLINK v1.07. Position refers to NCBI's build 36. Bonferroni corrected *P*-value threshold for significance was determined to be $< 7.9 \times 10^{-4}$. SNPs with nominal association (*P*-value < 0.05) are shown in bold. All the reported odds ratios are with reference to minor allele vs. major allele.
Chr – chromosome, MAF – Minor allele frequency, OR – Odds ratio, CI – Confidence interval.

c) Association at 10p15

At the 10p15 locus, significant associations were observed after correction for multiple testing (*P*-value < 7.9 x 10^{-4}), for three SNPs: rs4147359, *P*-value = 2.6 x 10^{-4}, $OR_{A\ vs\ G}$ = 1.20, 95% CI (1.09 - 1.33); rs706778, *P*-value = 4.3 x 10^{-4}, $OR_{T\ vs\ C}$ = 1.19, 95% CI (1.08 - 1.31); and rs7090530, *P*-value = 7.0 x 10^{-4}, $OR_{C\ vs\ A}$ = 0.84, 95% CI (0.76 - 0.93).

A nominal association (*P*-value < 0.05) was seen for an additional 10 SNPs at this locus (Table 6.6). A regional association plot for all the SNPs analysed at 10p15 locus is shown in Figure 6.1 (page 128).

Table 6.6. Association results for SNPs analysed at 10p15 locus.

Chr	Locus	SNP	Position	Allele	MAF (cases/controls)	*P*-value (UK cohort)	OR (95% CI) (UK cohort)
10	***IL2RA***	**rs4147359**	**6,148,445**	**A/G**	**0.39/0.34**	**2.6E-04**	**1.20 (1.09 - 1.33)**
10	***IL2RA***	**rs706778**	**6,138,955**	**T/C**	**0.44/0.40**	**4.3E-04**	**1.19 (1.08 - 1.31)**
10	***IL2RA***	**rs7090530**	**6,150,881**	**C/A**	**0.36/0.40**	**7.0E-04**	**0.84 (0.76 - 0.93)**
10	***IL2RA***	**rs10905718**	**6,154,862**	**G/A**	**0.35/0.31**	**1.0E-03**	**1.18 (1.07 - 1.31)**
10	***IL2RA***	**rs10905669**	**6,132,099**	**T/C**	**0.27/0.23**	**1.6E-03**	**1.20 (1.07 - 1.33)**
10	***IL2RA***	**rs11594656**	**6,162,015**	**A/T**	**0.21/0.24**	**2.8E-03**	**0.84 (0.74 - 0.94)**
10	***IL2RA***	**rs10905716**	**6,154,016**	**T/C**	**0.19/0.22**	**4.2E-03**	**0.84 (0.74 - 0.95)**
10	***IL2RA***	**rs706779**	**6,138,830**	**T/C**	**0.44/0.47**	**8.8E-03**	**0.88 (0.80 - 0.97)**
10	***IL2RA***	**rs7090512**	**6,150,835**	**C/T**	**0.27/0.30**	**0.011**	**0.87 (0.78 - 0.97)**
10	***IL2RA***	**rs706781**	**6,126,391**	**C/T**	**0.28/0.26**	**0.021**	**1.14 (1.02 - 1.27)**
10	***IL2RA***	**rs4749924**	**6,122,402**	**C/A**	**0.30/0.33**	**0.025**	**0.89 (0.80 - 0.99)**
10	***IL2RA***	**rs11256456**	**6,120,718**	**C/T**	**0.23/0.21**	**0.032**	**1.14 (1.01 - 1.28)**
10	***IL2RA***	**rs2104286**	**6,139,051**	**C/T**	**0.26/0.28**	**0.044**	**0.89 (0.80 - 1.00)**
10	*IL2RA*	rs2476491	6,135,416	T/A	0.27/0.29	0.053	0.90 (0.81 - 1.00)
10	*IL2RA*	rs11256497	6,127,800	A/G	0.35/0.37	0.072	0.91 (0.82 - 1.01)
10	*IL2RA*	rs12722563	6,109,567	A/G	0.10/0.12	0.085	0.87 (0.75 - 1.02)
10	*IL2RA*	rs2256774	6,137,171	C/T	0.32/0.34	0.11	0.92 (0.83 - 1.02)
10	*IL2RA*	rs12251307	6,163,501	T/C	0.11/0.12	0.11	0.88 (0.76 - 1.03)
10	*IL2RA*	rs12359875	6,091,113	T/C	0.23/0.25	0.14	0.92 (0.82 - 1.03)
10	*IL2RA*	rs6602392	6,118,085	A/C	0.10/0.09	0.18	1.12 (0.95 - 1.31)
10	*IL2RA*	rs791587	6,128,705	A/G	0.45/0.46	0.21	0.94 (0.85 - 1.04)
10	*IL2RA*	rs9663421	6,095,610	T/C	0.27/0.28	0.22	0.93 (0.84 - 1.04)

10	*IL2RA*	rs11598648	6,124,031	A/G	0.44/0.43	0.22	1.06 (0.96 - 1.17)
10	*IL2RA*	rs11256457	6,120,800	G/C	0.38/0.40	0.25	0.94 (0.86 - 1.04)
10	*IL2RA*	rs2076846	6,103,259	G/A	0.37/0.36	0.40	1.04 (0.94 - 1.15)
10	*IL2RA*	rs4749955	6,158,972	C/T	0.44/0.45	0.44	0.96 (0.87 - 1.06)
10	*IL2RA*	rs12722596	6,096,300	C/T	0.10/0.10	0.55	1.05 (0.90 - 1.23)
10	*IL2RA*	rs12722588	6,100,439	T/C	0.19/0.19	0.68	1.03 (0.91 - 1.16)
10	*IL2RA*	rs7093069	6,103,325	T/C	0.19/0.19	0.68	1.03 (0.91 - 1.16)
10	*IL2RA*	rs12722489	6,142,018	T/C	0.16/0.16	0.72	0.98 (0.86 - 1.11)
10	*IL2RA*	rs12244380	6,093,380	G/A	0.43/0.43	0.81	1.01 (0.92 - 1.12)

Association results for all the SNPs analysed in the UK cohort (992 PSC cases and 5162 controls) at 10p15 locus are shown. Logistic regression analysis was performed in PLINK v1.07. Position refers to NCBI's build 36. Bonferroni corrected *P*-value threshold for significance was determined to be < 7.9×10^{-4}. SNPs with nominal association (*P*-value < 0.05) are shown in bold. All the reported odds ratios are with reference to minor allele vs. major allele. Chr – chromosome, MAF – Minor allele frequency, OR – Odds ratio, CI – Confidence interval.

Meta-analysis

Results from the UK cohort were combined with summary statistics from three other published genetic association studies (150, 151, 173) for four SNPs (at non-HLA loci) associated previously with PSC and selected SNPs at 4q27 and 10p15 loci where data were available.

Results of the combined analysis for SNPs included in the meta-analysis are shown in Table 6.7. Genome-wide significance (*P*-value < 5×10^{-8}) was observed for rs3197999 at 3p21 ($P_{combined} = 3.8 \times 10^{-12}$, $OR_{A\ vs\ G} = 1.38$, 95% CI (1.26 - 1.51)) and for rs4147359 at 10p15 ($P_{combined} = 1.5 \times 10^{-8}$, $OR_{A\ vs\ G} = 1.25$, 95% CI (1.16 - 1.36)). The other previously reported non-HLA SNPs at 2q13; 2q35 and 13q31 were not associated in the UK cohort or combined analysis (Table 6.7).

At the 4q27 locus, the association at rs6822844 did not reach the significance threshold (robust for multiple corrections) in the UK cohort. However, in a recent study of 1,186 northern European PSC patients, it was found to be associated with PSC. So I performed a combined analysis using published summary statistics from the study by Janse *et al.* and the UK cohort, which

strengthened the association at rs6822844 (P-value$_{combined}$ = 6.9 x 10^{-6}; $OR_{T\ vs\ G}$ = 0.77, 95% CI (0.69 – 0.86)) (Table 6.7).

Table 6.7. Results for SNPs included in the meta-analysis.

Chr	SNP	Locus	Alleles (minor/ major)	MAF (case/ control)	*P*-value (UK cohort)	OR (95% CI) (UK cohort)	Meta-Analysis (random effects)	
							OR (combined)	*P*-value (combined)
2	rs6720394*	*BCL2L11*	G/T	0.13/0.12	0.40	1.07 (0.92 – 1.23)	1.28 (1.03-1.60)	0.024
2	rs12612347§	*GPBAR1*	G/A	0.48/0.50	0.21	0.94 (0.85 – 1.04)	0.88 (0.74-1.04)	0.12
3	**rs3197999¶**	***MST1***	**A/G**	**0.33/0.28**	**1.9 x 10^{-6}**	**1.28 (1.16 – 1.42)**	**1.38 (1.26-1.51)**	**3.8 x 10^{-12}**
4	rs6822844λ	*IL2/IL21*	T/G	0.15/0.17	1.9 x 10^{-3}	0.81 (0.70 – 0.92)	0.77 (0.69-0.86)	6.9 x 10^{-6}
10	**rs4147359¶**	***IL2RA***	**A/G**	**0.39/0.34**	**2.6 x 10^{-4}**	**1.20 (1.09 – 1.33)**	**1.25 (1.16-1.36)**	**1.5 x 10^{-8}**
10	rs706778¶	*IL2RA*	T/C	0.44/0.40	4.3 x 10^{-4}	1.19 (1.08 – 1.31)	1.24 (1.14-1.35)	3.4 x 10^{-7}
13	rs9524260§	*GPC6*	A/G	0.40/0.40	0.77	0.99 (0.89 – 1.09)	0.82 (0.68-0.98)	0.034

Association results are shown for SNPs associated after correction for multiple testing (Bonferroni corrected threshold for significance *P*-value < 7.9 x 10^{-4}). Meta-analysis for selected SNPs¶ was performed by combining summary statistics from a recent PSC GWAS (151), using the Metagen (inverse variance method) package in R statistical software package. A random effects model was used for meta-analysis even though a test for heterogeneity was not statistically significant for any of the associated SNPs. All the reported odds ratios are with reference to minor allele vs. major allele. Chr – Chromosome, MAF – minor allele frequency, OR – odds ratio, CI – confidence interval.

Conditional analysis

a) At 4q27 locus

I performed logistic regression analyses at 4q27 locus, conditioned on the two most associated SNPs rs12511287 and rs6822844, which were not in linkage disequilibrium with each other (pairwise r^2 = 0). When conditioned on rs12511287 (most associated SNP at 4q27), the strength of association of all the remaining SNPs remained unchanged when compared to the original results (figure 6.2 A; page 129). However, conditioning on rs6822844 weakened the effect of all other SNPs except for rs12511287, suggesting an independent effect of these two SNPs at this locus (figure 6.2 B; page 129).

b) At 10p15 locus

A logistic regression analysis conditioned on the three most associated SNPs (rs4147359; rs706778; and rs7090530) at 10p15 locus was performed. Pairwise linkage disequilibrium between these three SNPs is shown in Table 6.8. For each of the conditional analysis performed, the strength of association weakened for all other SNPs, suggesting the association to be dependent on

each other. Results (*P*-value) of the conditional analysis are shown in Table 6.9. An association plot for each of the conditional analyses is shown in Figure 6.3 (A - C) (page 130).

Table 6.8. Pairwise LD for SNPs at 10p15 locus included in the conditional analysis.

SNP1	SNP2	r^2
rs4147359	rs706778	0.68
rs706778	rs7090530	0.42
rs7090530	rs4147359	0.34

Table 6.9. Association analysis results conditioned on the three most associated SNPs at 10p15 locus.

Chr	SNP	*P*-value (UK cohort)	*P*-value (conditioned on rs4147359)	*P*-value (conditioned on rs706778)	*P*-value (conditioned on rs7090530)
10	**rs4147359**	**2.60E-04**	**N/A**	0.09254	0.04597
10	**rs706778**	**4.30E-04**	0.6068	**N/A**	0.09858
10	**rs7090530**	**7.00E-04**	0.1246	0.1216	**N/A**
10	rs10905718	1.00E-03	0.7835	0.4598	0.1009
10	rs10905669	1.60E-03	0.4527	0.2738	0.06342
10	rs11594656	2.80E-03	0.09713	0.09589	0.3156
10	rs10905716	4.20E-03	0.04862	0.02835	0.5427
10	rs706779	8.80E-03	0.3666	0.9962	0.647
10	rs7090512	0.011	0.3716	0.3292	0.8246
10	rs706781	0.021	0.767	0.394	0.1071
10	rs4749924	0.025	0.1769	0.1085	0.2634
10	rs11256456	0.032	0.8255	0.5931	0.1983
10	rs2104286	0.044	0.5799	0.7143	0.04148
10	rs2476491	0.053	0.4958	0.6038	0.8746
10	rs11256497	0.072	0.3225	0.546	0.8598
10	rs12722563	0.085	0.3031	0.3874	0.6156
10	rs12251307	0.11	0.4167	0.3379	0.5672
10	rs2256774	0.11	0.5235	0.77	0.8992
10	rs12359875	0.14	0.7532	0.7859	0.9697
10	rs6602392	0.18	0.2379	0.242	0.2152
10	rs791587	0.21	0.6126	0.4668	0.9935
10	rs11598648	0.22	0.7524	0.9171	0.2835
10	rs9663421	0.22	0.7164	0.7707	0.8993
10	rs11256457	0.25	0.3905	0.273	0.7457
10	rs2076846	0.4	0.5463	0.8806	0.4255

10	rs4749955	0.44	0.2125	0.5727	0.3115
10	rs12722596	0.55	0.6957	0.8599	0.3992
10	rs12722588	0.68	0.09227	0.03735	0.6468
10	rs7093069	0.68	0.09172	0.03758	0.6472
10	rs12722489	0.72	0.3499	0.3675	0.1245
10	rs12244380	0.81	0.4509	0.3691	0.9039

Results of conditional analysis at 10p15 locus are shown. The three most associated SNPs in the UK cohort at this locus are highlighted in bold. When conditioned on one another, these SNPs lose their significance suggesting an association that is dependent on each another at this locus.

Phenotype based sub-group analyses

I performed a phenotype based, sub-group analysis within the cohort for the following sub-groups of patients based on:

a) Autoimmune disease phenotype

63 SNPs were analysed in 875 cases and 5162 controls using logistic regression in PLINK v 1.07. A significant association (*P*-value < 7.9×10^{-4}) was observed for rs3197999 at 3p21 (*MST1*), *P*-value $_{(AID\ excluded)} = 1.2 \times 10^{-6}$, $OR_{A\ vs\ G} = 1.31$, 95% CI (1.17 – 1.46); and rs12511287 at 4q27 (*IL2/IL21*), *P*-value$_{(AID\ excluded)} = 4.7 \times 10^{-4}$, $OR_{A\ vs\ T} = 1.22$, 95% CI (1.09 – 136). None of the SNPs at 10p15 (*IL2RA*) locus reached the threshold for significance.

Results for all SNPs that reached nominal significance (*P*-value $_{(AID\ excluded)} < 0.05$) in this sub-group analysis are shown in Supplementary Table 6.2.

b) IBD phenotype

I performed a logistic regression analysis in PLINK comparing PSC cases with (PSC + IBD = 625) or without (PSC - IBD = 367) IBD against 5162 controls. A significant association (*P*-value < 7.9×10^{-4}) was observed only for rs3197999 at 3p21 (*MST1*) in PSC patients with IBD, *P*-value $_{(PSC\ with\ IBD)} = 1.2 \times 10^{-5}$, $OR_{A\ vs\ T} = 1.32$ (1.16 – 1.49); *P*-value $_{(PSC\ without\ IBD)} = 7.2 \times 10^{-3}$, $OR_{A\ vs\ T} = 1.24$ (1.06 – 1.45). Results for all SNPs that reached nominal significance (*P*-value $_{(PSC\ with\ IBD)\ or\ (PSC\ without\ IBD)} < 0.05$) in this sub-group analysis are shown in Supplementary Table 6.3.

Discussion

This study identifies for the first time an association between PSC and SNPs at a 10p15 locus with genome wide significance and confirms an association at 3p21 and 4q27. The most plausible candidate genes at these loci are *IL2RA*, *MST1* and *IL-2/IL-21* respectively.

The *MST1* SNP, rs3197999, encodes a non-synonymous change p.Arg689Cys in *MST1* and is also associated with IBD, suggesting shared genetic risk factors between PSC and IBD (189). *MST1* is highly expressed in the liver and encodes macrophage-stimulating protein (MSP). It plays a role in cell proliferation, leucocyte adhesion and chemotaxis, and has a critical role in attenuating the inflammatory response by inhibition of lipopolysaccharide-induced inflammatory mediators (190).

At the 10p15 locus, association signal peaks at rs4147359, which is located in the intergenic region between *IL2RA* and RNA binding motif protein-17 (*RBM17*). This SNP has been associated previously with type-1 diabetes (180). Other type-1 diabetes risk alleles at this locus have been associated with reduced serum concentrations of soluble IL-2RA (180), making *IL2RA* a plausible candidate gene, as well as suggesting a possible mechanism for the genetic association. Of note, spontaneous mutation of *IL2RA* in humans causes systemic autoimmunity and multi-organ inflammation (191).

At 4q27 locus, the most associated SNP (rs12511287) is located upstream of the *IL-21* gene. The second most associated SNP at this locus, rs6822844, located in a noncoding region upstream of IL2 and downstream of IL21, has previously been associated with other autoimmune diseases and is considered as a general autoimmune risk locus (192). I have for the first time shown two independent association signals at this locus marked by rs12511287 and rs6822844 in PSC.

IL-2 is a likely candidate gene at 4q27 (173), especially given the strong PSC association with its receptor (IL2RA). IL-2 is a cytokine produced

predominantly by activated T cells and exhibits its actions both in an autocrine and paracrine fashion by binding to the IL-2 receptor (IL-2R). It plays a crucial role in modulating immune responses by promoting proliferation, differentiation, and function of activated T cells as well as maintaining the homeostasis and functioning of $CD4^{+}CD25^{+}$ $Foxp3^{+}$ T regulatory (Treg) cells (193). A defect in Treg cell production is believed to be the main reason for autoimmunity associated with IL-2/IL-2R deficiency as seen in murine models and humans (191, 194).

Another gene encoded at 4q27 locus, which also appears to be a plausible candidate gene at this locus is *IL-21*. IL-21 plays an important role in immunoglobulin production and T-cell function, has pro-apoptotic actions on B-cells and drives the terminal differentiation of B-cells to plasma cells (195). IL-21 protein expression is increased in colonic tumours and it has been shown in animal models that $IL21^{-/-}$ knockout mice are resistant to colon cancer (196). IL-21 may thus have a role in amplifying the gut inflammatory milieu in patients with ulcerative colitis, thereby promoting the growth of colitis-associated cancer. PSC further increases the risk of colorectal cancer in IBD patients but the mechanism is unknown (197). It is possible that PSC patients with IBD may have dysregulated IL-21 function, making it an interesting candidate gene at this locus.

Genetic variants at 4q27 (*IL-2/IL-21*) and 10p15 (*IL2RA*) are associated with other autoimmune diseases, such as coeliac disease, rheumatoid arthritis, type 1 diabetes mellitus, multiple sclerosis and psoriasis (174, 198, 199).

There is widespread sharing of genetic risk loci (including variants at 4q27 (*IL2/IL21*) and 10p15 (*IL2RA*)) between these diseases (174), making them non-organ specific autoimmune disease susceptibility loci. PSC is associated with an increased prevalence of other autoimmune diseases and its association with genetic variants at 4q27 and 10p15 loci could imply not only a shared genetic risk with other autoimmune diseases but also allow us to identify the shared immunological pathways favouring disease development.

The present results refine the known genetic architecture of PSC by confirming *MST1*, *IL2RA* and *IL-2/IL-21* locus associations, suggesting a role of both innate and adaptive immune responses in PSC pathogenesis. Despite successfully replicating and confirming the associations at these loci, I was not able to refine the association signal to a specific causal variant. It is important to emphasise that genetic association studies (GWAS or candidate gene studies) have limited utility in identifying true causal genetic variants associated with complex diseases such as PSC. They do allow identification of the most likely biological pathways that may be involved in disease pathogenesis and serve as good starting points to design robust functional studies based on the implicated pathways (200). PSC is believed to be a multi-factorial disease and genetic polymorphisms not yet discovered are likely to be important along with potential environmental and epigenetic risk factors. Functional studies at each of the associated locus are now required to identify the true causative gene or genes to facilitate rapid translation to the discovery of novel therapeutics.

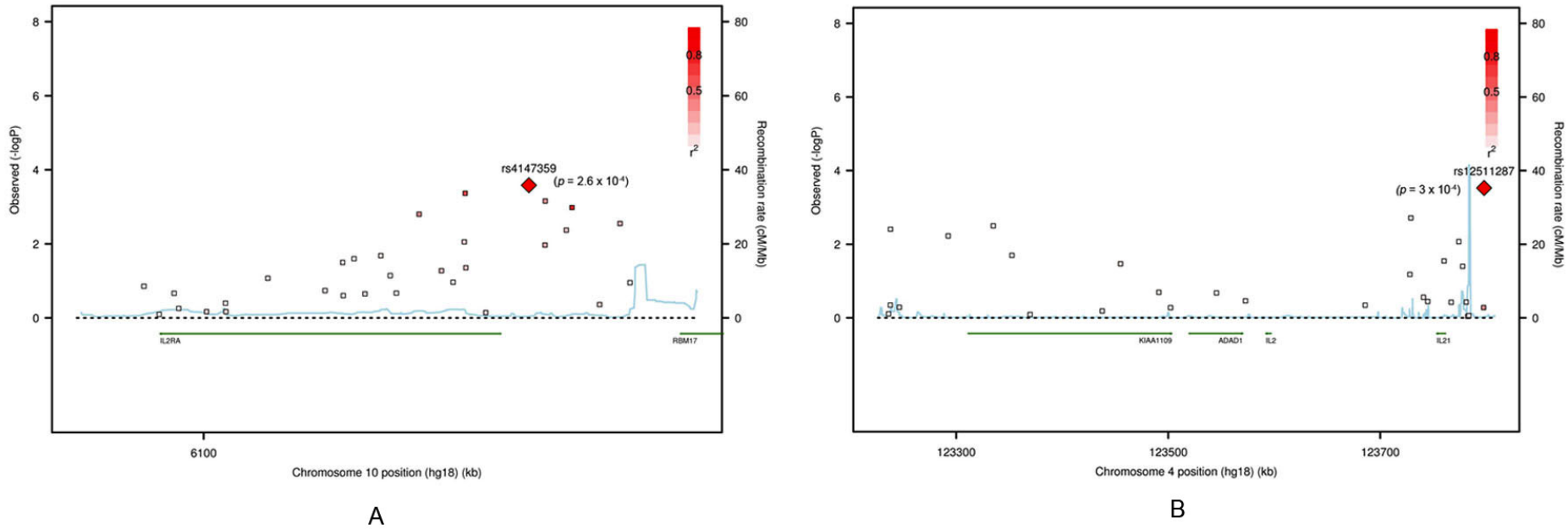


Figure 6.1 Regional association plots at 10p15 (IL2RA) and 4q27 (IL-2/IL-21). The plots show association results at 10p15 (IL2RA) (Panel A) and 4q27 (IL-2/IL-21) (Panel B) for all the SNPs analyszed in the study. Results are shown as –log of the p-values plotted against the marker position (NCBI B36). All the SNPs are denoted by a square with red colour gradient as the measure of LD in relation to the SNP with the lowest p-value (denoted by a diamond) in each panel. The blue lines denote the rate of recombination along the chromosomal region based on data from the HapMap project. The plots were created using SNAP (201). LD = linkage disequilibrium.

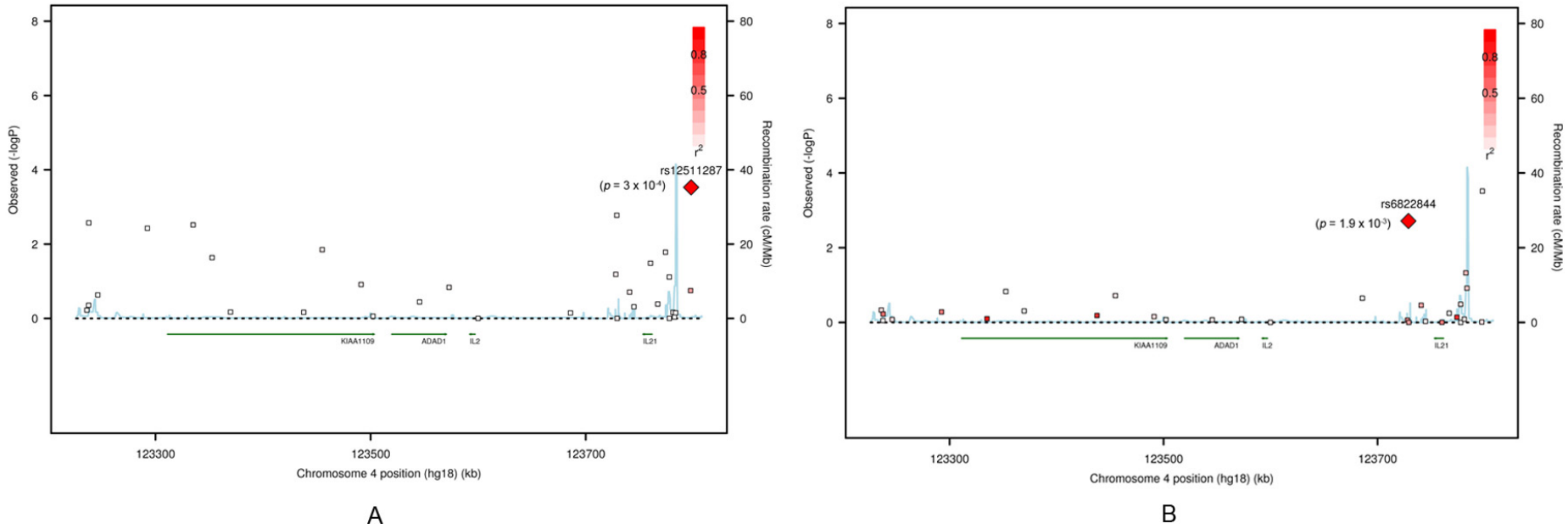


Figure 6.2. **Regional association plots for all the SNPs at 4q27 (*IL2/IL21*) conditioned on rs12511287 (A) and rs6822844 (B).** The plots show association results from a conditional analysis of the two most associated SNPs at 4q27 locus. Panel A: When conditioned on rs12511287 (most associated SNP at 4q27), the strength of association of all the remaining SNPs remains unchanged as compared to the original results. Panel B: Conditional analysis on rs6822844 weakens the effect of all other SNPs except for rs12511287 suggesting an independent effect of these two SNPs at this locus. All the SNPs are denoted by a square with red colour gradient as the measure of LD with the lead SNP (denoted by a diamond) in each panel. The blue lines denote the rate of recombination along the chromosomal region based on data from the HapMap project. The blue peak just upstream of *IL-21* gene denotes the point of recombination peak at this locus. The plots were created using SNAP.

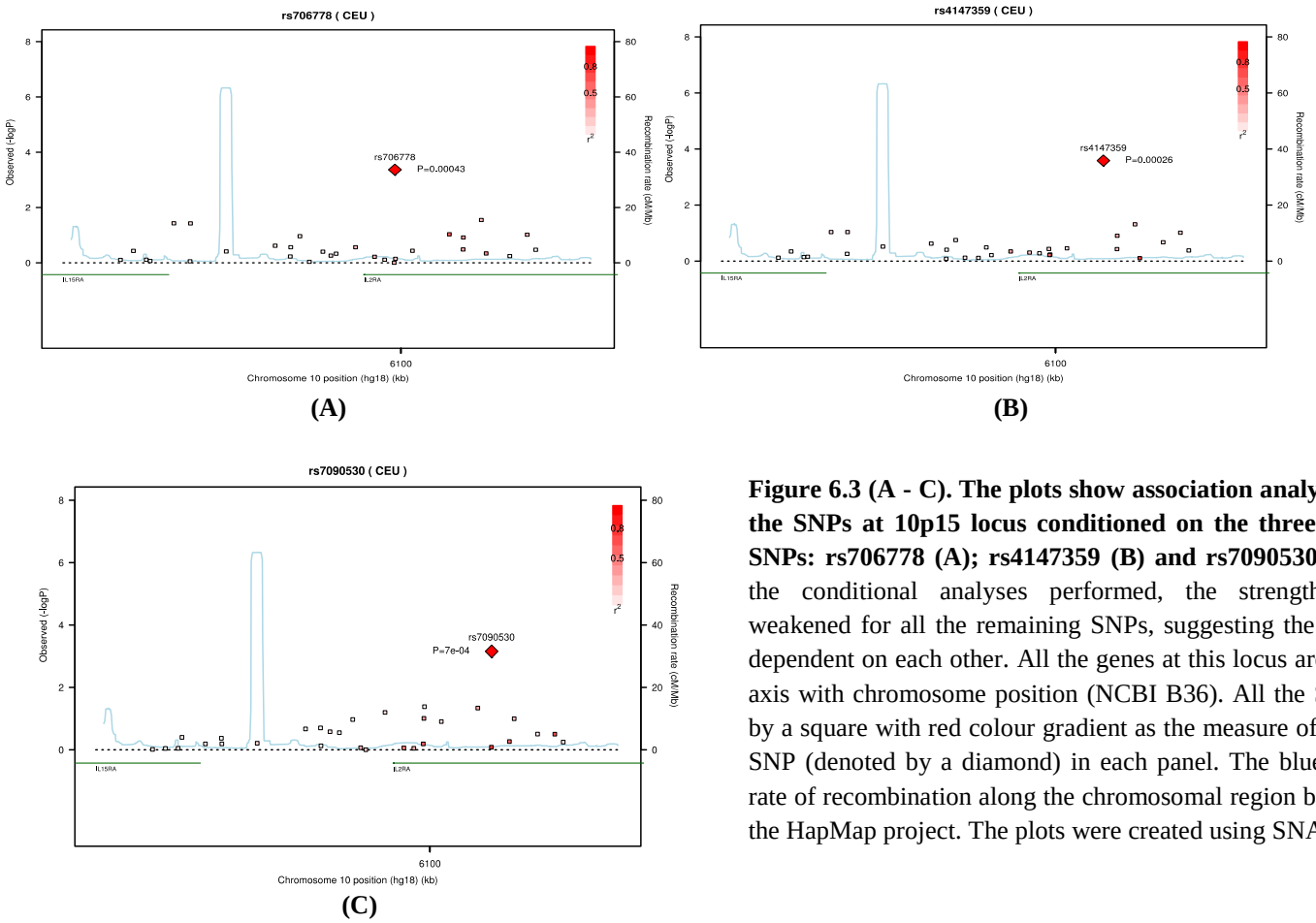


Figure 6.3 (A - C). The plots show association analysis results for all the SNPs at 10p15 locus conditioned on the three most associated SNPs: rs706778 (A); rs4147359 (B) and rs7090530 (C). For each of the conditional analyses performed, the strength of association weakened for all the remaining SNPs, suggesting the association to be dependent on each other. All the genes at this locus are shown on the x-axis with chromosome position (NCBI B36). All the SNPs are denoted by a square with red colour gradient as the measure of LD with the lead SNP (denoted by a diamond) in each panel. The blue lines denote the rate of recombination along the chromosomal region based on data from the HapMap project. The plots were created using SNAP.

Chapter 7: Genome wide association study in the UK PSC cohort.

A genome wide association study (GWAS) was performed in a well-characterised cohort of UK PSC patients in 2011. At the time of undertaking this study, it was the largest GWAS in PSC patients.

I collected and plated all the DNA samples, which were genotyped at the Wellcome Trust Sanger Institute. I performed the statistical analysis for this study, under the supervision of Dr George Mells (Research fellow, Academic department of medical genetics, Cambridge) and Dr Carl Anderson (Genetic and Genomic Epidemiology Unit, Wellcome Trust Sanger Institute, Cambridge).

Aims and objectives

The main aim of this study was to find genetic markers associated with PSC and to identify potential susceptibility loci that could be implicated in PSC pathogenesis.

Methods

Case ascertainment and description

A total of 1020 PSC cases were included in the discovery panel and cases were recruited from the UK including all transplant centres. All PSC cases were diagnosed using standard diagnostic criteria as described in chapter 2 (page 59), and were mostly of Caucasian ethnicity. The characteristics of the cohort are shown in Table 7.1.

Table 7.1. Characteristics of patients included in the GWAS (n = 1020).

Ethnic origin	100% Caucasian (97.5% British)
Median age at recruitment	58 (range, 16 – 86) years
Gender	Males = 646, Females = 374 (M:F = 1.7:1)
Inflammatory bowel disease	Total = 652 (64%); UC = 568 (87%), CD = 80 (12%), IC = 4 (1%)
Liver transplant recipients	252 (24.5%)

UC – Ulcerative colitis; CD – Crohn's disease; IC – Indeterminate colitis

Controls

The control population in the discovery panel comprised 2930 individuals from the 1958 British Birth Cohort (BBC) and 2737 individuals from the National Blood Service (NBS) cohort. Their DNA samples were genotyped as part of the Wellcome Trust Case Control Consortium 2 (WTCCC2) project. A total of 5667 controls were available for the discovery panel.

DNA preparation

DNA preparation and plating steps were as described in chapter 2 (page 62).

Genotyping platform

PSC cases were genotyped using the Illumina HumanOmni2.5-8 BeadChip (Illumina, San Diego, CA, USA) and called using the GenCall algorithm implemented in GenomeStudio. Genotyping was done at the Wellcome Trust Sanger Institute (Hinxton, UK).

We did not genotype the control population. Instead, control genotype data were obtained for 5667 individuals from the WTCCC2, who had previously been genotyped using the Illumina Human1M-Duo array (Illumina, San Diego, California).

Statistical analysis

Quality control

Cases and Controls:

Stringent quality control steps were applied to the cases and controls in the discovery panel to identify and remove individuals with:

a) Discordant sex information
b) Genotype failure rate ≥ 0.03
c) Heterozygosity rate (mean ± 3 standard deviations (s.d.))
d) Identity by descent > 10% (i.e. related), or
e) Divergent or non-European ancestry (using principal component analysis (PCA)).

Ancestry clustering was done using HapMap 3 reference samples, into three population groups (Figure 7.1a):

a) CEU (European) - Utah residents with Northern and Western European ancestry from the CEPH collection
b) CHB + JPT (Asian) - Han Chinese in Beijing, China + Japanese in Tokyo, Japan
c) YRI (African) - Yoruba in Ibadan, Nigeria

Marker (SNPs):

SNPs with a call rate < 97%, HWE p-value < 1×10^{-6} and MAF < 0.05 were removed.

Association analysis

Case-control association tests were conducted using 1-degree-of-freedom Cochran-Armitage trend tests implemented in PLINK v1.07. Cluster plots for

all SNPs with p-value < 1×10^{-4} were examined to check the accuracy of genotype calling.

Replication

Replication cohort

This comprised a cohort of well-characterised 332 Scandinavian and 383 German patients with PSC who had been genotyped on Affymetrix Genome-Wide Human SNP Array 6.0 (Affymetrix) for a previous GWAS study (151).

Replication analysis

As part of the replication analysis, we did not genotype the most associated SNPs (outside of the HLA region) from the discovery panel in an independent validation cohort. Instead, we used summary statistics from a GWAS study done in an independent European (Scandinavian and German) PSC cohort, and performed an *in-silico* replication analysis to identify SNPs associated with PSC. Details regarding the *in-silico* replication analysis are described below.

Selection of SNPs for in-silico replication analysis

All SNPs with p-value < 1×10^{-4} were identified and susceptibility locus defined for each associated region by finding a 0.1 cM interval on either side of the most associated SNP. Melum *et al.* provided summary statistics for these SNPs, for *in-silico* replication and meta-analysis.

Allele/strand matching of SNPs between UK and replication data

The GWAS data from UK PSC and replication cohorts were merged using PLINK. When merging two datasets, it is important that the two sets of SNPs are concordant in terms of positive or negative strand. We flipped the strands for SNPs with a mismatch and recoded the allele information (in PLINK) to have full concordance between the discovery panel and replication datasets.

Meta-analysis

Meta-analysis was performed using GWAS data and summary statistics from the replication cohort using Metagen (inverse variance method) package in the R statistical software package (R v 2.13.1). To account for differences between the population groups, we used a random effects model even though a test for heterogeneity was not statistically significant for any of the associated SNPs. In the combined meta-analysis, genome wide significance threshold (p-value $< 5 \times 10^{-8}$) was used to confirm disease association.

We did not perform any sub-group analysis comparing IBD risk loci because of small number of patients in each sub-group cohort, which would reduce the power of such type of study to detect true association.

Results

QC results

Individual QC

After QC steps, 79/1020 cases were removed. 143/2930 controls were removed from the BBC controls and 111/2737 controls were removed from the NBS controls (total number of controls removed = 254 out of 5667). The discovery panel thus comprised 941 PSC cases and 5413 healthy controls. Table 7.2 lists individuals removed for each of the QC steps in cases and controls.

Figure 7.1 (a) (page 142) shows the ancestry clustering of PSC cases based on the genome wide data and HapMap 3 reference samples. Figure 7.1 (b) (page 142) shows the thresholds applied to remove PSC cases, based on the first and second principal component scores.

Supplementary figures 7.1 (a) and (b), show the thresholds applied to remove BBC and NBS controls, based on the first and second principal component scores respectively.

Table 7.2. Individuals failing quality control thresholds.

Sample	Quality control criteria					Total*
	Gender	Genotype failure rate	Heterozygosity rate	Relatedness (IBD)	Non-European	
PSC cases	1	3	17	17	67	79
BBC controls	5	83	118	1	92	143
NBS controls	12	58	69	2	72	111

IBD – identity by descent; BBC – British birth cohort; NBS – national blood service.
*Some individuals failed quality control for more than one criteria.

Marker (SNP) QC

After applying QC thresholds in each of the case and control cohorts, 465,662 SNPs were taken forward for analysis.

Supplementary figures 7.2 (a - c) show the histograms of missing genotype data across all PSC cases and controls (BBC and NBS cohorts) that passed individual quality control.

Discovery panel results

Following QC, 465,662 SNPs were analysed in 941 PSC cases and 5413 controls. The strongest genome-wide significant association was seen at the HLA loci. Figure 7.2 (page 143) shows the Manhattan plot of GWAS results in the discovery panel.

Outside of the known HLA association, **26 loci** (across 16 chromosomes) were identified, where one or more markers showed at least suggestive (p-value $\leq 1 \times 10^{-4}$) association in the discovery panel (supplementary table 7.1 shows the list of all SNPs with suggestive association). Regional association plots were

plotted using SNAP (201) for each of the 26 suggestive loci, to examine the distribution of association across all the SNPs at each locus and identify any spurious associations (supplementary figures 7.3(a – z)).

Replication results

A total of 96 SNPs meeting criteria for suggestive association (p-value < 1 x 10^{-4}), were considered for *in-silico* replication analysis (supplementary table 7.2). As the genotyping platforms used in the discovery panel (Illumina HumanOmni2.5-8 BeadChip) and replication cohort (Affymetrix) were different, genotype data were available for 86/96 SNPs identified for *in-silico* replication. Table 7.3 shows the list of all SNPs for which summary statistics from the replication cohort were available and analysed in the meta-analysis.

Meta-analysis results

The meta-analysis of summary statistics from the UK-PSC data and summary statistics provided by Melum, *et al.* (151) identified genome wide significant association ($p < 5 \times 10^{-8}$) at three loci: 3p21, 10p15 and 12q24 in the combined analysis (Table 7.4). The direction of effect of the minor allele was same in both the cohorts. Test for heterogeneity at these loci using the Cochrane's Q statistic (in PLINK) was not significant (i.e. p-value > 0.05).

Table 7.4 shows the PSC risk loci that reached genome-wide significance in the combined meta-analysis. Supplementary table 7.3 shows the combined analysis results for all the 86 SNPs that were analysed in the meta-analysis.

Melum, *et al*. reported the association at 3p21, with *MST1* as the most likely candidate gene (151). At 10p15, *IL2RA* is the most likely candidate gene, which I have reported in chapter 6 (page 109). There are two plausible candidate genes at 12q24: *SH2B3* and *ATXN2* and are discussed below.

Table 7.3. List of all SNPs (n = 86) taken forward for *in-silico* replication analysis.

CHR	SNP	Position (GRCh37.5); build 19	CHR	SNP	Position (GRCh37.5); build 19
1	rs11121203	8648240	10	rs7090530	6110875
1	rs1953827	8679848	10	rs10905669	6092093
1	rs4908760	8526142	10	rs7072793	6106266
1	rs7553298	8770883	10	rs3763700	6181709
1	rs10492965	8598127	10	rs12412095	6113523
1	rs10492966	8598005	10	rs2387015	6215257
1	rs11121182	8522553	10	rs3814195	6213960
1	rs7520572	8871690	10	rs11255072	5806248
2	rs16823732	145204758	11	rs836141	34537316
2	rs12105918	145208193	11	rs11221265	128236936
2	rs10192562	145184316	11	rs7928323	38841059
2	rs13389578	145216048	11	rs2068461	38831942
2	rs11883829	145519623	11	rs11034972	38860022
2	rs13032840	145506108	11	rs10837055	38845174
3	rs13070798	48705934	11	rs7936255	38998451
3	rs9836462	48712791	11	rs7116065	38979245
3	rs12107418	48689787	11	rs2912573	38950513
3	rs11719291	48735706	12	rs10774625	111910219
3	rs13063312	48661985	12	rs1265566	111716376
3	rs13324142	48669447	12	rs11065987	112072424
3	rs2276852	48666923	12	rs17696736	112486818
4	rs13132933	123010587	12	rs11066320	112906415
4	rs6822844	123509421	12	rs11066301	112871372
4	rs13151961	123115502	12	rs770460	78132040
4	rs11938795	123073009	14	rs17119456	84485390
4	rs11734090	123228113	14	rs1257641	99480395
4	rs7684187	123341159	14	rs1015277	84438074
4	rs10027390	123368516	14	rs17119553	84509670
4	rs345367	86754638	14	rs1824343	98920773
4	rs4693142	87252259	16	rs8060332	2892770
4	rs7689808	87254477	16	rs3810801	2892370
4	rs434193	86253489	16	rs8047221	2897372
5	rs419119	6021846	18	rs7229974	28890717
5	rs1567520	118159871	18	rs9952617	28881801
5	rs17132677	118175674	18	rs2114270	28885116
5	rs17648108	177831556	20	rs1885082	17593984
5	rs6874399	118225616	20	rs4814628	17593315
5	rs2029036	118275869	20	rs13734	17594729
5	rs12109252	118267633	20	rs2064726	50715685
8	rs2617094	4456167	21	rs1893592	43855067
8	rs17070773	4463359	21	rs378108	40469520
9	rs7027092	95812707	21	rs2836878	40465534
10	rs3118470	6101713	21	rs2836881	40466299

Table 7.4. Results for SNPs showing suggestive association in the UK-PSC cohort and genome-wide significance in the meta-analysis.

Chr	SNP	Candidate gene	Alleles (minor/ major)	MAF (case/ control)	*P*-value (UK cohort)	OR (95% CI) (UK cohort)	*P*-value (replication cohort)	OR (95% CI) (replication cohort)	Meta-analysis (random effects)	
									OR (combined)	*P*-value (combined)
3	rs13070798	*MST1*	C/T	0.14/0.10	1.21×10^{-6}	1.39 (1.21 - 1.59)	1.9×10^{-3}	1.32 (1.10 – 1.57)	1.36 (1.22 – 1.51)	1.07×10^{-8}
3	rs9836462	*MST1*	G/A	0.14/0.10	3.63×10^{-6}	1.36 (1.19 -1.56)	1.8×10^{-3}	1.33 (1.11 – 1.58)	1.35 (1.21 – 1.50)	2.73×10^{-8}
10	rs3118470	*IL2RA*	C/T	0.36/0.31	1.65×10^{-5}	1.23 (1.12 - 1.35)	2.01×10^{-5}	1.33 (1.16 – 1.51)	1.26 (1.17 – 1.36)	2.27×10^{-9}
12	rs10774625	*SH2B3/ ATXN2*	G/A	0.43/0.49	1.22×10^{-7}	0.78 (0.71 - 0.85)	1.2×10^{-3}	0.81 (0.71 – 0.92)	0.79 (0.73 – 0.85)	6.79×10^{-10}
12	rs1265566	*SH2B3/ ATXN2*	C/T	0.26/0.31	6.06×10^{-6}	0.79 (0.71 – 0.87)	2.0×10^{-3}	0.79 (0.69 – 0.92)	0.79 (0.73 – 0.86)	4.46×10^{-8}

Meta-analysis was performed by combining summary statistics from a recent PSC GWAS (151), using PLINK. A random effects model was used for meta-analysis even though a test for heterogeneity was not statistically significant for any of the associated SNPs. All the reported odds ratios are with reference to minor allele vs. major allele. Chr – Chromosome, MAF – minor allele frequency, OR – odds ratio, CI – confidence interval.

Discussion

This genome wide association study confirms the reported genome-wide association at 3p21 (*MST1*) and the association at 10p15 (*IL2RA*). In addition, this study has identified a novel association at 12q24 in a relatively modest size cohort of PSC patients.

Melum *et al.* (151) and Folseraas *et al.* (152) have reported genome-wide association at 2q13 (BCL2L11) and 1p36 (*MMEL1, TNFRSF14*) respectively. However, we did not find significant associations at these loci in our study.

I have discussed the plausible candidate genes at 3p21 (*MST1*) and 10p15 (*IL2RA)* and their functional role in chapter 6 (page 125).

The association at 12q24 is novel and confirmed in a homogeneous cohort of UK PSC patients. There are two plausible candidate genes at this locus: *SH2B3* and *ATXN2*.

SH2B3 is a protein-coding gene for a member of the SH2B adaptor family of proteins involved in a range of signaling activities by growth factor and cytokine receptors. The encoded protein is a key negative regulator of cytokine signaling pathways, T cell activation, tumour necrosis factor and Janus kinase 2 and 3 (JAK2/3) signaling and plays a critical role in hematopoiesis. Interestingly, an association at 12q24 (*SH2B3)* is also seen in primary biliary cholangitis (PBC) (202) and autoimmune hepatitis (AIH) (203).

In addition, mutations in this gene are linked strongly to myeloproliferative disorders, autoimmune diseases such as coeliac disease and insulin-dependent diabetes mellitus and inflammatory syndromes, where both immune and vascular cells display a role (204).

The functional effect of *SH2B3* (rs3184504) genotype on inflammatory cytokine response, in relation to stimulation by lipopolysaccharide (LPS) and muramyl dipeptide (MDP - which is a specific ligand of the pattern-recognition receptor NOD2) - has recently been studied in patients with coeliac disease (205). The cells isolated from individuals homozygous for the *SH2B3* risk allele, showed a significantly increased cytokine production, thereby implicating an inhibitory effect of the *SH2B3* protein on the MDP-NOD2-RIP signalling pathway, so that *SH2B3* may be an important autoimmune locus, especially in immune-mediated liver diseases.

ATXN2 is involved in regulating mRNA translation through its interactions with the poly(A)-binding protein. It is involved in the formation of stress granules and P-bodies, which also plays a role in RNA regulation (206). Loss-of-function mutations in this gene may be associated with susceptibility to type I diabetes, obesity and hypertension.

The genetic risk loci found thus far in PSC, could suggest involvement of an immune mediated process in pathogenesis of PSC. However, unlike GWAS in other diseases (e.g. primary biliary cholangitis) (207), it has not yet been possible to identify any clear biological pathways that can be implicated in

PSC causation. Further immunological studies are needed to elucidate gain or loss of function of the implicated candidate genes in patients with PSC, matched with healthy and disease controls (e.g. PBC or AIH). Until then, the role of such candidate genes remains speculative.

The main limitation of this study was the relatively small discovery and replication sample size. Due to the low incidence and prevalence of PSC, there is a need for large-scale genome-wide association studies to be undertaken in collaboration with other international cohorts to identify additional susceptibility loci. This could shed light on yet unknown disease causing pathways in PSC, which could be targeted for functional studies and therapy.

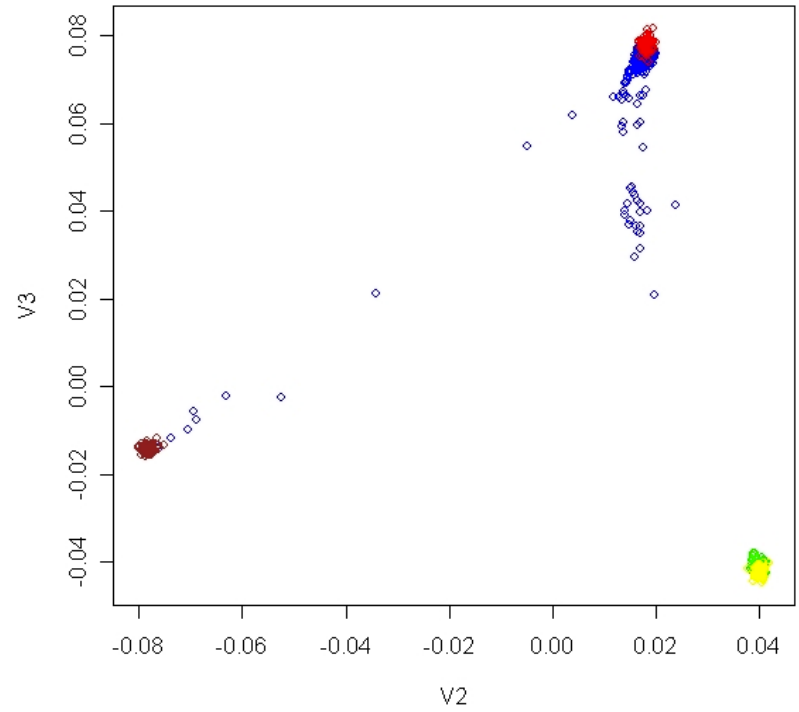


Figure 7.1(a). Ancestry clustering based on genome-wide association data in all PSC cases (blue circles). HapMap 3 reference samples: CEU (red), CHB+JPT (green/yellow) and YRI (brown).

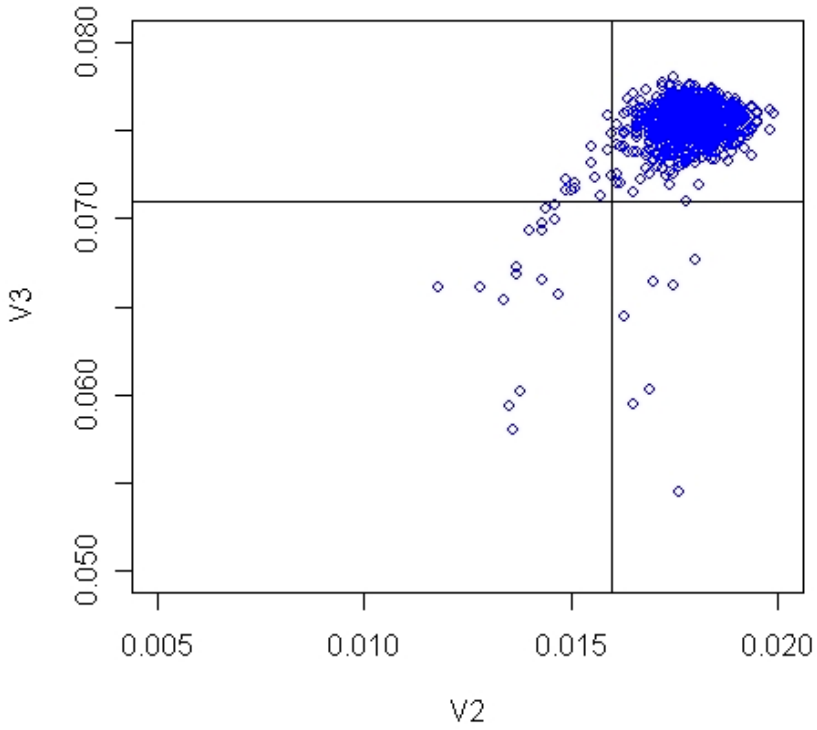


Figure 7.1(b). PSC cases (blue circles) with first principal component score (V2) < 0.16 (vertical line) and second principal component score (V3) < 0.072 (horizontal line) were removed.

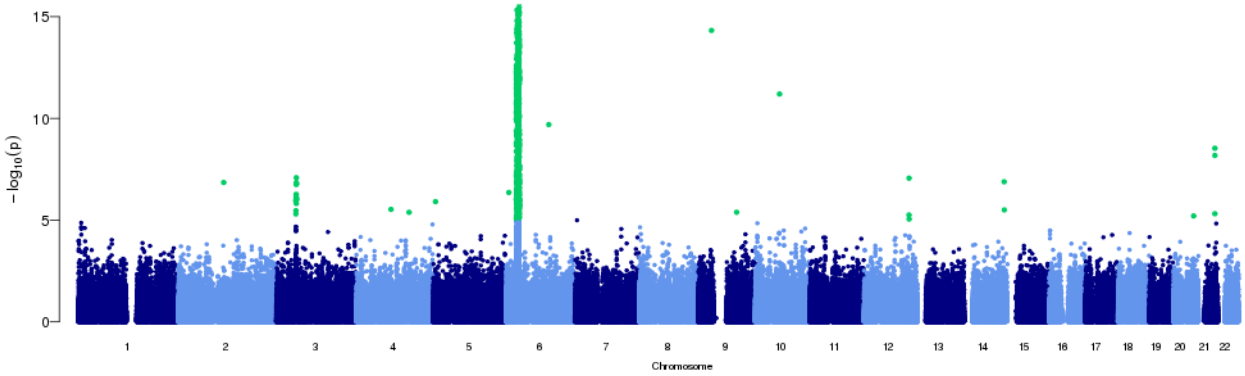


Figure 7.2. Manhattan plot of simple $\chi 2$ allelic test of association p-values in the discovery panel. The plot shows –log10 p-values for each SNP against each chromosomal location (blue and green dots). Green points indicate SNPs with p-values $\leq 1 \times 10^{-5}$.

Chapter 8: Discussion.

The data presented in the previous chapters have already been discussed at the end of each chapter.

The main focus of this chapter is to discuss the:

1) International collaborative efforts to identify additional risk loci in PSC
2) Summary of genetic risk in PSC
3) Role of genetic risk factors in predicting disease outcome
4) Limitations of genetic association studies

1) International collaboration to identify additional susceptibility loci in PSC

One of the strategies to increase the power of detecting additional susceptibility loci in genome wide association studies is to have a large sample size of cases and controls. Commonly occurring diseases such as inflammatory bowel disease (IBD) and type 1 diabetes mellitus, have been able to recruit > 10,000 cases for such studies.

Due to the low prevalence of PSC, it has not been possible to undertake large-scale genetic association studies in PSC. During the period of my MD thesis this has driven international groups to collaborate and undertake studies, combining data from the PSC cohorts across Europe and U.S.A.

The phenotype and genotype data from the UK PSC cohort is an invaluable resource for such collaborative and other downstream studies. It represents the largest homogeneous cohort of PSC patients from a single nation; two studies undertaken in collaboration with other European ancestry PSC cohorts from Europe and USA have identified 13 novel loci at genome-wide significance (208, 209). These studies are discussed briefly below.

Please note that the lead authors of respective studies undertook the primary analyses described in studies below. UK PSC and I contributed by recruiting PSC patients from the UK, providing genetic and phenotypic data and were involved in writing the final drafts.

A) Immunochip study

Immunochip is a custom-made Illumina Infinium genotyping chip, containing 196,524 polymorphisms (718 small insertion deletions, 195,806 SNPs) spread across 186 known immune-mediated disease loci (210). It was designed to map the genetic overlap and variance between various immune-mediated diseases and fine-map the established genetic risk loci.

Liu *et al.* compared 3,789 European ancestry PSC cases (including 1,033 UK PSC cases) to 25,079 healthy population controls across 130,422 SNPs, which were genotyped using the Immunochip (208). 72% of PSC patients had concomitant IBD (UC = 58%; CD = 11%; Indeterminate = 3%). Twelve (non-HLA) genome-wide significant loci were identified, of which nine were novel associations (Table 8.1). The most significantly associated SNP at seven of the nine loci, was the same SNP or in strong linkage disequilibrium ($r^2 > 0.80$) with the original association reports for another immune-mediated disease.

Identifying candidate genes at associated loci

Various methods can be used to prioritise candidate genes in GWAS, such as functional annotation of risk loci using data from the ENCODE project, construction of networks based on functional similarity measures known as protein-protein interactions (DAPPLE) and published literature (GRAIL) as well as expression quantitative loci (eQTL) analysis. The functional role of the plausible candidate genes identified in PSC is highlighted in Box 2 (page 161) and discussed further in the section: ‘summary of genetic risk in PSC’ (page 150).

Table 8.1. Association results for twelve non-HLA genome-wide significant associations for PSC in the Immunochip study.

Chr	SNP	Candidate gene	Risk allele	*P*-value (UK cohort)	OR (95% CI) (UK cohort)
1p36	rs3748816	*MMEL1, TNFRSF14*	A	7.41E-12	1.21 (1.14 – 1.27)
2q33	**rs7426056**	***CD28***	**A**	**1.89E-20**	**1.30 (1.23 – 1.37)**
3p21	rs3197999	*MST1*	A	2.45E-26	1.33 (1.26 – 1.40)
4q27	**rs13140464**	***IL2/IL21***	**C**	**8.87E-13**	**1.30 (1.21 – 1.40)**
6q15	**rs56258221**	***BACH2***	**G**	**8.36E-12**	**1.23 (1.16 – 1.31)**
10p15	rs4147359	*IL2RA*	A	8.19E-17	1.24 (1.18 – 1.30)
11q23	**rs7937682**	***SIK2***	**G**	**3.17E-09**	**1.17 (1.11 – 1.24)**
12q13	**rs11168249**	***HDAC7***	**G**	**5.49E-09**	**1.15 (1.10 – 1.21)**
12q24	**rs3184504**	***SH2B3, ATXN2***	**A**	**5.91E-11**	**1.18 (1.12 – 1.24)**
18q22	**rs1788097**	***CD226***	**A**	**3.06E-08**	**1.15 (1.10 – 1.21)**
19q13	**rs60652743**	***PRKD2, STRN4***	**A**	**6.51E-10**	**1.25 (1.16 – 1.34)**
21q22	**rs2836883**	***PSMG1***	**G**	**3.19E-17**	**1.28 (1.21 – 1.36)**

Chr – Chromosome; OR – odds ratio; CI – 95% confidence interval; Nine novel SNPs associated with PSC in the Immunochip study are highlighted in bold.

The majority of the patients in this cohort had concomitant IBD (72%). However, despite that, only six of the twelve PSC risk loci were found to be associated with IBD in the International IBD Genetics Consortium (IIBDGC) Immunochip analysis (211). Upon comparing the genetic architecture of PSC and IBD risk loci using the Immunochip study data, no particular functional sub cluster of PSC genes was associated with IBD and vice versa. This suggested that the genetic susceptibility to PSC is not defined entirely by the risk factors related to IBD and that the new immune mediated risk loci may have an important role in PSC pathophysiology.

B) GWAS in PSC and quantification of genetic relationship with IBD

PSC is strongly associated with IBD (mainly UC), but the cause of this phenotypic association has not been clearly identified. Data from the Immunochip study did not show any significant genetic overlap between PSC-

IBD, at least at risk loci, which are associated with other immune mediated diseases.

Ji *et al.* (209) performed the largest GWAS in PSC to date and also studied the genome-wide genetic correlation between PSC and IBD (UC and CD). The discovery panel comprised of 2,871 PSC cases (1,227 UK-PSC cases) and 12,019 population controls. IBD was present in 65% of cases (UC = 53%; CD = 9% and IC = 3%).

7,891,602 SNPs were tested in the discovery panel and forty SNPs were taken forward for replication study in an independent cohort of 1,925 PSC cases and 7,936 population controls. In the replication panel, the prevalence of IBD was 51% (UC = 39%; CD = 9% and IC = 3%). IBD status was not known in 14% and 27% of discovery and replication panels respectively.

Association was confirmed at fifteen of the previously known loci in PSC and four novel loci were identified at 3p13, 11q13, 16p13 and 21q22, with the most likely candidate genes as *FOXP1, CCDC88B, CLEC16A* and *UBASH3A* respectively (table 8.2). The associated SNPs at 11q13, 16p13 and 21q22, were found to be in high linkage disequilibrium with variants associated with other immune-mediated diseases, suggesting a shared immune genetic risk with PSC. This could potentially explain the increased prevalence of some forms of immune-mediated disease in PSC. However, there has been no prior reported association at 3p13 (*FOXP1*) for any other immune mediated disease.

The most associated SNPs at each locus were evaluated for their functional role using eQTL databases and Genome Wide Annotation of Variants (GWAVA) online tool (212). The most associated SNP at 21q22 (rs1893592) was the most strongly associated eQTL of *UBASH3A* gene, which is involved in regulation of T-cell signaling. The C allele was found to be associated with a reduced risk of PSC (frequency in controls = 27.8%).

Table 8.2. GWAS summary statistics of previous and new loci (highlighted in bold) associated with PSC (adapted from Ji *et al*. (209)).

SNP	Risk Allele	Chr	Gene	GWAS			Replication			GWAS + Replication		
				RAF	OR	P-value	RAF	OR	P-value	RAF	OR	P-value
rs3748816	A	1p36	*MMEL1*	0.67	1.18	3.54E-07	0.66	1.23	2.27E-07	0.66	1.20	5.17E-13
rs72837826	T	2q13	*BCL2L11*	0.11	1.35	1.26E-09	0.12	1.20	1.30E-03	0.12	1.29	2.36E-11
rs7426056	A	2q33	*CD28*	0.25	1.21	3.26E-08	0.24	1.31	4.26E-10	0.25	1.25	2.12E-16
rs3749171	T	2q37	*GPR35*	0.18	1.16	3.44E-04	NA	NA	NA	0.18	1.16	3.47E-04
rs80060485	**C**	**3p13**	***FOXP1***	**0.07**	**1.41**	**8.54E-09**	**0.07**	**1.49**	**4.67E-08**	**0.07**	**1.44**	**2.62E-15**
rs3197999	A	3p21	*MST1*	0.29	1.30	2.60E-13	0.31	1.37	1.59E-14	0.30	1.33	5.11E-26
rs13140464	G	4q27	*IL2-IL21*	0.84	1.28	5.17E-09	0.84	1.27	4.53E-06	0.84	1.28	1.19E-13
rs56258221	C	6q15	*BACH2*	0.19	1.23	6.21E-07	0.18	1.18	4.80E-04	0.18	1.21	1.41E-09
rs4147359	A	10p15	*IL2RA*	0.36	1.24	4.06E-13	0.37	1.18	2.04E-05	0.36	1.22	7.54E-17
rs663743	**G**	**11q13**	***CCDC88B***	**0.67**	**1.18**	**8.42E-08**	**0.66**	**1.22**	**4.44E-07**	**0.66**	**1.20**	**2.24E-13**
rs7937682	C	11q23	*SIK2*	0.28	1.13	1.85E-04	0.27	1.16	6.66E-04	0.28	1.14	4.77E-07
rs11168249	C	12q13	*HDAC7*	NA	NA	NA	0.48	1.12	3.36E-03	0.48	1.12	3.33E-03
rs3184504	T	12q24	*SH2B3*	0.50	1.20	5.05E-10	0.48	1.15	1.20E-04	0.50	1.18	4.27E-13
rs725613	**T**	**16p13**	***CLEC16A***	**0.65**	**1.22**	**5.50E-10**	**0.66**	**1.17**	**9.52E-05**	**0.65**	**1.20**	**3.59E-13**
rs1452787	A	18q21	*TCF4*	NA	NA	NA	0.72	1.02	6.88E-01	0.72	1.02	6.87E-01
rs1788097	T	18q22	*CD226*	0.49	1.19	9.73E-07	0.49	1.20	1.41E-06	0.49	1.19	6.58E-12
rs60652743	A	19q13	*FUT2, PRKD2, STRN4*	0.84	1.30	1.01E-07	0.84	1.27	3.91E-06	0.84	1.28	1.99E-12
rs2836883	G	21q22	*PSMG1*	0.74	1.35	5.40E-14	0.73	1.12	7.77E-03	0.74	1.23	4.21E-13
rs1893592	**A**	**21q22**	***UBASH3A***	**0.73**	**1.22**	**1.90E-07**	**0.72**	**1.22**	**2.42E-06**	**0.73**	**1.22**	**2.19E-12**

Genetic relationship with IBD

The genetic relationship between PSC and IBD was studied across eighteen established PSC risk loci, to identify loci with evidence of shared or independent causal variants between the two phenotypes. Summary statistics from previous IBD association studies (from the IBD genetics consortium) were obtained for 17,647 UC cases, 20,550 CD cases and 48,485 controls of European ancestry (213).

Of the eighteen PSC risk loci, four (*BCL211*, *FOXP1*, *SIK2* and *UBASH3A*) have not previously been found to be associated at genome-wide significance with IBD. Of the remaining fourteen loci found previously to be associated with IBD, the lead SNPs at four loci (*IL2RA*, *CCDC88B*, *CLEC16A* and *PRKD2*) in the PSC and IBD studies, were found to be in low linkage disequilibrium ($r^2 < 0.2$), suggesting an independent association at these loci between the two diseases.

This would imply that even though there may be association at the same region of the genome for diseases occurring together, the causal variant might be different. Six of the fourteen loci associated with PSC and IBD displayed strong evidence of shared causal variant with UC, CD or both (*MST1*, *IL21*, *HDAC7*, *SH2B3*, *CD226* and *PSMG1)*.

The observed genetic correlation in this study between PSC and IBD is small and does not explain the extent of comorbidity between the two diseases, implicating other environmental factors and/or rare genetic variants not captured by this GWAS. The genetic dissimilarity also supports the notion that PSC-IBD might be an entity distinct from UC.

2) Summary of the genetic risk in PSC and role of identified candidate genes

Other than the known HLA associations, genetic association studies in PSC have now confirmed 22 susceptibility loci with genome-wide significance. Table 8.3 shows the timeline of all the non-HLA susceptibility loci associated with PSC.

Table 8.3. Timeline of all non-HLA susceptibility loci associated with PSC.

Year	Chr	Lead SNP	Candidate gene	UK PSC data included	Ref
2011	2q13	rs6720394	*BCL2L11*	No	(151)
	3p21	rs3197999	*MST1*	No	
2012	10p15	rs4147359	*IL2RA*	**Yes**	(214)
2012	1p36	rs3748816	*MMEL1, TNFRSF14*	No	(152)
2013	2q33	rs7426056	*CD28, CTLA4*	**Yes**	(208)
	4q27	rs13140464	*IL2, IL21*		
	6q15	rs56258221	*BACH2*		
	11q23	rs7937682	*SIK2*		
	12q13	rs11168249	*HDAC7*		
	12q24	rs3184504	*SH2B3, ATXN2*		
	18q22	rs1788097	*CD226*		
	19q13	rs60652743	*FUT2, PRKD2, STRN4*		
	21q22	rs2836883	*PSMG1*		
2013	2q37	rs3749171	*GPR35*	No	(215)
	18q21	rs1452787	*TCF4*		
2016	2q36	rs7556897	*CCL20*	No	(216)
	4q24	rs3774937	*NFKB1*		
	12q23	rs12369214	*RFX4, RIC8B*		
2017	3p13	rs80060485	*FOXP1*	**Yes**	(209)
	11q13	rs663743	*CCDC88B*		
	16p13	Rs725613	*CLEC16A, SOCS1*		
	21q22	rs1893592	*UBASH3A*		

Associations between PSC and the HLA complex were covered in the chapter 1 (page 51). The significance and pathophysiological implications of non-HLA susceptibility loci are discussed in the section below.

Non-HLA susceptibility loci in PSC

There are a total of 29 plausible candidate genes across the 22 loci found to be associated with PSC with genome-wide significance. Putative functions of these candidate genes are shown in Box 2 (page 161). A significant proportion of PSC susceptibility loci are shared with other autoimmune diseases, particularly at 4q27 (*Il2/IL21*), 6p15 (*BACH2*), 10p15 (*IL2RA*) and 12q24 (*SH2B3*). The pleiotropic nature of some of these loci could imply shared biological pathways in the development of immune-mediated diseases. It is important to note that often there is more than one plausible candidate gene at each associated loci and it is not possible to identify the exact causal variant at each locus due to the presence of high linkage disequilibrium across the genome.

Loci/Genes involved in immunological pathway

Most of the plausible candidate genes identified in PSC play a role in both innate and adaptive immune responses, supporting the hypothesis that PSC is an immune mediated disease.

Cftr$^{-/-}$ mice develop biliary inflammation via the LPS–TLR4–NF κ B axis, an innate immune pathway. *PRDX5*, *TGR5* and *PSMG1* might have a role in humans in regulating bile duct inflammatory changes in response to LPS along this pathway (217).

There is suggestive evidence of involvement of the IL-2 signalling pathway in PSC. IL2RA (CD25) forms part of the IL2 receptor and is constitutively expressed by the regulatory T cells (T_{reg}). Impairment of T_{reg} function is associated with autoimmunity and it is possible, that it could also cause immune dysregulation in PSC.

In a recent study, Sebode *et al.* (218) investigated the frequency of T_{reg} cells in patients with PSC and the association between their numbers with gene polymorphisms. Interestingly, the number of T_{reg} cells in peripheral blood as well as in the liver, was decreased in PSC (in comparison to PBC). Furthermore, the reduced peripheral T_{reg} numbers were significantly associated with homozygosity for the major allele of the SNP 'rs10905718' in the IL2RA gene. This study lends further support to implicate impaired T_{reg} function as a possible cause for immune dysregulation in PSC.

The association at 19q23, where *FUT2* is a potential candidate gene, is interesting. *FUT2* encodes the enzyme galactoside 2-alpha-L-fucosyltransferase 2, which regulates the type 1 oligosaccharide chains and is involved in antigen expression in the ABO blood group. In addition, it has a role in modifying the composition of the gut microbiome by affecting carbohydrate metabolism in the gut and has been be associated with Crohn's disease. This makes *FUT2* a potentially relevant candidate gene in PSC.

The association at 12q24 (*SH2B3)* is also seen in PBC and AIH, suggesting its predominant role in liver related immune functions and has been discussed in detail in the previous chapter (page 140). Association at *TNFRSF14* in PSC could imply a role for TNF signalling pathways in PSC, but further functional studies are needed.

The overwhelming association of the immune risk genes with PSC, as well as the high rate of PSC recurrence following a liver transplant, supports the role of peri-biliary immune dysregulation as one of the likely initial events in PSC development. It is plausible that certain environmental factors presented via the gut, act as the initial trigger for this immune activation, but these are yet to be identified. In addition to the potential role of innate and adaptive immune responses, PSC risk genes also have been found to have a role in apoptosis (*BCL2L11, UBASH3A*), autophagy (*CLEC16A*) and metabolic function (*ATXN2*).

The functional role of genetic risk factors identified in the genetic association studies in PSC, is at best speculative. However, it forms the basis to undertake functional work at each associated risk loci, to determine gain or loss of putative gene function, identify biological pathways and develop targeted therapeutic interventions.

3) Role of genetic risk factors in predicting disease outcome

The association between genetic risk variants and disease severity in complex disease, including PSC has not been well studied. Using the Immunochip genotype data, Alberts *et al.* (219) studied the role of genetic variants with risk of progression and complications in PSC.

The phenotypic data collected included: sex, date of birth, PSC subtype (small or large duct), date of PSC diagnosis, intrahepatic and/or extrahepatic disease, dominant strictures, concomitant IBD and type of IBD, date of IBD diagnosis and smoking status. Follow-up data were also collected for: date and cause of death, date and indication of liver transplantation, occurrence and date of diagnosis of cholangiocarcinoma, colorectal cancer and gallbladder carcinoma. After quality control, a total of 130,422 SNPs were analysed in 3,402 PSC cases (UK PSC cases = 1,033). Clinical characteristics of the cohort including sub-phenotypes are shown in table 8.4.

Genome-wide multivariable Cox proportional hazards regression analysis was performed to estimate the effect of genetic variants on time-to-event analysis. The time-to-event liver transplant-free survival was defined as the time between PSC diagnosis and the composite endpoint of either liver transplantation or PSC-related death.

Table 8.4. Clinical characteristics of the PSC cohort (n = 3402).

Variable	Groups	Number (%)
Age at PSC diagnosis (median)		38.6 years old (28.0–50.1)
Sex	Male	2185 (64.7)
	Female	1193 (35.3)
Main diagnosis	PSC	3159 (94.6)
	Small duct PSC	75 (2.2)
	PSC with AIH overlap	107 (3.2)
Liver transplantation	Yes	874 (26.3)
	No	2444 (73.7)
	Missing	84 (2.5)
IBD	No IBD	816 (25.5)
	Ulcerative colitis	1940 (60.5)
	Crohn's disease	357 (11.1)
	IBD-U	93 (2.9)
Cholangiocarcinoma	Yes	188 (5.6)
	No	3147 (94.4)
Colorectal carcinoma	Yes	127 (4.3)
	No	2822 (95.7)
Gall bladder carcinoma	Yes	30 (1.0)
	No	2977 (99.0)
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)

In the sub-group genetic association analysis, no genome wide significant genetic associations were identified for the following binary PSC sub-phenotypes – small-duct PSC, cholangiocarcinoma or gall bladder carcinoma. However, upon comparing 107 patients with PSC and autoimmune hepatitis (PSC/AIH) with 3159 patients with PSC but without AIH overlap, a strong genetic association in the *HLA-DQB1* gene was identified for rs3891175 (p-value = 4.6×10^{-11}, OR = 2.41).

Interestingly, in the time to event analysis, genome-wide significant association was identified for rs853974 (p-value = 6.07×10^{-9}), which is located on chromosome 6. The homozygous AA genotype had a 2.14 (95% CI 1.66 to 2.76) increased hazard, indicating a 2.14 larger relative risk for need

for liver transplantation or for PSC-related death compared with GG homozygous genotype.

The most likely candidate gene at this locus is the *R-Spondin 3 (RSPO3)* gene, and RNA sequencing on healthy and cholestatic cholangiocytes (from normal C57BL/6 mice), showed an increased *RSPO3* expression in the cholangiocytes (healthy > cholestatic), compared to any other organs.

Interestingly, neither the Immunochip study, nor the GWAS, has identified association at this locus in PSC. It is possible that this locus may have a predominant role in disease progression rather than disease susceptibility or interacts with other associated genes (epistasis), to exert functional effect.

As a proof-of-concept, this study shows that it is possible to determine genetic variants associated with disease progression. The main utility of such a study would be to identify biological pathways that may play a role in disease progression, and develop targeted therapy. This would ultimately pave way for stratified medicine, in which individuals can be prioritised into high-risk groups based on their respective phenotype and genotype, to determine most appropriate therapeutic intervention.

4) Limitations of genetic association studies

The genetic risk factors identified in complex diseases typically account for < 10% of disease susceptibility, often referred to as the *missing heritability*. In addition, only a handful of the common variants have been validated to have a functional role in disease biology. As a result, the common disease common variant (CDCV) hypothesis, which forms the basis for genome wide association studies, has come under increased scrutiny.

There is now emphasis on the role played by rare variants (MAF 0.002 - 0.01) with large effect-size (rare allele model), which are not identified by conventional genotyping platforms and GWAS. Studies to identify rare

variants require a considerably large sample size (more than 15,000 - 25,000 cases), making such studies difficult or even unfeasible in the context of PSC.

There may be a role for the environmental (e.g. gut microbiome) and epigenetic (glossary Box 1; page 160) risk factors in disease causation, which are not captured by conventional GWAS. These should now be studied in greater detail, alongside studies involving finer resolution of the genome (e.g. using partial or whole genome sequencing).

Chapter 9: Summary and Conclusion.

Summary of the UK PSC studies

The UK PSC cohort has recruited more than 2000 patients, from district general hospitals, university hospitals and transplant centres. It represents a unique cohort of ethnically homogeneous patients with PSC in the UK and the largest independent single nation cohort in the world.

In the first phase of the study, I recruited patients to the UK PSC cohort and collected samples to develop a PSC DNA biobank. This was followed by detailed phenotypic characterisation of the UK PSC cohort, which firmly established the cohort as a true representation of patients with PSC, with demographic and phenotype features similar to those reported in previous studies. Only 6 patients out of 1230, who had an underlying diagnosis of PBC instead, were recruited incorrectly, giving further confidence in our robust patient recruitment process.

One of the key findings to emerge from the phenotype study, and not reported previously, was the association between the extent of bile duct involvement and patient related outcome of liver transplant or death. Patients with both intrahepatic and extrahepatic bile duct involvement with PSC, have a significantly reduced transplant-free survival in comparison to those with only intrahepatic disease.

This cohort is not a true representation of the malignant risk in PSC, especially cholangiocarcinoma (CCA), as these patients have an extremely poor median survival of approximately 6 months, and thus will not be captured by such a study. In addition, the data collected were retrospective in nature and full laboratory data were not collected at the time of recruitment.

The laboratory and missing clinical data at initial diagnosis and follow-up, was subsequently collected by one of my collaborators and another UK PSC

research fellow (Dr Elizabeth Goode), which allowed a more detailed assessment of risk factors associated with clinical outcome (liver transplantation or death) in PSC, and development of the UK PSC risk score (awaiting peer review): http://www.uk-psc.com/resources/the-uk-psc-risk-scores/. When NHS Digital comes on stream these data will be collated with ease.

A recent multi-centre study by Weismuller *et al.* (46) evluated basic demographic and clinical parameters in more than 7000 patients with PSC and found younger age, female sex and IBD phenotype (Crohn's disease) to be associated with improved transplant-free survival or death. Information on the extent of bile duct involvement was not available for participants, so that cholangiographic changes could not be included.

In the UK PSC study, we did not oberve any difference in outocme according to sex or subtype of IBD. It is possible that our cohort of 1,452 patients was underpowered to detect any such effect.

However, we have shown that it is now possible to estimate short term and long term risk for patients presenting at different stages of the disease. In addition, there is evidence now to suggest that younger, female patients without IBD (or with Crohn's disease) and with intrahepatic disease have a very favourable long term outcome in an otherwise debilitating disease. This allows patients to be risk-stratified earlier in the course of their disease in clinical practice and will reduce the risk of recruitment bias in clinical trials, by categorising patients appropriately based on the disease severity.

In the second phase of the study, I performed a replication/candidate gene study and a genome wide association study in a modest cohort of 1,030 and 1,020 patients with PSC respectively. We identified two novel associations at 10p15 and 12q24 at genome-wide significance, with the most likely candidate genes as *Il2RA* and *SH2B3/ATXN2*. The plausible candidate genes at both loci play an important role in adaptive immune pathways and support the notion that immune dysregulation in and around bile ducts may have role in PSC pathophysiology.

The international collaborative studies in PSC have led to two large scale genetic association studies, with UK PSC consortium contributing the largest number of patients. These studies have confirmed the associations described above and identified a total of 22 PSC risk loci (table 8.3) and in addition, mapped the genetic correlation beween PSC and IBD risk loci (208, 209). Furthermore, genetic association has been confirmed for a locus on chromosome 6 (*RSPO3*) with the risk of liver transplantation or death.

Conclusion

In this thesis, I have outlined the development of the UK-PSC cohort, to which I recruited over 1,400 patients from the UK. I have described the phenotypic characteristics of the UK PSC cohort and identified the association between the extent of cholangiography changes and clinical outcome. I also performed two genetic association studies, and identified genome wide significant association at two novel immune risk loci.

Future directions

Set up in 2008 as a Clinical research network (CLRN) portfolio study, the UK PSC study has now recruited more than 2000 patients. The UK PSC cohort has been expanded recently to include children in the study.

The main priorities of the study are to now undertake more detailed sub-phenotyping of patient groups, evaluate the impact of PSC on patients quality of life and identify biomarkers associated with disease severity and cholangiocarcinoma, which is a significant cause of mortality in PSC.

One of the major challenges and unmet need in PSC is the complete lack of effective medical therapy that can alter disease progression. The long-term vision of the UK PSC study would be to develop a national PSC bio repository and undertake translational work to identify genes or biological pathways that can be targeted for therapy, as well as determine the role of ‘yet unknown’ environmental variables in disease causation.

Box 1. Glossary of terms.

Call rate (genotype)	The proportion of genotypes per marker with non-missing data.
Candidate gene	Any gene that has a priori possibility to cause a disease, either due to its location in a particular chromosome region suspected of being involved in the disease or its protein product.
Confounding	A type of bias in statistical analysis causing spurious or distorted findings, due to the existence of factors that are associated with disease risk as well as the exposure of interest.
Epigenetics	Modifications that occur in a chromosome without alterations to the underlying DNA structure, and contribute to phenotypic changes (e.g. DNA methylation and Histone modification).
Genetic variance	The contribution of genotypic differences among individuals to phenotypic variation.
Genotype-by-environment interactions	(G×E interactions). Refers to the situation in which the effect of the genotype is conditional on the environment, which may include abiotic (temperature), biotic (viral exposure) and cultural/behavioural influences.
Genotype-by-genotype interactions	(G×G interactions). Otherwise known as epistasis, this refers to the situation in which the effect of one genotype is conditional on genotypes at one or more other unlinked loci.
Genotype relative risk (GRR)	The ratios of the risk of disease between individuals with and without the genotype. A ratio of 1.1 equates to a 10% increase in risk.
Heritability	The proportion of the phenotypic variance in a population that is due to genotypic differences among individuals.
Penetrance	Describes the proportion of individuals with a mutation or risk variant that have the disease.
Pleiotropy	Association of one gene with two or more seemingly unrelated phenotypes.
Population stratification	A situation of confounding in genetic studies, where cases and controls are not selected from the same population (i.e. different ethnic origin), and in which the subpopulations differ in relation to the allele frequencies of the genetic variants under study and the prevalence of disease.
Power	The probability of a study to obtain a significant result if this result is true in the underlying population from which the study subjects were sampled.
Type 1 error	The probability of rejecting the null hypothesis of no effect of exposure on disease when in fact the null hypothesis is true. For genetic association studies, type I errors reflect false positive findings of associations between allele/genotype and disease.

Box 2. Overview of putative candidate gene function at PSC associated risk loci (217, 220).

Candidate gene	Putative gene function
BCL2L11	*BCL2L11* encodes BCL2-Like 11 (BIM), which is an apoptotic activator critical for apoptosis in B and T cells, macrophages and granulocytes
MST1	The MST1 encoded protein MSP has an inhibitory function towards macrophages during inflammation
IL2RA	*IL2RA* encodes the alpha chain of the IL-2 receptor (consisted of α, β and γ chains), which is responsible for capturing IL-2 cytokine and mediating its signalling effects. PSC-associated variants within IL2RA lead to reduced expression of IL2RA, which have implications for induction of immune tolerance. Il2ra−/− mice spontaneously develop intestinal and biliary inflammation.
MMEL1	MMEL1 is a member of the membrane metallo-endopeptidase family. Little is known about its function.
TNFRSF14	The TNFRSF14 protein is expressed on CD4 and CD8 T lymphocytes, B cells, monocytes, neutrophils, dendritic cells and mucosal epithelium. It acts as a molecular switch modulating T cell activation via binding to its ligand LIGHT [73], but can also propagate inhibitory signals through the immunoglobulin superfamily member BTLA (B and T lymphocyte attenuator). This dual role of TNFRSF14 is thought to regulate immune tolerance
CD28	CD28 is expressed on T lymphocytes and provides co-stimulatory signals that are essential for T cell activation, survival and proliferation.
CTLA4	CTLA-4, also known as CD152, is a major negative regulator of T-cell activation by binding to CD80 and CD86 in competition with CD28.
IL2	IL-2 is the most investigated interleukin with a broad role in regulation of the immune system. IT is essential for T-cell proliferation and regulatory T (T_{reg}) cell homeostasis.
IL21	IL-21 is an inflammatory cytokine mainly expressed by activated type 1 and type 17 T helper cells, and overproduction leads to tissue damage and destruction.
BACH2	BACH2 encodes a transcription factor, which regulates B cell differentiation. Downregulation of BACH2 has been demonstrated to attenuate activation of genes involved in antiviral innate immune response after nucleic-acid triggering.
SIK2	SIK2 is a serine/threonine protein kinase belonging to the AMP-activated protein kinase (AMPK) family. It regulates IL-10 in macrophages and Nur77 in leukocytes.
HDAC7	HDAC7 is a class IIa deacetylase, which has essential roles in the vasculature

	development and cytotoxic T lymphocyte function.
SH2B3	SH2B3 is a key regulator of cytokine signalling, and genetic variation within SH2B3 has been demonstrated to affect the production of pro-inflammatory cytokines in response to stimulation with bacterial cell wall peptidoglycans, pointing to a role for SH2B3 in innate immunity and in the protection against bacterial infection.
ATXN2	ATXN2 is a stress-regulated protein expressed in specific neuron populations and hepatocytes, and might have a role in obesity and insulin resistance.
CD226	CD226 is a co-stimulatory adhesion molecule constitutively expressed on immune and endothelial cells. It can enhance cytotoxic function of natural killer cells and T cells.
PRKD2	PRKD2 belongs to the protein kinase D (PKD) family, which are involved in the regulation of cell proliferation and cytokine production.
STRN4	STRN4 is associated with protein phosphatases and protein kinases and involved in cell proliferation, differentiation, apoptosis and transformation.
PSMG1	PSMG1 encodes a chaperone protein. Little is known about the function of PSMG1, but the protein has been hypothesized to be involved in microbial processing in Crohn's disease
GPR35	GPR35 is one of the "metabolite-sensing" G-protein-coupled receptors that expressed on immune cells and some gut epithelial cells. Proteins in this family generally mediate an anti-inflammatory effect.
TCF4	TCF4 is a transcription factor affecting cell proliferation and differentiation. It is important in T-cell, B- cell and plasmacytoid dendritic cell development.
CCL20	CCL20 is the ligand of chemokine receptor CCR6 and strongly attracts immature dendritic cells and memory T cells.
NFKB1	NFKB1 belongs to NF-κB family that controls genes regulating a broad range of biological processes including inflammation, cell growth, tumorigenesis and apoptosis.
RFX4	RFX4 is a transcription factor belonging to regulatory factor X (RFX) family, which potentially contributions to brain development and disease.
RIC8B	RIC8B has been identified as a G alpha(s)-binding protein, which catalyzes cAMP production and regulates many physiological aspects.
FOXP1	FOXP1, which belongs to subfamily P of the forkhead box (FOX) transcription factor family, regulates lymphocyte expansion and differentiation processes, and might act as a tumour suppressor.
CCDC88B	CCDC88B is a member of the hook-related protein family and acts as a novel

	regulator of maturation and effector functions of T cells during pathological inflammation.
CLEC16A	CLEC16A is a novel regulator of mitophagy that maintains glucose homeostasis.
SOCS1	SOCS1 is a negative regulator of cytokine signalling. It has an important role in Treg-cell integrity and function.
UBASH3A	UBASH3A belongs to a new type of protein tyrosine phosphatase family that negatively regulates T-cell signalling by promoting apoptosis.

References

1. Chapman RW, Arborgh BA, Rhodes JM, Summerfield JA, Dick R, Scheuer PJ, et al. Primary sclerosing cholangitis: a review of its clinical features, cholangiography, and hepatic histology. Gut. 1980 Oct;21(10):870-7.
2. Schwartz SI, Dale WA. Primary sclerosing cholangitis; review and report of six cases. AMA Arch Surg. 1958 Sep;77(3):439-51.
3. Boberg KM, Aadland E, Jahnsen J, Raknerud N, Stiris M, Bell H. Incidence and prevalence of primary biliary cirrhosis, primary sclerosing cholangitis, and autoimmune hepatitis in a Norwegian population. Scand J Gastroenterol. 1998 Jan;33(1):99-103.
4. Bambha K, Kim WR, Talwalkar J, Torgerson H, Benson JT, Therneau TM, et al. Incidence, clinical spectrum, and outcomes of primary sclerosing cholangitis in a United States community. Gastroenterology. 2003 Nov;125(5):1364-9.
5. Card TR, Solaymani-Dodaran M, West J. Incidence and mortality of primary sclerosing cholangitis in the UK: a population-based cohort study. J Hepatol. 2008 Jun;48(6):939-44.
6. Schrumpf E, Boberg KM. Epidemiology of primary sclerosing cholangitis. Best practice & research Clinical gastroenterology. [Review]. 2001 Aug;15(4):553-62.
7. Boonstra K, Weersma RK, van Erpecum KJ, Rauws EA, Spanier BW, Poen AC, et al. Population-based epidemiology, malignancy risk, and outcome of primary sclerosing cholangitis. Hepatology. [Multicenter Study Observational Study
Research Support, Non-U.S. Gov't]. 2013 Dec;58(6):2045-55.
8. Karlsen TH, Schrumpf E, Boberg KM. Update on primary sclerosing cholangitis. Digestive and liver disease : official journal of the Italian Society of Gastroenterology and the Italian Association for the Study of the Liver. [Review]. 2010 Jun;42(6):390-400.
9. Warren KW, Athanassiades S, Monge JI. Primary sclerosing cholangitis. A study of forty-two cases. American journal of surgery. 1966 Jan;111(1):23-38.
10. Takikawa H, Manabe T. Primary sclerosing cholangitis in Japan--analysis of 192 cases. Journal of gastroenterology. 1997 Feb;32(1):134-7.
11. Tanaka A, Takikawa H. Geoepidemiology of primary sclerosing cholangitis: a critical review. Journal of autoimmunity. [Review]. 2013 Oct;46:35-40.
12. Wiesner RH, LaRusso NF. Clinicopathologic features of the syndrome of primary sclerosing cholangitis. Gastroenterology. 1980 Aug;79(2):200-6.
13. Broome U, Olsson R, Loof L, Bodemar G, Hultcrantz R, Danielsson A, et al. Natural history and prognostic factors in 305 Swedish patients with primary sclerosing cholangitis. Gut. [Research Support, Non-U.S. Gov't]. 1996 Apr;38(4):610-5.
14. Farrant JM, Hayllar KM, Wilkinson ML, Karani J, Portmann BC, Westaby D, et al. Natural history and prognostic variables in primary sclerosing cholangitis. Gastroenterology. 1991 Jun;100(6):1710-7.
15. Ponsioen CY, Vrouenraets SM, Prawirodirdjo W, Rajaram R, Rauws EA, Mulder CJ, et al. Natural history of primary sclerosing cholangitis and prognostic

value of cholangiography in a Dutch population. Gut. [Multicenter Study]. 2002 Oct;51(4):562-6.
16. Tischendorf JJ, Hecker H, Kruger M, Manns MP, Meier PN. Characterization, outcome, and prognosis in 273 patients with primary sclerosing cholangitis: A single center study. The American journal of gastroenterology. 2007 Jan;102(1):107-14.
17. Weismuller TJ, Wedemeyer J, Kubicka S, Strassburg CP, Manns MP. The challenges in primary sclerosing cholangitis--aetiopathogenesis, autoimmunity, management and malignancy. Journal of hepatology. [Research Support, Non-U.S. Gov't
Review]. 2008;48 Suppl 1:S38-57.
18. Rudolph G, Gotthardt D, Kloters-Plachky P, Kulaksiz H, Rost D, Stiehl A. Influence of dominant bile duct stenoses and biliary infections on outcome in primary sclerosing cholangitis. Journal of hepatology. 2009 Jul;51(1):149-55.
19. Chapman R, Fevery J, Kalloo A, Nagorney DM, Boberg KM, Shneider B, et al. Diagnosis and management of primary sclerosing cholangitis. Hepatology. [Practice Guideline]. 2010 Feb;51(2):660-78.
20. Kotei CN, Baddeley H, Salmon PR. The role of endoscopic retrograde cholangiography in the diagnosis and management of patients with primary sclerosing cholangitis. Clinical radiology. [Case Reports]. 1979 Jul;30(4):445-50.
21. MacCarty RL, LaRusso NF, Wiesner RH, Ludwig J. Primary sclerosing cholangitis: findings on cholangiography and pancreatography. Radiology. [Comparative Study]. 1983 Oct;149(1):39-44.
22. Majoie CB, Reeders JW, Sanders JB, Huibregtse K, Jansen PL. Primary sclerosing cholangitis: a modified classification of cholangiographic findings. AJR American journal of roentgenology. 1991 Sep;157(3):495-7.
23. Rajaram R, Ponsioen CY, Majoie CB, Reeders JW, Lameris JS. Evaluation of a modified cholangiographic classification system for primary sclerosing cholangitis. Abdom Imaging. [Evaluation Studies]. 2001 Jan-Feb;26(1):43-7.
24. Freeman ML, Nelson DB, Sherman S, Haber GB, Herman ME, Dorsher PJ, et al. Complications of endoscopic biliary sphincterotomy. The New England journal of medicine. [Multicenter Study
Research Support, Non-U.S. Gov't]. 1996 Sep 26;335(13):909-18.
25. Angulo P, Pearce DH, Johnson CD, Henry JJ, LaRusso NF, Petersen BT, et al. Magnetic resonance cholangiography in patients with biliary disease: its role in primary sclerosing cholangitis. Journal of hepatology. [Comparative Study]. 2000 Oct;33(4):520-7.
26. Dave M, Elmunzer BJ, Dwamena BA, Higgins PD. Primary sclerosing cholangitis: meta-analysis of diagnostic performance of MR cholangiopancreatography. Radiology. 2010 Aug;256(2):387-96.
27. Burak KW, Angulo P, Lindor KD. Is there a role for liver biopsy in primary sclerosing cholangitis? The American journal of gastroenterology. [Comparative Study
Research Support, Non-U.S. Gov't]. 2003 May;98(5):1155-8.
28. Hov JR, Boberg KM, Karlsen TH. Autoantibodies in primary sclerosing cholangitis. World J Gastroenterol. 2008 Jun 28;14(24):3781-91.
29. Terjung B, Spengler U. Role of auto-antibodies for the diagnosis of chronic cholestatic liver diseases. Clinical reviews in allergy & immunology. [Review]. 2005 Apr;28(2):115-33.

30. Terjung B, Spengler U. Atypical p-ANCA in PSC and AIH: a hint toward a "leaky gut"? Clin Rev Allergy Immunol. 2009 Feb;36(1):40-51.
31. Terjung B, Muennich M, Gottwein J, Soehne J, et al. . Identification of myeloid-specific tubulin-beta isotype 5 as target antigen of antineutrophil cytoplasmic antibodies in autoimmune liver disorders. Hepatology. 2005;42(Suppl 1).
32. Wiesner RH, Grambsch PM, Dickson ER, Ludwig J, MacCarty RL, Hunter EB, et al. Primary sclerosing cholangitis: natural history, prognostic factors and survival analysis. Hepatology. 1989 Oct;10(4):430-6.
33. Karlsen TH, Schrumpf E, Boberg KM. Primary sclerosing cholangitis. Best practice & research Clinical gastroenterology. [Review]. 2010 Oct;24(5):655-66.
34. Fausa O, Schrumpf E, Elgjo K. Relationship of inflammatory bowel disease and primary sclerosing cholangitis. Semin Liver Dis. 1991 Feb;11(1):31-9.
35. Loftus EV, Jr., Harewood GC, Loftus CG, Tremaine WJ, Harmsen WS, Zinsmeister AR, et al. PSC-IBD: a unique form of inflammatory bowel disease associated with primary sclerosing cholangitis. Gut. 2005 Jan;54(1):91-6.
36. Rasmussen HH, Fallingborg JF, Mortensen PB, Vyberg M, Tage-Jensen U, Rasmussen SN. Hepatobiliary dysfunction and primary sclerosing cholangitis in patients with Crohn's disease. Scandinavian journal of gastroenterology. 1997 Jun;32(6):604-10.
37. Lunder AK, Hov JR, Borthne A, Gleditsch J, Johannesen G, Tveit K, et al. Prevalence of Sclerosing Cholangitis Detected by Magnetic Resonance Cholangiography in Patients With Long-term Inflammatory Bowel Disease. Gastroenterology. [Research Support, Non-U.S. Gov't]. 2016 Oct;151(4):660-9 e4.
38. Verdonk RC, Dijkstra G, Haagsma EB, Shostrom VK, Van den Berg AP, Kleibeuker JH, et al. Inflammatory bowel disease after liver transplantation: risk factors for recurrence and de novo disease. Am J Transplant. 2006 Jun;6(6):1422-9.
39. Schrumpf E, Fausa O, Kolmannskog F, Elgjo K, Ritland S, Gjone E. Sclerosing cholangitis in ulcerative colitis. A follow-up study. Scandinavian journal of gastroenterology. 1982 Jan;17(1):33-9.
40. Lamberts LE, Janse M, Haagsma EB, van den Berg AP, Weersma RK. Immune-mediated diseases in primary sclerosing cholangitis. Digestive and liver disease : official journal of the Italian Society of Gastroenterology and the Italian Association for the Study of the Liver. [Research Support, Non-U.S. Gov't]. 2011 Oct;43(10):802-6.
41. Saarinen S, Olerup O, Broome U. Increased frequency of autoimmune diseases in patients with primary sclerosing cholangitis. The American journal of gastroenterology. [Research Support, Non-U.S. Gov't]. 2000 Nov;95(11):3195-9.
42. Nikolaidis NL, Giouleme OI, Tziomalos KA, Patsiaoura K, Kazantzidou E, Voutsas AD, et al. Small-duct primary sclerosing cholangitis. A single-center seven-year experience. Digestive diseases and sciences. 2005 Feb;50(2):324-6.
43. Ludwig J. Small-duct primary sclerosing cholangitis. Semin Liver Dis. 1991 Feb;11(1):11-7.
44. Bjornsson E, Olsson R, Bergquist A, Lindgren S, Braden B, Chapman RW, et al. The natural history of small-duct primary sclerosing cholangitis. Gastroenterology. 2008 Apr;134(4):975-80.

45. Bjornsson E, Boberg KM, Cullen S, Fleming K, Clausen OP, Fausa O, et al. Patients with small duct primary sclerosing cholangitis have a favourable long term prognosis. Gut. [Research Support, Non-U.S. Gov't]. 2002 Nov;51(5):731-5.
46. Weismuller TJ, Trivedi PJ, Bergquist A, Imam M, Lenzen H, Ponsioen CY, et al. Patient Age, Sex, and Inflammatory Bowel Disease Phenotype Associate With Course of Primary Sclerosing Cholangitis. Gastroenterology. [Comparative Study Multicenter Study
Research Support, N.I.H., Extramural
Research Support, Non-U.S. Gov't]. 2017 Jun;152(8):1975-84 e8.
47. Stone JH, Zen Y, Deshpande V. IgG4-related disease. The New England journal of medicine. [Review]. 2012 Feb 9;366(6):539-51.
48. Boonstra K, Culver EL, de Buy Wenniger LM, van Heerde MJ, van Erpecum KJ, Poen AC, et al. Serum immunoglobulin G4 and immunoglobulin G1 for distinguishing immunoglobulin G4-associated cholangitis from primary sclerosing cholangitis. Hepatology. [Research Support, Non-U.S. Gov't]. 2014 May;59(5):1954-63.
49. Webster GJ, Pereira SP, Chapman RW. Autoimmune pancreatitis/IgG4-associated cholangitis and primary sclerosing cholangitis--overlapping or separate diseases? J Hepatol. 2009 Aug;51(2):398-402.
50. Sah RP, Chari ST. Serologic issues in IgG4-related systemic disease and autoimmune pancreatitis. Curr Opin Rheumatol. [Review]. 2011 Jan;23(1):108-13.
51. Culver EL, Chapman RW. IgG4-related hepatobiliary disease: an overview. Nature reviews Gastroenterology & hepatology. [Review]. 2016 Oct;13(10):601-12.
52. Mendes FD, Jorgensen R, Keach J, Katzmann JA, Smyrk T, Donlinger J, et al. Elevated serum IgG4 concentration in patients with primary sclerosing cholangitis. The American journal of gastroenterology. [Comparative Study]. 2006 Sep;101(9):2070-5.
53. Hirano K, Kawabe T, Yamamoto N, Nakai Y, Sasahira N, Tsujino T, et al. Serum IgG4 concentrations in pancreatic and biliary diseases. Clinica chimica acta; international journal of clinical chemistry. 2006 May;367(1-2):181-4.
54. Hirschfield GM, Karlsen TH, Lindor KD, Adams DH. Primary sclerosing cholangitis. Lancet. 2013 Jun 27.
55. Bjornsson E. Immunoglobulin G4-associated cholangitis. Current opinion in gastroenterology. [Review]. 2008 May;24(3):389-94.
56. Boberg KM, Chapman RW, Hirschfield GM, Lohse AW, Manns MP, Schrumpf E. Overlap syndromes: the International Autoimmune Hepatitis Group (IAIHG) position statement on a controversial issue. J Hepatol. 2011 Feb;54(2):374-85.
57. Abdalian R, Dhar P, Jhaveri K, Haider M, Guindi M, Heathcote EJ. Prevalence of sclerosing cholangitis in adults with autoimmune hepatitis: evaluating the role of routine magnetic resonance imaging. Hepatology. [Evaluation Studies]. 2008 Mar;47(3):949-57.
58. Lewin M, Vilgrain V, Ozenne V, Lemoine M, Wendum D, Paradis V, et al. Prevalence of sclerosing cholangitis in adults with autoimmune hepatitis: a prospective magnetic resonance imaging and histological study. Hepatology. 2009 Aug;50(2):528-37.

59. Gregorio GV, Portmann B, Karani J, Harrison P, Donaldson PT, Vergani D, et al. Autoimmune hepatitis/sclerosing cholangitis overlap syndrome in childhood: a 16-year prospective study. Hepatology. [Research Support, Non-U.S. Gov't]. 2001 Mar;33(3):544-53.
60. Fevery J, Henckaerts L, Van Oirbeek R, Vermeire S, Rutgeerts P, Nevens F, et al. Malignancies and mortality in 200 patients with primary sclerosering cholangitis: a long-term single-centre study. Liver international : official journal of the International Association for the Study of the Liver. 2012 Feb;32(2):214-22.
61. Lazaridis KN, Gores GJ. Primary sclerosing cholangitis and cholangiocarcinoma. Seminars in liver disease. [Review]. 2006 Feb;26(1):42-51.
62. Chalasani N, Baluyut A, Ismail A, Zaman A, Sood G, Ghalib R, et al. Cholangiocarcinoma in patients with primary sclerosing cholangitis: a multicenter case-control study. Hepatology. [Multicenter Study Research Support, Non-U.S. Gov't]. 2000 Jan;31(1):7-11.
63. Ahrendt SA, Pitt HA, Nakeeb A, Klein AS, Lillemoe KD, Kalloo AN, et al. Diagnosis and management of cholangiocarcinoma in primary sclerosing cholangitis. Journal of gastrointestinal surgery : official journal of the Society for Surgery of the Alimentary Tract. 1999 Jul-Aug;3(4):357-67; discussion 67-8.
64. Fevery J, Verslype C, Lai G, Aerts R, Van Steenbergen W. Incidence, diagnosis, and therapy of cholangiocarcinoma in patients with primary sclerosing cholangitis. Digestive diseases and sciences. [Case Reports]. 2007 Nov;52(11):3123-35.
65. Blechacz BR, Sanchez W, Gores GJ. A conceptual proposal for staging ductal cholangiocarcinoma. Current opinion in gastroenterology. [Editorial Research Support, N.I.H., Extramural]. 2009 May;25(3):238-9.
66. Krones E, Graziadei I, Trauner M, Fickert P. Evolving concepts in primary sclerosing cholangitis. Liver Int. 2011 Aug 17.
67. Rosen CB, Nagorney DM, Wiesner RH, Coffey RJ, Jr., LaRusso NF. Cholangiocarcinoma complicating primary sclerosing cholangitis. Annals of surgery. [Clinical Trial Randomized Controlled Trial]. 1991 Jan;213(1):21-5.
68. Vazquez-Elizondo G, Mucino-Bermejo J, Mendez-Sanchez N. Gallbladder disease in patients with primary sclerosing cholangitis. Ann Hepatol. 2008 Apr-Jun;7(2):182-3.
69. Leung UC, Wong PY, Roberts RH, Koea JB. Gall bladder polyps in sclerosing cholangitis: does the 1-cm rule apply? ANZ J Surg. 2007 May;77(5):355-7.
70. Karlsen TH, Schrumpf E, Boberg KM. Gallbladder polyps in primary sclerosing cholangitis: not so benign. Curr Opin Gastroenterol. 2008 May;24(3):395-9.
71. Dorudi S, Chapman RW, Kettlewell MG. Carcinoma of the gallbladder in ulcerative colitis and primary sclerosing cholangitis. Report of two cases. Dis Colon Rectum. 1991 Sep;34(9):827-8.
72. Brandt DJ, MacCarty RL, Charboneau JW, LaRusso NF, Wiesner RH, Ludwig J. Gallbladder disease in patients with primary sclerosing cholangitis. AJR Am J Roentgenol. 1988 Mar;150(3):571-4.

73. Said K, Glaumann H, Bergquist A. Gallbladder disease in patients with primary sclerosing cholangitis. Journal of hepatology. [Comparative Study]. 2008 Apr;48(4):598-605.
74. Eaton JE, Thackeray EW, Lindor KD. Likelihood of Malignancy in Gallbladder Polyps and Outcomes Following Cholecystectomy in Primary Sclerosing Cholangitis. The American journal of gastroenterology. 2011 Oct 25.
75. Ito H, Hann LE, D'Angelica M, Allen P, Fong Y, Dematteo RP, et al. Polypoid lesions of the gallbladder: diagnosis and followup. J Am Coll Surg. 2009 Apr;208(4):570-5.
76. Broome U, Bergquist A. Primary sclerosing cholangitis, inflammatory bowel disease, and colon cancer. Seminars in liver disease. [Review]. 2006 Feb;26(1):31-41.
77. Lindstrom L, Lapidus A, Ost A, Bergquist A. Increased risk of colorectal cancer and dysplasia in patients with Crohn's colitis and primary sclerosing cholangitis. Diseases of the colon and rectum. [Research Support, Non-U.S. Gov't]. 2011 Nov;54(11):1392-7.
78. Broome U, Lofberg R, Veress B, Eriksson LS. Primary sclerosing cholangitis and ulcerative colitis: evidence for increased neoplastic potential. Hepatology. [Research Support, Non-U.S. Gov't]. 1995 Nov;22(5):1404-8.
79. Jess T, Loftus EV, Jr., Velayos FS, Winther KV, Tremaine WJ, Zinsmeister AR, et al. Risk factors for colorectal neoplasia in inflammatory bowel disease: a nested case-control study from Copenhagen county, Denmark and Olmsted county, Minnesota. The American journal of gastroenterology. [Research Support, Non-U.S. Gov't]. 2007 Apr;102(4):829-36.
80. Bergquist A, Ekbom A, Olsson R, Kornfeldt D, Loof L, Danielsson A, et al. Hepatic and extrahepatic malignancies in primary sclerosing cholangitis. Journal of hepatology. [Research Support, Non-U.S. Gov't]. 2002 Mar;36(3):321-7.
81. Borchers AT, Shimoda S, Bowlus C, Keen CL, Gershwin ME. Lymphocyte recruitment and homing to the liver in primary biliary cirrhosis and primary sclerosing cholangitis. Semin Immunopathol. 2009 Sep;31(3):309-22.
82. Ponsioen CY, Kuiper H, Ten Kate FJ, van Milligen de Wit M, van Deventer SJ, Tytgat GN. Immunohistochemical analysis of inflammation in primary sclerosing cholangitis. European journal of gastroenterology & hepatology. 1999 Jul;11(7):769-74.
83. Bo X, Broome U, Remberger M, Sumitran-Holgersson S. Tumour necrosis factor alpha impairs function of liver derived T lymphocytes and natural killer cells in patients with primary sclerosing cholangitis. Gut. [Research Support, Non-U.S. Gov't]. 2001 Jul;49(1):131-41.
84. Spengler U, Moller A, Jung MC, Messer G, Zachoval R, Hoffmann RM, et al. T lymphocytes from patients with primary biliary cirrhosis produce reduced amounts of lymphotoxin, tumor necrosis factor and interferon-gamma upon mitogen stimulation. Journal of hepatology. [Research Support, Non-U.S. Gov't]. 1992 May;15(1-2):129-35.
85. Terjung B, Sohne J, Lechtenberg B, Gottwein J, Muennich M, Herzog V, et al. p-ANCAs in autoimmune liver disorders recognise human beta-tubulin isotype 5 and cross-react with microbial protein FtsZ. Gut. [Research Support, Non-U.S. Gov't]. 2010 Jun;59(6):808-16.
86. O'Mahony CA, Vierling JM. Etiopathogenesis of primary sclerosing cholangitis. Seminars in liver disease. [Review]. 2006 Feb;26(1):3-21.

87. Eaton JE, Talwalkar JA, Lazaridis KN, Gores GJ, Lindor KD. Pathogenesis of primary sclerosing cholangitis and advances in diagnosis and management. Gastroenterology. [Review
Video-Audio Media]. 2013 Sep;145(3):521-36.
88. Grant AJ, Lalor PF, Salmi M, Jalkanen S, Adams DH. Homing of mucosal lymphocytes to the liver in the pathogenesis of hepatic complications of inflammatory bowel disease. Lancet. 2002 Jan 12;359(9301):150-7.
89. Grant AJ, Lalor PF, Hubscher SG, Briskin M, Adams DH. MAdCAM-1 expressed in chronic inflammatory liver disease supports mucosal lymphocyte adhesion to hepatic endothelium (MAdCAM-1 in chronic inflammatory liver disease). Hepatology. 2001 May;33(5):1065-72.
90. Smit JJ, Schinkel AH, Oude Elferink RP, Groen AK, Wagenaar E, van Deemter L, et al. Homozygous disruption of the murine mdr2 P-glycoprotein gene leads to a complete absence of phospholipid from bile and to liver disease. Cell. [Research Support, Non-U.S. Gov't]. 1993 Nov 5;75(3):451-62.
91. Ohishi Y, Nakamura M, Iio N, Higa S, Inayoshi M, Aiba Y, et al. Single-nucleotide polymorphism analysis of the multidrug resistance protein 3 gene for the detection of clinical progression in Japanese patients with primary biliary cirrhosis. Hepatology. [Research Support, Non-U.S. Gov't]. 2008 Sep;48(3):853-62.
92. Cullen SN, Chapman RW. Review article: current management of primary sclerosing cholangitis. Alimentary pharmacology & therapeutics. [Review]. 2005 Apr 15;21(8):933-48.
93. Olsson R, Broome U, Danielsson A, Hagerstrand I, Jarnerot G, Loof L, et al. Colchicine treatment of primary sclerosing cholangitis. Gastroenterology. [Clinical Trial
Randomized Controlled Trial
Research Support, Non-U.S. Gov't]. 1995 Apr;108(4):1199-203.
94. Bharucha AE, Jorgensen R, Lichtman SN, LaRusso NF, Lindor KD. A pilot study of pentoxifylline for the treatment of primary sclerosing cholangitis. The American journal of gastroenterology. [Clinical Trial
Comparative Study]. 2000 Sep;95(9):2338-42.
95. Hommes DW, Erkelens W, Ponsioen C, Stokkers P, Rauws E, van der Spek M, et al. A double-blind, placebo-controlled, randomized study of infliximab in primary sclerosing cholangitis. J Clin Gastroenterol. 2008 May-Jun;42(5):522-6.
96. Paumgartner G, Beuers U. Ursodeoxycholic acid in cholestatic liver disease: mechanisms of action and therapeutic use revisited. Hepatology. [Review]. 2002 Sep;36(3):525-31.
97. Beuers U, Spengler U, Kruis W, Aydemir U, Wiebecke B, Heldwein W, et al. Ursodeoxycholic acid for treatment of primary sclerosing cholangitis: a placebo-controlled trial. Hepatology. 1992 Sep;16(3):707-14.
98. Olsson R, Boberg KM, de Muckadell OS, Lindgren S, Hultcrantz R, Folvik G, et al. High-dose ursodeoxycholic acid in primary sclerosing cholangitis: a 5-year multicenter, randomized, controlled study. Gastroenterology. [Multicenter Study
Randomized Controlled Trial
Research Support, Non-U.S. Gov't]. 2005 Nov;129(5):1464-72.
99. Lindor KD, Kowdley KV, Luketic VA, Harrison ME, McCashland T, Befeler AS, et al. High-dose ursodeoxycholic acid for the treatment of primary sclerosing cholangitis. Hepatology. 2009 Sep;50(3):808-14.

100. Triantos CK, Koukias NM, Nikolopoulou VN, Burroughs AK. Meta-analysis: ursodeoxycholic acid for primary sclerosing cholangitis. Alimentary pharmacology & therapeutics. [Meta-Analysis]. 2011 Oct;34(8):901-10.
101. Lindstrom L, Boberg KM, Wikman O, Friis-Liby I, Hultcrantz R, Prytz H, et al. High dose ursodeoxycholic acid in primary sclerosing cholangitis does not prevent colorectal neoplasia. Aliment Pharmacol Ther. 2012 Jan 5.
102. Singh S, Khanna S, Pardi DS, Loftus EV, Jr., Talwalkar JA. Effect of ursodeoxycholic acid use on the risk of colorectal neoplasia in patients with primary sclerosing cholangitis and inflammatory bowel disease: a systematic review and meta-analysis. Inflammatory bowel diseases. [Review]. 2013 Jul;19(8):1631-8.
103. EASL Clinical Practice Guidelines: management of cholestatic liver diseases. Journal of hepatology. [Practice Guideline]. 2009 Aug;51(2):237-67.
104. Fickert P, Hirschfield GM, Denk G, Marschall HU, Altorjay I, Farkkila M, et al. norUrsodeoxycholic acid improves cholestasis in primary sclerosing cholangitis. Journal of hepatology. [Clinical Trial, Phase II
Randomized Controlled Trial
Research Support, Non-U.S. Gov't]. 2017 Sep;67(3):549-58.
105. Halilbasic E, Fiorotto R, Fickert P, Marschall HU, Moustafa T, Spirli C, et al. Side chain structure determines unique physiologic and therapeutic properties of norursodeoxycholic acid in Mdr2-/- mice. Hepatology. [Research Support, N.I.H., Extramural
Research Support, Non-U.S. Gov't]. 2009 Jun;49(6):1972-81.
106. Floreani A, Franceschet I, Perini L, Cazzagon N, Gershwin ME, Bowlus CL. New Therapies for Primary Biliary Cirrhosis. Clinical reviews in allergy & immunology. 2014 Oct 21.
107. Hirschfield GM, Chazouilleres O, Drenth JP, Thorburn D, Harrison SA, Landis CS, et al. Effect of NGM282, an FGF19 analogue, in primary sclerosing cholangitis: A multicenter, randomized, double-blind, placebo-controlled phase II trial. Journal of hepatology. 2019 Mar;70(3):483-93.
108. Trauner M, Halilbasic E, Kazemi-Shirazi L, Kienbacher C, Staufer K, Traussnigg S, et al. Therapeutic role of bile acids and nuclear receptor agonists in fibrosing cholangiopathies. Digestive diseases. [Research Support, Non-U.S. Gov't]. 2014;32(5):631-6.
109. Barnabas A, Chapman RW. Primary sclerosing cholangitis: is any treatment worthwhile? Curr Gastroenterol Rep. 2012 Feb;14(1):17-24.
110. Singh S, Talwalkar JA. Primary sclerosing cholangitis: diagnosis, prognosis, and management. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association. [Review]. 2013 Aug;11(8):898-907.
111. Ponsioen CY, Arnelo U, Bergquist A, Rauws EA, Paulsen V, Cantu P, et al. No Superiority of Stents vs Balloon Dilatation for Dominant Strictures in Patients With Primary Sclerosing Cholangitis. Gastroenterology. [Multicenter Study
Randomized Controlled Trial
Research Support, Non-U.S. Gov't]. 2018 Sep;155(3):752-9 e5.
112. Kashyap R, Safadjou S, Chen R, Mantry P, Sharma R, Patil V, et al. Living donor and deceased donor liver transplantation for autoimmune and cholestatic liver diseases--an analysis of the UNOS database. Journal of gastrointestinal

surgery : official journal of the Society for Surgery of the Alimentary Tract. [Comparative Study]. 2010 Sep;14(9):1362-9.
113. Boberg KM, Rocca G, Egeland T, Bergquist A, Broome U, Caballeria L, et al. Time-dependent Cox regression model is superior in prediction of prognosis in primary sclerosing cholangitis. Hepatology. [Research Support, Non-U.S. Gov't]. 2002 Mar;35(3):652-7.
114. Ricci P, Therneau TM, Malinchoc M, Benson JT, Petz JL, Klintmalm GB, et al. A prognostic model for the outcome of liver transplantation in patients with cholestatic liver disease. Hepatology. [Research Support, U.S. Gov't, P.H.S.]. 1997 Mar;25(3):672-7.
115. Kim WR, Therneau TM, Wiesner RH, Poterucha JJ, Benson JT, Malinchoc M, et al. A revised natural history model for primary sclerosing cholangitis. Mayo Clinic proceedings. [Comparative Study
Multicenter Study
Research Support, U.S. Gov't, P.H.S.]. 2000 Jul;75(7):688-94.
116. de Vries EM, de Krijger M, Farkkila M, Arola J, Schirmacher P, Gotthardt D, et al. Validation of the prognostic value of histologic scoring systems in primary sclerosing cholangitis: An international cohort study. Hepatology. [Multicenter Study
Validation Studies
Research Support, Non-U.S. Gov't]. 2017 Mar;65(3):907-19.
117. Eaton JE, Vesterhus M, McCauley BM, Atkinson EJ, Schlicht EM, Juran BD, et al. Primary Sclerosing Cholangitis Risk Estimate Tool (PREsTo) Predicts Outcomes of the Disease: A Derivation and Validation Study Using Machine Learning. Hepatology. 2018 May 9.
118. Heimbach JK, Gores GJ, Haddock MG, Alberts SR, Pedersen R, Kremers W, et al. Predictors of disease recurrence following neoadjuvant chemoradiotherapy and liver transplantation for unresectable perihilar cholangiocarcinoma. Transplantation. 2006 Dec 27;82(12):1703-7.
119. Carbone M, Neuberger J. Liver transplantation in PBC and PSC: indications and disease recurrence. Clinics and research in hepatology and gastroenterology. [Review]. 2011 Jun;35(6-7):446-54.
120. Vera A, Gunson BK, Ussatoff V, Nightingale P, Candinas D, Radley S, et al. Colorectal cancer in patients with inflammatory bowel disease after liver transplantation for primary sclerosing cholangitis. Transplantation. 2003 Jun 27;75(12):1983-8.
121. Graziadei IW, Wiesner RH, Marotta PJ, Porayko MK, Hay JE, Charlton MR, et al. Long-term results of patients undergoing liver transplantation for primary sclerosing cholangitis. Hepatology. [Research Support, Non-U.S. Gov't]. 1999 Nov;30(5):1121-7.
122. Campsen J, Zimmerman MA, Trotter JF, Wachs M, Bak T, Steinberg T, et al. Clinically recurrent primary sclerosing cholangitis following liver transplantation: a time course. Liver transplantation : official publication of the American Association for the Study of Liver Diseases and the International Liver Transplantation Society. 2008 Feb;14(2):181-5.
123. Alexander J, Lord JD, Yeh MM, Cuevas C, Bakthavatsalam R, Kowdley KV. Risk factors for recurrence of primary sclerosing cholangitis after liver transplantation. Liver transplantation : official publication of the American

Association for the Study of Liver Diseases and the International Liver Transplantation Society. [Research Support, N.I.H., Extramural
Research Support, Non-U.S. Gov't]. 2008 Feb;14(2):245-51.
124. Cholongitas E, Shusang V, Papatheodoridis GV, Marelli L, Manousou P, Rolando N, et al. Risk factors for recurrence of primary sclerosing cholangitis after liver transplantation. Liver transplantation : official publication of the American Association for the Study of Liver Diseases and the International Liver Transplantation Society. 2008 Feb;14(2):138-43.
125. Alabraba E, Nightingale P, Gunson B, Hubscher S, Olliff S, Mirza D, et al. A re-evaluation of the risk factors for the recurrence of primary sclerosing cholangitis in liver allografts. Liver Transpl. 2009 Mar;15(3):330-40.
126. Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. Science. [Review]. 2008 Nov 7;322(5903):881-8.
127. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, et al. Complement factor H polymorphism in age-related macular degeneration. Science. [Research Support, Non-U.S. Gov't
Research Support, U.S. Gov't, P.H.S.]. 2005 Apr 15;308(5720):385-9.
128. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic acids research. [Research Support, N.I.H., Extramural
Research Support, Non-U.S. Gov't]. 2014 Jan;42(Database issue):D1001-6.
129. Gulamhusein AF, Juran BD, Lazaridis KN. Genome-Wide Association Studies in Primary Biliary Cirrhosis. Seminars in liver disease. [Research Support, N.I.H., Extramural
Research Support, Non-U.S. Gov't
Review]. 2015 Nov;35(4):392-401.
130. Karlsen TH, Schrumpf E, Boberg KM. Genetic epidemiology of primary sclerosing cholangitis. World journal of gastroenterology : WJG. [Editorial Review]. 2007 Nov 7;13(41):5421-31.
131. Bergquist A, Montgomery SM, Bahmanyar S, Olsson R, Danielsson A, Lindgren 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 Aug;6(8):939-43.
132. Horton R, Wilming L, Rand V, Lovering RC, Bruford EA, Khodiyar VK, et al. Gene map of the extended human MHC. Nat Rev Genet. [Research Support, Non-U.S. Gov't
Research Support, U.S. Gov't, P.H.S.
Review]. 2004 Dec;5(12):889-99.
133. Chapman RW, Varghese Z, Gaul R, Patel G, Kokinon N, Sherlock S. Association of primary sclerosing cholangitis with HLA-B8. Gut. 1983 Jan;24(1):38-41.
134. Schrumpf E, Fausa O, Forre O, Dobloug JH, Ritland S, Thorsby E. HLA antigens and immunoregulatory T cells in ulcerative colitis associated with hepatobiliary disease. Scandinavian journal of gastroenterology. [Research Support, Non-U.S. Gov't]. 1982 Mar;17(2):187-91.
135. Candore G, Lio D, Colonna Romano G, Caruso C. Pathogenesis of autoimmune diseases associated with 8.1 ancestral haplotype: effect of multiple gene interactions. Autoimmun Rev. [Research Support, Non-U.S. Gov't
Review]. 2002 Feb;1(1-2):29-35.

136. Donaldson PT, Norris S. Evaluation of the role of MHC class II alleles, haplotypes and selected amino acid sequences in primary sclerosing cholangitis. Autoimmunity. 2002 Dec;35(8):555-64.
137. Chapman R, Cullen S. Etiopathogenesis of primary sclerosing cholangitis. World J Gastroenterol. 2008 Jun 7;14(21):3350-9.
138. Mells GF, Kaser A, Karlsen TH. Novel insights into autoimmune liver diseases provided by genome-wide association studies. Journal of autoimmunity. [Research Support, Non-U.S. Gov't
Review]. 2013 Oct;46:41-54.
139. Bowlus CL, Li CS, Karlsen TH, Lie BA, Selmi C. Primary sclerosing cholangitis in genetically diverse populations listed for liver transplantation: unique clinical and human leukocyte antigen associations. Liver Transpl. 2010 Nov;16(11):1324-30.
140. Glazier AM, Nadeau JH, Aitman TJ. Finding genes that underlie complex traits. Science. [Research Support, Non-U.S. Gov't
Review]. 2002 Dec 20;298(5602):2345-9.
141. Donaldson PT, Norris S, Constantini PK, Bernal W, Harrison P, Williams R. The interleukin-1 and interleukin-10 gene polymorphisms in primary sclerosing cholangitis: no associations with disease susceptibility/resistance. Journal of hepatology. 2000 Jun;32(6):882-6.
142. Pauli-Magnus C, Kerb R, Fattinger K, Lang T, Anwald B, Kullak-Ublick GA, et al. BSEP and MDR3 haplotype structure in healthy Caucasians, primary biliary cirrhosis and primary sclerosing cholangitis. Hepatology. 2004 Mar;39(3):779-91.
143. Wiencke K, Boberg KM, Donaldson P, Harbo H, Ling V, Schrumpf E, et al. No major effect of the CD28/CTLA4/ICOS gene region on susceptibility to primary sclerosing cholangitis. Scandinavian journal of gastroenterology. [Research Support, Non-U.S. Gov't]. 2006 May;41(5):586-91.
144. Melum E, Karlsen TH, Broome U, Thorsby E, Schrumpf E, Boberg KM, et al. The 32-base pair deletion of the chemokine receptor 5 gene (CCR5-Delta32) is not associated with primary sclerosing cholangitis in 363 Scandinavian patients. Tissue Antigens. [Comparative Study
Research Support, Non-U.S. Gov't]. 2006 Jul;68(1):78-81.
145. Girodon E, Sternberg D, Chazouilleres O, Cazeneuve C, Huot D, Calmus Y, et al. Cystic fibrosis transmembrane conductance regulator (CFTR) gene defects in patients with primary sclerosing cholangitis. Journal of hepatology. [Research Support, Non-U.S. Gov't]. 2002 Aug;37(2):192-7.
146. Wiencke K, Louka AS, Spurkland A, Vatn M, Schrumpf E, Boberg KM. Association of matrix metalloproteinase-1 and -3 promoter polymorphisms with clinical subsets of Norwegian primary sclerosing cholangitis patients. Journal of hepatology. [Research Support, Non-U.S. Gov't]. 2004 Aug;41(2):209-14.
147. Karlsen TH, Hampe J, Wiencke K, Schrumpf E, Thorsby E, Lie BA, et al. Genetic polymorphisms associated with inflammatory bowel disease do not confer risk for primary sclerosing cholangitis. The American journal of gastroenterology. [Research Support, Non-U.S. Gov't]. 2007 Jan;102(1):115-21.
148. Bowlus CL, Karlsen TH, Broome U, Thorsby E, Vatn M, Schrumpf E, et al. Analysis of MAdCAM-1 and ICAM-1 polymorphisms in 365 Scandinavian patients with primary sclerosing cholangitis. Journal of hepatology. [Research Support, Non-U.S. Gov't]. 2006 Nov;45(5):704-10.

149. Vogel A, Liermann H, Harms A, Strassburg CP, Manns MP, Obermayer-Straub P. Autoimmune regulator AIRE: evidence for genetic differences between autoimmune hepatitis and hepatitis as part of the autoimmune polyglandular syndrome type 1. Hepatology. [Research Support, Non-U.S. Gov't]. 2001 May;33(5):1047-52.
150. Karlsen TH, Franke A, Melum E, Kaser A, Hov JR, Balschun T, et al. Genome-wide association analysis in primary sclerosing cholangitis. Gastroenterology. 2010 Mar;138(3):1102-11.
151. Melum E, Franke A, Schramm C, Weismuller TJ, Gotthardt DN, Offner FA, et al. Genome-wide association analysis in primary sclerosing cholangitis identifies two non-HLA susceptibility loci. Nat Genet. [Research Support, Non-U.S. Gov't]. 2011 Jan;43(1):17-9.
152. Folseraas T, Melum E, Rausch P, Juran BD, Ellinghaus E, Shiryaev A, et al. Extended analysis of a genome-wide association study in primary sclerosing cholangitis detects multiple novel risk loci. Journal of hepatology. [Research Support, N.I.H., Extramural
Research Support, Non-U.S. Gov't]. 2012 Aug;57(2):366-75.
153. Vanderpump MP, Tunbridge WM, French JM, Appleton D, Bates D, Clark F, et al. The incidence of thyroid disorders in the community: a twenty-year follow-up of the Whickham Survey. Clin Endocrinol (Oxf). [Research Support, Non-U.S. Gov't]. 1995 Jul;43(1):55-68.
154. Leeds JS, Hopper AD, Sanders DS. Coeliac disease. Br Med Bull. [Research Support, Non-U.S. Gov't
Review]. 2008;88(1):157-70.
155. Talwalkar JA, Lindor KD. Natural history and prognostic models in primary sclerosing cholangitis. Best practice & research Clinical gastroenterology. [Review]. 2001 Aug;15(4):563-75.
156. Levy C, Lindor KD. Primary sclerosing cholangitis: epidemiology, natural history, and prognosis. Seminars in liver disease. [Review]. 2006 Feb;26(1):22-30.
157. Al Mamari S, Djordjevic J, Halliday JS, Chapman RW. Improvement of serum alkaline phosphatase to <1.5 upper limit of normal predicts better outcome and reduced risk of cholangiocarcinoma in primary sclerosing cholangitis. Journal of hepatology. 2013 Feb;58(2):329-34.
158. Mahid SS, Minor KS, Soto RE, Hornung CA, Galandiuk S. Smoking and inflammatory bowel disease: a meta-analysis. Mayo Clin Proc. [Meta-Analysis Research Support, Non-U.S. Gov't]. 2006 Nov;81(11):1462-71.
159. Chapman MH, Webster GJ, Bannoo S, Johnson GJ, Wittmann J, Pereira SP. Cholangiocarcinoma and dominant strictures in patients with primary sclerosing cholangitis: a 25-year single-centre experience. European journal of gastroenterology & hepatology. [Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't]. 2012 Sep;24(9):1051-8.
160. Kennedy TM, Jones RH. Epidemiology of cholecystectomy and irritable bowel syndrome in a UK population. The British journal of surgery. [Research Support, Non-U.S. Gov't]. 2000 Dec;87(12):1658-63.
161. Harnois DM, Angulo P, Jorgensen RA, Larusso NF, Lindor KD. High-dose ursodeoxycholic acid as a therapy for patients with primary sclerosing cholangitis. The American journal of gastroenterology. [Clinical Trial Randomized Controlled Trial]. 2001 May;96(5):1558-62.

162. Wiencke K, Boberg KM. Current consensus on the management of primary sclerosing cholangitis. Clinics and research in hepatology and gastroenterology. 2011 Dec;35(12):786-91.
163. Stiehl A, Rudolph G, Sauer P, Theilmann L. Biliary secretion of bile acids and lipids in primary sclerosing cholangitis. Influence of cholestasis and effect of ursodeoxycholic acid treatment. Journal of hepatology. 1995 Sep;23(3):283-9.
164. Mitchell SA, Bansi DS, Hunt N, Von Bergmann K, Fleming KA, Chapman RW. A preliminary trial of high-dose ursodeoxycholic acid in primary sclerosing cholangitis. Gastroenterology. [Clinical Trial
Comparative Study
Randomized Controlled Trial]. 2001 Oct;121(4):900-7.
165. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. Bmj. [Research Support, Non-U.S. Gov't]. 2009 Jun 29;338:b2393.
166. Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, et al. Identification of the cystic fibrosis gene: genetic analysis. Science. [Research Support, Non-U.S. Gov't
Research Support, U.S. Gov't, P.H.S.]. 1989 Sep 8;245(4922):1073-80.
167. MacDonald ME, Novelletto A, Lin C, Tagle D, Barnes G, Bates G, et al. The Huntington's disease candidate region exhibits many different haplotypes. Nat Genet. [Research Support, Non-U.S. Gov't
Research Support, U.S. Gov't, P.H.S.
Review]. 1992 May;1(2):99-103.
168. Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science. [Research Support, Non-U.S. Gov't
Research Support, U.S. Gov't, P.H.S.]. 1993 Aug 13;261(5123):921-3.
169. Bush WS, Moore JH. Chapter 11: Genome-wide association studies. PLoS Comput Biol. [Research Support, N.I.H., Extramural
Review]. 2012;8(12):e1002822.
170. The International HapMap Project. Nature. [Multicenter Study
Research Support, Non-U.S. Gov't
Research Support, U.S. Gov't, P.H.S.]. 2003 Dec 18;426(6968):789-96.
171. Hong EP, Park JW. Sample size and statistical power calculation in genetic association studies. Genomics Inform. 2012 Jun;10(2):117-22.
172. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. G3 (Bethesda). 2011 Nov;1(6):457-70.
173. Janse M, Lamberts LE, Franke L, Raychaudhuri S, Ellinghaus E, Muri Boberg K, et al. Three ulcerative colitis susceptibility loci are associated with primary sclerosing cholangitis and indicate a role for IL2, REL, and CARD9. Hepatology. [Multicenter Study
Research Support, N.I.H., Extramural
Research Support, Non-U.S. Gov't]. 2011 Jun;53(6):1977-85.
174. Cotsapas C, Voight BF, Rossin E, Lage K, Neale BM, Wallace C, et al. Pervasive sharing of genetic effects in autoimmune disease. PLoS genetics. [Meta-Analysis]. 2011 Aug;7(8):e1002254.
175. Hsu W, Zhang W, Tsuneyama K, Moritoki Y, Ridgway WM, Ansari AA, et al. Differential mechanisms in the pathogenesis of autoimmune cholangitis versus

inflammatory bowel disease in interleukin-2Ralpha(-/-) mice. Hepatology. [Research Support, N.I.H., Extramural
Research Support, Non-U.S. Gov't]. 2009 Jan;49(1):133-40.
176. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005 Jan 15;21(2):263-5.
177. Cavanillas ML, Alcina A, Nunez C, de las Heras V, Fernandez-Arquero M, Bartolome M, et al. Polymorphisms in the IL2, IL2RA and IL2RB genes in multiple sclerosis risk. European journal of human genetics : EJHG. [Research Support, Non-U.S. Gov't]. 2010 Jul;18(7):794-9.
178. Qu HQ, Bradfield JP, Belisle A, Grant SF, Hakonarson H, Polychronakos C. The type I diabetes association of the IL2RA locus. Genes and immunity. [Research Support, N.I.H., Extramural
Research Support, Non-U.S. Gov't]. 2009 Dec;10 Suppl 1:S42-8.
179. Hafler DA, Compston A, Sawcer S, Lander ES, Daly MJ, De Jager PL, et al. Risk alleles for multiple sclerosis identified by a genomewide study. The New England journal of medicine. [Research Support, N.I.H., Extramural
Research Support, Non-U.S. Gov't]. 2007 Aug 30;357(9):851-62.
180. Lowe CE, Cooper JD, Brusko T, Walker NM, Smyth DJ, Bailey R, et al. Large-scale genetic fine mapping and genotype-phenotype associations implicate polymorphism in the IL2RA region in type 1 diabetes. Nat Genet. [Research Support, Non-U.S. Gov't]. 2007 Sep;39(9):1074-82.
181. Stahl EA, Raychaudhuri S, Remmers EF, Xie G, Eyre S, Thomson BP, et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nat Genet. [Meta-Analysis
Research Support, N.I.H., Extramural
Research Support, N.I.H., Intramural
Research Support, Non-U.S. Gov't]. 2010 Jun;42(6):508-14.
182. van Heel DA, Franke L, Hunt KA, Gwilliam R, Zhernakova A, Inouye M, et al. A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. Nat Genet. [Research Support, Non-U.S. Gov't]. 2007 Jul;39(7):827-9.
183. Marquez A, Orozco G, Martinez A, Palomino-Morales R, Fernandez-Arquero M, Mendoza JL, et al. Novel association of the interleukin 2-interleukin 21 region with inflammatory bowel disease. The American journal of gastroenterology. [Research Support, Non-U.S. Gov't]. 2009 Aug;104(8):1968-75.
184. Zhernakova A, Alizadeh BZ, Bevova M, van Leeuwen MA, Coenen MJ, Franke B, et al. Novel association in chromosome 4q27 region with rheumatoid arthritis and confirmation of type 1 diabetes point to a general risk locus for autoimmune diseases. American journal of human genetics. [Research Support, Non-U.S. Gov't]. 2007 Dec;81(6):1284-8.
185. Festen EA, Goyette P, Scott R, Annese V, Zhernakova A, Lian J, et al. Genetic variants in the region harbouring IL2/IL21 associated with ulcerative colitis. Gut. [Meta-Analysis
Research Support, Non-U.S. Gov't]. 2009 Jun;58(6):799-804.
186. Liu Y, Helms C, Liao W, Zaba LC, Duan S, Gardner J, et al. A genome-wide association study of psoriasis and psoriatic arthritis identifies new disease loci. PLoS genetics. 2008 Mar;4(3):e1000041.

187. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol. [Research Support, N.I.H., Extramural]. 2010 Dec;34(8):816-34.
188. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics. [Research Support, N.I.H., Extramural
Research Support, Non-U.S. Gov't]. 2007 Sep;81(3):559-75.
189. Goyette P, Lefebvre C, Ng A, Brant SR, Cho JH, Duerr RH, et al. Gene-centric association mapping of chromosome 3p implicates MST1 in IBD pathogenesis. Mucosal immunology. [Multicenter Study
Research Support, N.I.H., Extramural
Research Support, Non-U.S. Gov't]. 2008 Mar;1(2):131-8.
190. Wang MH, Zhou YQ, Chen YQ. Macrophage-stimulating protein and RON receptor tyrosine kinase: potential regulators of macrophage inflammatory activities. Scand J Immunol. [Research Support, U.S. Gov't, P.H.S.
Review]. 2002 Dec;56(6):545-53.
191. Sharfe N, Dadi HK, Shahar M, Roifman CM. Human immune disorder arising from mutation of the alpha chain of the interleukin-2 receptor. Proceedings of the National Academy of Sciences of the United States of America. [Case Reports]. 1997 Apr 1;94(7):3168-71.
192. Maiti AK, Kim-Howard X, Viswanathan P, Guillen L, Rojas-Villarraga A, Deshmukh H, et al. Confirmation of an association between rs6822844 at the Il2-Il21 region and multiple autoimmune diseases: evidence of a general susceptibility locus. Arthritis Rheum. [Meta-Analysis
Research Support, N.I.H., Extramural
Research Support, Non-U.S. Gov't]. 2010 Feb;62(2):323-9.
193. Letourneau S, Krieg C, Pantaleo G, Boyman O. IL-2- and CD25-dependent immunoregulatory mechanisms in the homeostasis of T-cell subsets. J Allergy Clin Immunol. [Review]. 2009 Apr;123(4):758-62.
194. Malek TR. The biology of interleukin-2. Annu Rev Immunol. [Research Support, N.I.H., Extramural
Research Support, Non-U.S. Gov't
Review]. 2008;26:453-79.
195. Spolski R, Leonard WJ. Interleukin-21: a double-edged sword with therapeutic potential. Nat Rev Drug Discov. [Research Support, N.I.H., Intramural Review]. 2014 May;13(5):379-95.
196. Stolfi C, Rizzo A, Franze E, Rotondi A, Fantini MC, Sarra M, et al. Involvement of interleukin-21 in the regulation of colitis-associated colon cancer. The Journal of experimental medicine. [Research Support, Non-U.S. Gov't]. 2011 Oct 24;208(11):2279-90.
197. Boberg KM, Lind GE. Primary sclerosing cholangitis and malignancy. Best Pract Res Clin Gastroenterol. 2011 Dec;25(6):753-64.
198. Karlsen TH, Kaser A. Deciphering the genetic predisposition to primary sclerosing cholangitis. Seminars in liver disease. [Research Support, Non-U.S. Gov't
Review]. 2011 May;31(2):188-207.
199. Rai E, Wakeland EK. Genetic predisposition to autoimmunity--what have we learned? Semin Immunol. [Review]. 2011 Apr;23(2):67-83.

200. Karlsen TH, Melum E, Franke A. The utility of genome-wide association studies in hepatology. Hepatology. [Review]. 2010 May;51(5):1833-42.
201. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics. [Research Support, N.I.H., Extramural
Research Support, N.I.H., Intramural]. 2008 Dec 15;24(24):2938-9.
202. Liu JZ, Almarri MA, Gaffney DJ, Mells GF, Jostins L, Cordell HJ, et al. Dense fine-mapping study identifies new susceptibility loci for primary biliary cirrhosis. Nat Genet. [Research Support, N.I.H., Extramural
Research Support, Non-U.S. Gov't]. 2012 Oct;44(10):1137-41.
203. de Boer YS, van Gerven NM, Zwiers A, Verwer BJ, van Hoek B, van Erpecum KJ, et al. Genome-wide association study identifies variants associated with autoimmune hepatitis type 1. Gastroenterology. [Multicenter Study
Research Support, Non-U.S. Gov't]. 2014 Aug;147(2):443-52 e5.
204. Devalliere J, Charreau B. The adaptor Lnk (SH2B3): an emerging regulator in vascular cells and a link between immune and inflammatory signaling. Biochemical pharmacology. [Research Support, Non-U.S. Gov't
Review]. 2011 Nov 15;82(10):1391-402.
205. Zhernakova A, Elbers CC, Ferwerda B, Romanos J, Trynka G, Dubois PC, et al. Evolutionary and functional analysis of celiac risk loci reveals SH2B3 as a protective factor against bacterial infection. American journal of human genetics. [Research Support, N.I.H., Extramural
Research Support, Non-U.S. Gov't]. 2010 Jun 11;86(6):970-7.
206. Orr HT. Cell biology of spinocerebellar ataxia. The Journal of cell biology. [Research Support, N.I.H., Extramural
Review]. 2012 Apr 16;197(2):167-77.
207. Mells GF, Floyd JA, Morley KI, Cordell HJ, Franklin CS, Shin SY, et al. Genome-wide association study identifies 12 new susceptibility loci for primary biliary cirrhosis. Nat Genet. [Meta-Analysis
Research Support, N.I.H., Extramural
Research Support, Non-U.S. Gov't]. 2011 Mar 13;43(4):329-32.
208. Liu JZ, Hov JR, Folseraas T, Ellinghaus E, Rushbrook SM, Doncheva NT, et al. Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis. Nat Genet. [Meta-Analysis
Research Support, Non-U.S. Gov't]. 2013 Jun;45(6):670-5.
209. Ji SG, Juran BD, Mucha S, Folseraas T, Jostins L, Melum E, et al. Genome-wide association study of primary sclerosing cholangitis identifies new risk loci and quantifies the genetic relationship with inflammatory bowel disease. Nat Genet. 2017 Feb;49(2):269-73.
210. Cortes A, Brown MA. Promise and pitfalls of the Immunochip. Arthritis Res Ther. [Review]. 2011 Feb 1;13(1):101.
211. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. [Meta-Analysis
Research Support, N.I.H., Extramural
Research Support, Non-U.S. Gov't]. 2012 Nov 1;491(7422):119-24.
212. Ritchie GR, Dunham I, Zeggini E, Flicek P. Functional annotation of noncoding sequence variants. Nat Methods. [Research Support, Non-U.S. Gov't]. 2014 Mar;11(3):294-6.

213. Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat Genet. [Meta-Analysis
Research Support, Non-U.S. Gov't]. 2015 Sep;47(9):979-86.
214. Srivastava B, Mells GF, Cordell HJ, Muriithi A, Brown M, Ellinghaus E, et al. Fine mapping and replication of genetic risk loci in primary sclerosing cholangitis. Scandinavian journal of gastroenterology. [Meta-Analysis
Research Support, Non-U.S. Gov't]. 2012 Jul;47(7):820-6.
215. Ellinghaus D, Folseraas T, Holm K, Ellinghaus E, Melum E, Balschun T, et al. Genome-wide association analysis in primary sclerosing cholangitis and ulcerative colitis identifies risk loci at GPR35 and TCF4. Hepatology. [Multicenter Study
Research Support, Non-U.S. Gov't]. 2013 Sep;58(3):1074-83.
216. Ellinghaus D, Jostins L, Spain SL, Cortes A, Bethune J, Han B, et al. Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. Nature genetics. [Research Support, N.I.H., Extramural
Research Support, Non-U.S. Gov't]. 2016 May;48(5):510-8.
217. Jiang X, Karlsen TH. Genetics of primary sclerosing cholangitis and pathophysiological implications. Nature reviews Gastroenterology & hepatology. [Review]. 2017 May;14(5):279-95.
218. Sebode M, Peiseler M, Franke B, Schwinge D, Schoknecht T, Wortmann F, et al. Reduced FOXP3(+) regulatory T cells in patients with primary sclerosing cholangitis are associated with IL2RA gene polymorphisms. Journal of hepatology. [Research Support, Non-U.S. Gov't]. 2014 May;60(5):1010-6.
219. Alberts R, de Vries EMG, Goode EC, Jiang X, Sampaziotis F, Rombouts K, et al. Genetic association analysis identifies variants associated with disease progression in primary sclerosing cholangitis. Gut. [Multicenter Study
Research Support, Non-U.S. Gov't]. 2018 Aug;67(8):1517-24.
220. Folseraas T, Liaskou E, Anderson CA, Karlsen TH. Genetics in PSC: what do the "risk genes" teach us? Clinical reviews in allergy & immunology. [Review]. 2015 Jun;48(2-3):154-64.

Supplementary material

Chapter 4: Factors associated with clinical outcome and the development of a risk scoring system for patients with PSC.

Supplementary table 4.1. Univariate analysis using un-imputed data and multivariate analysis using 10 imputed data sets, of factors at diagnosis associated with 10-year risk of transplantation or death.

	Univariable Analysis		Multivariable Analysis	
Factor	**HR (95% CI)**	**p-value**	**HR (95% CI)**	**p-value**
Female	0.83 (0.64,1.07)	0.157		
Age at diagnosis	1.01 (1.01,1.02)	0.001	1.02 (1.01,1.03)	0.001
Extra-hepatic disease	1.76 (1.34,2.31)	<0.001	1.45 (1.09,1.92)	0.010
IBD presence	0.94 (0.65,1.37)	0.766		
UC				
CD				
IC				
Autoimmune disease	1.16 (0.84,1.60)	0.353		
Smoker	1.03 (0.79,1.34)	0.846		
Bilirubin				
35 -49	2.92 (1.67,5.11)	<0.001	1.92 (1.07,3.45)	0.030
50+	4.88 (3.52,6.75)	<0.001	2.50 (1.66,3.75)	0.000
ALP ratio				
1.5 - <2.5	1.15 (0.72,1.82)		1.15 (0.75,1.78)	0.519
2.5+	2.04 (1.46,2.85)	<0.001	1.70 (1.16,2.50)	0.007
ALT	1.02 (1.00,1.03)	0.030		
Albumin	0.92 (0.89,0.94)	<0.001	0.97 (0.94,1.00)	0.075
Haemoglobin	0.85 (0.79,0.92)	<0.001	0.99 (0.98,1.00)	0.037
Platelets group				
150 - 199	0.37 (0.21,0.66)	0.001	0.63 (0.36,1.11)	0.106
200 - 399	0.24 (0.16,0.37)	<0.001	0.40 (0.25,0.63)	0.000
400+	0.24 (0.14,0.41)	<0.001	0.24 (0.13,0.45)	0.000
Eosinophils	1.00 (0.82,1.21)	0.964		
Sodium	0.91 (0.87,0.96)	0.001		
Creatinine >120	1.75 (0.77,3.96)	0.181		
IgG	1.03 (0.95,1.12)	0.478		

Supplementary Box 4.1

Short-term UK-PSC Risk Score (R_{ST}) = 0.745 (Bili_t_0 Group 1 [0/1] + 1.613 (Bili_t_0 Group 2 [0/1]) – 0.061 (Alb_t_0 [g/l]) - 0.012 (Hb_t_0 [g/l]) -0.476 (Plts_t_0 Group 1 [0/1]) – 0.698 (Plts_t_0 Group 2 [0/1]) -0.962 (Plts_t_0 Group 3 [0/1]).

Long-term UK-PSC Risk Score (RS_{LT}) = 0.015 (Age_t_0[yrs]) + 1.022 (Bili_t_2 Group 1 [0/1]) + 1.156 (Bili_t_2 Group 2 [0/1]) + 0.462 (ALP_t_2 Group 1 [0/1]) + 0.490 (ALP_t_2 Group 2 [0/1]) +- 0.060 (Alb_t_2[g/l]) – 0.414 (Plts_t_2 Group 1) – 0.420 (Plts_t2 group 2) – 0.623 (Plts_t_2 Group 3 [0/1]) + 0.538 (disease type_t_0 [0/1]) + 0.987 (variceal bleed_t_2 [0/1]).

***Bili_t_0/t_2 group 1**; 0= Bili_t_0<35 µmol/l or >50 µmol/l, 1= 35 to ≤ 50 µmol/l*

***Bili_t_0/t_2 group 2**; 0=Bili_t_0 < 50 µmol/l, 1=Bili_t_0 ≥50 µmol/l*

***Plts_t_0/t_2 group 1**; 0=Plts_t_0<150×10^9/l, or ≥200×10^9/l, 1= Plts_t_0 150 to <200×10^9/l*

***Plts t_0/t_2 group 2**; 0=<200 or ≥ 400×10^9/l, 1= 200 to <400×10^9/l*

***Plts_t_0/t_2 group 3**; 0=<400×10^9/l, 1= ≥ 400×10^9/l*

***ALP_t_2 group 1**; 0=ALP_t_2<1.5×ULN or ≥2.5×ULN, 1=1.5 to <2.5,*

***ALP_t_2 group 2**; 0= ALP_t_2<2.5×ULN, 1=≥2.5×ULN,*

***Disease type_t_0**; 0=no extra-hepatic disease, 1=presence of extra-hepatic disease Variceal bleed_t_2; 0= no bleed by t_2, 1=bleed by t_2.*

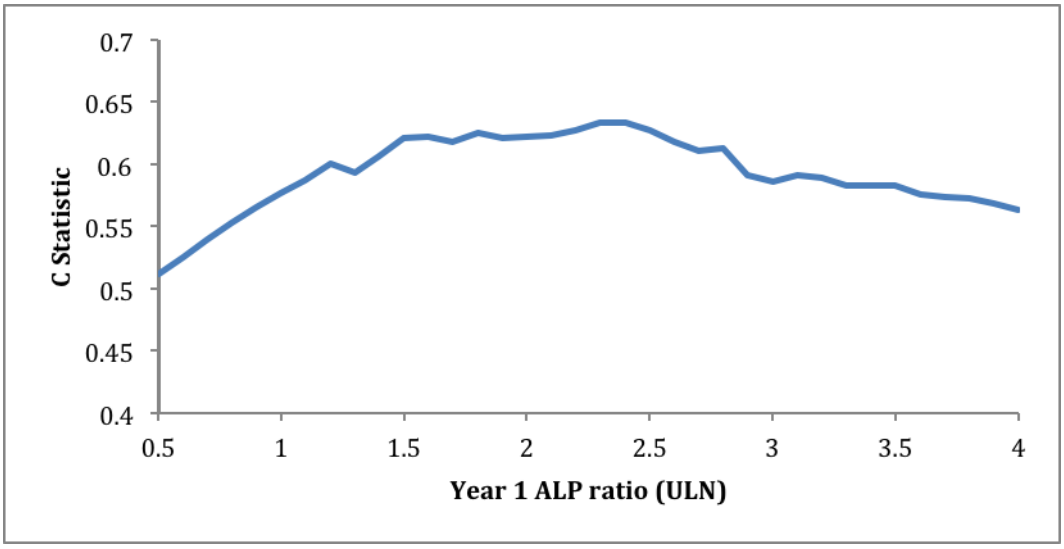
Predicted survival rate at time t = (Baseline survival at time t) ^ exp ($RS_{ST\ or\ LT}$),

RS_{ST} baseline survival at time t; 1 year: 0.0096612, 2 years: 0.0001109

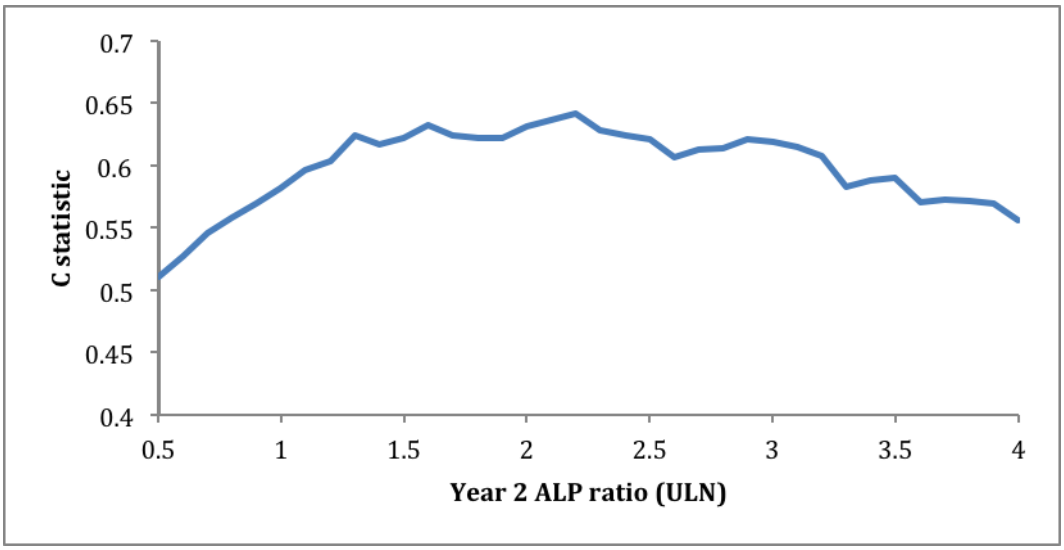
RS_{LT} baseline survival at time t; 1 year: 0.8913371, 2 years: 0.7695844, 5 years: 0.4129811, 8 years: 0.0218214.

Example

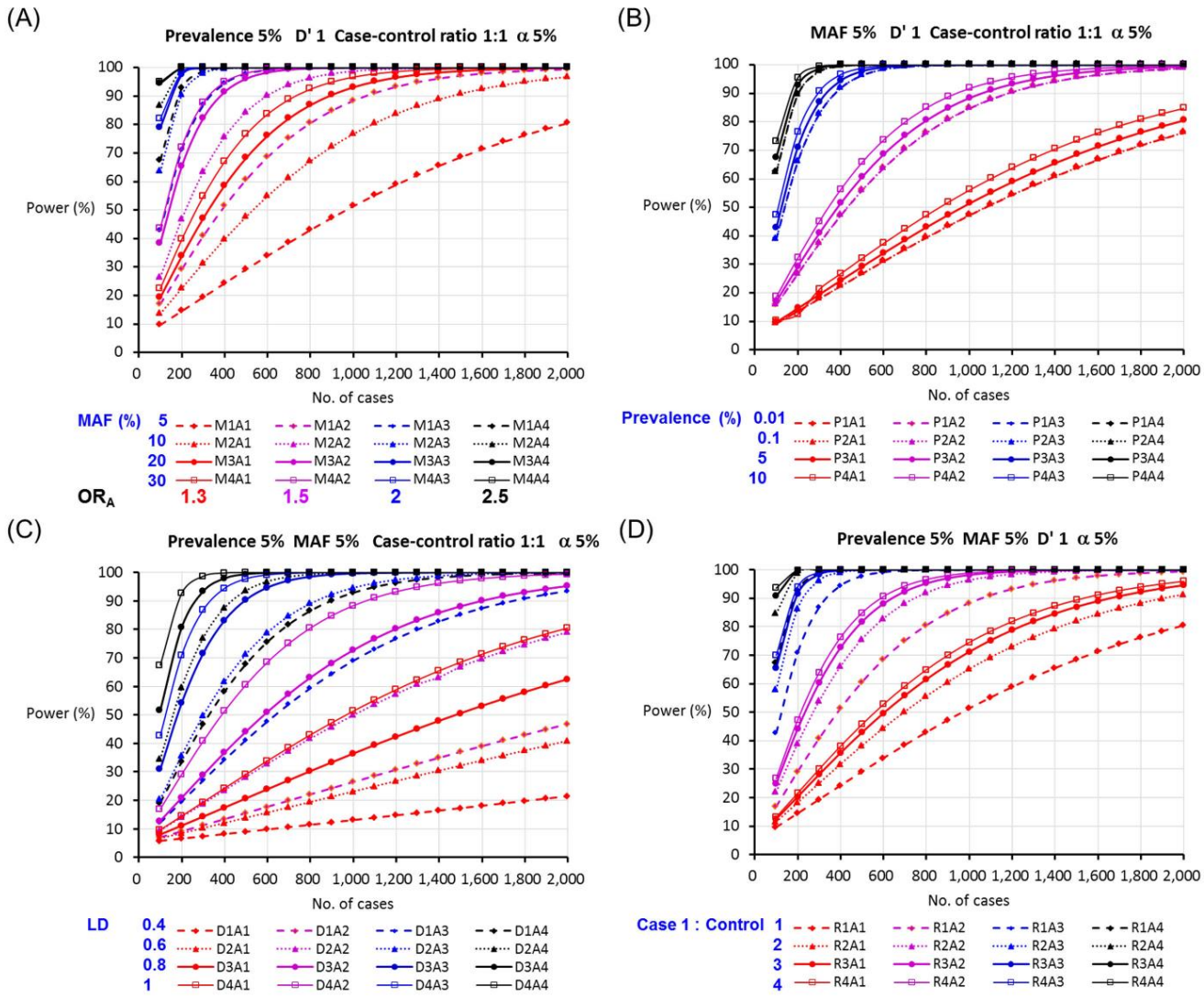
An individual diagnosed at age 47 and with no evidence of extra-hepatic disease at diagnosis with the following biochemistry at t_0: Bili 37μmol/l, Alb 34 g/l, Hb 130 g/l and Plts 245×10^9/l, and the following biochemistry at t_2: Bili 24μmol/l, ALP 2×ULN, Alb 30 g/l, Plts 152×10^9/l and no variceal bleed by t_2, would score would score: RS_{ST}= (0.745 ×1) – (0.061 ×34)-(0.012 ×130)- (0.698×1)= -3.587. Predicted event free survival rate at 2 years = (0.0001109)^exp (-3.587)= 0.78= 78%. RS_{LT}=(0.015 × 47) + 0.462 + (- 0.060 × 30)– 0.414 = -1.047. Predicted event free survival rate at 5 years = 0.413 ^ exp (-1.047) = 0.73 =73%.



Supplementary figure 4.1a): Harrell's C- statistic per ALP cut-point at year 1 for 10-year hazard of outcome.



Supplementary figure 4.1b): Harrell's C- statistic per ALP cut-point at year 2 for 10-year hazard of outcome.



Supplementary figure 5.1. The statistical power for the allelic test in a case-control study according to (A) minor allele frequency (MAF), (B) disease prevalence, (C) linkage disequilibrium (LD), and (D) case-to-control ratio (M, MAF; P, prevalence; D, LD; R, case-control ratio; A1=1.3, A2=1.5, A3=2, and A4=2.5 for heterozygous odds ratios) (figure shown is from Hong *et al* (158)).

Supplementary Table 6.1. Association results for all the studied SNPs.

Chr	Locus	SNP	Position	Alleles	MAF (cases/controls)	*P*-value (UK cohort)	OR (95% CI) (UK cohort)
2	*BCL2L11*	rs6720394	111,705,843	G/T	0.13/0.12	0.40	1.07 (0.92 - 1.23)
2	*GPBAR1*	rs12612347	218,765,583	G/A	0.48/0.50	0.21	0.94 (0.85 - 1.04)
3	***MST1***	**rs3197999**	**49,696,536**	**A/G**	**0.33/0.28**	**1.9E-06**	**1.28 (1.16 - 1.42)**
13	*GPC6*	rs9524260	93,311,791	A/G	0.40/0.40	0.77	0.99 (0.89 - 1.09)
4	***IL-2/IL-21***	**rs12511287**	**123,797,981**	**A/T**	**0.31/0.27**	**3.0E-04**	**1.21 (1.09 - 1.35)**
4	***IL-2/IL-21***	**rs6822844**	**123,728,871**	**T/G**	**0.15/0.17**	**1.9E-03**	**0.81 (0.70 - 0.92)**
4	***IL-2/IL-21***	**rs13151961**	**123,334,952**	**G/A**	**0.15/0.17**	**3.2E-03**	**0.82 (0.71 - 0.93)**
4	***IL-2/IL-21***	**rs13125087**	**123,238,043**	**T/A**	**0.16/0.19**	**3.9E-03**	**0.82 (0.72 - 0.94)**
4	***IL-2/IL-21***	**rs11938795**	**123,292,459**	**C/T**	**0.23/0.26**	**5.9E-03**	**0.85 (0.76 - 0.95)**
4	***IL-2/IL-21***	**rs6840978**	**123,774,157**	**T/C**	**0.18/0.21**	**8.5E-03**	**0.85 (0.75 - 0.96)**
4	***IL-2/IL-21***	**rs6827444**	**123,352,628**	**G/C**	**0.22/0.19**	**0.02**	**1.15 (1.02 - 1.30)**
4	***IL-2/IL-21***	**rs13143866**	**123,760,208**	**A/G**	**0.26/0.28**	**0.028**	**0.89 (0.79 - 0.99)**
4	***IL-2/IL-21***	**rs7693745**	**123,455,033**	**G/A**	**0.29/0.26**	**0.034**	**1.12 (1.01 - 1.25)**
4	***IL-2/IL-21***	**rs2137497**	**123,777,704**	**T/G**	**0.43/0.40**	**0.040**	**1.11 (1.01 - 1.22)**
4	*IL-2/IL-21*	rs12642902	123,727,951	A/G	0.32/0.34	0.066	0.91 (0.82 - 1.01)
4	*IL-2/IL-21*	rs11722421	123,491,118	C/G	0.35/0.33	0.20	1.07 (0.97 - 1.18)
4	*IL-2/IL-21*	rs6849146	123,545,541	C/T	0.39/0.38	0.21	1.06 (0.97 - 1.17)
4	*IL-2/IL-21*	rs975404	123,740,742	C/T	0.35/0.36	0.28	0.94 (0.85 - 1.05)
4	*IL-2/IL-21*	rs17454584	123,572,882	G/A	0.23/0.22	0.34	1.06 (0.94 - 1.19)
4	*IL-2/IL-21*	rs1512971	123,744,785	T/C	0.28/0.27	0.36	1.05 (0.94 - 1.17)
4	*IL-2/IL-21*	rs6534355	123,781,100	T/C	0.24/0.24	0.37	1.05 (0.94 - 1.18)
4	*IL-2/IL-21*	rs4295278	123,766,991	C/T	0.06/0.05	0.38	1.10 (0.89 - 1.34)
4	*IL-2/IL-21*	rs4833810	123,237,840	T/C	0.39/0.38	0.45	1.04 (0.94 - 1.15)
4	*IL-2/IL-21*	rs4833834	123,685,801	G/A	0.09/0.10	0.45	0.94 (0.79 - 1.11)
4	*IL-2/IL-21*	rs6534338	123,246,319	T/C	0.30/0.29	0.52	1.04 (0.93 - 1.15)
4	*IL-2/IL-21*	rs13110000	123,797,510	C/T	0.44/0.43	0.52	1.03 (0.94 - 1.14)
4	*IL-2/IL-21*	rs7678445	123,502,222	T/G	0.08/0.07	0.53	1.06 (0.89 - 1.27)
4	*IL-2/IL-21*	rs13119723	123,437,763	G/A	0.15/0.15	0.65	0.97 (0.85 - 1.11)
4	*IL-2/IL-21*	rs11931332	123,236,177	C/T	0.34/0.34	0.78	1.01 (0.92 - 1.12)
4	*IL-2/IL-21*	rs6848868	123,369,736	C/T	0.08/0.08	0.81	0.98 (0.82 - 1.16)
4	*IL-2/IL-21*	rs6419221	123,783,569	T/C	0.37/0.37	0.87	0.99 (0.90 - 1.10)
4	*IL-2/IL-21*	rs1022234	123,782,528	G/A	0.31/0.32	0.91	0.99 (0.90 - 1.10)
10	***IL2RA***	**rs4147359**	**6,148,445**	**A/G**	**0.39/0.34**	**2.6E-04**	**1.20 (1.09 - 1.33)**
10	***IL2RA***	**rs706778**	**6,138,955**	**T/C**	**0.44/0.40**	**4.3E-04**	**1.19 (1.08 - 1.31)**
10	***IL2RA***	**rs7090530**	**6,150,881**	**C/A**	**0.36/0.40**	**7.0E-04**	**0.84 (0.76 - 0.93)**
10	***IL2RA***	**rs10905718**	**6,154,862**	**G/A**	**0.35/0.31**	**1.0E-03**	**1.18 (1.07 - 1.31)**
10	***IL2RA***	**rs10905669**	**6,132,099**	**T/C**	**0.27/0.23**	**1.6E-03**	**1.20 (1.07 - 1.33)**
10	***IL2RA***	**rs11594656**	**6,162,015**	**A/T**	**0.21/0.24**	**2.8E-03**	**0.84 (0.74 - 0.94)**
10	***IL2RA***	**rs10905716**	**6,154,016**	**T/C**	**0.19/0.22**	**4.2E-03**	**0.84 (0.74 - 0.95)**
10	***IL2RA***	**rs706779**	**6,138,830**	**T/C**	**0.44/0.47**	**8.8E-03**	**0.88 (0.80 - 0.97)**
10	***IL2RA***	**rs7090512**	**6,150,835**	**C/T**	**0.27/0.30**	**0.011**	**0.87 (0.78 - 0.97)**
10	***IL2RA***	**rs706781**	**6,126,391**	**C/T**	**0.28/0.26**	**0.021**	**1.14 (1.02 - 1.27)**
10	***IL2RA***	**rs4749924**	**6,122,402**	**C/A**	**0.30/0.33**	**0.025**	**0.89 (0.80 - 0.99)**

10	***IL2RA***	**rs11256456**	**6,120,718**	**C/T**	**0.23/0.21**	**0.032**	**1.14 (1.01 - 1.28)**
10	***IL2RA***	**rs2104286**	**6,139,051**	**C/T**	**0.26/0.28**	**0.044**	**0.89 (0.80 - 1.00)**
10	*IL2RA*	rs2476491	6,135,416	T/A	0.27/0.29	0.053	0.90 (0.81 - 1.00)
10	*IL2RA*	rs11256497	6,127,800	A/G	0.35/0.37	0.072	0.91 (0.82 - 1.01)
10	*IL2RA*	rs12722563	6,109,567	A/G	0.10/0.12	0.085	0.87 (0.75 - 1.02)
10	*IL2RA*	rs2256774	6,137,171	C/T	0.32/0.34	0.11	0.92 (0.83 - 1.02)
10	*IL2RA*	rs12251307	6,163,501	T/C	0.11/0.12	0.11	0.88 (0.76 - 1.03)
10	*IL2RA*	rs12359875	6,091,113	T/C	0.23/0.25	0.14	0.92 (0.82 - 1.03)
10	*IL2RA*	rs6602392	6,118,085	A/C	0.10/0.09	0.18	1.12 (0.95 - 1.31)
10	*IL2RA*	rs791587	6,128,705	A/G	0.45/0.46	0.21	0.94 (0.85 - 1.04)
10	*IL2RA*	rs9663421	6,095,610	T/C	0.27/0.28	0.22	0.93 (0.84 - 1.04)
10	*IL2RA*	rs11598648	6,124,031	A/G	0.44/0.43	0.22	1.06 (0.96 - 1.17)
10	*IL2RA*	rs11256457	6,120,800	G/C	0.38/0.40	0.25	0.94 (0.86 - 1.04)
10	*IL2RA*	rs2076846	6,103,259	G/A	0.37/0.36	0.40	1.04 (0.94 - 1.15)
10	*IL2RA*	rs4749955	6,158,972	C/T	0.44/0.45	0.44	0.96 (0.87 - 1.06)
10	*IL2RA*	rs12722596	6,096,300	C/T	0.10/0.10	0.55	1.05 (0.90 - 1.23)
10	*IL2RA*	rs12722588	6,100,439	T/C	0.19/0.19	0.68	1.03 (0.91 - 1.16)
10	*IL2RA*	rs7093069	6,103,325	T/C	0.19/0.19	0.68	1.03 (0.91 - 1.16)
10	*IL2RA*	rs12722489	6,142,018	T/C	0.16/0.16	0.72	0.98 (0.86 - 1.11)
10	*IL2RA*	rs12244380	6,093,380	G/A	0.43/0.43	0.81	1.01 (0.92 - 1.12)

Association results for all the 63 SNPs analysed in the UK cohort (992 PSC cases and 5162 controls) are shown with the method of genotyping used in the control population. All the 63 SNPs were genotyped in cases. Logistic regression analysis was performed in PLINK v1.07. Position refers to NCBI's build 36. Bonferroni corrected *P*-value threshold for significance was determined to be $< 7.9 \times 10^{-4}$. SNPs with nominal association (*P*-value < 0.05) are shown in boldface.

All the reported odds ratios are with reference to minor allele vs. major allele.

Chr – chromosome, MAF – Minor allele frequency, OR – Odds ratio, CI – Confidence interval.

Supplementary Table 6.2. Association results for all the SNPs (nominally associated) analysed after excluding patients with auto-immune disease.

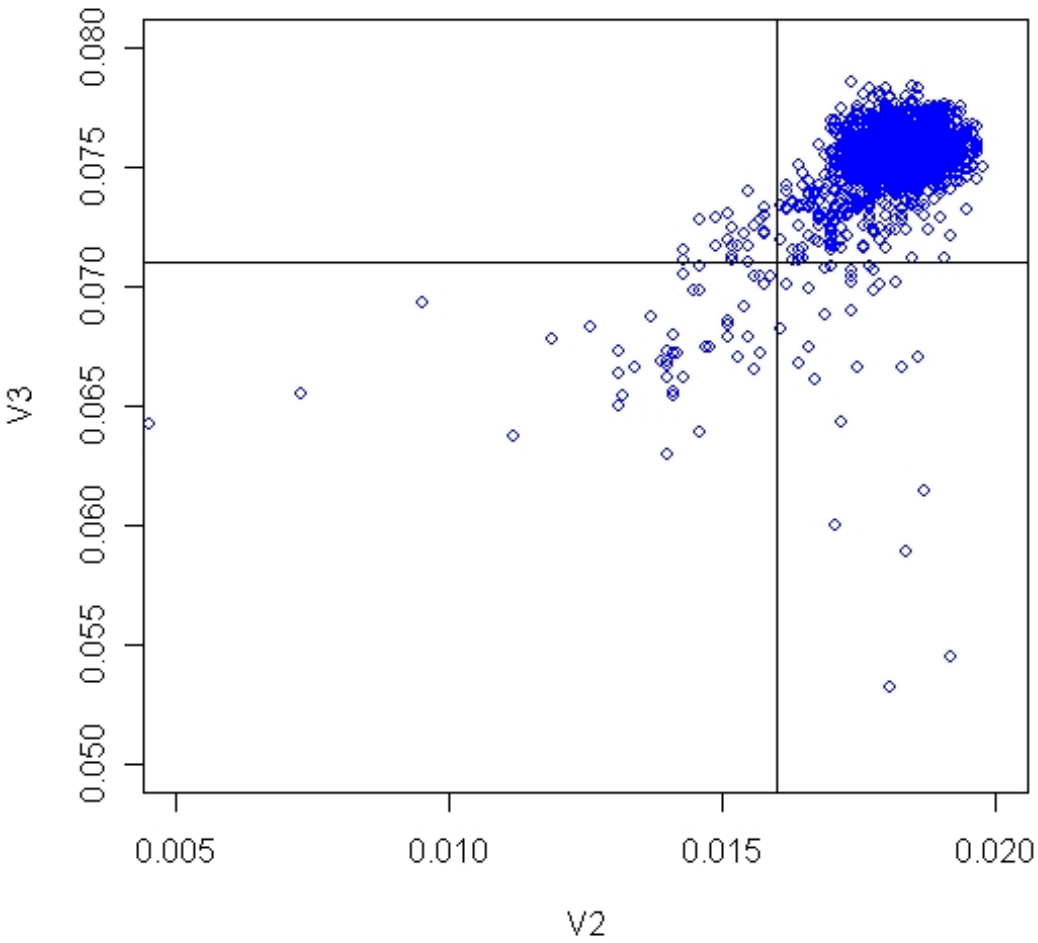
Chr	Locus	SNP	BP	Alleles	MAF (cases/controls)	*P*-value (UK cohort)	OR (95% CI) (UK cohort)	*P*-value (AID excluded)	OR (95% CI) (AID excluded)
3	***MST1***	**rs3197999**	**49,696,536**	**A/G**	**0.33/0.28**	**1.90E-06**	**1.28 (1.16 - 1.42)**	**1.20E-06**	**1.31 (1.17 - 1.46)**
4	***IL-2/IL-21***	**rs12511287**	**123,797,981**	**A/T**	**0.31/0.27**	**3.00E-04**	**1.21 (1.09 - 1.35)**	**4.70E-04**	**1.22 (1.09 - 1.36)**
4	*IL-2/IL-21*	rs6822844	123,728,871	T/G	0.15/0.17	1.90E-03	0.81 (0.70 - 0.92)	6.50E-03	0.82 (0.71 - 0.95)
4	*IL-2/IL-21*	rs13125087	123,238,043	T/A	0.16/0.19	3.90E-03	0.82 (0.72 - 0.94)	7.00E-03	0.83 (0.72 - 0.95)
4	*IL-2/IL-21*	rs13151961	123,334,952	G/A	0.15/0.17	3.20E-03	0.82 (0.71 - 0.93)	0.01	0.83 (0.72 - 0.96)
4	*IL-2/IL-21*	rs11938795	123,292,459	C/T	0.23/0.26	5.90E-03	0.85 (0.76 - 0.95)	0.023	0.87 (0.77 - 0.98)
4	*IL-2/IL-21*	rs6840978	123,774,157	T/C	0.18/0.21	8.50E-03	0.85 (0.75 - 0.96)	0.012	0.85 (0.74 - 0.96)
4	*IL-2/IL-21*	rs6827444	123,352,628	G/C	0.22/0.19	0.02	1.15 (1.02 - 1.30)	0.017	1.17 (1.03 - 1.32)
4	*IL-2/IL-21*	rs13143866	123,760,208	A/G	0.26/0.28	0.028	0.89 (0.79 - 0.99)	0.042	0.89 (0.79 - 1.00)
4	*IL-2/IL-21*	rs2137497	123,777,704	T/G	0.43/0.40	0.04	1.11 (1.01 - 1.22)	0.018	1.13 (1.02 - 1.26)
10	*IL2RA*	rs10905669	6,132,099	T/C	0.27/0.23	1.60E-03	1.20 (1.07 - 1.33)	5.10E-03	1.18 (1.05 - 1.33)
10	*IL2RA*	rs4147359	6,148,445	A/G	0.39/0.34	2.60E-04	1.20 (1.09 - 1.33)	1.20E-03	1.19 (1.07 - 1.32)
10	*IL2RA*	rs706778	6,138,955	T/C	0.44/0.40	4.30E-04	1.19 (1.08 - 1.31)	2.90E-03	1.17 (1.05 - 1.29)
10	*IL2RA*	rs7090530	6,150,881	C/A	0.36/0.40	7.00E-04	0.84 (0.76 - 0.93)	2.10E-03	0.85 (0.76 - 0.94)
10	*IL2RA*	rs10905718	6,154,862	G/A	0.35/0.31	1.00E-03	1.18 (1.07 - 1.31)	3.00E-03	1.18 (1.06 -1.31)
10	*IL2RA*	rs11594656	6,162,015	A/T	0.21/0.24	2.80E-03	0.84 (0.74 - 0.94)	7.10E-03	0.84 (0.75 - 0.95)
10	*IL2RA*	rs10905716	6,154,016	T/C	0.19/0.22	4.20E-03	0.84 (0.74 - 0.95)	0.015	0.85 (0.75 - 0.97)
10	*IL2RA*	rs706779	6,138,830	T/C	0.44/0.47	8.80E-03	0.88 (0.80 - 0.97)	0.019	0.88 (0.80 - 0.98)
10	*IL2RA*	rs7090512	6,150,835	C/T	0.27/0.30	0.011	0.87 (0.78 - 0.97)	0.024	0.88 (0.78 - 0.98)
10	*IL2RA*	rs706781	6,126,391	C/T	0.28/0.26	0.021	1.14 (1.02 - 1.27)	0.031	1.13 (1.01 - 1.27)

A phenotype based sub-group analysis was performed against 5162 controls for PSC cases without autoimmune disease (AID) (n = 875). Results of this sub-group analysis is shown in the column marked as: *P*-value$_{(AID\ excluded)}$. Only SNPs with nominal association (*P*-value$_{(AID\ excluded)}$ < 0.05) are shown. Bonferroni corrected *P*-value threshold for significance was determined to be < 7.9×10^{-4} and SNPs reaching this threshold are shown in boldface. Position refers to NCBI's build 36. All the reported odds ratios are with reference to minor allele vs. major allele. Chr – chromosome, MAF – Minor allele frequency, OR – Odds ratio, CI – Confidence interval.

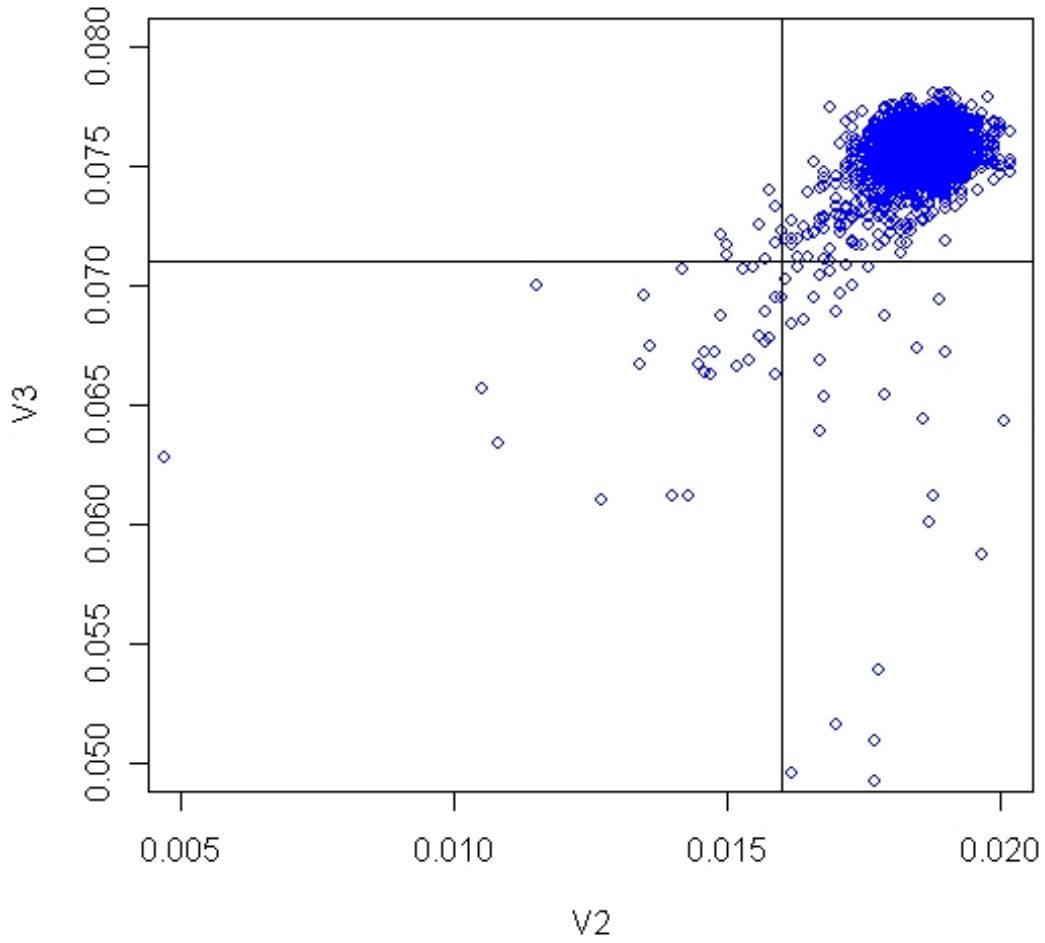
Supplementary Table 6.3. Association results for all the SNPs (nominally associated) analysed for patients with or without IBD.

Chr	Locus	SNP	Position	Alleles	MAF (cases/controls)	*P*-value (UK cohort)	OR (95% CI) (UK cohort)	*P*-value (PSC with IBD)	*P*-value (PSC without IBD)
3	***MST1***	**rs3197999**	**49,696,536**	**A/G**	**0.33/0.28**	**1.9E-06**	**1.28 (1.16 - 1.42)**	1.2E-05	7.2E-03
4	*IL-2/IL-21*	rs12511287	123,797,981	A/T	0.31/0.27	3.0E-04	1.21 (1.09 - 1.35)	4.9E-03	9.0E-03
4	*IL-2/IL-21*	rs6822844	123,728,871	T/G	0.15/0.17	1.9E-03	0.81 (0.70 - 0.92)	0.039	7.3E-03
4	*IL-2/IL-21*	rs13151961	123,334,952	G/A	0.15/0.17	3.2E-03	0.82 (0.71 - 0.93)	0.038	0.016
4	*IL-2/IL-21*	rs13125087	123,238,043	T/A	0.16/0.19	3.9E-03	0.82 (0.72 - 0.94)	0.019	0.04
4	*IL-2/IL-21*	rs6840978	123,774,157	T/C	0.18/0.21	8.5E-03	0.85 (0.75 - 0.96)	0.10	0.017
4	*IL-2/IL-21*	rs13143866	123,760,208	A/G	0.26/0.28	0.028	0.89 (0.79 - 0.99)	0.19	0.046
10	*IL2RA*	rs4147359	6,148,445	A/G	0.39/0.34	2.6E-04	1.20 (1.09 - 1.33)	1.3E-03	0.031
10	*IL2RA*	rs10905669	6,132,099	T/C	0.27/0.23	1.6E-03	1.20 (1.07 - 1.33)	2.8E-03	0.077
10	*IL2RA*	rs706778	6,138,955	T/C	0.44/0.40	4.3E-04	1.19 (1.08 - 1.31)	4.6E-03	0.020
10	*IL2RA*	rs10905718	6,154,862	G/A	0.35/0.31	1.0E-03	1.18 (1.07 - 1.31)	4.9E-03	0.036
10	*IL2RA*	rs706781	6,126,391	C/T	0.28/0.26	0.021	1.14 (1.02 - 1.27)	5.3E-03	0.55
10	*IL2RA*	rs7090530	6,150,881	C/A	0.36/0.40	7.0E-04	0.84 (0.76 - 0.93)	0.041	1.9E-03
10	*IL2RA*	rs11594656	6,162,015	A/T	0.21/0.24	2.8E-03	0.84 (0.74 - 0.94)	0.086	4.7E-03
10	*IL2RA*	rs10905716	6,154,016	T/C	0.19/0.22	4.2E-03	0.84 (0.74 - 0.95)	0.085	7.6E-03
10	*IL2RA*	rs706779	6,138,830	T/C	0.44/0.47	8.8E-03	0.88 (0.80 - 0.97)	0.11	0.016
10	*IL2RA*	rs7090512	6,150,835	C/T	0.27/0.30	0.011	0.87 (0.78 - 0.97)	0.25	4.8E-03
10	*IL2RA*	rs4749924	6,122,402	C/A	0.30/0.33	0.025	0.89 (0.80 - 0.99)	0.30	0.017
10	*IL2RA*	rs11256456	6,120,718	C/T	0.23/0.21	0.032	1.14 (1.01 - 1.28)	0.024	0.36
10	*IL2RA*	rs2104286	6,139,051	C/T	0.26/0.28	0.044	0.89 (0.80 - 1.00)	0.017	0.74

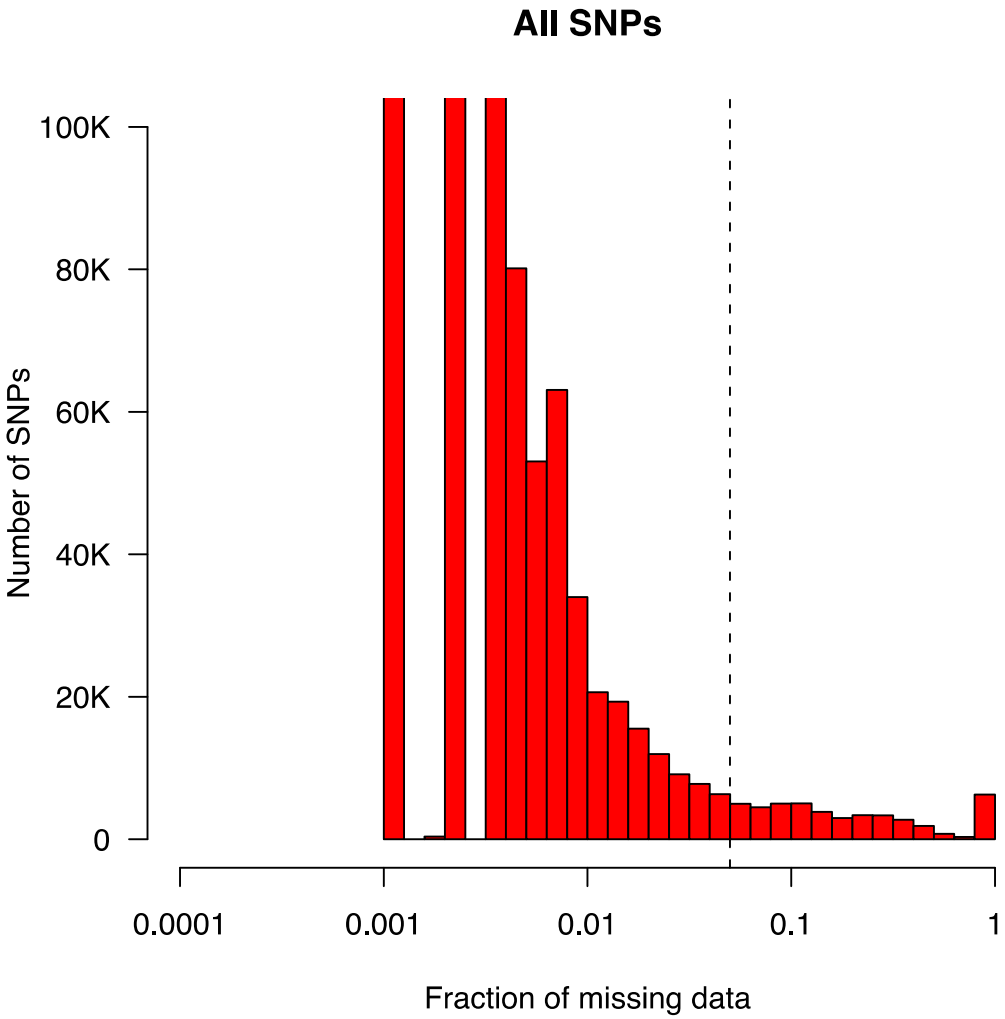
A phenotype based sub-group analysis was performed against 5162 controls for PSC cases with IBD (n = 625); and PSC cases without IBD (n = 367). Results (*P*-values only) of these sub-group analyses are given in the last two columns marked as: $P\text{-value}_{(\text{PSC with IBD})}$; and $P\text{-value}_{(\text{PSC without IBD})}$ respectively. Only SNPs with nominal association ($P\text{-value}_{(\text{AID excluded})} < 0.05$) are shown. Bonferroni corrected *P*-value threshold for significance was determined to be < 7.9×10^{-4} and the only SNP reaching this threshold is shown in boldface. Position refers to NCBI's build 36. All the reported odds ratios are with reference to minor allele vs. major allele. Chr – chromosome, MAF – Minor allele frequency, OR – Odds ratio, CI – Confidence interval.



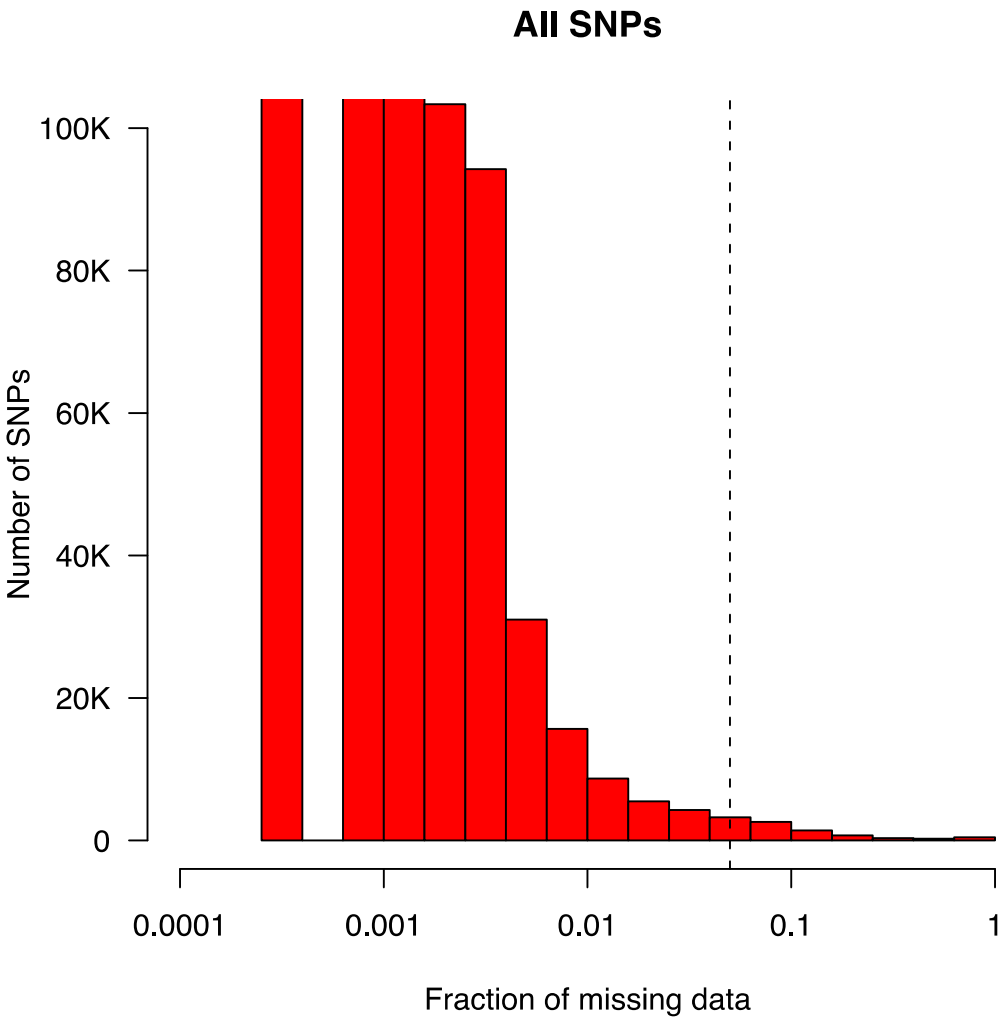
Supplementary Figure 7.1 (a). Individuals in the BBC control group with first principal component score < 0.16 and second principal component score < 0.072 were removed.



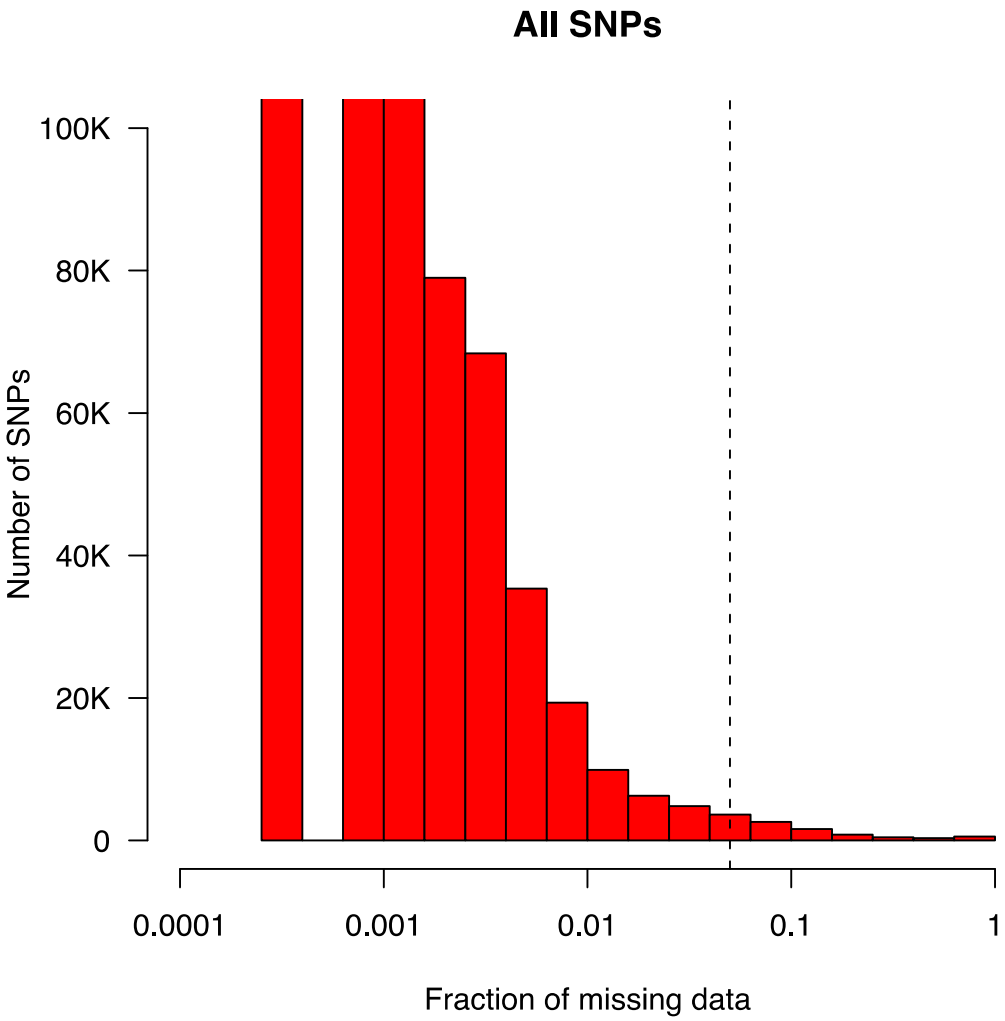
Supplementary Figure 7.1 (b). Individuals in the NBS control group with first principal component score < 0.16 and second principal component score < 0.072 were removed.



Supplementary Figure 7.2 (a). Histogram of missing data rate across all PSC cases. The vertical dashed line represents the threshold (3%) at which SNPs were removed from further analysis because of an excess failure rate.



Supplementary Figure 7.2 (b). Histogram of missing data rate across all BBC controls. The vertical dashed line represents the threshold (3%) at which SNPs were removed from further analysis because of excessive failure rate.



Supplementary Figure 7.2 (c). Histogram of missing data rate across all NBS controls. The vertical dashed line represents the threshold (3%) at which SNPs were removed from further analysis because of excessive failure rate.

Supplementary Table 7.1. Genetic association results for suggestive associations (p-value < 1 x 10^{-4}) in the discovery panel.

CHR	SNP	BP	A1	F_A	F_U	A2	CHISQ	P	OR	L95	U95
1	rs10492966	8598005	A	0.2241	0.1846	G	18.81	1.45E-05	1.276	1.143	1.424
1	rs11121182	8522553	A	0.2234	0.1849	C	17.88	2.35E-05	1.268	1.136	1.416
1	rs10492965	8598127	A	0.2221	0.1837	G	17.8	2.46E-05	1.268	1.135	1.416
1	rs9435739	17342924	G	0.2937	0.3388	A	17.24	3.30E-05	0.8113	0.735	0.8956
1	rs11121203	8648240	G	0.2226	0.185	A	17.03	3.68E-05	1.261	1.129	1.408
1	rs7553298	8770883	G	0.2361	0.1985	A	16.05	6.16E-05	1.247	1.119	1.39
1	rs12087224	84130384	A	0.122	0.09476	G	15.51	8.21E-05	1.328	1.153	1.53
2	rs1516637	112634339	G	0.2933	0.3491	A	25.92	3.56E-07	0.7739	0.7011	0.8543
2	rs13389578	145216048	G	0.1063	0.08101	A	15.2	9.66E-05	1.349	1.16	1.568
3	rs3197999	49721532	A	0.3336	0.2772	G	29.14	6.75E-08	1.306	1.185	1.439
3	rs1131095	49714225	G	0.3398	0.2845	A	27.49	1.58E-07	1.294	1.175	1.425
3	rs11130213	49712297	A	0.3395	0.2845	G	27.17	1.86E-07	1.292	1.173	1.424
3	rs4283605	49678651	G	0.3398	0.285	A	27.05	1.98E-07	1.291	1.173	1.422
3	rs11706370	49441091	A	0.3521	0.2987	G	25.02	5.67E-07	1.276	1.16	1.404
3	rs17080528	49389842	A	0.3516	0.2988	G	24.49	7.48E-07	1.273	1.157	1.4
3	rs10640	49454277	A	0.3516	0.2991	G	24.21	8.63E-07	1.271	1.155	1.398
3	rs6997	49453834	A	0.3515	0.2991	G	24.02	9.53E-07	1.27	1.154	1.397
3	rs11715915	49455330	A	0.3516	0.2993	G	23.99	9.66E-07	1.27	1.154	1.397
3	rs6766131	49538932	G	0.3494	0.2975	A	23.74	1.10E-06	1.268	1.153	1.396
3	rs4625	49572140	G	0.3516	0.2996	A	23.7	1.12E-06	1.268	1.152	1.395
3	rs13070798	48705934	G	0.1402	0.1049	A	23.56	1.21E-06	1.391	1.217	1.591
3	rs2286652	48689192	A	0.1396	0.1044	G	23.47	1.27E-06	1.391	1.217	1.591
3	rs3811697	49590770	A	0.3538	0.3025	G	22.95	1.66E-06	1.262	1.147	1.389
3	rs9836462	48712791	G	0.1427	0.1086	A	21.45	3.63E-06	1.367	1.197	1.56
3	rs12107418	48689787	G	0.142	0.1085	A	20.67	5.46E-06	1.36	1.191	1.554
3	rs13063312	48661985	A	0.1251	0.09551	G	18.08	2.11E-05	1.354	1.177	1.558
3	rs11719291	48735706	G	0.151	0.1186	A	17.98	2.23E-05	1.321	1.161	1.503
3	rs13324142	48669447	A	0.1247	0.09557	G	17.49	2.89E-05	1.348	1.171	1.551
3	rs2276850	48669648	A	0.1247	0.09567	G	17.36	3.09E-05	1.346	1.17	1.549

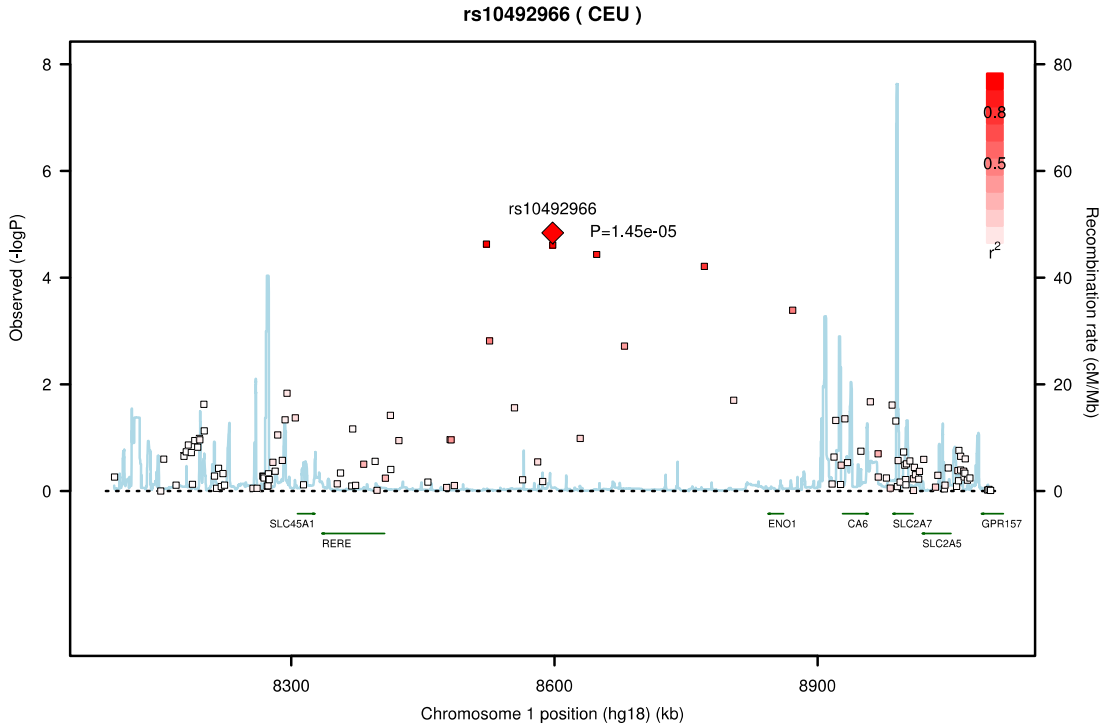
3	rs3821876	48668394	A	0.1251	0.09608	C	17.32	3.16E-05	1.345	1.169	1.548
3	rs2276852	48666923	G	0.1305	0.1009	A	17.2	3.36E-05	1.337	1.165	1.534
3	rs2953130	128170341	A	0.1182	0.09041	G	16.72	4.34E-05	1.348	1.168	1.557
4	rs345367	86754638	A	0.2612	0.3112	G	22.04	2.67E-06	0.7828	0.7067	0.8672
4	rs7665833	131381117	A	0.1523	0.1174	G	21.1	4.36E-06	1.351	1.188	1.537
4	rs11735471	189536624	G	0.4165	0.4657	A	18.23	1.96E-05	0.819	0.7472	0.8977
4	rs2088131	11511291	A	0.3556	0.4006	G	15.94	6.54E-05	0.8254	0.7512	0.907
4	rs4862251	184687644	A	0.2834	0.2435	G	15.88	6.76E-05	1.229	1.11	1.36
4	rs11734090	123228113	G	0.2206	0.2598	A	15.23	9.52E-05	0.8063	0.7235	0.8985
5	rs419119	6021846	A	0.172	0.2182	C	23.92	1.00E-06	0.7443	0.661	0.8381
5	rs2029036	118275869	C	0.1049	0.07897	A	16.44	5.03E-05	1.367	1.175	1.591
5	rs17648108	177831556	G	0.3128	0.2715	A	15.92	6.62E-05	1.222	1.107	1.348
5	rs1567520	118159871	A	0.08692	0.06392	G	15.6	7.81E-05	1.394	1.181	1.645
7	rs10951953	5510159	G	0.4512	0.5032	A	20.26	6.77E-06	0.8115	0.7409	0.8889
7	rs10260121	115110126	G	0.1431	0.1117	A	17.76	2.50E-05	1.328	1.163	1.515
7	rs2023702	114945308	G	0.08304	0.0604	A	15.9	6.69E-05	1.409	1.19	1.669
7	rs1477216	155235978	A	0.4342	0.3894	G	15.65	7.63E-05	1.203	1.098	1.319
8	rs9657390	2546665	G	0.3216	0.3677	A	17.22	3.32E-05	0.8152	0.7401	0.8979
8	rs17070773	4463359	A	0.06409	0.09053	G	16.67	4.45E-05	0.688	0.5744	0.824
8	rs2617094	4456167	A	0.09991	0.1298	G	15.31	9.10E-05	0.744	0.6413	0.8632
9	rs12237858	33439738	A	0.1107	0.05898	C	78.67	7.36E-19	1.987	1.703	2.318
9	rs7027092	95812707	G	0.2722	0.3223	A	21.82	2.99E-06	0.7863	0.7107	0.8699
9	rs16931895	117609919	C	0.1493	0.1188	A	15.94	6.55E-05	1.301	1.143	1.481
10	rs436207	61087914	C	0.2885	0.3647	A	47.28	6.15E-12	0.7062	0.6394	0.7801
10	rs3118470	6101713	G	0.363	0.3164	A	18.56	1.65E-05	1.231	1.12	1.354
10	rs953355	124515295	A	0.2125	0.1748	G	17.85	2.39E-05	1.274	1.138	1.425
10	rs11255072	5806248	A	0.163	0.1302	G	17.12	3.51E-05	1.301	1.148	1.475
10	rs990394	71347276	A	0.2524	0.2129	C	17.07	3.61E-05	1.249	1.124	1.388
10	rs10885476	115330721	A	0.1308	0.1655	G	16.8	4.16E-05	0.759	0.6649	0.8663
10	rs1658495	60280153	A	0.4965	0.4496	C	16.56	4.72E-05	1.207	1.102	1.322
10	rs7099881	60316211	G	0.4908	0.444	A	16.49	4.88E-05	1.207	1.102	1.321

10	rs1649048	60324282	G	0.496	0.4494	A	16.42	5.09E-05	1.206	1.102	1.321
10	rs1649031	60279742	A	0.4956	0.449	G	16.37	5.20E-05	1.206	1.101	1.32
10	rs1904694	52905494	G	0.4451	0.4001	A	15.69	7.45E-05	1.203	1.098	1.318
10	rs7072793	6106266	G	0.4535	0.4084	A	15.61	7.78E-05	1.202	1.097	1.316
10	rs12252820	71351990	A	0.1646	0.1332	G	15.44	8.54E-05	1.282	1.132	1.452
10	rs1413611	36466891	G	0.4214	0.3774	A	15.3	9.16E-05	1.202	1.096	1.317
10	rs12412095	6113523	G	0.3528	0.3107	A	15.26	9.38E-05	1.209	1.099	1.33
10	rs1451193	36461567	G	0.4583	0.5033	A	15.19	9.74E-05	0.8348	0.7623	0.9142
10	rs1343065	54676411	G	0.3827	0.3398	A	15.15	9.94E-05	1.204	1.097	1.323
11	rs11221265	128236936	A	0.14	0.1106	G	15.81	6.99E-05	1.309	1.146	1.496
11	rs836141	34537316	A	0.4421	0.3975	G	15.4	8.72E-05	1.201	1.096	1.316
11	rs7936255	38998451	A	0.294	0.3365	G	15.26	9.39E-05	0.8212	0.7439	0.9066
12	rs10774625	111910219	G	0.4362	0.4974	A	28	1.22E-07	0.7818	0.7136	0.8566
12	rs1265566	111716376	G	0.2676	0.316	A	20.47	6.06E-06	0.7911	0.7146	0.8757
12	rs11065987	112072424	G	0.4692	0.4186	A	19.53	9.93E-06	1.228	1.121	1.345
12	rs7134542	102339665	A	0.2572	0.2183	G	16.3	5.41E-05	1.241	1.117	1.378
12	rs11066320	112906415	A	0.4736	0.428	G	15.84	6.91E-05	1.203	1.098	1.317
12	rs11066301	112871372	G	0.4749	0.4292	A	15.83	6.95E-05	1.203	1.098	1.317
12	rs770460	78132040	A	0.4855	0.4399	G	15.66	7.60E-05	1.201	1.097	1.315
12	rs17696736	112486818	G	0.4772	0.4322	A	15.38	8.79E-05	1.199	1.095	1.313
14	rs1824343	98920773	C	0.294	0.2392	A	30.05	4.22E-08	1.325	1.198	1.465
14	rs1257641	99480395	A	0.08253	0.1163	G	21.76	3.09E-06	0.6834	0.5819	0.8026
14	rs1015277	84438074	C	0.2076	0.1731	A	15.17	9.83E-05	1.252	1.118	1.402
16	rs8047221	2897372	C	0.3041	0.3495	A	17.11	3.53E-05	0.8134	0.7374	0.8971
16	rs3810801	2892370	A	0.3007	0.3458	C	17	3.74E-05	0.8136	0.7375	0.8975
16	rs8060332	2892770	A	0.3409	0.3871	G	16.94	3.85E-05	0.819	0.7447	0.9008
17	rs4405612	66733514	A	0.07784	0.055	G	17.52	2.84E-05	1.45	1.217	1.727
17	rs10775405	45188090	A	0.3525	0.3098	G	15.62	7.75E-05	1.213	1.102	1.334
18	rs470549	74747488	A	0.1078	0.05378	G	91.4	1.17E-21	2.125	1.815	2.488
18	rs9952617	28881801	A	0.4947	0.4472	G	17.05	3.65E-05	1.21	1.105	1.325
19	rs4897966	389873	A	0.2669	0.2275	G	16.18	5.77E-05	1.236	1.115	1.371

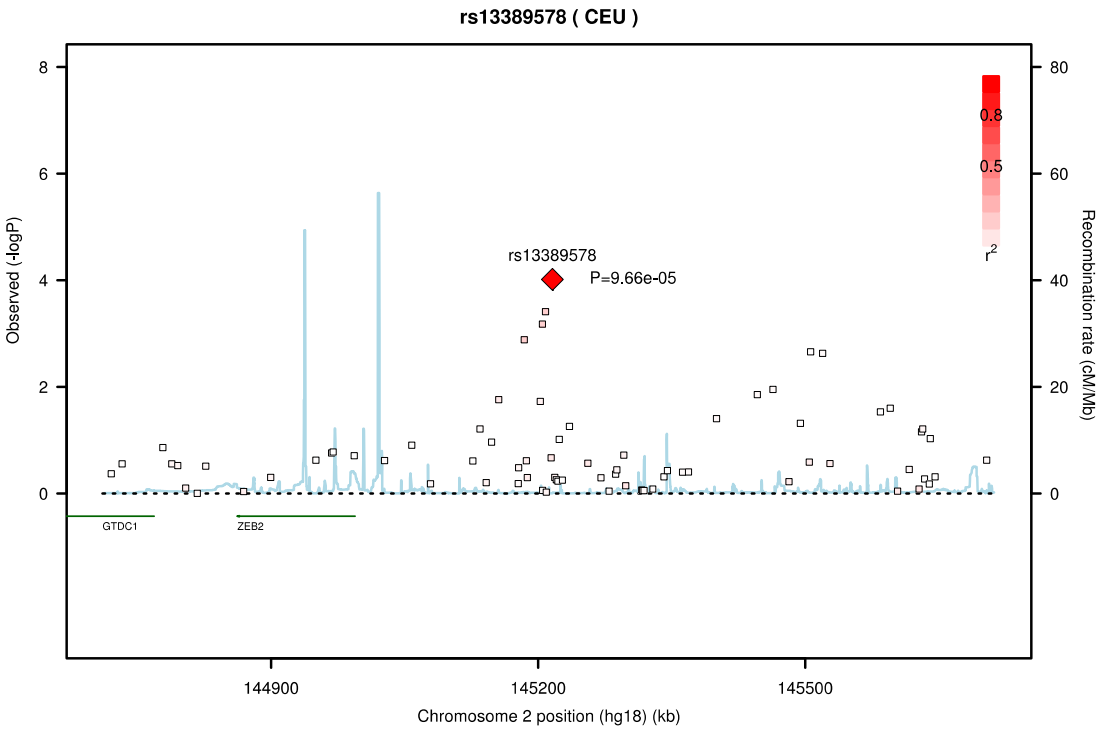
20	rs2064726	50715685	A	0.277	0.3258	G	20.55	5.82E-06	0.793	0.7172	0.8768
20	rs1885082	17593984	G	0.2557	0.2178	A	15.45	8.46E-05	1.234	1.111	1.371
21	rs2836881	40466299	A	0.2094	0.2699	C	35.51	2.54E-09	0.7166	0.642	0.7999
21	rs2836878	40465534	A	0.2109	0.2699	G	33.89	5.82E-09	0.7229	0.6479	0.8066
21	rs378108	40469520	G	0.5316	0.4789	A	20.76	5.22E-06	1.235	1.128	1.352
21	rs1893592	43855067	C	0.2537	0.2997	A	19.11	1.24E-05	0.7946	0.7167	0.881

A1 = minor allele; A2 = major allele; F_A = frequency of minor allele in cases; F_U = frequency of minor allele in controls; OR = odds ratio; L95 = lower limit of 95% confidence interval; U95 = upper limit of 95% confidence interval

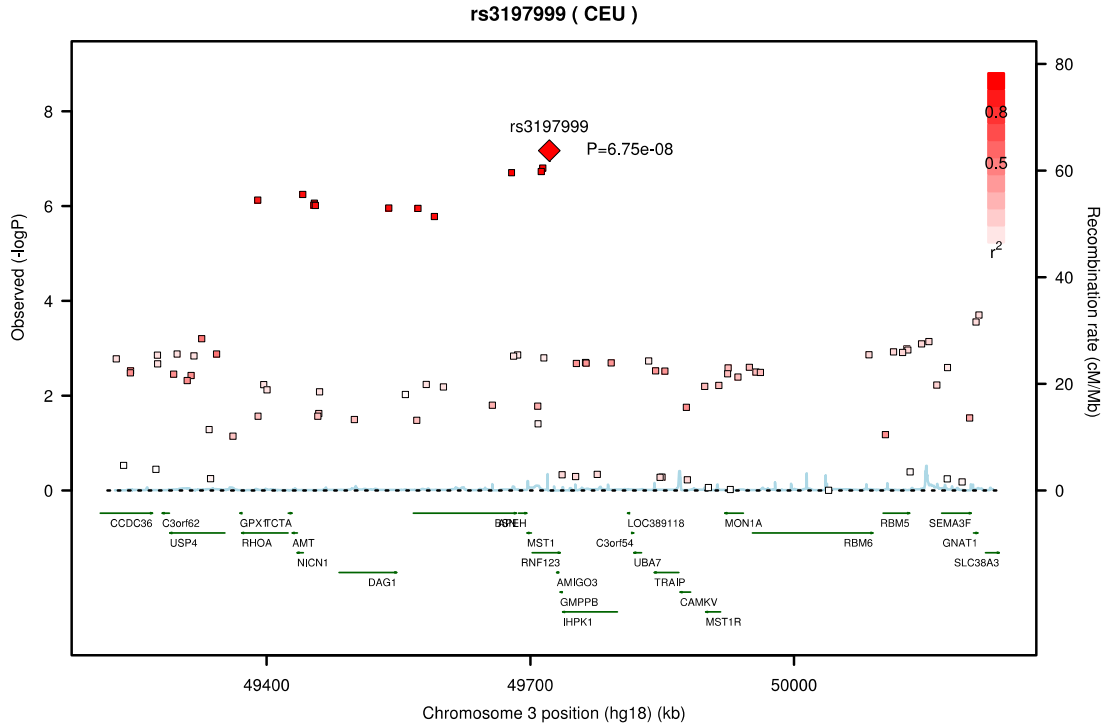
Supplementary Figures 7.3 (a – z). Regional association plots for all SNPs showing suggestive or genome wide significant association in the discovery panel. Plots were created using SNAP except for rs10774625 (Figure 7.3r) which was plotted using Locus zoom plot.



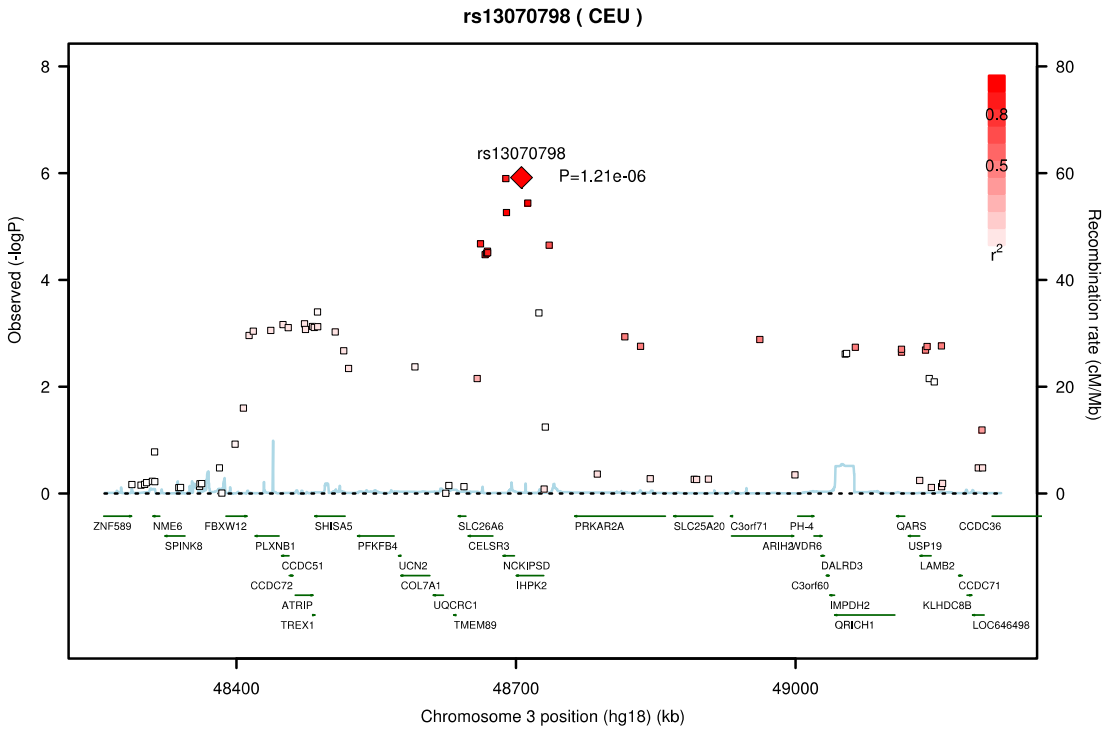
Supplementary Figure 7.3a.



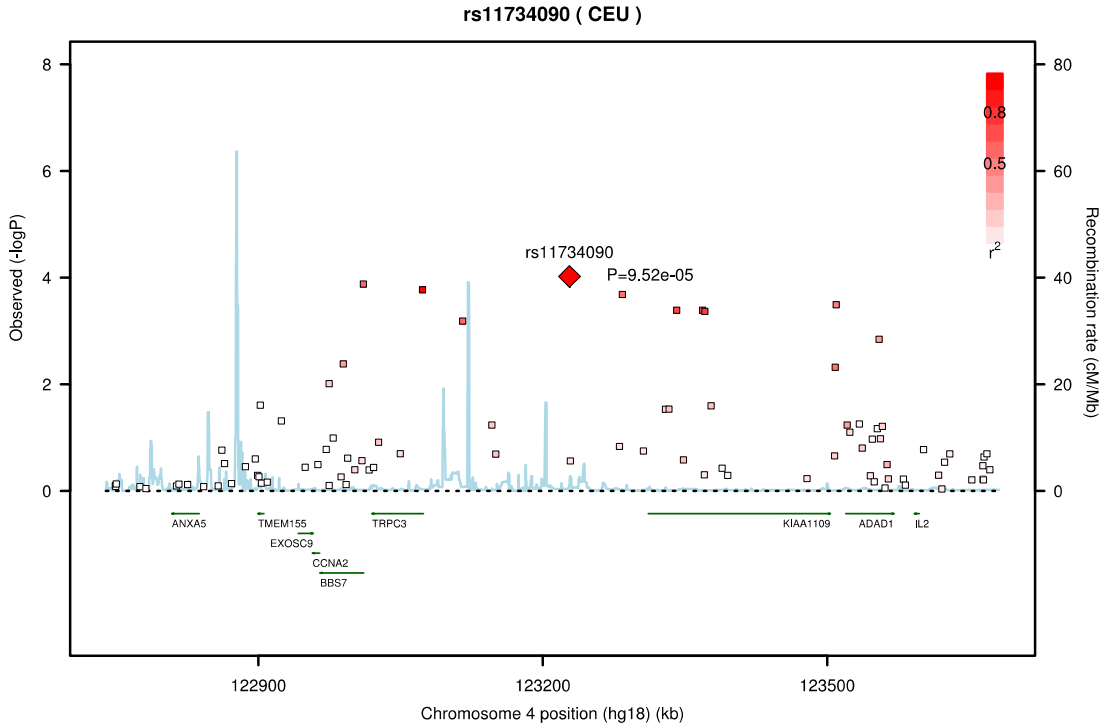
Supplementary Figure 7.3b.



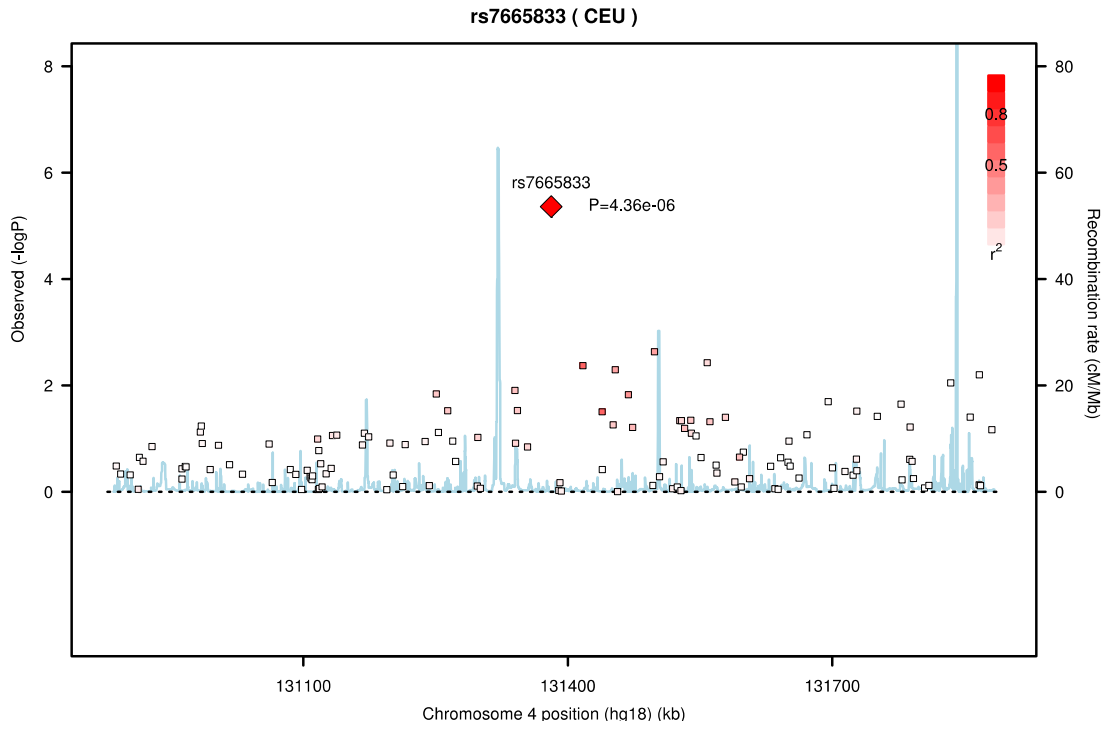
Supplementary Figure 7.3c.



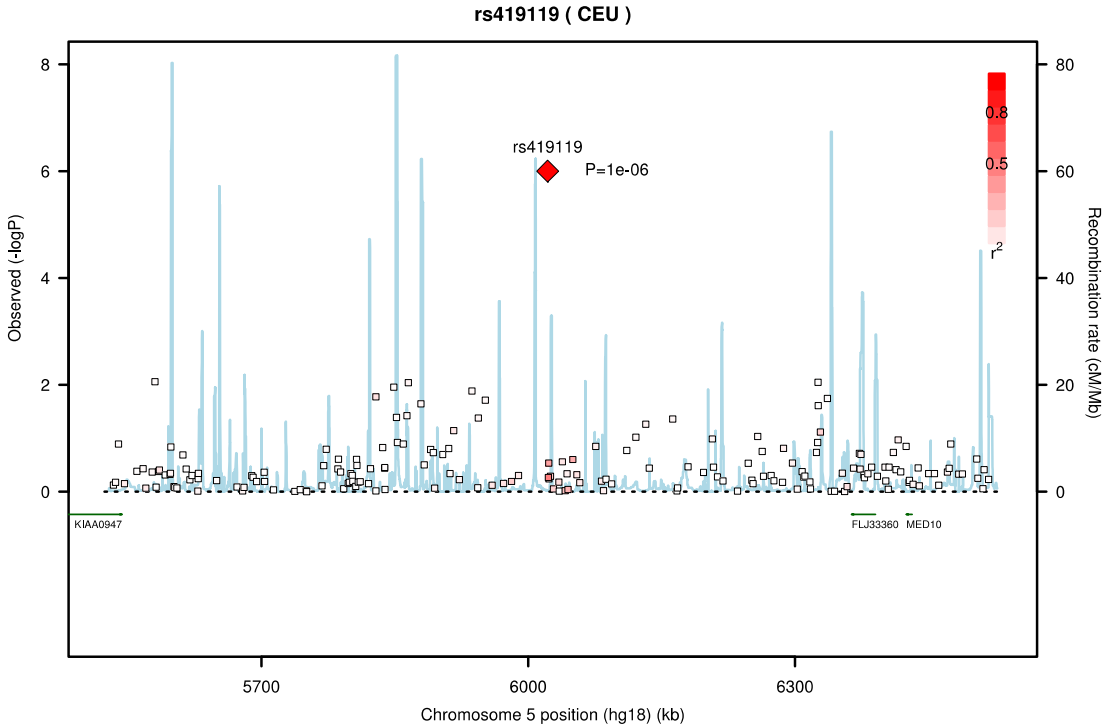
Supplementary Figure 7.3d.



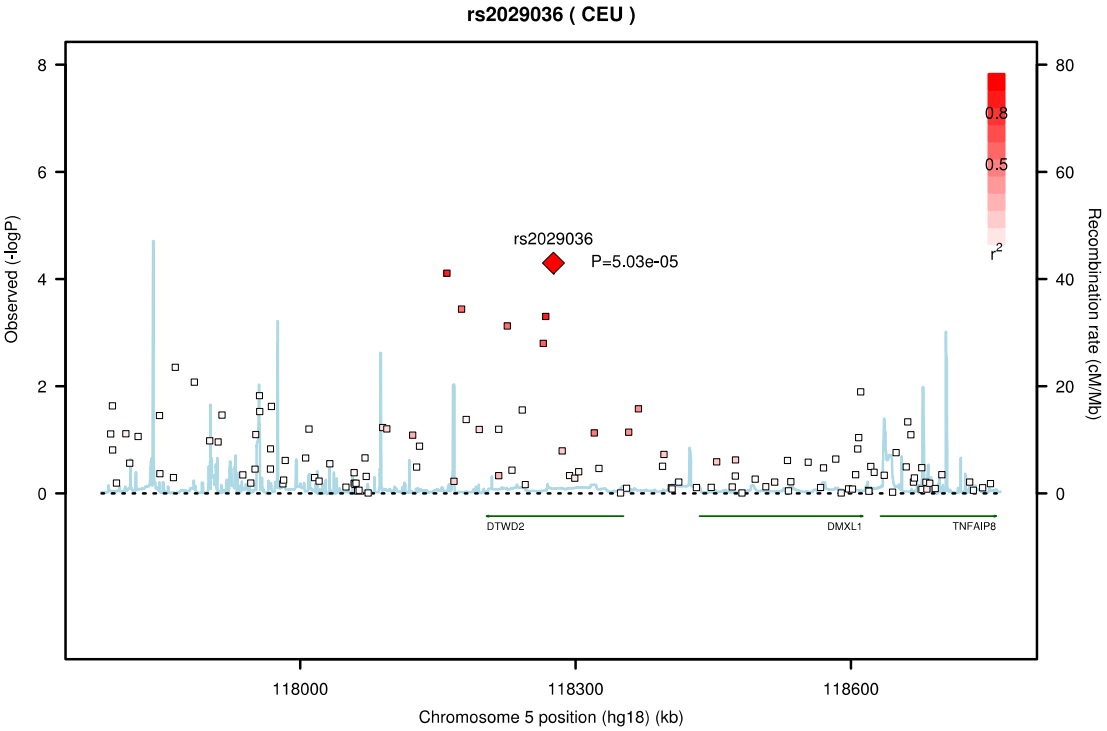
Supplementary Figure 7.3e.



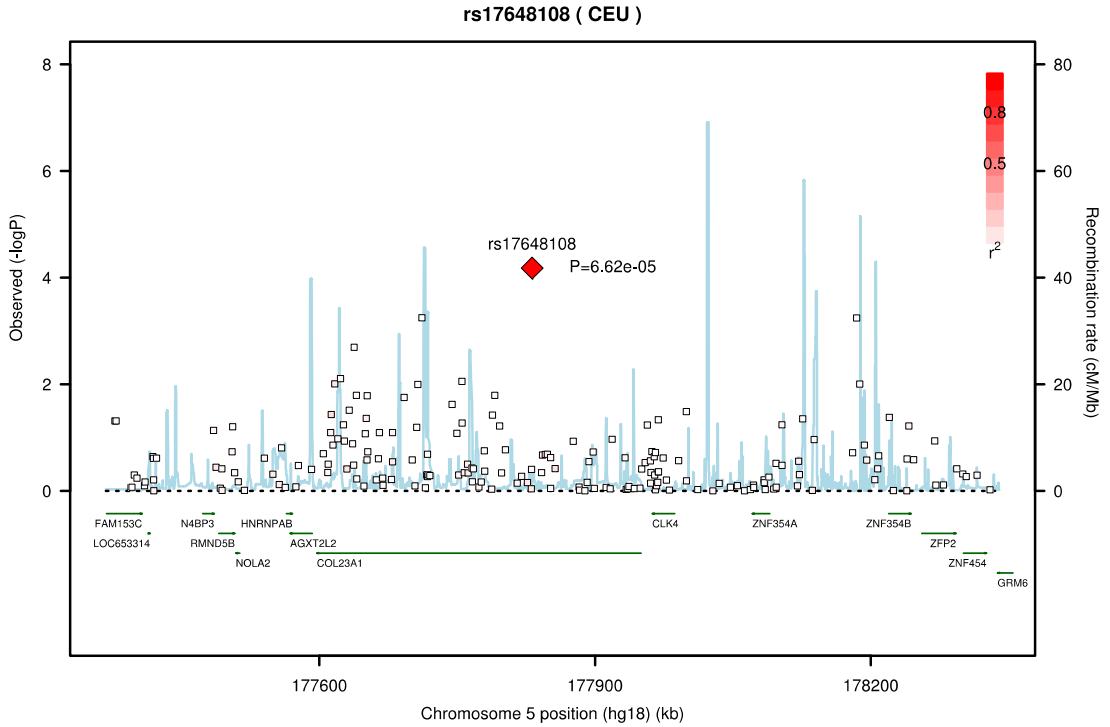
Supplementary Figure 7.3f.



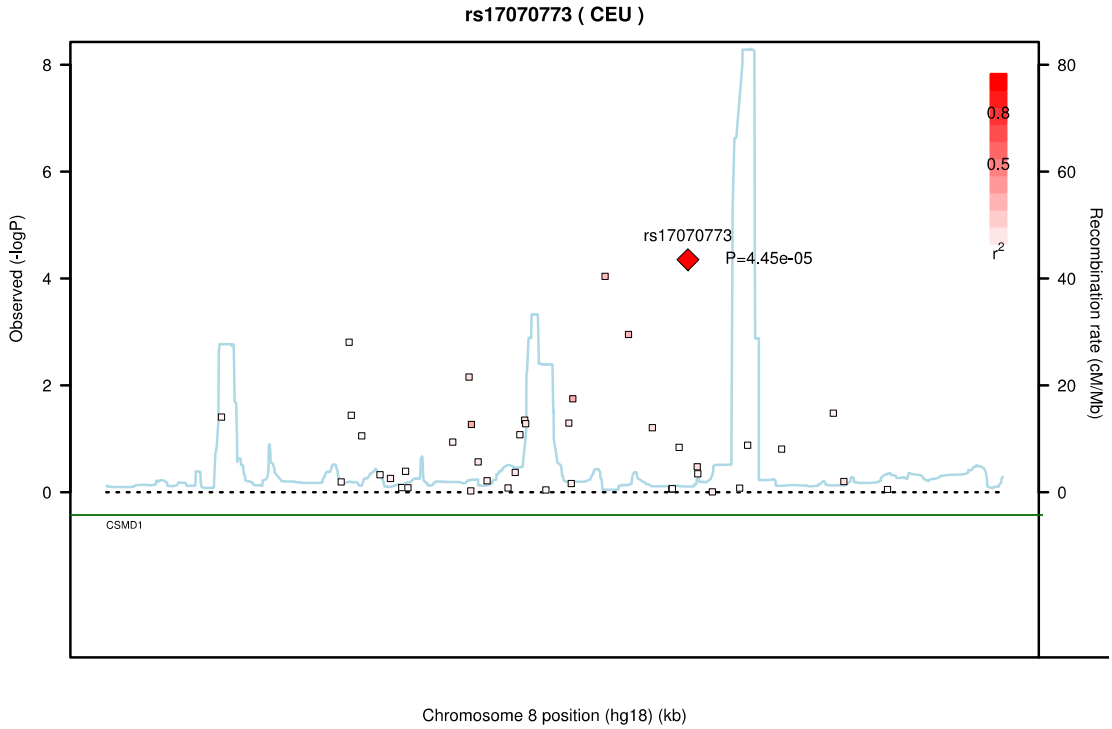
Supplementary Figure 7.3g.



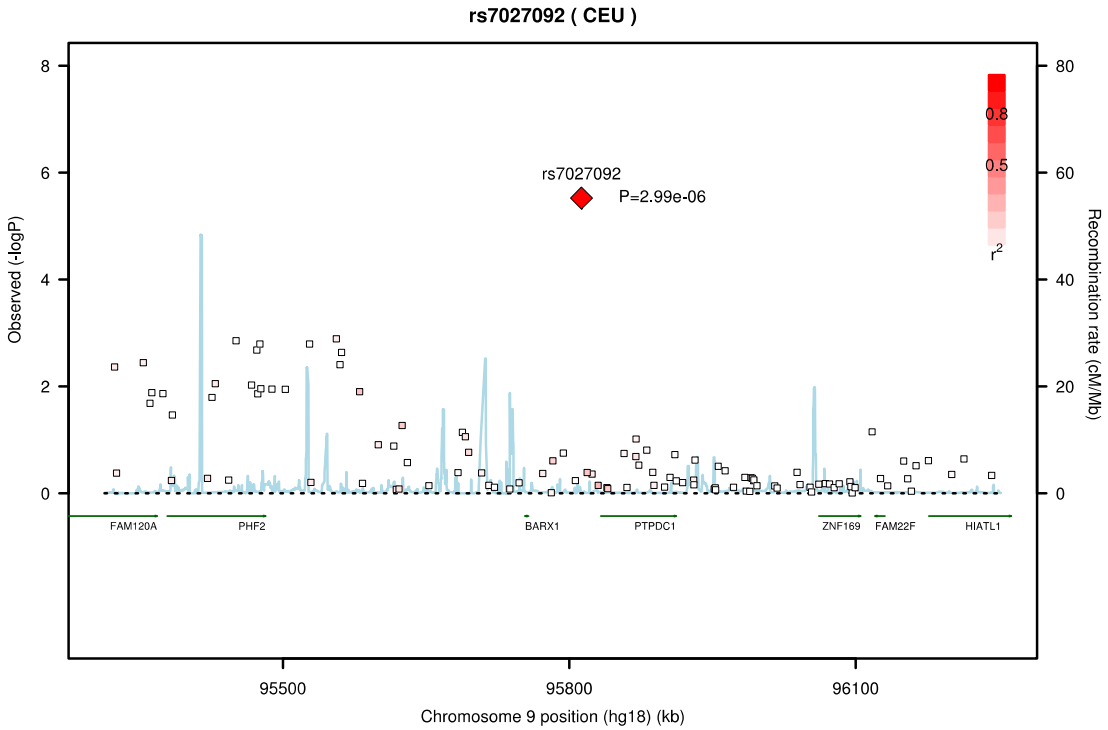
Supplementary Figure 7.3h.



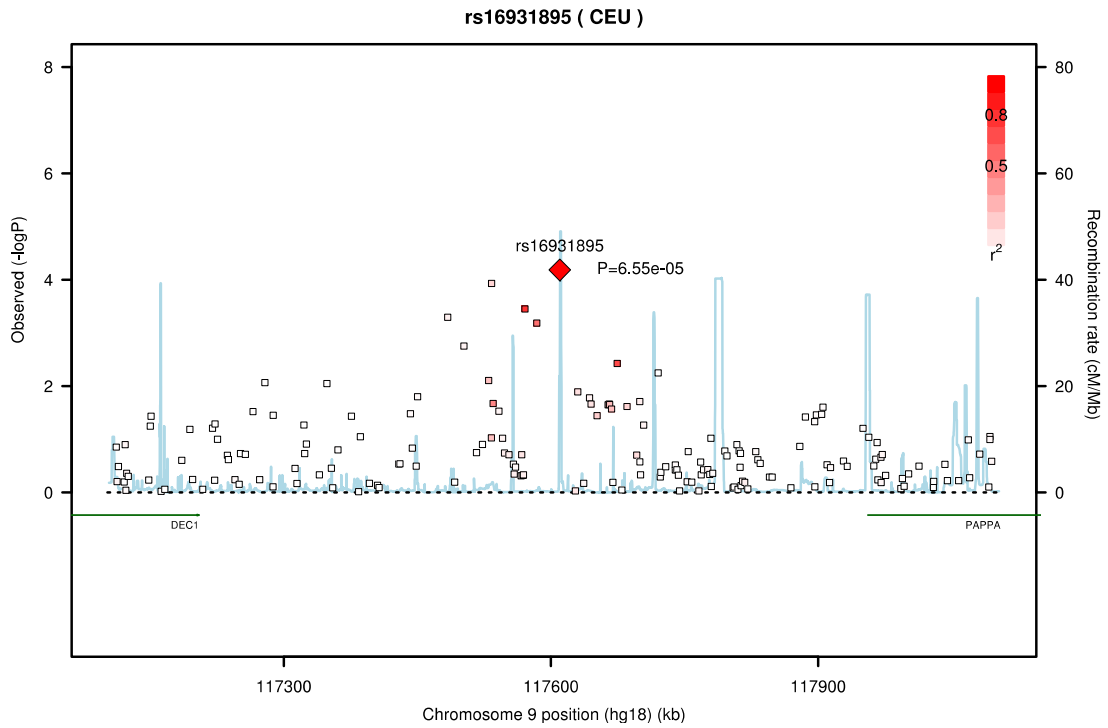
Supplementary Figure 7.3i.



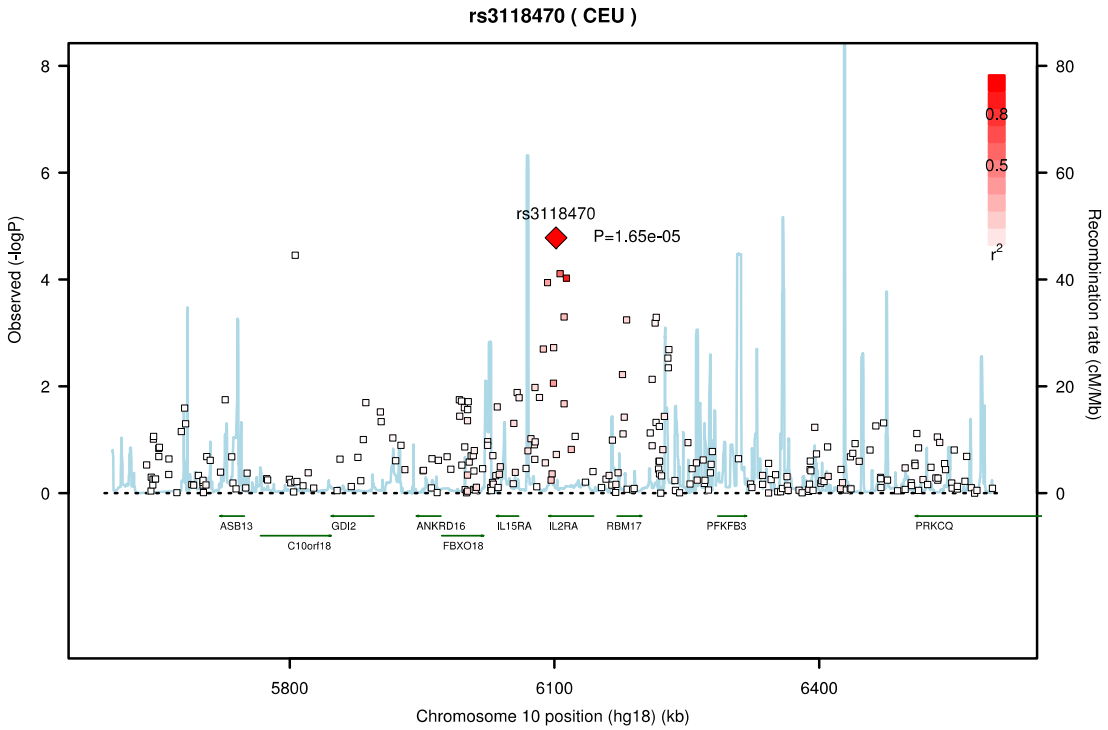
Supplementary Figure 7.3j.



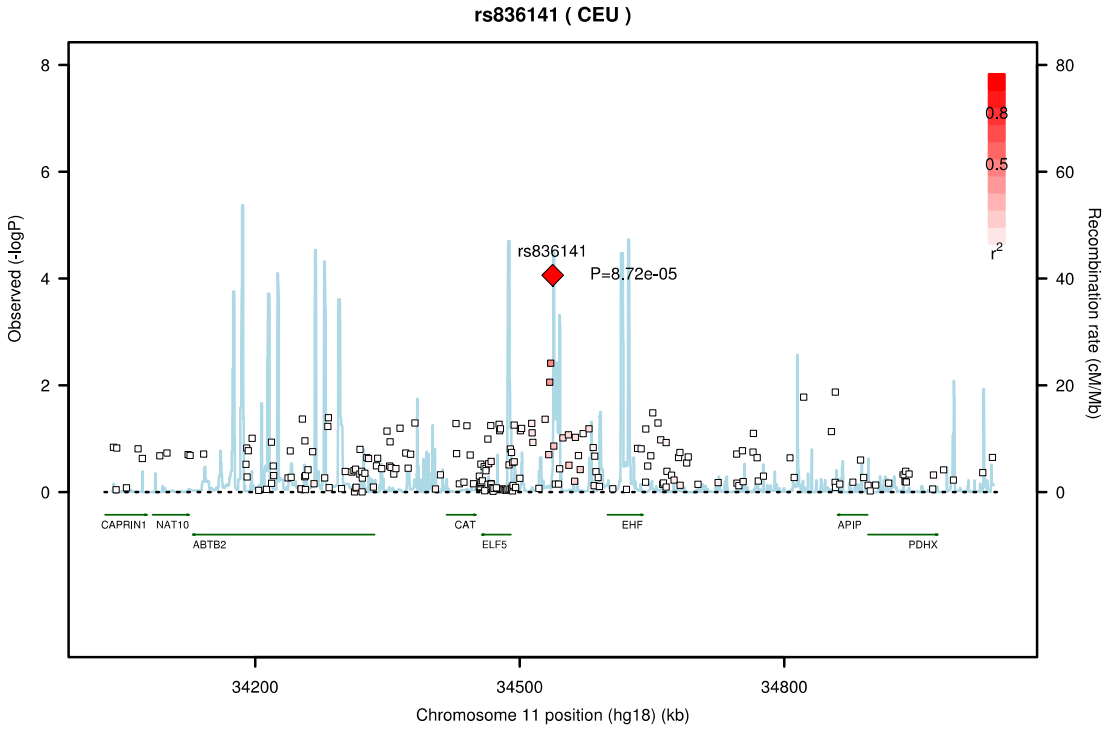
Supplementary Figure 7.3k.



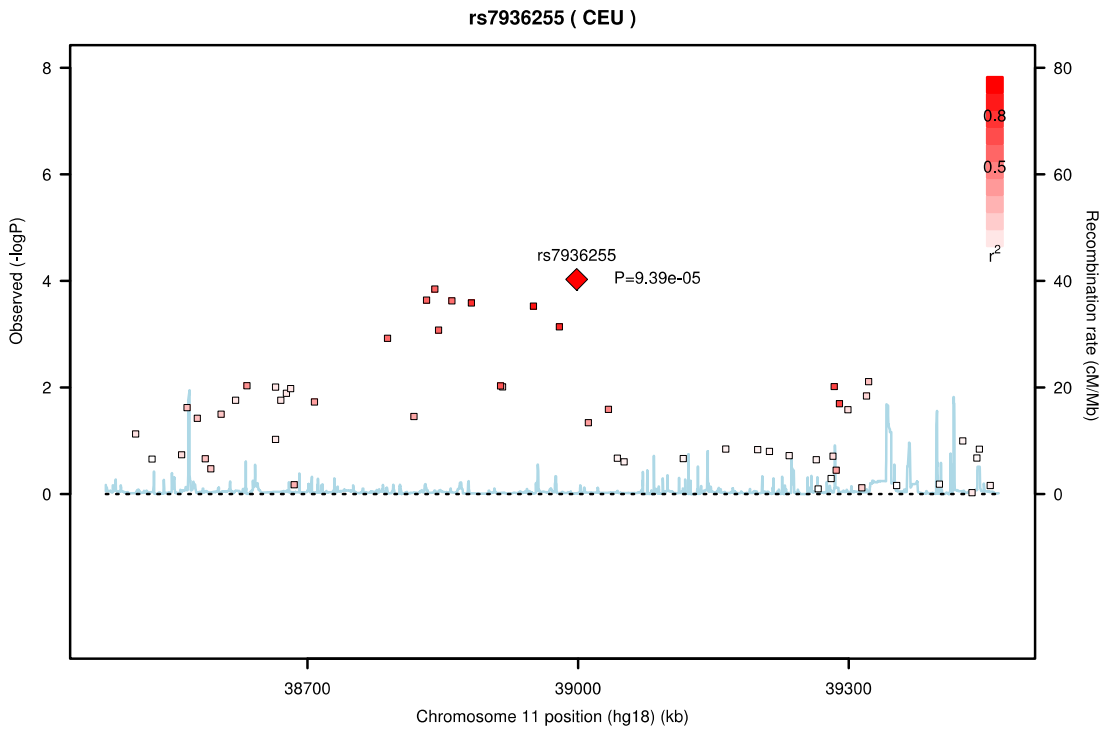
Supplementary Figure 7.3l.



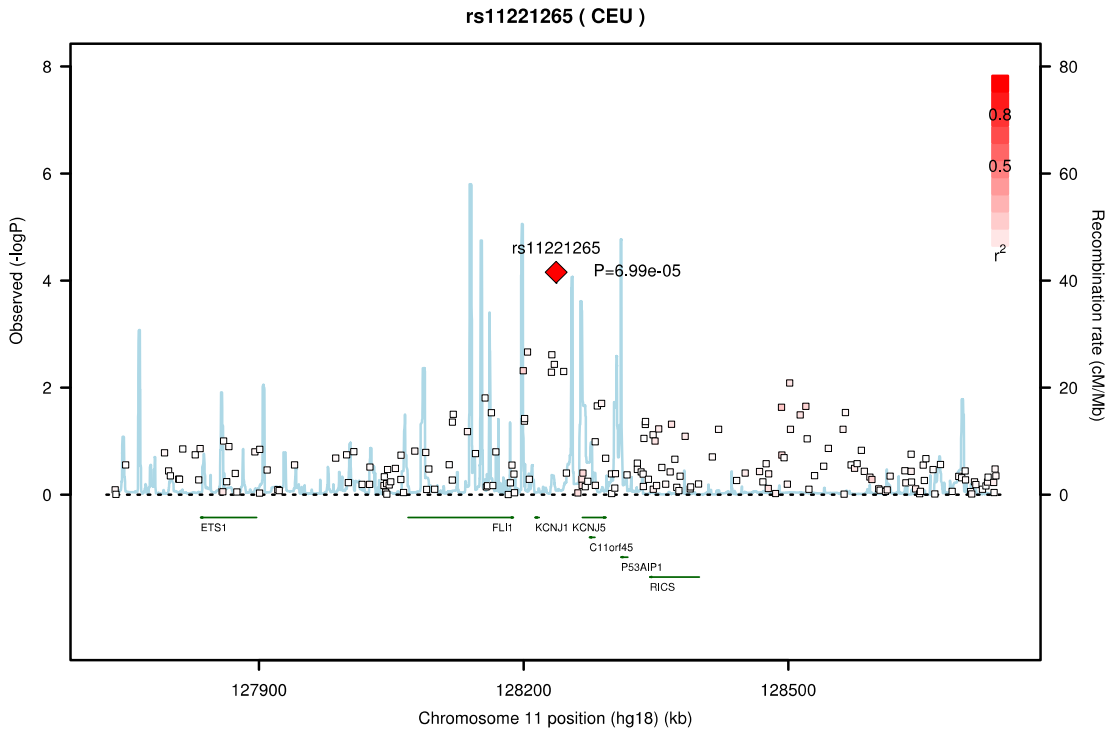
Supplementary Figure 7.3m.



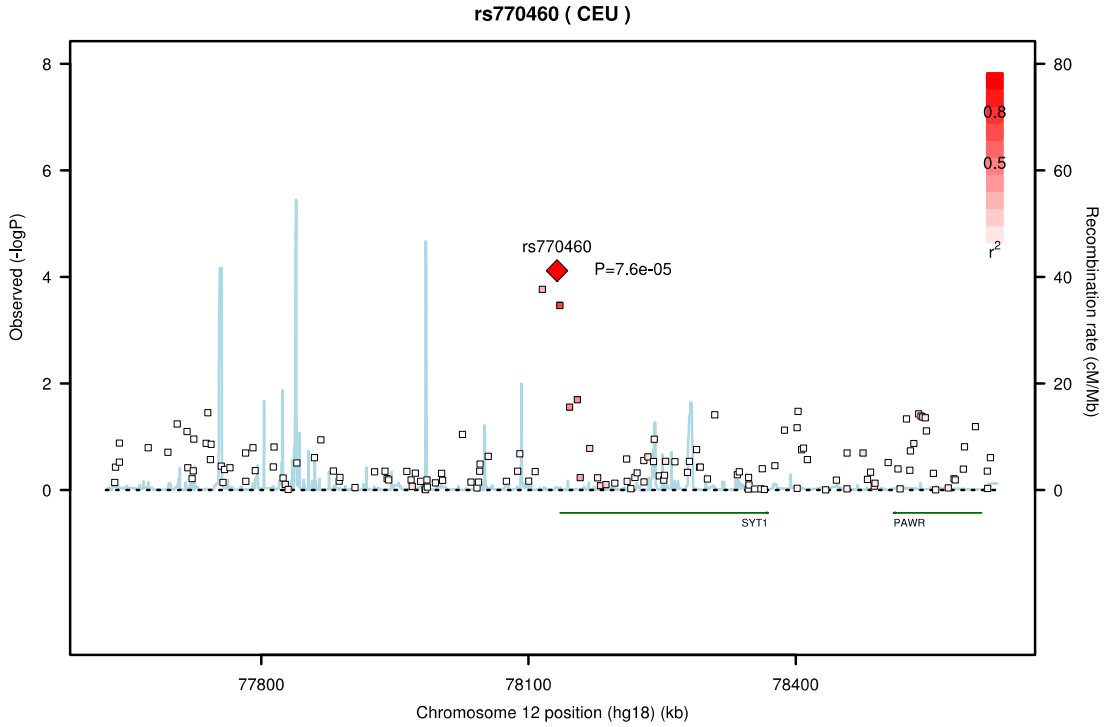
Supplementary Figure 7.3n.



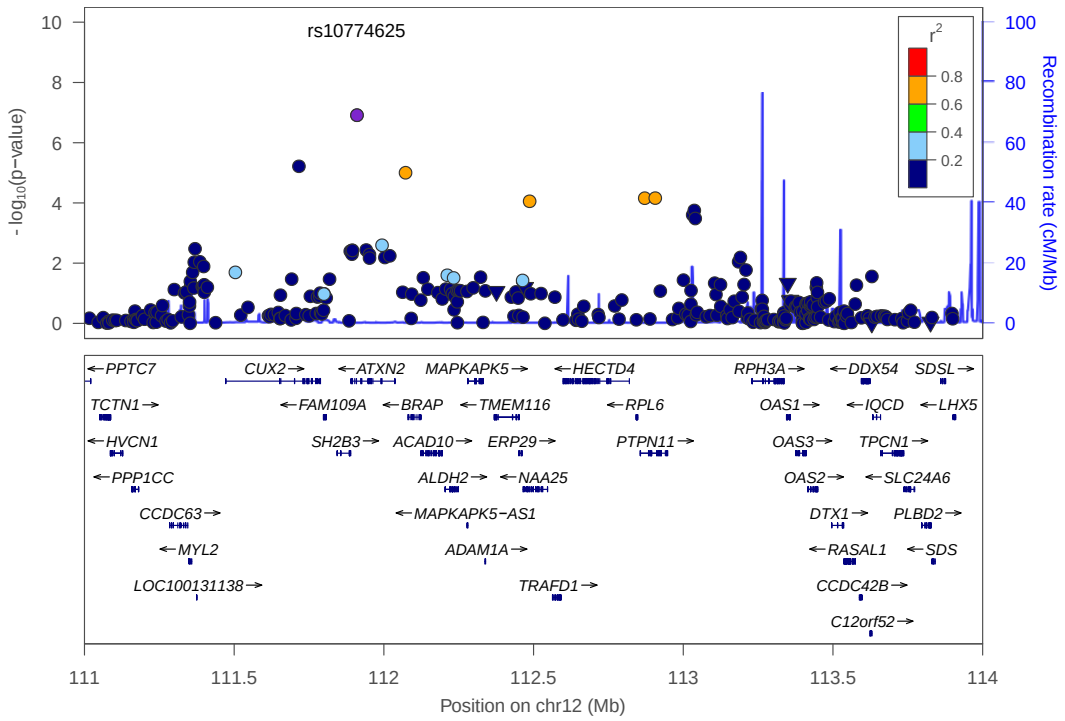
Supplementary Figure 7.3o.



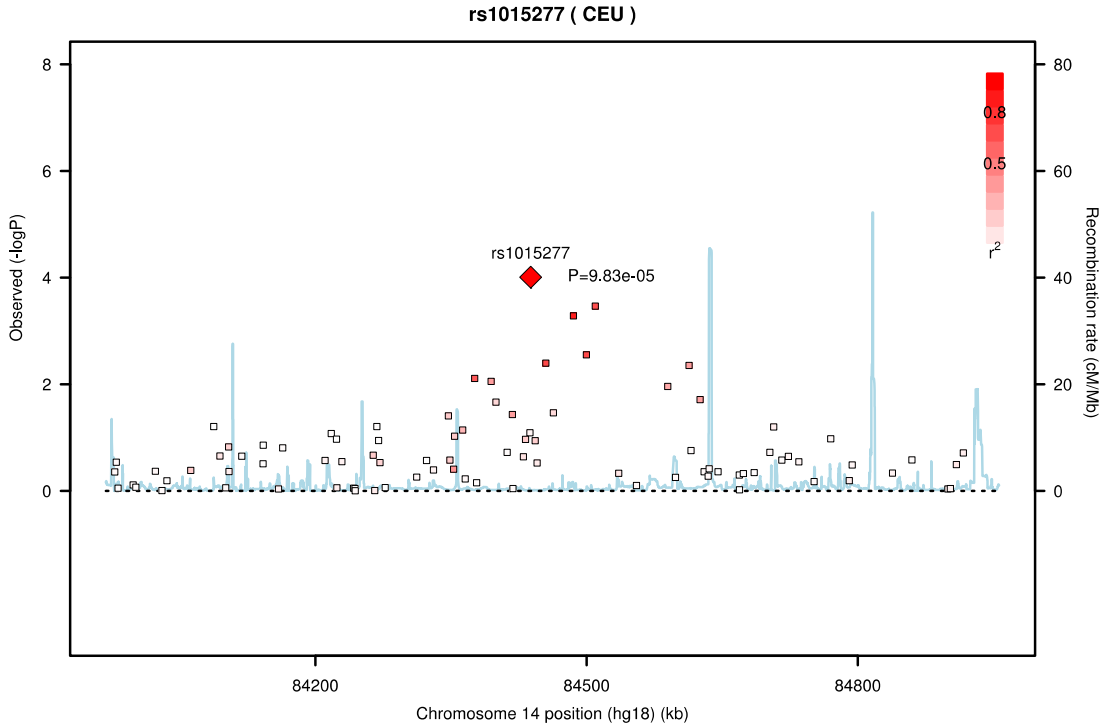
Supplementary Figure 7.3p.



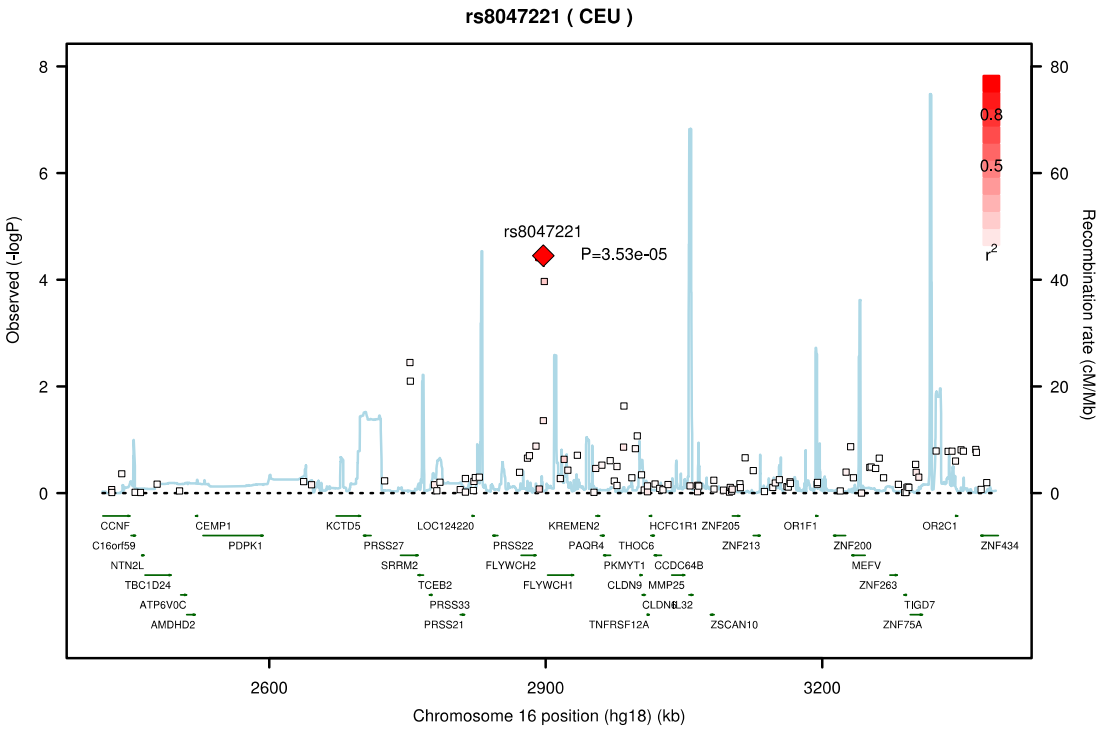
Supplementary Figure 7.3q.



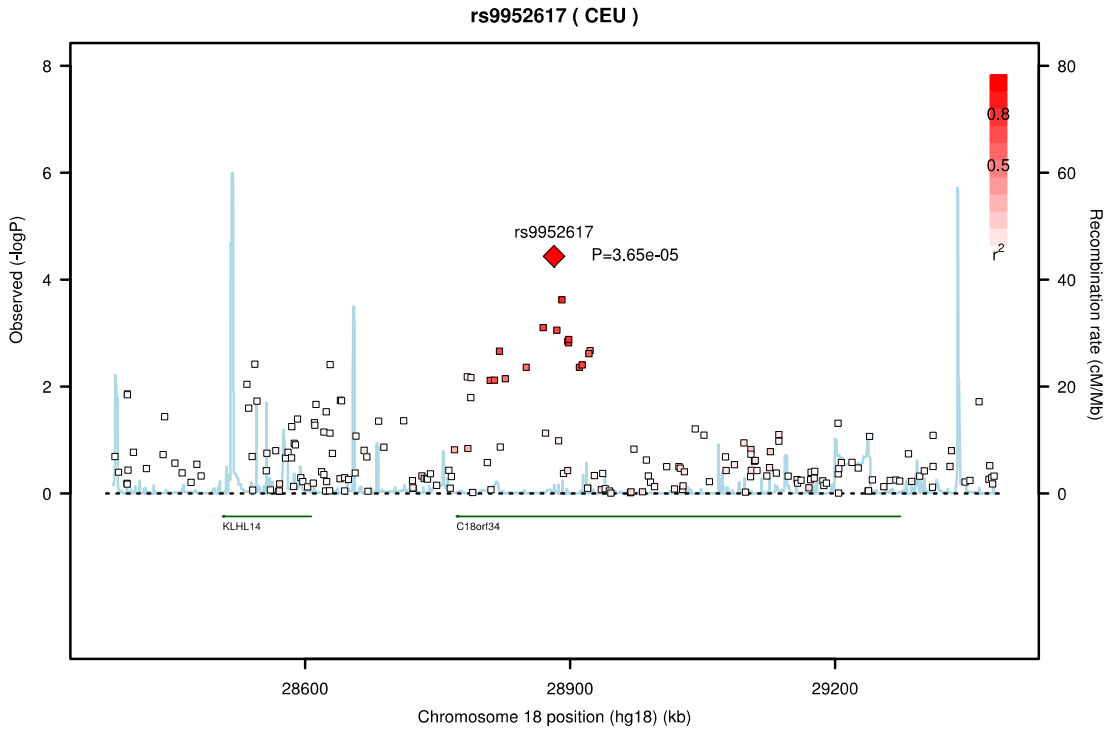
Supplementary Figure 7.3r.



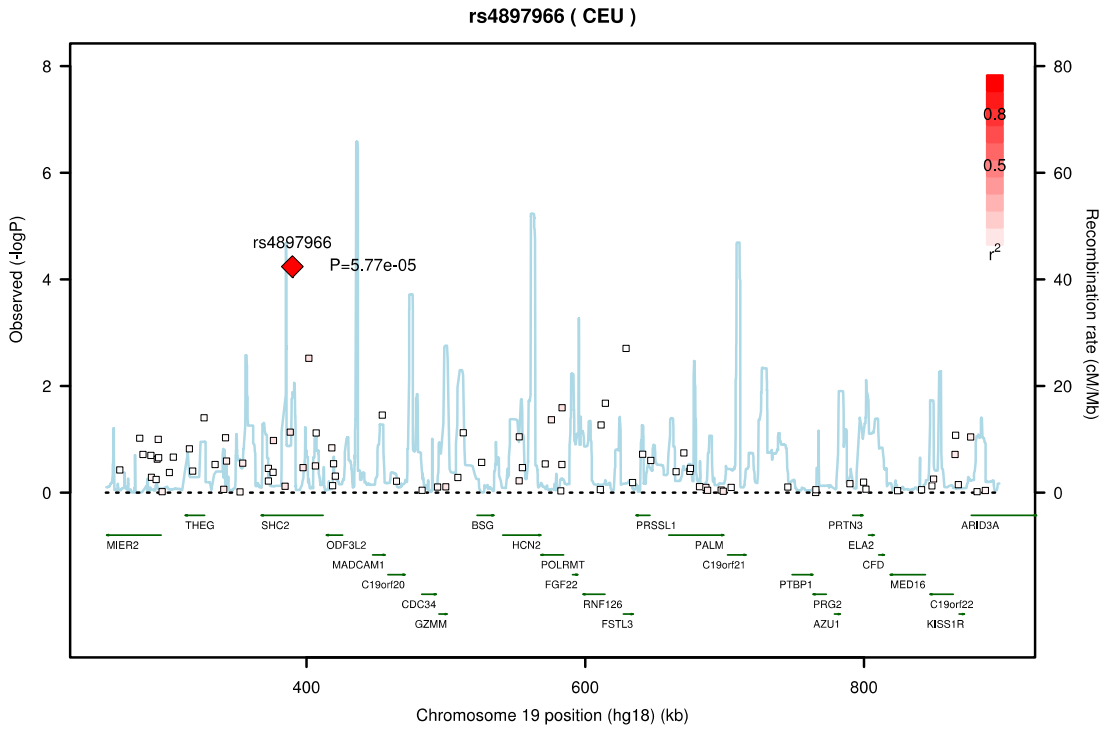
Supplementary Figure 7.3s.



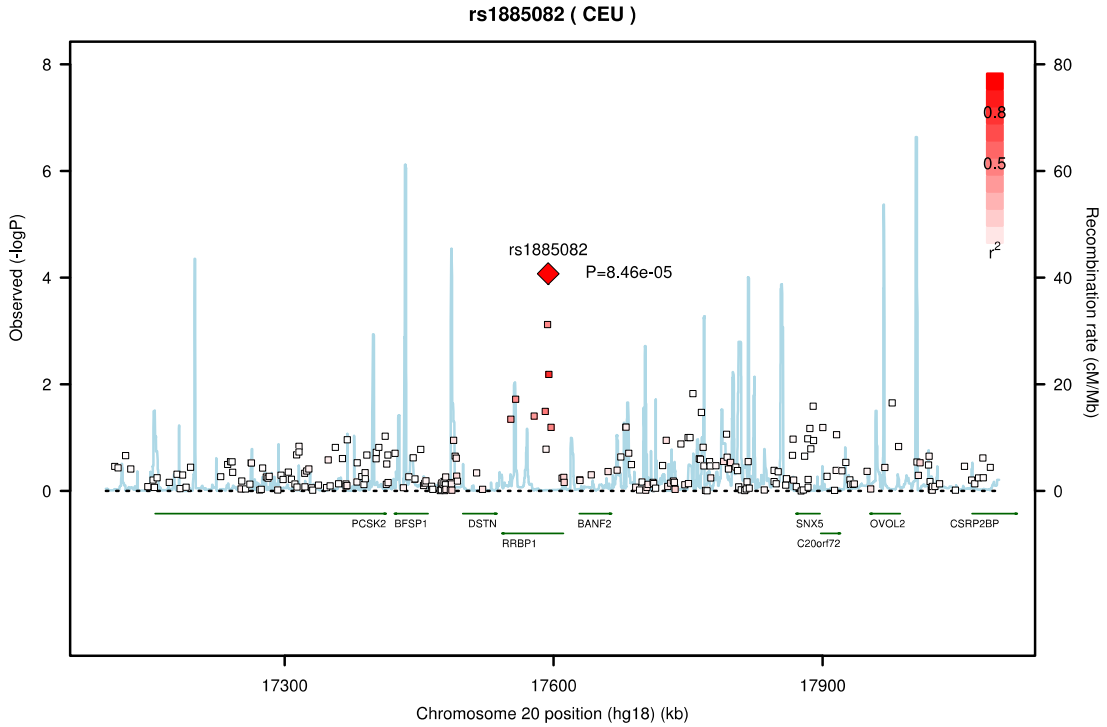
Supplementary Figure 7.3t.



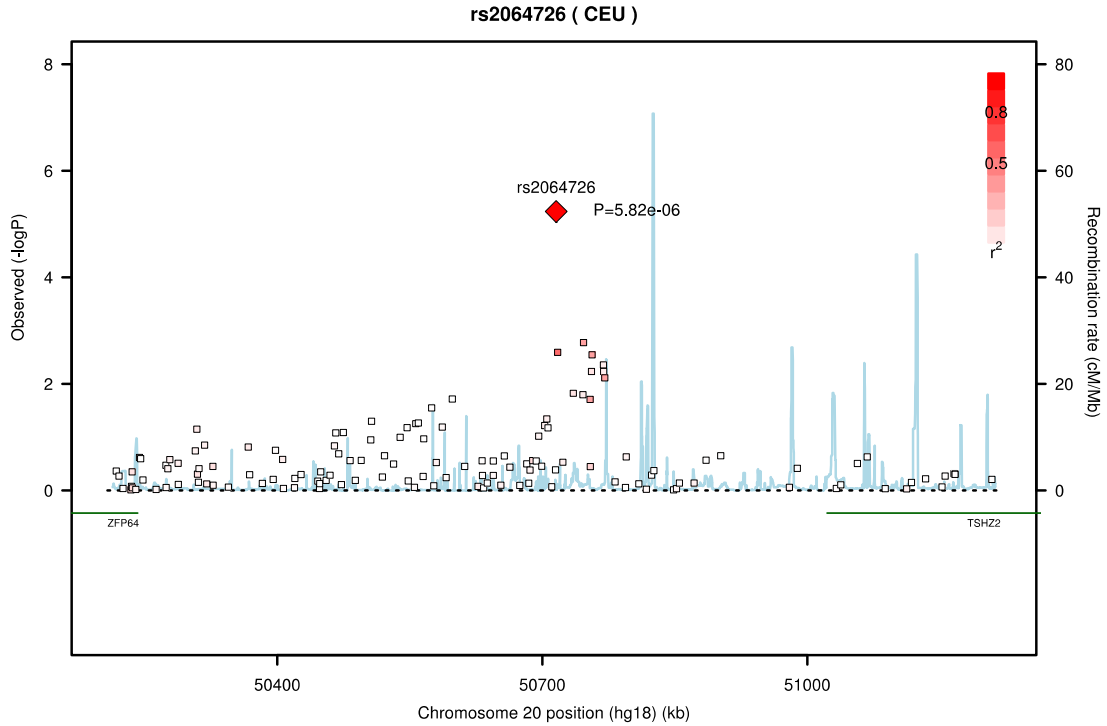
Supplementary Figure 7.3u.



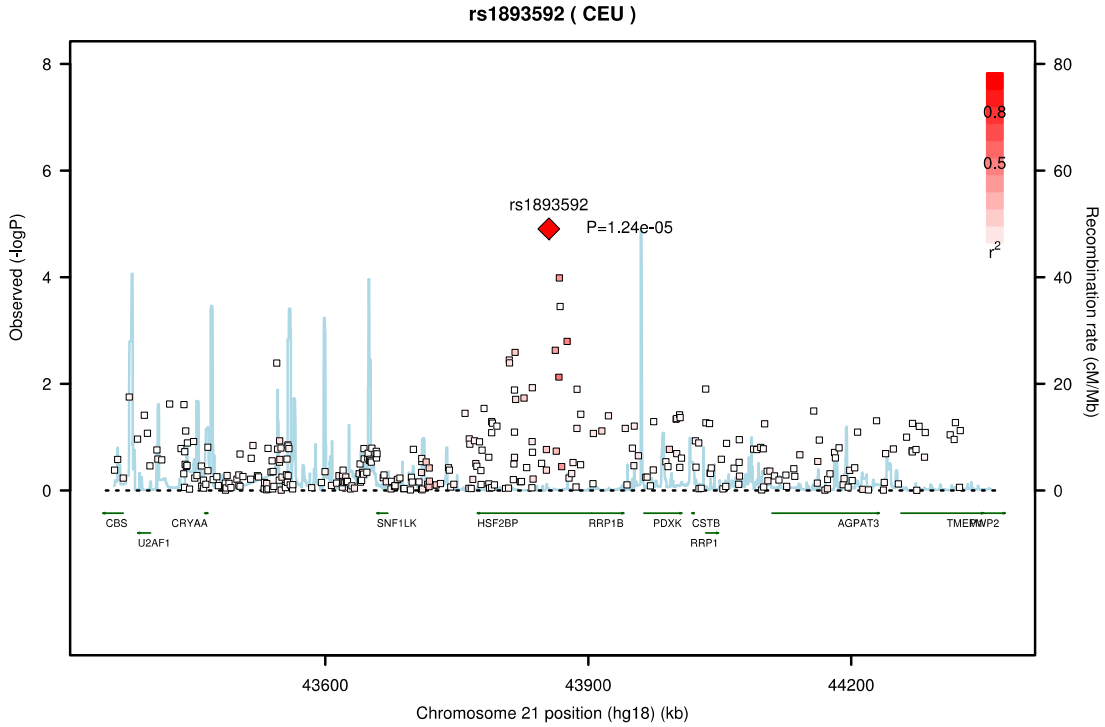
Supplementary Figure 7.3v.



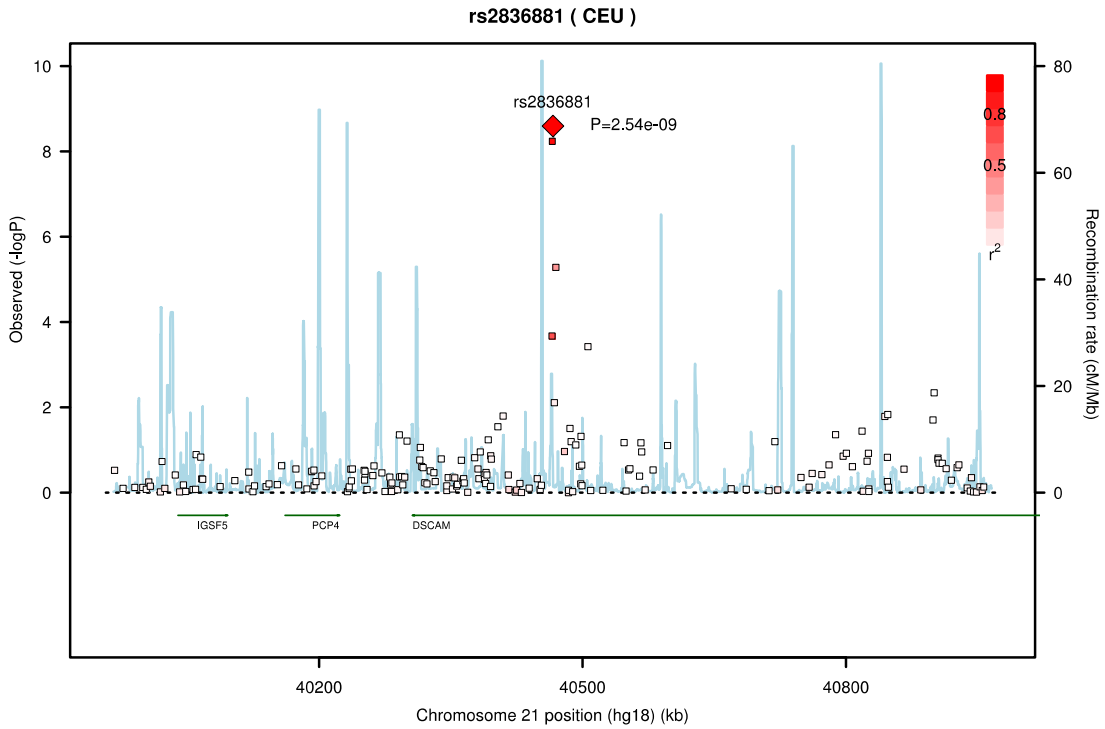
Supplementary Figure 7.3w.



Supplementary Figure 7.3x.



Supplementary Figure 7.3y.



Supplementary Figure 7.3z.

Supplementary Table 7.2. List of all SNPs considered for in-silico replication analysis (n = 96).

CHR	SNP	Position (GRCh37.5); build 19	CHR	SNP	Position (GRCh37.5); build 19
1	rs11121182	8522553	10	rs3118470	6101713
1	rs7553298	8770883	10	rs10905669	6092093
1	rs10492966	8598005	10	rs2387015	6215257
1	rs10492965	8598127	10	rs3814195	6213960
1	rs11121203	8648240	10	rs3763700	6181709
1	rs7520572	8871690	10	rs11255072	5806248
1	rs4908760	8526142	10	rs7072793	6106266
1	rs1953827	8679848	10	rs12412095	6113523
2	rs13389578	145216048	10	rs7090530	6110875
2	rs16823732	145204758	11	rs7928323	38841059
2	rs11883829	145519623	11	rs11034972	38860022
2	rs12105918	145208193	11	rs2068461	38831942
2	rs10192562	145184316	11	rs10837055	38845174
2	rs13032840	145506108	11	rs836141	34537316
3	rs13070798	48705934	11	rs7936255	38998451
3	rs9836462	48712791	11	rs2555524	38881682
3	rs13324142	48669447	11	rs2912573	38950513
3	rs12107418	48689787	11	rs7116065	38979245
3	rs13063312	48661985	11	rs11221265	128236936
3	rs2276852	48666923	12	rs1265566	111716376
3	rs2286652	48689192	12	rs17696736	112486818
3	rs11719291	48735706	12	rs770460	78132040
3	rs2276850	48669648	12	rs10774625	111910219
3	rs3821876	48668394	12	rs11065987	112072424
4	rs434193	86253489	12	rs11066320	112906415
4	rs4693142	87252259	12	rs11066301	112871372
4	rs7684187	123341159	14	rs1257641	99480395
4	rs7689808	87254477	14	rs1015277	84438074
4	rs345367	86754638	14	rs17119456	84485390
4	rs11734090	123228113	14	rs1824343	98920773
4	rs13132933	123010587	14	rs17119553	84509670
4	rs11938795	123073009	16	rs8047221	2897372
4	rs34135604	123283993	16	rs3810801	2892370
4	rs6822844	123509421	16	rs8060332	2892770
4	rs10027390	123368516	16	rs11076866	2898356
4	rs11575812	123371049	18	rs9952617	28881801
4	rs13151961	123115502	18	rs7229974	28890717
5	rs17648108	177831556	18	rs11663134	28869498
5	rs17132677	118175674	18	rs2114270	28885116
5	rs12109252	118267633	19	rs4897966	389873

5	rs6874399	118225616	20	rs2064726	50715685
5	rs419119	6021846	20	rs1885082	17593984
5	rs2029036	118275869	20	rs4814628	17593315
5	rs1567520	118159871	20	rs13734	17594729
7	rs10951953	5510159	21	rs378108	40469520
8	rs17070773	4463359	21	rs1893592	43855067
8	rs2617094	4456167	21	rs2836881	40466299
9	rs7027092	95812707	21	rs2836878	40465534

Supplementary table 7.3. Meta-analysis results for all SNPs (n = 86). SNPs reaching genome-wide significance (p-value < 5 x 10^{-8}) are highlighted in bold.

CHR	BP	SNP	A1	A2	P-value (combined); random effects	OR (combined) ; random effects	Q	I
1	8648240	rs11121203	C	T	0.0002113	1.212	0.2668	18.9
1	8679848	rs1953827	A	G	0.0004828	0.8734	0.648	0
1	8526142	rs4908760	A	G	0.0006915	0.8772	0.4838	0
1	8770883	rs7553298	A	G	0.001383	0.8378	0.2304	30.49
1	8598127	rs10492965	A	G	0.03827	1.1808	0.0963	63.85
1	8598005	rs10492966	C	T	0.04297	0.8447	0.0837	66.57
1	8522553	rs11121182	A	C	0.04574	1.1784	0.0885	65.53
1	8871690	rs7520572	A	G	0.07293	1.1304	0.1205	58.53
2	145204758	rs16823732	A	G	0.0006241	1.2897	0.3153	0.83
2	145208193	rs12105918	C	T	0.001232	1.297	0.284	12.89
2	145184316	rs10192562	A	G	0.001258	0.7837	0.3301	0
2	145216048	rs13389578	C	T	0.008451	1.2567	0.1881	42.27
2	145519623	rs11883829	A	G	0.5935	0.9468	0.014	83.42
2	145506108	rs13032840	A	G	0.6143	1.0544	0.0118	84.22
3	**48705934**	**rs13070798**	**C**	**T**	**1.07E-08**	**1.3648**	**0.6451**	**0**
3	**48712791**	**rs9836462**	**A**	**G**	**2.73E-08**	**0.7406**	**0.7701**	**0**
3	48689787	rs12107418	A	G	6.48E-08	0.7453	0.7364	0
3	48735706	rs11719291	A	G	2.10E-07	0.7599	0.917	0
3	48661985	rs13063312	A	G	1.30E-06	1.3165	0.5185	0
3	48669447	rs13324142	C	T	1.75E-06	0.7621	0.538	0
3	48666923	rs2276852	A	G	1.95E-06	0.7656	0.5831	0
4	123010587	rs13132933	C	T	3.04E-07	0.7666	0.5399	0
4	123509421	rs6822844	G	T	1.28E-05	1.3292	0.2397	27.65
4	123115502	rs13151961	A	G	9.61E-05	1.322	0.2001	39.09
4	123073009	rs11938795	C	T	0.0003023	0.8421	0.2929	9.61
4	123228113	rs11734090	C	T	0.003339	0.8451	0.2124	35.69
4	123341159	rs7684187	A	G	0.0116	1.1504	0.2073	37.1
4	123368516	rs10027390	C	T	0.01177	0.8691	0.2061	37.45
4	86754638	rs345367	A	G	0.0956	0.8552	0.0306	78.61
4	87252259	rs4693142	A	G	0.1232	0.9091	0.1789	44.65
4	87254477	rs7689808	C	T	0.1235	1.1001	0.1781	44.84
4	86253489	rs434193	C	T	0.1613	0.9154	0.1367	54.84
5	6021846	rs419119	A	C	0.03127	0.813	0.0596	71.82
5	118159871	rs1567520	C	T	0.5916	0.8871	0.0042	87.78
5	118175674	rs17132677	C	T	0.6138	0.9146	0.0074	86.08
5	177831556	rs17648108	C	T	0.652	1.0667	0.003	88.62
5	118225616	rs6874399	C	T	0.6595	0.9255	0.008	85.76
5	118275869	rs2029036	G	T	0.7511	1.081	0.0012	90.51
5	118267633	rs12109252	C	T	0.7683	0.9333	0.0025	89.04
8	4456167	rs2617094	A	G	0.4139	0.8733	0.0072	86.17
8	4463359	rs17070773	A	G	0.6219	0.8824	0.0006	91.56
9	95812707	rs7027092	A	G	0.6021	1.089	0.0017	89.89

10	**6101713**	**rs3118470**	**C**	**T**	**2.27E-09**	**1.2646**	**0.341**	**0**
10	6110875	rs7090530	A	C	8.63E-07	1.2125	0.3597	0
10	6092093	rs10905669	C	T	1.08E-06	0.8108	0.8171	0
10	6106266	rs7072793	C	T	2.07E-06	1.247	0.2338	29.47
10	6181709	rs3763700	C	T	9.45E-06	0.8398	0.7899	0
10	6113523	rs12412095	A	G	0.0003783	0.7761	0.089	65.42
10	6215257	rs2387015	C	T	0.01449	0.8473	0.2063	37.39
10	6213960	rs3814195	C	T	0.01532	0.847	0.2099	36.39
10	5806248	rs11255072	C	T	0.4958	0.8964	0.004	87.93
11	34537316	rs836141	A	G	0.01898	1.1483	0.1806	44.21
11	128236936	rs11221265	A	G	0.4373	1.1307	0.0149	83.12
11	38841059	rs7928323	A	C	0.7299	0.9532	0.0006	91.44
11	38831942	rs2068461	C	T	0.7399	1.0461	0.0009	91.01
11	38860022	rs11034972	C	T	0.7483	0.9571	0.0007	91.2
11	38845174	rs10837055	A	C	0.7734	1.0374	0.0017	89.89
11	38998451	rs7936255	A	G	0.8812	0.9743	0	93.99
11	38979245	rs7116065	C	T	0.9041	0.9818	0.0003	92.42
11	38950513	rs2912573	A	G	0.9222	1.0166	0.0001	93.71
12	**111910219**	**rs10774625**	**A**	**G**	**6.79E-10**	**1.2645**	**0.6726**	**0**
12	**111716376**	**rs1265566**	**C**	**T**	**4.46E-08**	**0.7932**	**0.929**	**0**
12	112072424	rs11065987	A	G	3.08E-07	0.8232	0.694	0
12	112486818	rs17696736	A	G	2.79E-06	0.8382	0.8478	0
12	112906415	rs11066320	A	G	3.20E-06	1.1935	0.767	0
12	112871372	rs11066301	A	G	3.24E-06	0.8379	0.7705	0
12	78132040	rs770460	C	T	0.4487	0.9222	0.0073	86.13
14	84485390	rs17119456	A	G	0.05121	0.8712	0.1494	51.88
14	99480395	rs1257641	A	G	0.07703	0.7802	0.0361	77.23
14	84438074	rs1015277	A	C	0.08445	0.862	0.0883	65.57
14	84509670	rs17119553	A	G	0.1138	0.8796	0.0977	63.54
14	98920773	rs1824343	G	T	0.5856	1.1063	0.0001	93.49
16	2892770	rs8060332	A	G	0.02782	0.8644	0.1401	54.07
16	2892370	rs3810801	A	C	0.09511	0.873	0.081	67.15
16	2897372	rs8047221	A	C	0.1404	1.1362	0.0514	73.65
18	28890717	rs7229974	A	G	0.6075	0.9426	0.0037	88.14
18	28881801	rs9952617	C	T	0.6761	0.9442	0.0006	91.57
18	28885116	rs2114270	G	T	0.6823	0.9547	0.0049	87.38
20	17593984	rs1885082	A	G	3.62E-05	0.833	0.3604	0
20	17593315	rs4814628	C	T	0.0001539	0.8234	0.7811	0
20	17594729	rs13734	A	G	0.000516	1.1792	0.8332	0
20	50715685	rs2064726	C	T	0.608	1.0831	0.0003	92.29
21	43855067	rs1893592	A	C	1.19E-05	1.217	0.298	7.66
21	40469520	rs378108	A	G	0.0003878	0.8418	0.2169	34.41
21	40465534	rs2836878	A	G	0.0314	0.7973	0.0301	78.75
21	40466299	rs2836881	G	T	0.03502	1.2594	0.0246	80.22

Appendix 1

NHS

National Research Ethics Service

Cambridgeshire 4 Research Ethics Committee
c/o Research Ethics Committees Office
The Norfolk & Norwich University Hospital NHS Trust
Room 2.08, 1st Floor
Aldwych House
57 Bethel Street
NORWICH
NR2 1NR

Telephone: 01603 289813
Facsimile: 01603 286573

27 August 2008

Dr Simon Matthew Rushbrook
Senior Clinical Fellow in Hepatology
Addenbrookes Hospital
Hills Road
Cambridge
Cambridgeshire
CB2 0QQ

Dear Dr Rushbrook

Full title of study: **A UK Collaborative Study to Determine the Genetic Basis of Primary Sclerosing Cholangitis (UK-PSC)**

REC reference number: **08/H0305/45**

Thank you for your undated letter (postmark 08 August 2008, received in the office on 11 August 2008), responding to the Committee's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chair Dr Leslie Gelling.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised, subject to the conditions specified below.

Ethical review of research sites

The Committee has designated this study as exempt from site-specific assessment (SSA). The favourable opinion for the study applies to all sites involved in the research. There is no requirement for other Local Research Ethics Committees to be informed or SSA to be carried out at each site.

Conditions of the favourable opinion

The favourable opinion is subject to the following conditions being met prior to the start of the study.

Management permission or approval must be obtained from each host organisation prior to the start of the study at the site concerned.

Management permission at NHS sites ("R&D approval") should be obtained from the relevant care organisation(s) in accordance with NHS research governance arrangements. Guidance on applying for NHS permission is available in the Integrated Research Application System or at http://www.rdforum.nhs.uk.

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	*Version*	*Date*
Application	5.6, NHS REC	
Investigator CV	Chief Investigator Dr S M Rushbrook	
Protocol	1	01 June 2008
Covering Letter	CI, Dr Simon Rushbrook	05 June 2008
Peer Review	Peer Review Form V5 (February 2008) Cambridge University Hospitals NHS Foundation Trust	04 June 2008
Questionnaire: Clinician's (Post Liver Transplant)	1	01 June 2008
Questionnaire: Clinician's (Pre-Transplant)	1	01 June 2008
Questionnaire: Participant	2	05 August 2008
Advertisement	1	05 August 2008
Letter of invitation to participant	1	01 June 2008
GP/Consultant Information Sheets	1	05 August 2008
Participant Information Sheet	2	05 August 2008
Participant Consent Form	2	05 August 2008
Response to Request for Further Information	Letter	
E-mail to confirm response to favourable opinion	Chief Investigator	18 August 2008
Confirmation of arrangements for stored samples	Letter from Rebecca Treacy, Acting Head of Molecular Genetics, Addenbrooke's Hospital	24 July 2008
Funding Information	Letter from Roger Chapman, Consultant Gastroenterologist, Oxford Radcliffe Hospitals	28 July 2008
confirmation of checklist	e-mail, Simon Rushbrook	12 June 2008
Applicant's Checklist (completed by hand)	5.6, NHS REC lock code AB/113523/1	
Letter explaining indemnity	CI, Dr Simon Rushbrook	05 June 2008
Recruitment Pack Letter	1	01 June 2008

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Now that you have completed the application process please visit the National Research Ethics Website > After Review

You are invited to give your view of the service that you have received from the National Research Ethics Service and the application procedure. If you wish to make your views known please use the feedback form available on the website.

The attached document "After ethical review – guidance for researchers" gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Progress and safety reports
- Notifying the end of the study

The NRES website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

We would also like to inform you that we consult regularly with stakeholders to improve our service. If you would like to join our Reference Group please email referencegroup@nres.npsa.nhs.uk.

08/H0305/45	**Please quote this number on all correspondence**

With the Committee's best wishes for the success of this project

Yours sincerely

Dr Leslie Gelling PhD RGN FRSA MICR
Chair

Email: janette.guymer@nnuh.nhs.uk

Enclosures: "After ethical review – guidance for researchers" *[SL- AR2 for other studies]*

Copy to: *Dr Claudia Rizzini, Cambridge University Hospital NHS Foundation Trust*

Appendix 2

The UK PSC Consortium

Chief Investigator:

Prof Gideon Hirschfield, University Hospitals Birmingham NHS Trust

Steering committee group members:

Dr Graeme Alexander Cambridge University Hospitals NHS Trust

Dr Richard Sandford, Cambridge University hospital NHS Trust

Dr George Mells, Cambridge University hospital NHS Trust

Dr Roger Chapman, John Radcliffe Hospital NHS Trust

Dr Mark Hudson, Newcastle upon Tyne Hospitals NHS Trust

Prof David Jones, Newcastle upon Tyne Hospitals NHS Trust

Dr Douglas Thorburn, Royal Free London NHS Trust

Professor Massimo Pinzani, Royal Free London NHS Trust

Dr Simon Rushbrook, Norfolk and Norwich University Hospital NHS Trust

Martine Walsmsley, PSC Support group, Trustee and Chair

Principal investigators at each recruitment site:

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 (**Dr John 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 Graeme Alexander**), 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, Dr Simon Rushbrook**), 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**), Oxford Radcliffe Hospitals NHS Trust (**Dr Roger Chapman**), 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**).

Appendix 3

Participant Questionnaire (Version TWO)

Research study into the genetic causes of Primary Sclerosing Cholangitis

[Patient name]
[Patient address]

1) What is your date of birth? _____

2) Sometimes the genes involved in causing a disease like PSC are different in different ethnic groups. Knowing your ethnic category will help us to analyse the results of this study. Please indicate your ethnic category by ticking the most appropriate box.

NATIONAL CODE		YES
WHITE		
A	BRITISH	
B	IRISH	
C	ANY OTHER WHITE BACKGROUND	
MIXED		
D	WHITE AND BLACK CARIBBEAN	
E	WHITE AND BLACK AFRICAN	
F	WHITE AND ASIAN	
G	ANY OTHER MIXED BACKGROUND	
ASIAN OR ASIAN BRITISH		
H	INDIAN	
J	PAKISTANI	
K	BANGLADESHI	
L	ANY OTHER ASIAN BACKGROUND	
BLACK OR BLACK BRITISH		
M	CARIBBEAN	
N	AFRICAN	
P	ANY OTHER BLACK BACKGROUND	
OTHER ETHNIC GROUPS		
R	CHINESE	

S	ANY OTHER ETHNIC GROUP	
Z	NOT STATED	

3) What is your sex (Please circle) Male
Female

4) Do you have Inflammatory Bowel Disease? Yes ☐ No ☐ **Not Sure** ☐

5) If yes do you have: Ulcerative Colitis Crohn's **Not Sure**
(Please circle your response)

6) Do you or have you ever smoked? ☐ Yes ☐ No

7) If you have smoked then please list the dates between which you smoked tobacco

8) Have you lived in accommodation that is occupied with pets? Yes ☐ No ☐

If yes please could you indicate which pets and at what ages of your life?

Please answer question 9 and 10 if possible, and in which format you are more familiar with.

9) What is your current weight? (Either answer in stones and pounds or kilograms).

……
…….

10) How tall are you? (Either answer in meters/cms or feet and inches).

……
……

11) Please indicate the areas that you have ever lived in

Country of Birth
…….

Areas Lived	**Location (Post Code)**	**Duration of stay in that area**

12) If you had symptoms from your PSC how long did it take for you to be diagnosed?

----

13) Have you had adequate information regarding your condition **Yes No Not sure**
(Please circle)

14a) Do other members of your family have PSC? **Yes No Not sure**
(Please circle)
14b) If yes, please indicate which other members of your family have PSC.

RELATIVE	YES
GRANDMOTHER	
GRANDFATHER	
FATHER	
MOTHER	
MATERNAL AUNT	
MATERNAL UNCLE	
PATERNAL AUNT	
PATERNAL UNCLE	
BROTHER	
SISTER	
OTHERS	

14c) How many brothers and sisters do you have **(alive or dead)**?

_____

15a) Do other members of your family have Inflammatory Bowel Disease?

Yes ☐ No ☐ **Not Sure** ☐

15b) If yes, please indicate which other members of your family have Inflammatory Bowel disease

RELATIVE	YES
GRANDMOTHER	
GRANDFATHER	
FATHER	
MOTHER	
MATERNAL AUNT	
MATERNAL UNCLE	
PATERNAL AUNT	
PATERNAL UNCLE	
BROTHER	
SISTER	
OTHERS	

16) How many children do you have?

DAUGHTERS (Please indicate how many)	**SONS (Please indicate how many)**

17) Do they have any illnesses? If so please describe them all in the box below**:**

Illnesses your children have please detail:

18) Do you suffer from any of the following medical conditions (you'll know if you have it!)?

	YES

A) SYSTEMIC LUPUS ERYTHEMATOSIS (SLE)	
B) OVERACTIVE OR UNDERACTIVE THYROID DISEASE	
C) INSULIN DEPENDANT DIABETES STARTING FROM YOUNG AGE	
D) SJOGREN'S SYNDROME	
E) SCLERODERMA	
F) COELIAC DISEASE	

19) Have you had any operations (If so please detail)

20) How old were you when you were first told you had PSC?

_____

(Please leave blank if not sure)

21) If you have inflammatory bowel disease how old were you when you first had it?
(Please leave blank if not sure)

---

22) When you were first told you had PSC, did you have any of the following symptoms?

	YES
A) ITCHING	
B) EXCESSIVE TIREDNESS	
C) DISCOMFORT IN THE LIVER AREA (THE RIGHT-SIDED, UPPER PART OF THE TUMMY)	
D) ACHING OF THE BONES	
E) ASCITES (FLUID INSIDE THE TUMMY)	
F) BLEEDING FROM VARICES (SWOLLEN VEINS AT THE BOTTOM END OF THE GULLET)	
G) JAUNDICE (YELLOW DISCOLOURATION AFFECTING THE WHITE OF THE EYE)	
H) HEPATIC ENCEPHALOPATHY (CONFUSION OWING TO LIVER DISEASE)	
I) NO SYMPTOMS (ONLY THE LIVER TESTS WERE ABNORMAL)	
J) OTHERS	

23a) Have you had a liver transplant? Yes ☐ No ☐

23b) If you have had a liver transplant, when was it performed? _____

Questions 24, 25 & 26: please only answer these questions if you have *not* had a liver transplant.

24) If you have not had a liver transplant, do you have any of the following symptoms now? How long have you had them?

	YES	DURATION
A) ITCHING		
B) EXCESSIVE TIREDNESS		
C) DISCOMFORT IN THE LIVER AREA (THE RIGHT-SIDED, UPPER PART OF THE TUMMY)		
D) ACHING OF THE BONES		
E) ASCITES (FLUID INSIDE THE TUMMY)		
F) BLEEDING FROM VARICES (SWOLLEN VEINS AT THE BOTTOM END OF THE GULLET)		
G) JAUNDICE (YELLOW DISCOLOURATION AFFECTING THE WHITE OF THE EYE)		
H) HEPATIC ENCEPHALOPATHY (CONFUSION OWING TO LIVER DISEASE)		
I) NO SYMPTOMS		
J) OTHERS		

25) Are you receiving any of the following medications for PSC**?**

MEDICATION	YES	
A) URSODEOXYCHOLIC ACID (URSO)?		
B) CHOLESTYRAMINE?		
C) RIFAMPICIN?		
D) HOW MUCH URSODEOXYCHOLIC ACID DO YOU TAKE? (PLEASE WRITE DOSE IN YES BOX IF TAKING IT)		

26) Are you waiting for a liver transplant**?** **Yes** ☐ **No** ☐

27) Have you ever from suffered from a cancer**?**

If yes please detail

No more questions. Thank-you for completing the questionnaire…

Clinician's Questionnaire (Version 1)

Research study into the genetic causes of Primary Sclerosing Cholangitis

Dear [collaborating clinician]

Re: [Patient name, date of birth, address]

This patient has agreed to participate in a research study into the genetic causes of primary sclerosing cholangitis. As part of the study, please provide the following information about the patient:

1) Is the patient under follow-up at [hospital]? Yes ☐ No ☐

2) Is the patient known to have PSC? Yes ☐ No ☐

3) Mode of diagnosis

3.1) Investigations

	YES	NO	DATE
3.1 A) MRCP			
3.1 B) ERCP			
3.1 C) LIVER BIOPSY			

3.2) If ERCP or MRCP did the patient have

	YES	NO	NOT CLEAR
3.2 A) INTRAHEPATIC CHOLANGIOGRAPHIC CHANGES			
3.2 B) EXTRAHEPATIC CHOLANGIOGRAPHIC CHANGES			
3.2 C) BOTH INTRA AND EXTRAHEPATIC			

3.3) Does the patient have Inflammatory Bowel Disease YES NO
Please circle

IF YES, PLEASE INDICATE WHICH TYPE	YES	NO
3.3 A) ULCERATIVE COLITIS		
3.3 B) INDETERMINATE COLITIS		
3.3 C) CHRONS COLITIS		

3.4) Regarding Liver biopsy

	YES	DATE
3.4 A) LIVER HISTOLOGY COMPATIBLE WITH PSC		
3.4 B) LIVER HISTOLOGY NOT COMPATIBLE WITH PSC		
3.4 C) LIVER BIOPSY NOT UNDERTAKEN		NA

4) Liver Biochemistry

Please provide *any* abnormal liver biochemistry result
Please ensure:

1) The blood test was undertaken at the time of diagnosis, or since the diagnosis was made.
2) When the blood test was undertaken, the patient had no intercurrent illness (e.g. sepsis).

	RESULT	DATE OF TEST
4.A) TOTAL BILIRUBIN		
4.B) AST		
4.C) ALT		
4.D) ALP		
4.E) ALBUMIN		

5) Does the patient suffer from any of the following co-morbid hepatic disorders?

	YES
5.A) CHRONIC HEPATITIS B	
5.B) CHRONIC HEPATITIS C	
5.C) ALCOHOLIC LIVER DISEASE	
5.D) NON-ALCOHOLIC LIVER DISEASE	
5.E) AUTOIMMUNE HEPATITIS (AUTOIMMUNE OVERLAP)	
5.F) PRIMARY BILIARY CIRRHOSIS	
5.G) HAEMACHROMATOSIS	
5.H) OTHER CO-MORBID LIVER DISORDER (PLEASE SPECIFY)	

6) What is the patient's ethnicity? _____

No more questions. Thank-you for completing the questionnaire.

Dr Gideon Hirschfield
The UK PSC Study

Recruitment centre: XXXX

Research Centre:
Dept. of Medical Genetics
Lv 6, Addenbrooke's Treatment Centre
Cambridge Biomedical Campus
Hills Rd
Cambridge
CB2 0QQ

Research study into the genetic causes of Primary Sclerosing Cholangitis

Clinician Questionnaire

Question 1

1.1) What is the patient's date of birth?.....

1.2) What is the patient's gender? Male ☐ Female ☐

1.3) Is the patient under follow-up at your hospital? Yes ☐ No ☐

1.4) What is the name of your hospital?
…….

1.5) Was the patient under another NHS Trust before you started looking after them?

Yes ☐ No ☐

1.6) If the answer to 1.5 was yes, please name the trust and hospital that was looking after the patient

.....
.....

Question 2

2.1) Is the patient known to have PSC? Yes ☐ No ☐

2.2) Date of Diagnosis of PSC (i.e confirmation on biopsy or cholangiography, if known)

……...

2.3) Was the patient initially **diagnosed** with another liver condition?

Yes ☐ No ☐

2.4) If yes, what was the name of this condition?
……..

2.5) What date was this diagnosis made?.....

Question 3

3.1) Is the patient still alive? Yes ☐ No ☐

If yes, please go to Q.4.

3.2) What is the cause of death documented on the death certificate or in the hospital notes (if known)?

Ia…….....
.....

Ib…….....
.....

Ic…….....
.....

3.3) What was the date of death?
.....

Question 4

4.1) Has the patient had a liver transplant? Yes ☐ No ☐

4.2) Is the patient on the liver transplant waiting list? Yes ☐ No ☐

If no, move straight to Question 5.

4.3) Date the patient was **listed** for liver transplant (if known):
……...

4.4) Date of actual liver transplant (if known):
..……..……...

4.5) Name of transplant centre (if known)…….……...

☐ ☐

4.6) Does the patient have recurrent PSC post-transplant? Yes No

Question 5

5.1) Please provide the blood results **at the time of the diagnosis**

	RESULT	REFERENCE RANGE	DATE OF TEST
TOTAL BILIRUBIN			
AST			
ALT			
ALP			
ALBUMIN			
CREATININE			
SODIUM			
HAEMOGLOBIN			
PLATELETS			
WHITE CELL COUNT			
EOSINOPHIL COUNT			
PROTHROMBIN TIME			
APTT			
INR			

5.2) Please provide the results of the patient's blood tests **one year after initial diagnosis**

	RESULT	REFERENCE RANGE	DATE OF TEST
TOTAL BILIRUBIN			
AST			
ALT			
ALP			
ALBUMIN			
CREATININE			
SODIUM			
HAEMOGLOBIN			
PLATELETS			
WHITE CELL COUNT			
EOSINOPHIL COUNT			
PROTHROMBIN TIME			
APTT			
INR			

5.3) Please provide the blood results **two years after initial diagnosis**

	RESULT	REFERENCE RANGE	DATE OF TEST
TOTAL BILIRUBIN			
AST			
ALT			
ALP			
ALBUMIN			
CREATININE			
SODIUM			
HAEMOGLOBIN			
PLATELETS			
WHITE CELL COUNT			
EOSINOPHIL COUNT			
PROTHROMBIN TIME			
APTT			
INR			

5.4) Please provide the results of the patient's **most recent** blood tests

*****IF THE PATIENT IS A TRANSPLANT RECIPIENT, THEN PLEASE PROVIDE BLOOD TEST RESULTS IMMEDIATELY PRIOR TO TRANSPLANTATION*****

	RESULT	REFERENCE RANGE	DATE OF TEST
TOTAL BILIRUBIN			
AST			
ALT			
ALP			
ALBUMIN			
CREATININE			
SODIUM			
HAEMOGLOBIN			
PLATELETS			
WHITE CELL COUNT			
EOSINOPHIL COUNT			
PROTHROMBIN TIME			
APTT			
INR			

Question 6

6.1) Please provide the following immunology results (if multiple results available, please state the earliest result, post-diagnosis)

	RESULT	REFERENCE RANGE	DATE OF TEST
P-ANCA *(POSITIVE OR NEGATIVE)*		N/A	
ANA *(POSITIVE OR NEGATIVE)*		N/A	
SMA* *(POSITIVE OR NEGATIVE)* *SMOOTH MUSCLE ANTIBODY	RESULT: TITRE:	N/A	
ANTI MITOCHONDRIAL ANTIBODY *(POSITIVE OR NEGATIVE)*		N/A	
LKM-1 ANTIBODY *(POSITIVE OR NEGATIVE)*		N/A	
IMMUNOGLOBULIN IGG *(LEVEL)*			
IMMUNOGLOBULIN IGA *(LEVEL)*			
IMMUNOGLOBULIN IGM *(LEVEL)*			
IGG4 *(LEVEL AND UNITS)*			
CA19-9 *(LEVEL)*			

Question 7

7.1) Please indicate which of the following was used as the **initial diagnostic** investigation:

	YES	NO	DATE
MRCP			
ERCP			
LIVER BIOPSY			

7.2) Was ERCP or MRCP compatible with PSC? Yes ☐ No ☐

7.3) If the patient has had an ERCP or MRCP, did it show:

	YES	NO	UNCLEAR
INTRAHEPATIC CHOLANGIOGRAPHIC CHANGES OF PSC			
EXTRAHEPATIC CHOLANGIOGRAPHIC CHANGES OF PSC			
BOTH INTRA AND EXTRAHEPATIC CHANGES OF PSC			
NORMAL (NO EVIDENCE OF PSC)			

(Please note that extrahepatic biliary tree involves up to the point of bifurcation of the left and right hepatic ducts i.e. everything up to, and including the common hepatic duct).

7.4) Has the patient had a **follow-up** MRCP or ERCP? Yes ☐ No ☐

If yes, please provide the date:……

7.5) Was this an ERCP or MRCP?:……..

7.6) Please indicate changes seen on **most recent** MRCP or ERCP:

	YES	NO	UNCLEAR
INTRAHEPATIC CHOLANGIOGRAPHIC CHANGES OF PSC			
EXTRAHEPATIC CHOLANGIOGRAPHIC CHANGES OF PSC			
BOTH INTRA AND EXTRAHEPATIC CHANGES OF PSC			
NORMAL (NO EVIDENCE OF PSC)			

7.7) Are digital or DICOM copies of ERCP/MRCP images available? Yes ☐ ☐ No

7.8) If the patient has had a Liver Biopsy, please indicate the following findings:

	YES	DATE
LIVER HISTOLOGY COMPATIBLE WITH PSC		
LIVER HISTOLOGY NON-DIAGNOSTIC		
LIVER HISTOLOGY NOT COMPATIBLE WITH PSC		
LIVER BIOPSY NOT UNDERTAKEN		NA

7.9) Which stage of liver disease did the patient have (if known)? (Stage 1-4) .....

7.10) Has the patient had a Fibroscan? Yes ☐ ☐
No

If yes, please document the date……..…….. kPa……...
f=……..

*****PLEASE INCLUDE AN ANONYMISED COPY OF THE MRCP/ERCP AND LIVER BIOPSY REPORTS CITED, STATING THE PATIENT UK-PSC STUDY ID*****

Question 8

8.1) Has the patient had a colonoscopy to screen for Inflammatory Bowel Disease (IBD)?

Yes ☐ No ☐

8.2) Does the patient have Inflammatory Bowel Disease?

Yes ☐ No ☐

If no, please proceed to Q. 8.6

If yes to 8.2, please indicate which type:

	YES	NO
ULCERATIVE COLITIS		
INDETERMINATE COLITIS		
CROHNS COLITIS		

8.3) Date of diagnosis of IBD (if known)

……..…….

8.4) If they have colitis is it: MACROSCOPIC ☐ MICROSCOPIC ☐

8.5) Distribution of the Colitis (Please tick all that apply)

TERMINAL ILEUM		CAECUM		ASCENDING COLON	
TRANSVERSE COLON		DESCENDING COLON		SIGMOID COLON	
RECTUM					

*****PLEASE INCLUDE AN ANONYMISED COPY OF THE INITIAL COLONOSCOPY REPORT STATING THE PATIENT UK-PSC STUDY ID*****

8.6) If patient has ever had dysplasia, was it; Low grade ☐ High grade ☐

8.7) Has the patient had a colectomy? Yes ☐ No ☐

8.8) If yes, what was the date of their colectomy (if known)?

……..

8.9) Specify type of colectomy (if known): Sub-total colectomy with ileo-anal pouch ☐

Pan-proctocolectomy with ileostomy ☐

8.10) What was the indication for the colectomy?

Cancer ☐ Inflammation ☐

Low grade Dysplasia ☐ High grade dysplasia ☐

Other (state)……..……...

8.11) If patient has had colon cancer, please list the site ……..…….……

Question 9

9.1) Has the patient had any of the following:

	Yes	Date of first diagnosis
Cholangiocarcinoma		
Gallbladder cancer		
Hepatocellular carcinoma		
Pancreatic cancer		
Cholecystectomy		
Gallbladder dysplasia (on cholecystectomy report)		
Colorectal Cancer		
Bile duct dysplasia (from ERCP or surgery)		

9.2) Has the patient had a bone density/DEXA scan? Yes ☐ No ☐

If no, proceed to Q.10

9.3) If yes, provide details of the most recent scan:

Date of the report …… T score of hip……..…… T score of spine……

Question 10

10.1) Please put the details of the **first** liver ultrasound findings the patient had:

Date of ultrasound ……..

Is liver heterogenous? Yes ☐ No ☐ Not sure ☐

Is capsule of liver irregular? Yes ☐ No ☐ Not sure ☐

Size of spleen (cm)…….

If no size recorded, was the spleen: Normal sized ☐ Enlarged ☐

Gallstones Yes ☐ No ☐

Gallbladder Polyps Yes ☐ No ☐

Ascites Yes ☐ No ☐

10.2) Please put the details of the **most recent** US the patient had

Date of ultrasound ……..

Is liver heterogenous? Yes ☐ No ☐ Not sure ☐

Is capsule of liver irregular? Yes ☐ No ☐ Not sure ☐

Size of spleen (cm)…….

If no size recorded, was the spleen: Normal sized ☐ Enlarged ☐

Gallstones Yes ☐ No ☐

Gallbladder Polyps Yes ☐ No ☐

Ascites Yes ☐ No ☐

Question 11

11.1) Does the patient suffer from any of the following co-morbid hepatic disorders?

	YES
CHRONIC HEPATITIS B	
CHRONIC HEPATITIS C	
ALCOHOLIC LIVER DISEASE	
NON-ALCOHOLIC FATTY LIVER DISEASE (NAFLD OR NASH)	
AUTOIMMUNE HEPATITIS (OR AUTOIMMUNE OVERLAP)	
PRIMARY BILIARY CIRRHOSIS	
HAEMOCHROMATOSIS	
OTHER CO-MORBID LIVER DISORDER (PLEASE SPECIFY)	

11.2) Has the patient had any of the following complications/diagnoses?

	YES	DATE OF FIRST EPISODE OR FIRST NOTED
VARICES ON SCREENING ENDOCSOPY		
VARICEAL BLEED		
ASCITES		
ENCEPHALOPATHY		
CHOLANGITIS (BILIARY SEPSIS)		
JAUNDICE		
CIRRHOSIS		

Question 12

12.1) Does the patient suffer from any of these additional diseases?

	YES
TYPE 1 DIABETES	
PSORIASIS	
RHEUMATOID ARTHRITIS	
ANKYLOSING SPONDYLITIS	
SARCOIDOSIS	
MYAESTHENIA GRAVIS	
MULTIPLE SCLEROSIS	
COELIAC DISEASE	
SYSTEMIC LUPUS ERYTHEMATOSUS	
SJOGRENS SYNDROME	
HYPOTHYROIDISM	
HYPERTHYROIDISM	

Question 13

13.1) Has the patient ever taken Ursodeoxycholic acid? Yes ☐ ☐ No

If no, proceed to Q.13.5

13.2) If yes to 13.1, what date did they start? .....

13.3) What dose of Ursodeoxycholic Acid did they take (dose mg)……. (frequency)…….

13.4) If the patient stopped taking Ursodeoxycholic acid, what date did they stop?.....

13.5) Has the patient ever taken Azathioprine? Yes ☐ No ☐

13.6) If yes to 13.5, what date did they start?
.....

13.7) Has the patient ever taken a 5-ASA compound (e.g. mesalazine,'pentasa', 'asacol' 'ipocol', 'mesovant', 'octasa', Balsalazide, 'colazide'and balsalazide, olsalazine)?

☐ Yes No ☐

13.8) What is the patient's most recent weight (please state units)?
……..

No more questions. Thank you for completing the questionnaire.

FINAL CHECKLIST

Have you included an anonymised copy of the following, stating the patient's UK PSC Study number?

ERCP Report ☐

MRCP Report ☐

Liver Biopsy Report ☐

Colonoscopy report ☐

*If you have any **questions** or **queries** regarding the **completion** of this questionnaire please contact:*

Bridget Bell (B.F.Bell@bham.ac.uk; Tel: 0121 371 8101).

Please return completed questionnaires to:
The UK PSC Study,
Dept. of Medical Genetics
Level 6 Addenbrooke's Treatment Centre,
Cambridge Biomedical Campus,
Hills Rd,
Cambridge, CB2 0QQ

Published abstracts and original research papers