NIH Public Access
Author Manuscript
Genes Immun. Author manuscript; available in PMC 2012 June 02.

Published in final edited form as:
Genes Immun. 2012 June ; 13(4): 328–335. doi:10.1038/gene.2011.89.

Association of primary biliary cirrhosis with variants in the *CLEC16A, SOCS1, SPIB* and *SIAE* immunomodulatory genes

GM Hirschfield[1,2], **G Xie**[3], **E Lu**[4], **Y Sun**[3], **BD Juran**[5], **V Chellappa**[6], **C Coltescu**[1], **AL Mason**[7], **P Milkiewicz**[8], **RP Myers**[9], **JA Odin**[10], **VA Luketic**[11], **B Bacon**[12], **H Bodenheimer**[13], **V Liakina**[14], **C Vincent**[15], **C Levy**[16], **S Pillai**[6], **KN Lazaridis**[5], **CI Amos**[4], and **KA Siminovitch**[2,3,17]

[1]Liver Centre, Toronto Western Hospital, Toronto, Ontario, Canada [2]Department of Medicine, University of Toronto, Toronto, Ontario, Canada [3]Mount Sinai Hospital, Samuel Lunenfeld Research Institute and Toronto General Research Institute, Toronto, Ontario, Canada [4]Department of Epidemiology, University of Texas MD Anderson Cancer Center, Houston, TX, USA [5]Center for Basic Research in Digestive Diseases, Division of Gastroenterology and Hepatology, Mayo Clinic College of Medicine, Rochester, MN; USA [6]Cancer Center, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA [7]Department of Medicine, University of Alberta, Edmonton, Alberta, Canada [8]Liver Unit, Pomeranian Medical School, Szczecin, Poland [9]Liver Unit, University of Calgary, Calgary, Alberta, Canada [10]Division of Liver Diseases, Mount Sinai School of Medicine, New York, NY, USA [11]Department of Gastroenterology, Virginia Commonwealth University, Richmond, VA, USA [12]Division of Gastroenterology and Hepatology, Saint Louis University School of Medicine, St Louis, MO, USA [13]Department of Medicine, Albert Einstein College of Medicine, New York, NY, USA [14]Centre of Gastroenterology and Dietetics, Vilnius University, Vilnius, Lithuania [15]Université de Montréal Hospital Centre, Saint-Luc Hospital, Montréal, Quebec, Canada [16]Center for Liver Diseases Division of Hepatology, University of Miami School of Medicine, Miami, FL, USA [17]Departments of Immunology and Molecular Genetics, University of Toronto, Ontario, Canada

Abstract

We fine mapped two primary biliary cirrhosis (PBC) risk loci, *CLEC16A* (C-type lectin domain family 16 member A)–suppressor of cytokine signaling 1 (*SOCS1)* and Spi-B protein *(SPIB)* and sequenced a locus, sialic acid acetylesterase (*SIAE*), proposed to harbor autoimmunity-associated mutations. In all, 1450 PBC cases and 2957 healthy controls were genotyped for 84 single-nucleotide polymorphisms (SNPs) across the *CLEC16A-SOCS1* and *SPIB* loci. All 10 exons of the *SIAE* gene were resequenced in 381 cases and point substitutions of unknown significance assayed for activity and secretion. Fine mapping identified 26 SNPs across the *CLEC16A-SOCS1* and 11 SNPs across the *SPIB* locus with significant association to PBC, the strongest signals at the

© 2012 Macmillan Publishers Limited All rights reserved

Correspondence: Dr KA Siminovitch, Mount Sinai Hospital, Samuel Lunenfeld Research Institute and Toronto General Research Institute, 600 University Ave, Room 778D, Toronto, Ontario M5G 1X5, Canada. ksimin@mshri.on.ca.

Author contributions: GMH, GX and KAS designed the study. GX, EL, YS, VC, SP and CIA performed the genotyping, variant expression/functional analysis and data analysis. GMH, BDJ, CC, ALM, PM, RPM, JAO, VAL, BB, HB, VL, CV, CL, KNL and KAS developed the clinical network and sample collection and processing framework required for case and control accrual. GMH, GX and KAS wrote the manuscript. GMH, GX, SP, CIA and KAS vouch for the data and its analysis/interpretation. GMH and KAS decided in agreement with all other authors to publish this paper.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

Supplementary Information accompanies the paper on Genes and Immunity website (http://www.nature.com/gene)

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

CLEC16A-SOCS1 locus emanating from a *SOCS1* intergenic SNP (rs243325; $P = 9.91 \times 10^{-9}$) and at the *SPIB* locus from a *SPIB* intronic SNP (rs34944112; $P = 3.65 \times 10^{-9}$). Among the associated SNPs at the *CLEC16A-SOCS1* locus, two within the *CLEC16A* gene as well as one *SOCS1* SNP *(rs243325)* remained significant after conditional logistic regression and contributed independently to risk. Sequencing of the *SIAE* gene and functional assays of newly identified variants revealed six patients with functional non-synonymous *SIAE* mutations (Fisher's $P = 9 \times 10^{-4}$ vs controls) We demonstrate independent effects on risk of PBC for *CLEC16A, SOCS1* and *SPIB* variants, while identifying functionally defective *SIAE* variants as potential factors in risk for PBC.

Keywords

autoimmunity; fine mapping; sequence analysis; primary biliary cirrhosis

INTRODUCTION

Primary biliary cirrhosis (PBC) is a chronic and progressive biliary disease, characterized by a destructive granulomatous cholangitis and loss of tolerance to ubiquitous mitochondrial antigens.[1] Epidemiological data indicate important interactions between environmental and genetic factors in PBC predisposition and cumulative biological data identify an interplay of adaptive and innate immunity as integral to disease pathogenesis.[2] However, despite the well-recognized roles for immune cell populations in PBC pathophysiology, understanding of the molecular pathways underpinning PBC remains limited and needs to be improved.[3] Results of genome-wide association surveys (GWAS) confirmed association of PBC not only with the previously known human leukocyte antigen class II locus, but also many other novel non-human leukocyte antigen loci encompassing genes involved in immunity.[4] Thus, for example, the first GWAS data set and follow-up studies available for PBC identified not only the *IL12A* gene, but also genes encoding other components of the interleukin-12 (IL-12) signaling pathway (*IL12RB2* and *STAT4*) as well as the *interferon response factor 5* gene as PBC risk loci. These findings have since been replicated and extended.[5–8] Importantly, these data also identify many of the immune-related loci conferring risk for PBC to be shared with rheumatoid arthritis, celiac disease, systemic lupus erythematosus and other autoimmune diseases that appear with increased frequency in PBC patients and their families.

Although some of the non-human leukocyte antigen PBC risk loci containing immunologically relevant genes have been subjected to further analyses, several of the more recently identified susceptibility regions remain relatively uncharacterized. Two such immunologically relevant loci that we have prioritized, include the *CLEC16A* (C-type lectin domain family 16 member A)–*SOCS1* (suppressor of cytokine signaling 1) locus at chromosome 16p13 and the Spi-B protein *(SPIB)* locus at chromosome 19q13. At the *CLEC16A-SOCS1* locus, for example, *CLEC16A* (aka *KIAA0350*) encodes a tyrosine-based activation motif-containing protein with likely effects on immune cell signaling,[9] while the *SOCS1* gene product has integral roles in modulating cytokine and toll-like receptor-driven signaling pathways that regulate immune responsiveness.[10,11] The *CLEC16A-SOCS1* locus has already been implicated in risk for celiac disease,[12] multiple sclerosis,[13] Type 1 diabetes[9] and immunoglobulin A deficiency[14] and its association with PBC as well was initially flagged at a low level of significance by a Canadian PBC GWAS analysis and then formally statistically demonstrated in a UK-based PBC GWAS. However, in the former of these two studies, the single-nucleotide polymorphism (SNP) giving the strongest association signal was located just upstream of the *SOCS1* gene, while in the latter, the most significant SNP at this locus was located in the *CLEC16A* gene, which maps

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

over 70 kb upstream of *SOCS1*. This locus is genetically complex, and in multiple sclerosis a recent study based on linkage disequilibrium patterns in the region and logistic regression analysis, suggested that this region harbors independent disease risk loci.[15] *SPIB* is also a plausible candidate gene for PBC based on both GWAS data and data implicating the SPIB protein in both B-cell antigen receptor signaling and development of T cells and of plasmacytoid dendritic cells involved in inflammatory responses.[7,8] This locus has not, however, emerged as a risk locus for autoimmune diseases other than PBC.

Knowledge of specific risk relevant genes underpinning PBC remains limited and one factor thought to be important are contributions to risk from rare gene variants not amenable to discovery via conventional GWAS, but requiring costly sequencing efforts. This possibility is strongly supported by the recent discovery of rare variants in the sialic acid acetylesterase (*SIAE*) gene that occur significantly more frequently in patients with rheumatoid arthritis, systemic lupus erythematosus and multiple sclerosis than in healthy controls.[16] The SIAE enzyme, which removes 9-O-acetyl moieties from sialic acid, has been independently implicated in autoimmunity by the immunological phenotype of *SIAE*-deficient mice and by its potential to regulate B-cell immune responses via 9-O-acetylation of α-2,6-linked glycans that modulate activity of the CD22 B-cell inhibitory receptor.[17] This finding has not been extensively pursued, although a Met89Val variant was not associated with type 1 diabetes mellitus and Graves' disease in one study of a Polish population.[18]

Through high-density genotyping across the *CLEC16A-SOCS1* and *SPIB* loci and direct sequencing across the *SIAE* locus, we identify multiple disease variants across the *CLEC16A-SOCS1* and *SPIB* loci and show that variants within each of the *CLEC16A* and *SOCS1* gene contribute independently to the risk for PBC. We additionally demonstrate the presence of functional variants in the *SIAE* gene in patients with PBC.

RESULTS

Evidence for independent association of *CLEC16A* and *SOCS1* variants with risk for PBC

In our initial PBC GWAS analysis, we found suggestive, but not significant, evidence for PBC association with the *CLEC16A-SOCS1* locus, with a SNP (rs243324) located about 5.5 kb upstream of the *SOCS1* coding sequence giving the strongest association in the region. This locus failed to achieve genome-wide significance in our analysis, but was also identified in a subsequent UK-based PBC GWAS. In this latter study, the most significant SNP mapped to intron 19 of *CLEC16A*, about 166 kb upstream of the *SOCS1* gene. To better ascertain the source of the association signal at this locus, we genotyped our full study cohort (1450 cases and 2967 controls) for 57 SNPs spanning a 450-kb genomic region encompassing the *CLEC16A* and *SOCS1* genes as well as several very small genes, *transitional protein 2* and *protamine 1-3*, that map about 12 kb downstream of *SOCS1* (Table 1). From this analysis, 26 SNPs achieved the significance thresholds set for this study ($P<10^{-4}$), with the strongest association signal ($P = 9.91 \times 10^{-9}$; odds ratio (OR) = 1.33; 95% confidence interval (CI) = 1.21–1.47) deriving from the rs243325 SNP located about 5 kb upstream of the *SOCS1* gene start codon and about 65 kb downstream of *CLEC16A* (Table 1, Figure 1 and Supplementary Table 1). In all, 15 other SNPS flanking the *SOCS1* gene region, 7 SNPs in or near the *TPN2* and *protamine* genes and 10 SNPs within various *CLEC16A* gene introns also achieved nominal significance and analysis of the linkage disequilibrium patterns among the markers revealed these SNPs to constitute three major haplotype blocks, two across the *CLEC16A* gene and the third, a 27-kb region, encompassing the *SOCS1*, *transitional protein 2* and *protamine* genes (Supplementary Figure 1). Importantly, this analysis revealed the most strongly associated SNP in the single-point analysis (rs243325) to be in strong linkage disequilibrium with five other disease-associated SNPs mapping 3′ to *SOCS1* (pairwise r^2 ranging between 0.66 and 0.93 and D′

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

between 0.88 and 0.99). Within the two haplotype blocks across the *CLEC16A* gene, the strongest association signal ($P = 4.85 \times 10^{-7}$, OR = 1.46) came from the rs58102322 SNP located in intron 18. The rs12924729 SNP initially reported in the UK PBC GWAS[8] was also among the significant SNPs at this locus ($P = 6.73 \times 10^{-6}$, OR = 1.25), but shows negligible LD with rs58102322 and maps within a distinct haplotype block within the gene.

To investigate the extent to which SNPs within the *SOCS1* vs the *CLEC16A* haplotype block contribute to the disease association at this locus, conditional logistic regression analysis was undertaken, with each of the test SNPs conditioned on the effect of the other SNPs. Three SNPs, rs58102322 and rs12924129 (intronic variants in the *CLEC16A*) and rs243325 (*SOCS1* 5′ intergenic region) remained significant after conditioning sequentially on the most significant markers (Supplementary Table 2) strongly suggesting that there are multiple independent variants influencing PBC risk in this region. This conclusion is consistent with results of haplotype analyses incorporating these three risk loci (Table 2A), which reveal effects on risk for this locus (omnibus test: $P = 2.61 \times 10^{-14}$) and for a risk allele haplotype (AAG, $P = 5.36 \times 10^{-8}$; OR = 1.67) and a protective allele haplotype (GGA, $P = 3.93 \times 10^{-10}$; OR = 0.52) at the locus to be much stronger than for any single SNP or any other allele combination at these three loci. Thus at this locus, risk associated with PBC appears to be largely determined by complementary haplotypes, which might reflect effects from as-of-yet unidentified single variant(s) or from epistatic interactions among the three loci.

To further explore the relative influence of these variants on disease susceptibility, the combined contribution of the three risk alleles at the rs12924729 (G), rs58102322 (A) and rs243325 (A) loci was also explored using logistic regression analysis (Table 2B). For this analysis, the effect of carrying risk genotypes at the *CLEC16A* gene rs12924729 and rs58102322 loci was evaluated in the context of presence or absence of the risk genotypes at the *SOCS1* rs243325 locus. The results revealed the OR to be significantly increased by combining carriage of rs12924729 and rs58102322 risk alleles with (OR = 3.66; 95% CI = 1.58–8.46; $P = 0.002$) compared with without (OR = 2.60; 95% CI = 1.00–6.77; $P = 0.05$) rs243325 risk alleles. When compared with all the other allele combinations, the combination of all three risk alleles has an effect size, 1.57 (95% CI = 1.28–1.94), higher than that of any individual SNP at this locus. Thus, addition of the rs243325 risk allele to the other two risk alleles substantially increases risk for disease as is in keeping with independent effects of these alleles on risk.

Fine mapping of the PBC-associated *SPIB* locus

To refine and further validate the *SPIB* locus, we genotyped all cases and controls for 27 SNPs across an ~60-kb region encompassing the *SPIB* gene and two other genes, *polymerase delta 1* and *myosin-binding protein 2* that map 1 kb upstream and 3.9 kb downstream, respectively, of *SPIB*. This analysis identified 11 SNPs at this locus to be significantly associated with PBC, 3 of which achieved *P*-values $<5 \times 10^{-8}$ (Table 3, Supplementary Table 3; Figure 2) and almost all of which map within one of two haplotype blocks across the region (Supplementary Figure 2). The most significant SNP at this locus was an intronic polymorphism, rs34944112 ($P = 3.65 \times 10^{-9}$; OR = 1.49, 95% CI = 1.31–1.70), which was in strong LD ($r^2 = 0.96$; $D' = 0.98$) with another significantly associated SNP, the rs3745516 SNP identified as a PBC risk locus in the previous GWAS studies ($P = 1.93 \times 10^{-7}$; OR = 1.30). To further explore the contributions of these SNPs to the association signal, conditional logistic regression analysis was then performed again incorporating all test SNPs in the analysis. After conditioning on each marker, only the most significantly associated SNP in the region, rs34944112 remained strongly significant (Supplementary Table 4), the power for detecting an OR≥1.3 being >90%. These findings

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

suggest that this variant functions independently to confer much of the effect on risk at this locus.

Functionally deleterious *SIAE* gene mutations and PBC

To assess whether *SIAE* gene variants are relevant to risk for PBC, we resequenced all 10 exons of this gene in 381 unrelated individuals randomly chosen from our cohort of PBC cases. Results of this analysis revealed a total of 12 point substitutions among the patients, 6 of which were non-synonymous changes that were found in 10 of 381 cases but that occurred only once in 648 controls (Table 4); $P = 0.0003$. Among the non-synonymous polymorphisms, two encode SIAE protein species, F404S and T312M, previously shown to be catalytically defective and one encodes a species, M89V, reported to be catalytically intact (but not secreted). By contrast, three other polymorphisms detected in the patients encode variants, F199C, P356L and Q382R, that have not been previously described and have unknown functional effects.

To allow assessment of the frequency at which biologically relevant *SIAE* variants occur in PBC patients, we next explored the functional properties of the F199C, P365L and Q382R species. To this end, cDNAs encoding each of these proteins as well as the wild-type SIAE protein or a known catalytically dead SIAE mutant (S127A) were expressed in 293T cells and esterase activity of anti-SIAE immunoprecipitates from these cells then assayed. To screen for a secretory defect, SIAE levels were also assessed in culture supernatants and compared between the supernatants and lysates. Results of these analyses revealed catalytic activity to be severely reduced for the F199C variant, but essentially normal for the Q382R and P356L variant proteins (Figure 3). The F199C protein was also poorly secreted into culture supernatants. Thus, one of the three variants, F199C appears to be functionally defective. On the basis of these findings and the sequence data revealing presence of the F199C variant in 3 of our PBC cases, and variants encoding other functionally impaired *SIAE* species in 3 more subjects, it appears that 6 of the 381 cases tested here have a functionally defective *SIAE* variant, and 2 further patients are homozygote for a secretory defective variant. By contrast, in 648 healthy Caucasian controls of European descent, functionally relevant *SIAE* variants were identified in only two individuals and in 1351 samples of European American origin for which *SIAE* exome sequencing data are available (http://snp.gs.washington.edu/EVS/), one further control with a functionally relevant *SIAE* variant was identified. Therefore in contrast to our PBC patients, in whom 6 of 381 have functional mutations, only 3 of 1999 controls (648 from our study, 1351 from the Exome Sequencing Project data), carry functional *SIAE* variants ($P = 9 \times 10^{-4}$; Fisher's exact test).

DISCUSSION

Our current study in which we focus on further refining non-human leukocyte antigen association with PBC, builds on recent data pinpointing the *CLEC16A*-*SOCS1* and *SPIB* loci as risk loci for PBC and that identifying *SIAE* gene variants as risk alleles for several autoimmune diseases. Our findings confirm the presence of disease-associated SNPs across the *CLEC16A*-*SOCS1* and *SPIB* gene loci and identify a number of new potential PBC markers at each of these loci. We also demonstrate that 10 of the 381 PBC cases studied here carry non-synonymous point substitutions in the *SIAE* gene, 2 of which were previously shown to encode functionally defective proteins (F199C and T312M) and 1 of which was identified for the first time in this study and shown here to encode catalytically impaired protein (F199C). Other variants were also detected in the patients, including one other, which was the only variant found in the homozygous state and, which has been shown previously to encode a secretory defective, but functional protein (M89V). These results thus provide the first evidence for a potential role of *SIAE* variants in PBC.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

The current data confirm PBC association with the *CLEC16A-SOCS1* locus and the statistical data gleaned from conditional, haplotype and combination analyses provide compelling evidence for independent contributions of *CLEC16A* and *SOCS1* alleles to risk of developing PBC. This conclusion is consistent with the well-recognized role for *SOCS1* and likely role for *CLEC16A* in immune regulation. SOCS1 is a well-established immune cell signaling inhibitor that functions downstream of cytokine receptors to inhibit JAK-STAT-mediated induction of interferon-γ, IL-2, IL-6, IL-4 and tumor necrosis factor-α and many other proinflammatory cytokines.[19] Mice lacking SOCS1 protein thus exhibit cellular hyperreactivity to many such cytokines and rapidly develop a multi-organ inflammatory disease associated with profound activation of hepatic lymphocytes and death by age 3 weeks due to fulminating hepatitis.[20,21] Specific loss of SOCS1 in hepatocytes also renders mice more susceptible to ConA-induced hepatitis,[22] while *SOCS1*-specific deletion in T cells promotes Th1, but impairs Th17 cell development.[23] Thus, SOCS1 regulates T-cell differentiation pathways that are integral to maintenance of immune cell homeostasis. Roles for both *SOCS1* and *CLEC16A* variants in PBC are thus in keeping with genetic association data implicating this locus in other autoimmune diseases and with the cumulative biological understanding of the two proteins.

The *SPIB* locus was initially revealed as a likely PBC risk locus by our GWAS data showing an association signal that just missed genome-wide significance ($P = 9.12 \times 10^{-7}$; OR = 1.27) at the rs3745516 SNP locus within a *SPIB* gene intron.[6] Strong evidence for this gene's association with PBC was then provided by a follow-up meta-analysis of Italian and Canadian GWAS data[7] and a subsequent UK PBC GWAS.[8] Our current findings provide added evidence for *SPIB* involvement in PBC, revealing strong association of several intronic SNPs with risk and suggesting this locus's effect on risk to be primarily attributable to the most significant variant at this locus, rs34944112. Interestingly, another significant association signal from the region came from the rs11546996 SNP, which encodes an alanine to proline substitution in exon 4 of the *SPIB* gene. This conservative substitution, however, would be expected to have little functional relevance, a prediction consistent with the conditional analysis results showing no independent effect of this SNP on risk. Defining the disease-causal SNP at this locus will require further biological studies, but an effect of the *SPIB*-encoded Spi-B transcription factor on risk is in keeping with its prerequisite role in regulating both lineage commitment during haematopoiesis and differentiation of mature lymphocyte and dendritic cell populations.[24,25] The *SPIB* gene is also of particular potential interest as this locus has not, as of yet, been implicated in other autoimmune diseases and may prove to represent a PBC-specific risk locus. This possibility and the definition of the disease relevant allele at this locus, however, require further investigation through expanded genetic and functional studies.

In contrast to most of the gene loci now implicated in PBC susceptibility, *SIAE* emerged as a candidate gene for PBC in the context of a resequencing screen for *SIAE* variants in other autoimmune diseases. Involvement of this gene in such diseases was initially suggested by the enhanced B-cell activation and auto-antibody production observed in *SIAE*-deficient mice and was then confirmed by the demonstration of functionally deleterious *SIAE* gene mutations in patients with rheumatoid arthritis, systemic lupus erythematosus and multiple sclerosis.[16,17] One very limited replication study in type 1 diabetes and Graves disease was negative,[18] although our systematic *SIAE* resequencing and functional data obtained in this study do reveal a small number of the PBC patients tested to have loss-of-function *SIAE* missense mutations. As genotype data were not available for our full case population, genome-wide findings were not adjusted for population stratification. However, an association of *SIAE* variants with PBC, as suggested by our data, is consistent with the not infrequent occurrence in PBC patients and their families of other autoimmune diseases (rheumatoid arthritis and systemic lupus erythematosus) known to be associated with *SIAE*

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

variants. In our study, for example, 8 of the 10 PBC cases with a functional *SIAE* variant and/or a family member had one or more other autoimmune diseases. The concordance between PBC and other autoimmune diseases implies that observed *PBC*-gene associations may also reflect the presence of these other diseases. Sequencing studies of a much larger case/control cohort are needed to address this possibility and better delineate the frequency, repertoire and cellular consequences of rare variants in PBC and the interactions of such variants with more common variants in predisposing to PBC vs other autoimmune disease.

The *SIAE* locus has not been specifically implicated in PBC by prior reports. For this reason and also because there are little data on the contribution of rare variants to risk for autoimmune disease, we believe the demonstrated association of PBC with *SIAE* variants adds an important new dimension to the genetic understanding of such conditions. Similarly, while implicated in risk for PBC as well as other autoimmune diseases, the *CLEC16A-SOCS1* locus association with PBC had not been replicated before this study and there are no published data that refine this GWAS-defined risk locus. The data reported here thus provide new insights into the genetic basis of PBC and a framework for further studies using larger cohorts to address potential correlations of the gene variants reported here to PBC subphenotypes such as anti-mitochondrial antibodies status and severity.

Collectively, our current findings are in keeping with key roles for multiple immune-related genes in PBC and provide more evidence of overlap between some, but not all PBC risk alleles and those reported for other autoimmune diseases. Our data also begin the process of identifying contributory effects of rare missense mutations to risk for PBC and thus highlight the need to search for additional risk alleles by more extensive resequencing endeavors in extremely large case–control cohorts.

MATERIALS AND METHODS

Subjects

The subject cohort studied here included 1450 PBC cases and 2967 controls. Another 648 ‘historic’ controls served as a source for data on *SIAE* variant frequency in healthy controls.[16] Clinical characteristics of the patients have been previously reported,[5,6] but all were diagnosed with PBC according to the American Association for the Study of Liver Disease guidelines.[1] All study subjects as well as historic controls were of Caucasian European descent and were ascertained from either the Canada, USA or Europe. The Local Institutional Review Committee provided approval for conduct of this study, and all individuals gave their informed consent.

CLEC16A-SOCS1 and *SPIB* SNP selection and genotyping

SNPs initially tested across the 16p13 and 19q13 risk loci were selected based on at least one of the following criteria: (1) Hap Map Phase III data identifying the SNP as a Tag SNP with minor allele frequency >0.01. and r^2 threshold of 0.8) and/or (2) localization within 150 kb upstream or downstream of the SNPs at each locus most significantly associated with PBC in our GWAS and GWAS meta-analysis. Genotyping was performed using the Sequenom MassArray iPLEX platform (San Diego, CA, USA) with allele-extension products plated onto a SpectroCHIP and subjected to mass spectrometry and the genotypes then called using SpectroCALLER software (Sequenom). SNP annotation was based on NCBI dbSNP Build 129.

SIAE gene sequencing

DNA was prepared by the standard phenol/chloroform extraction from genomic DNA of each subject. All 10 exons of the *SIAE* gene were then amplified by PCR using previously

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

reported intronic primer pairs covering the 5′ and 3′ splice sites of each exon[16] and the amplicons then subjected to bidirectional automated sequencing using an Applied Biosystems 3730 DNA analyzer (Carlsbad, CA, USA). Sequence alignment and polymorphism identification were carried out using Sequencher 4.10.1 (Gene Codes Corporation, Ann Arbor, MI, USA). Only the 10 exons of the *SIAE* gene were sequenced, that is, the sequencing primers did not cover the promoter or other regulatory regions for the gene. Data on *SIAE* sequences in 1351 Caucasian controls included in the multiple Exome Sequencing Project were obtained from the Exome Variant Server, NHLBI Exome Sequencing Project, Seattle, WA, USA (URL: http://evs.gs.washington.edu/EVS/) (accessed November 2011).

SIAE activity assay

To assay functionality of *SIAE* variants, site-directed mutagenesis with Pfu Turbo-based DNA polymerase (Stratagene, Santa Clara, CA, USA) was used to introduce selected single-nucleotide changes into full-length, C-terminal Flag-tagged human *SIAE* cDNA subcloned in the pcDNA3.1. vector. These cDNAs as well as cDNAs encoding wild-type and the catalytically inert SIAE were transiently transfected into HEK 293T cells, the cells then lysed in 1% NP40 in phosphate-buffered saline and the lysates as well as culture supernatants subdivided into two equal aliquots from which the esterase was immunoprecipitated using monoclonal anti-Flag antibody (Sigma, St Louis, MO, USA) and protein A Sepharose. Catalytic activities were assessed in one aliquot of the immunoprecipitated SIAE using a fluorogenic substrate, 4-methylumbelliferyl acetate (UBS), a FluoroskanII instrument (Finstrument) and Spectroscope software as described previously,[16] and the enzymatic activity levels then normalized to protein levels. Relative quantity of the Flag-tagged protein was assessed in the second aliquot as reported previously[16] using a quantitative immunoblot assay, an IR Dye 800CW-labeled goat anti-mouse immunoglobulin G secondary antibody and the LI-COR Odyssey imaging system (Lincoln, NE, USA). The assay was performed three times for each cDNA.

Statistical analyses

Individual samples with genotype call rates of <95% and SNPs with call rates of <95%, minor allele frequencies of <1%, or deviation from the Hardy–Weinberg equilibrium at $P<0.001$. (as assessed in controls), were removed from the analysis. In total 1450 cases, 2957 controls and 84 SNPs were included in the statistical analyses. PLINK software (v1.07) was used to assess allelic associations, R-script (http://www.rproject.org/) to generate figures and haploview software, and Haploview software, version 4.1 (http://www.broad.mit.edu/mpg/haploview), to evaluate linkage disequilibrium between pairs of SNPs. Haplotype block structure was defined according to the criteria established by Gabriel and the pairwise estimates of standardized Lewontin's disequilibrium coefficient (D′), whereas the linkage disequilibrium among pairs of SNPs was characterized according to the square of the correlation coefficient (r^2). Conditional logistic regression analyses of multiple markers were performed using SAS v9.13 (SAS Institute Inc., Cary, NC, USA) with subsequent power analyses conducted using CATs software (http://www.sph.umich.edu/csg/abecasis/CaTS/index.html) and an additive model. χ^2 or Fisher's exact testing was performed to demonstrate association for sequence data. For allelic association tests, the threshold for declaring significance in the fine mapping studies was $P<0.05/84$ ($<6\times10^{-4}$).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Acknowledgments

This work was supported by grants from the Canadian Institutes for Health Research (MOP74621), the Ontario Research Fund (RE01-061) and the Canadian Primary Biliary Cirrhosis Society (GMH and KAS), the US National Institutes of Health (RO1 DK80670), the American Gastroenterological Association and the AJ and Sigismunda Palumbo Charitable Trust (KNL) and Alliance for Lupus Research, the Center for the Study of Inflammatory Bowel Disease at MGH and the NIH (AI 064930, AI 076505 and AR 058481) to SP KAS holds the Sherman Family Chair in Genomic Medicine and a Canada Research Chair award. We thank the NHLBI GO Exome Sequencing Project and its ongoing studies, which produced and provided exome variant calls for comparison: the Lung GO Sequencing Project (HL-102923), the WHI Sequencing Project (HL-102924), the Broad GO Sequencing Project (HL-102925), the Seattle GO Sequencing Project (HL-102926) and the Heart GO Sequencing Project (HL-103010).

References

1. Lindor KD, Gershwin ME, Poupon R, Kaplan M, Bergasa NV, Heathcote EJ. Primary biliary cirrhosis. Hepatology. 2009; 50:291–308. [PubMed: 19554543]
2. Hirschfield GM, Heathcote EJ, Gershwin ME. Pathogenesis of cholestatic liver disease and therapeutic approaches. Gastroenterology. 2010; 139:1481–1496. [PubMed: 20849855]
3. Hirschfield GM, Gershwin ME. Primary biliary cirrhosis: one disease with many faces. Isr Med Assoc J. 2011; 13:55–59. [PubMed: 21446239]
4. Hirschfield GM, Invernizzi P. Progress in the genetics of primary biliary cirrhosis. Semin Liver Dis. 2011; 31:147–156. [PubMed: 21538281]
5. Hirschfield GM, Liu X, Xu C, Lu Y, Xie G, Gu X, et al. Primary biliary cirrhosis associated with HLA, IL12A, and IL12RB2 variants. N Engl J Med. 2009; 360:2544–2555. [PubMed: 19458352]
6. Hirschfield GM, Liu X, Han Y, Gorlov IP, Lu Y, Xu C, et al. Variants at IRF5-TNPO3, 17q12-21 and MMEL1 are associated with primary biliary cirrhosis. Nat Genet. 2010; 42:655–657. [PubMed: 20639879]
7. Liu X, Invernizzi P, Lu Y, Kosoy R, Bianchi I, Podda M, et al. Genome-wide meta-analyses identify three loci associated with primary biliary cirrhosis. Nat Genet. 2010; 42:658–660. [PubMed: 20639880]
8. 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. 2011; 43:329–332. [PubMed: 21399635]
9. Hakonarson H, Grant SF, Bradfield JP, Marchand L, Kim CE, Glessner JT, et al. A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. Nature. 2007; 448:591–594. [PubMed: 17632545]
10. Palmer DC, Restifo NP. Suppressors of cytokine signaling (SOCS) in T cell differentiation, maturation, and function. Trends Immunol. 2009; 30:592–602. [PubMed: 19879803]
11. Ramanathan S, Dubois S, Gagnon J, Leblanc C, Mariathasan S, Ferbeyre G, et al. Regulation of cytokine-driven functional differentiation of CD8T cells by suppressor of cytokine signaling 1 controls autoimmunity and preserves their proliferative capacity toward foreign antigens. J Immunol. 2010; 185:357–366. [PubMed: 20519645]
12. Dubois PC, Trynka G, Franke L, Hunt KA, Romanos J, Curtotti A, et al. Multiple common variants for celiac disease influencing immune gene expression. Nat Genet. 2010; 42:295–302. [PubMed: 20190752]
13. 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. N Engl J Med. 2007; 357:851–862. [PubMed: 17660530]
14. Ferreira RC, Pan-Hammarstrom Q, Graham RR, Gateva V, Fontan G, Lee AT, et al. Association of IFIH1 and other autoimmunity risk alleles with selective IgA deficiency. Nat Genet. 2010; 42:777–780. [PubMed: 20694011]
15. Zuvich RL, Bush WS, McCauley JL, Beecham AH, De Jager PL, Ivinson AJ, et al. Interrogating the complex role of chromosome 16p13. 13 in multiple sclerosis susceptibility: Independent genetic signals in the CIITA-CLEC16A-SOCS1 gene complex. Hum Mol Genet. 2011; 20:3517–3524. [PubMed: 21653641]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

16. Surolia I, Pirnie SP, Chellappa V, Taylor KN, Cariappa A, Moya J, et al. Functionally defective germline variants of sialic acid acetylesterase in autoimmunity. Nature. 2010; 466:243–247. [PubMed: 20555325]

17. Cariappa A, Takematsu H, Liu H, Diaz S, Haider K, Boboila C, et al. B cell antigen receptor signal strength and peripheral B cell development are regulated by a 9-O-acetyl sialic acid esterase. J Exp Med. 2009; 206:125–138. [PubMed: 19103880]

18. Szymanski K, Skorka A, Szypowska A, Bednarczuk T, Ploski R. Functionally defective germline variant of sialic acid acetylesterase (Met89Val) is not associated with type 1 diabetes mellitus and Graves' disease in a Polish population. Tissue Antigens. 2011; 78:214–216. [PubMed: 21615338]

19. Fujimoto M, Naka T. SOCS1, a negative regulator of cytokine signals and TLR responses, in human liver diseases. Gastroenterol Res Pract. 201010.1155/2010/470468

20. Alexander WS, Starr R, Fenner JE, Scott CL, Handman E, Sprigg NS, et al. SOCS1 is a critical inhibitor of interferon gamma signaling and prevents the potentially fatal neonatal actions of this cytokine. Cell. 1999; 98:597–608. [PubMed: 10490099]

21. Marine JC, Topham DJ, McKay C, Wang D, Parganas E, Stravopodis D, et al. SOCS1 deficiency causes a lymphocyte-dependent perinatal lethality. Cell. 1999; 98:609–616. [PubMed: 10490100]

22. Torisu T, Nakaya M, Watanabe S, Hashimoto M, Yoshida H, Chinen T, et al. Suppressor of cytokine signaling 1 protects mice against concanavalin A-induced hepatitis by inhibiting apoptosis. Hepatology. 2008; 47:1644–1654. [PubMed: 18393318]

23. Tanaka K, Ichiyama K, Hashimoto M, Yoshida H, Takimoto T, Takaesu G, et al. Loss of suppressor of cytokine signaling 1 in helper T cells leads to defective Th17 differentiation by enhancing antagonistic effects of IFN-gamma on STAT3 and Smads. J Immunol. 2008; 180:3746–3756. [PubMed: 18322180]

24. DeKoter RP, Geadah M, Khoosal S, Xu LS, Thillainadesan G, Torchia J, et al. Regulation of follicular B cell differentiation by the related E26 transformation-specific transcription factors PU. 1, Spi-B, and Spi-C. J Immunol. 2010; 185:7374–7384. [PubMed: 21057087]

25. Nagasawa M, Schmidlin H, Hazekamp MG, Schotte R, Blom B. Development of human plasmacytoid dendritic cells depends on the combined action of the basic helix-loop-helix factor E2-2 and the Ets factor Spi-B. Eur J Immunol. 2008; 38:2389–2400. [PubMed: 18792017]

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

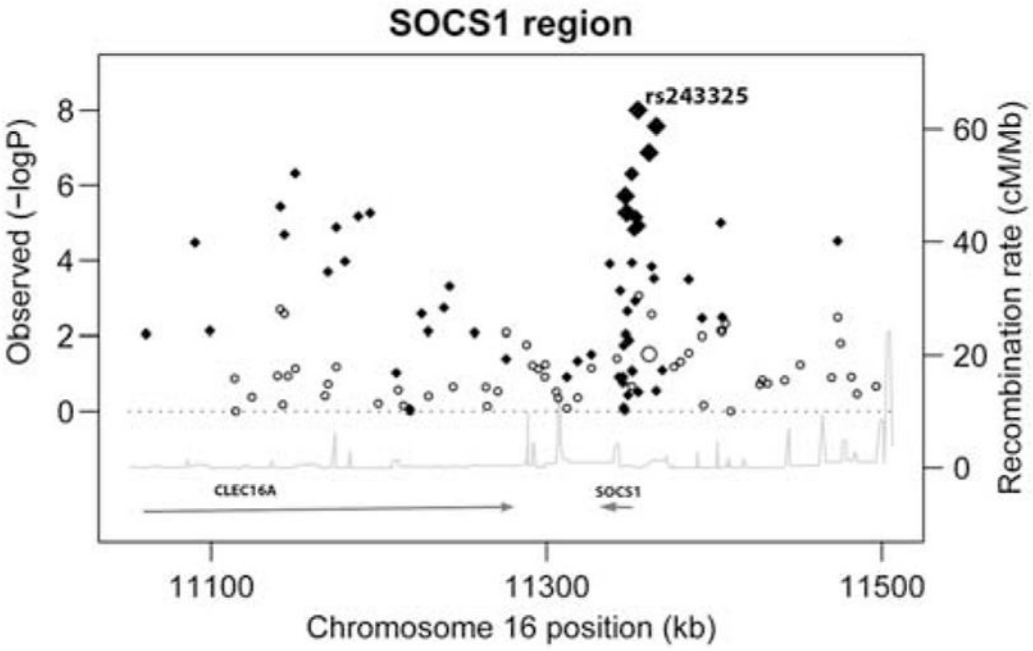

Figure 1.

Association plots for the *CLEC16A-SOCS1* locus. Strength of the associations and recombination rates estimated from HapMap data for genotyped SNPs are shown for the *SOCS1* loci. Both current fine-mapping data (diamonds) and genome-wide association data from a prior study of a subset of this cohort (circles) are shown. The extent of linkage disequilibrium with the most significant polymorphisms is indicated by the size of each data point; larger data points indicate stronger linkage disequilibrium. Chromosomal positions for each gene region are indicated by the arrows, with the arrow direction representing the orientation of translation. Linkage disequilibrium was calculated using observed data in PLINK.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

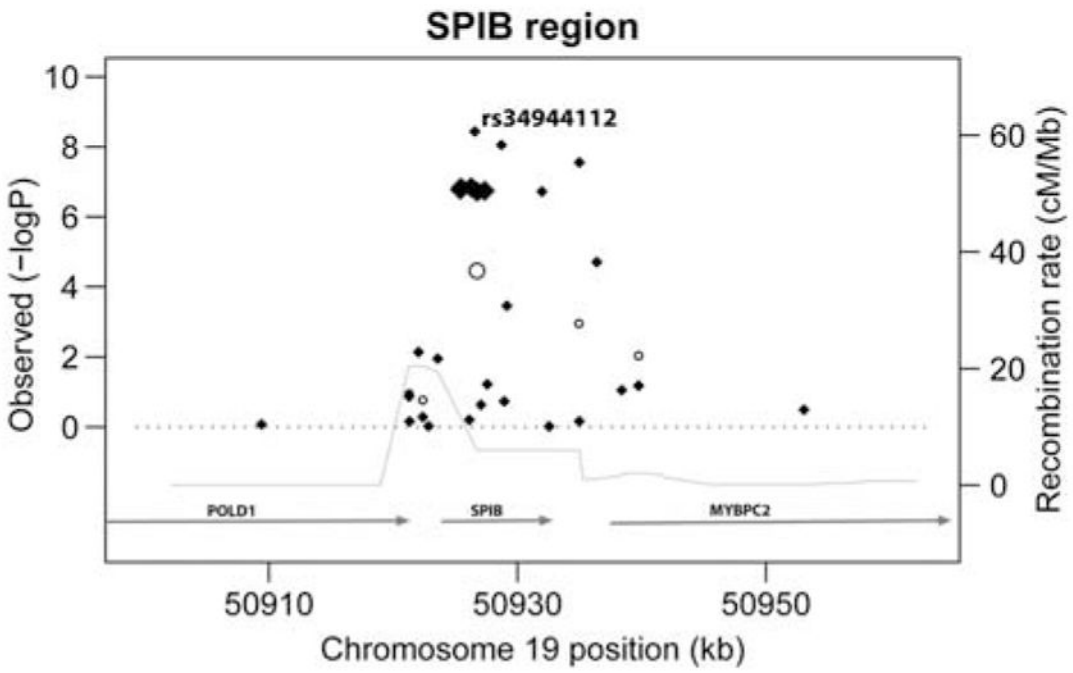

Figure 2.

Association plots for the *SPIB* locus. Strength of the associations and recombination rates estimated from HapMap data for genotyped SNPs are shown for the *SPIB* locus. Both fine-mapping data from this study (diamonds) and genome-wide association data from a prior study of a subset of this cohort (circles) are shown. The extent of linkage disequilibrium with the most significant polymorphisms is indicated by the size of each data point; larger data points indicate stronger linkage disequilibrium. Gene positions for each gene region are indicated by the arrows, with the arrow direction representing the orientation of translation. Linkage disequilibrium was calculated using observed data in PLINK.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

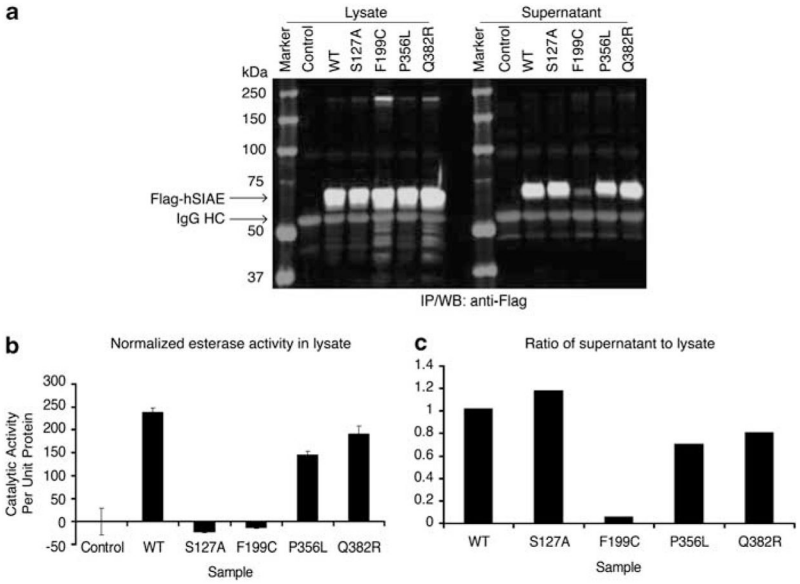

Figure 3.

Functional analysis of newly identified SIAE variants. *SIAE* cDNAs encoding Flag-tagged versions of wild-type SIAE, a catalytically inert SIAE mutation (S127A) and three *SIAE* variants of unknown functional significance (F199C, P356L and Q382R) were expressed in 293 cells, the cell lysates and culture supernatants harvested 48 h later and immunoprecipitated with anti-Flag antibody ('control' refers to untransfected cells). Immunoprecipitates were split into two equivalent aliquots with one used in immunoblotting to quantitate SIAE levels and the second used to assay for esterase activity on a fluorogenic substrate. Anti-Flag antibody immunoblotting analysis of cell lysates and supernatants from 293 transfected cells is shown in the upper panel. Bottom left panel shows esterase activity of each *SIAE* species normalized for lysate *SIAE* protein levels and bottom right panel shows ratios of *SIAE* levels in culture supernatant relative to lysate. Each variant was tested three times and the data are shown as mean±s.e.m.[16]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 1

Variants reaching significance across the *CLEC16A-SOCS1* locus

SNP	Position	Risk allele	RAF		OR (95% CI)	*P*-value	*Gene*
			Case	Control			
rs16957895	11090411	T	0.738	0.695	1.24 (1.12–1.37)	3.37×10^{-5}	*CLEC16A*[a]
rs16957976	11141321	C	0.127	0.094	1.39 (1.21–1.61)	3.71×10^{-6}	*CLEC16A*
rs876457	11143726	A	0.090	0.065	1.43 (1.21–1.69)	2.05×10^{-5}	*CLEC16A*
rs58102322	11150140	A	0.116	0.083	1.46 (1.26–1.69)	4.85×10^{-7}	*CLEC16A*
rs725613	11169683	A	0.670	0.629	1.20 (1.09–1.31)	2.05×10^{-4}	*CLEC16A*
rs2041670	11174652	C	0.710	0.663	1.24 (1.13–1.37)	1.32×10^{-5}	*CLEC16A*
rs12708716	11179873	A	0.672	0.630	1.20 (1.10–1.32)	1.07×10^{-4}	*CLEC16A*
rs12924729	11187783	G	0.707	0.658	1.25 (1.13–1.38)	6.73×10^{-6}	*CLEC16A*
rs12927355	11194771	C	0.712	0.663	1.25 (1.14–1.38)	5.48×10^{-6}	*CLEC16A*
rs12927046	11242191	G	0.922	0.900	1.34 (1.14–1.57)	4.95×10^{-4}	*CLEC16A*
rs11074956	11337721	G	0.614	0.570	1.19 (1.09–1.31)	1.23×10^{-4}	*Intergenic*
rs1111186	11347048	A	0.729	0.678	1.27 (1.15–1.40)	1.94×10^{-6}	*SOCS1*[b]
rs4780355	11347858	T	0.729	0.680	1.25 (1.14–1.38)	5.44×10^{-6}	*SOCS1*
rs33989964	11350813	CA[c]	0.757	0.705	1.30 (1.17–1.44)	4.98×10^{-7}	*SOCS1*
rs243330	11350991	G	0.570	0.526	1.19 (1.09–1.30)	1.17×10^{-4}	*SOCS1*
rs243329	11352313	T	0.563	0.514	1.22 (1.11–1.33)	1.50×10^{-5}	*SOCS1*
rs243327	11353304	T	0.564	0.513	1.23 (1.12–1.34)	7.06×10^{-6}	*SOCS1*
rs243325	11354497	A	0.730	0.668	1.33(1.21–1.47)	9.91×10^{-9}	*SOCS1*
rs243324	11354970	T	0.562	0.511	1.22 (1.12–1.34)	1.20×10^{-5}	*SOCS1*
rs243323	11361202	A	0.737	0.681	1.30 (1.18–1.44)	1.35×10^{-7}	*TNP2*
rs11640138	11362729	A	0.574	0.530	1.19 (1.09–1.30)	1.47×10^{-4}	*TNP2*
rs416603	11364079	A	0.583	0.541	1.18 (1.08–1.29)	3.09×10^{-4}	*TNP2*
rs367569	11365500	C	0.759	0.702	1.34 (1.21–1.48)	2.67×10^{-8}	*PRM3*
rs181694	11384829	G	0.831	0.799	1.24 (1.10–1.40)	3.25×10^{-4}	*PRM1*
rs12928822	11403893	C	0.863	0.826	1.34 (1.17–1.52)	9.97×10^{-6}	*Intergenic*
rs7203055	11473656	T	0.677	0.631	1.22 (1.11–1.35)	3.05×10^{-5}	*Intergenic*

Genes Immun. Author manuscript; available in PMC 2012 June 02.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Abbreviations: CI, confidence interval; OR, odds ratio; RAF, risk allele frequency; SNP, single-nucleotide polymorphism. A total of 1450 cases and 2957 controls were successfully genotyped; only variants reaching the pre-assigned significance level of $< 6 \times 10^{-4}$ are shown (see Supplementary Table 1 for full data set).

[a] All *CLEC16A* SNPs map within the gene's introns.

[b] All *SOCS1* SNPs map within 5 kb upstream or downstream of the *SOCS1* coding sequence.

[c] CA: deletion/insertion variation: –/CA.

Genes Immun. Author manuscript; available in PMC 2012 June 02.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 2

Haplotype and logistic regression association results for three independent SNPs at the *CLEC16A-SOCS1* locus, A (haplotype analysis), B (logistic regression analysis).

Haplotype	Frequency		*P*-value	OR
	Cases	Controls		
A				
OMNIBUS	NA	NA	2.61×10^{-14}	NA
GGA	0.068	0.104	$\mathbf{3.93 \times 10^{-10}}$	**0.52**
AGA	0.222	0.235	1.26×10^{-1}	0.91
GAG	0.025	0.021	1.60×10^{-1}	1.31
AAG	0.088	0.058	$\mathbf{5.36 \times 10^{-8}}$	**1.67**
GGG	0.176	0.206	3.58×10^{-4}	0.80
AGG	0.421	0.375	1.42×10^{-5}	1.24

Carriage of rs12924729 risk allele G	Carriage of rs58102322 risk allele A	Carriage of rs243325 risk allele A	Controls no. (%)	PBC patients no. (%)	OR (95% CI)	*P*-value
B						
AA	GG	GG	34 (1.2)	7 (0.5)	Reference	–
AA	GG	GA	147 (5.0)	34 (2.4)	1.12 (0.46–2.75)	0.80
AA	GG	AA	168 (5.8)	67 (4.8)	1.94 (0.82–4.58)	0.13
AA	AA+GA	GG	1 (0.0)	0 (0.0)	0.00	0.96
AA	AA+GA	GA	2 (0.1)	2 (0.1)	4.86 (0.58–40.6)	0.14
AA	AA+GA	AA	3 (0.1)	1 (0.1)	1.62 (0.15–17.9)	0.69
GG+GA	GG	GG	255 (8.8)	87 (6.2)	1.66 (0.71–3.87)	0.24
GG+GA	GG	GA	931 (32.0)	383 (27.4)	2.00 (0.88–4.55)	0.10
GG+GA	GG	AA	911 (31.3)	518 (37.1)	2.76 (1.22–6.27)	0.02
GG+GA	AA+GA	GG	43 (1.5)	23 (1.6)	2.60 (1.00–6.77)	0.05
GG+GA	AA+GA	GA	193 (6.6)	106 (7.6)	2.67 (1.14–6.22)	0.02
GG+GA	AA+GA	AA	223 (7.7)	168 (12.0)	3.66 (1.58–8.46)	0.002

Abbreviations: CI, confidence interval; OR, odds ratio; NA, not applicable; PBC, primary biliary cirrhosis; RAF, risk allele frequency; SNP, single-nucleotide polymorphism. (A) OMNIBUS and haplotype analysis were performed using PLINK (http://pngu.mgh.harvard.edu/purcell/plink). Three allele haplotype includes (in order) SNPs: rs243325, rs58102322 and rs12924729. The most significant risk (AAG) and protective (GGA) haplotype associations are shown in bold. (B) Association study of all possible gene combinations of the three independent PBC-associated variants at the *CLEC16A-SOCS1* locus. For calculation of OR, the reference was the allele combination with the lowest effect on risk.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 3

Significant variants associated with risk of PBC at the *SPIB* locus

SNP	Base pair	Risk allele	RAF		OR (95% CI)	*P*-value	Gene
			Case	Control			
rs62116060	50925395	G	0.303	0.249	1.30 (1.18–.44)	1.70×10^{-7}	*SPIB*
rs62116061	50925444	T	0.302	0.248	1.30 (1.18–1.44)	1.50×10^{-7}	*SPIB*
rs11546996	50926265	G	0.301	0.247	1.30 (1.18–1.44)	1.46×10^{-7}	*SPIB*[a]
rs34944112	50926565	T	0.149	0.105	1.49 (1.31–1.70)	3.65×10^{-9}	*SPIB*
rs3745516	50926742	A	0.300	0.247	1.30 (1.18–1.43)	1.93×10^{-7}	*SPIB*
rs8108811	50927358	G	0.303	0.249	1.30 (1.18–1.43)	1.82×10^{-7}	*SPIB*
rs34795041	50928711	C	0.146	0.103	1.48 (1.30–1.70)	8.85×10^{-9}	*SPIB*
rs56235776	50929142	G	0.395	0.355	1.18 (1.08–1.30)	3.48×10^{-4}	*SPIB*
rs1137895	50931964	C	0.266	0.215	1.31 (1.19–1.46)	1.91×10^{-7}	*SPIB*[b]
rs3810277	50934978	T	0.187	0.141	1.41 (1.25–1.58)	2.84×10^{-8}	*SPIB*[b]
rs3765069	50936378	A	0.468	0.419	1.22 (1.11–1.33)	1.97×10^{-7}	*MYBPC2*

Abbreviations: CI, confidence interval; OR, odds ratio; PBC, primary biliary cirrhosis; RAF, risk allele frequency; SNP, single-nucleotide polymorphism; UTR, untranslated region. A total of 1450 cases and 2957 controls were successfully genotyped; variants reaching the pre-assigned significance threshold of $<6 \times 10^{-4}$ are reported here (see Supplementary Table 3 for full data set).

[a]This risk variant encodes an alanine to proline substitution in *SPIB* exon.

[b]These variants map within the 3′UTR/intergenic regions of the *SPIB* gene. All other *SPIB* variants listed map with introns of the gene.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 4

SIAE sequence variants in patients with primary biliary cirrhosis

Genomic position (bp)	SNP	Amino-acid change	Mutation frequency		
			Cases (n =381)	Controls[a] Surolia et al.[16] (n =648)	NHLBI (n =1351)
124530664	265A>G	M89V/M89V	2	0	74
124519641	596T>G	F199C[b]	3	0	0
124517292	935C>T	T312M[c]	1	1	2
124509663	1067C>T	P356L[b]	1	0	0
124508703	1145A>G	Q382R[b]	1	0	0
124508547	1211T>C	F404S[c]	2	0	0

Abbreviations: ESP, Exome Sequencing Project; SIAE, sialic acid acetylesterase; SNP, single-nucleotide polymorphism.

[a] All control data were taken from historic controls whose *SIAE* gene was either sequenced specifically by Surolia *et al.*[16] or in the context of the NHLBI ESP, Seattle, WA, USA (URL: http://evs.gs.washington.edu/EVS/) (November 2011 accessed).

[b] Variants newly identified in this study.

[c] Previously identified catalytically defective *SIAE* variants.