UCSF

UC San Francisco Previously Published Works

Title

The cumulative effects of known susceptibility variants to predict primary biliary cirrhosis risk

Permalink

https://escholarship.org/uc/item/6183z52k

Journal

Genes & Immunity, 16(3)

ISSN

1466-4879

Authors

Tang, R
Chen, H
Miao, Q
et al.

Publication Date

2015-04-01

DOI

10.1038/gene.2014.76

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

HHS Public Access

Author manuscript

Genes Immun. Author manuscript; available in PMC 2017 August 11.

Published in final edited form as:

Genes Immun. 2015 ; 16(3): 193–198. doi:10.1038/gene.2014.76.

The Cumulative Effects of Known Susceptibility Variants to Predict Primary Biliary Cirrhosis Risk

Ruqi Tang[a,1], **Haoyan Chen**[a,b,1], **Qi Miao**[a], **Zhaolian Bian**[a], **Wansu Ma**[c], **Xiaomei Feng**[d], **Michael F. Seldin**[e,f], **Pietro Invernizzi**[g], **M. Eric Gershwin**[f], **Wilson Liao**[b], and **Xiong Ma**[a]

[a]State Key Laboratory for Oncogenes and Related Genes, Division of Gastroenterology and Hepatology, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai Cancer Institute, Shanghai Institute of Digestive Disease, China

[b]Department of Dermatology, University of California San Francisco, San Francisco, CA, USA

[c]Department of Anesthesia and Perioperative Care, University of California, San Francisco, San Francisco, CA, USA

[d]Division of Gastroenterology, The Second Xiangya Hospital of Central South University, Changsha, China

[e]Department of Biochemistry and Molecular Medicine, University of California at Davis, Davis, CA, USA

[f]Department of Medicine, Division of Rheumatology, Allergy and Clinical Immunology, University of California at Davis, Davis, CA, USA

[g]Clinical Immunology, IRCCS Istituto Clinico Humanitas, Rozzano, Milan, Italy

Abstract

Multiple genetic variants influence the risk for development of primary biliary cirrhosis (PBC). To explore the cumulative effects of known susceptibility loci on risk, we utilized a weighted genetic risk score (wGRS) to evaluate whether genetic information can predict susceptibility. The wGRS was created using twenty-six known susceptibility loci and investigated in 1840 UK PBC and 5164 controls. Our data indicate that the wGRS was significantly different between PBC and controls (*P*=1.61E-142). Moreover, we assessed predictive performance of wGRS on disease status by calculating the area under the receiver operator characteristic curve (AUC). The AUC for the purely genetic model was 0.72 and for gender plus genetic model was 0.82, with confidence limits

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

Correspondence to: M. Eric Gershwin, M.D., Division of Rheumatology, Allergy and Clinical Immunology, University of California at Davis School of Medicine, 451 Health Sciences Drive, Suite 6510, Davis, CA 95616; telephone: 530-752-2884; fax: 530-752-4669; megershwin@ucdavis.edu.

[1] These authors contributed equally to the work.

The Italian GWAS dataset used in this manuscript was obtained from the database of Genotypes and Phenotypes (dbGaP) found at http://www.ncbi.nlm.nih.gov/gap through dbGaP accession number phs000444.v1.p1. This study makes use of data generated by the Wellcome Trust Case Control Consortium. A full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk.

Conflicts of interest

The authors declare no financial conflict of interest.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

substantially above random predictions. The risk of PBC using logistic regression was estimated after dividing individuals into quartiles. Individuals in the highest risk group disclosed demonstrated a substantially increased risk for PBC, compared to the lowest risk group (OR: 9.3, *P*=1.91E-084). Finally, we validated our findings in an analysis of an Italian PBC cohort. Our data suggested that the wGRS utilizing genetic variants was significantly associated with increased risk for PBC with consistent discriminant ability. Our study is a first step toward risk prediction for PBC.

Keywords

Primary biliary cirrhosis (PBC); weighted genetic risk score (wGRS); genome-wide association study (GWAS); area under the curve (AUC); susceptibility

Introduction

Primary biliary cirrhosis (PBC) is the most common autoimmune liver disease, characterized by chronic nonsuppurative destructive cholangitis and the presence of anti-mitochondrial antibody, eventually resulting in cirrhosis and liver failure[1]. Although the etiology of PBC remains unclear, there has been increasing evidence that genetic, epigenetic and environmental factors contribute to the risk for breaking tolerance[2].

PBC is highly heritable, with a sibling relative risk of ~10, but the underlying genetics is complex[3]. Over the past number of years, major efforts have been made to uncover genetic causes of PBC, aiming to elucidate disease pathogenesis and improve therapeutic interventions. Before the advent of genome wide association studies, the risk loci within the major histocompatibility complex (MHC) were the only well-established candidates associated with PBC[4–7]. Recently, genome-wide association studies (GWAS) and related study designs not only confirmed the predominant role of the HLA region, but also identified additional 27 risk loci outside the MHC[8–13]. While these findings have considerably expanded our understanding of the genetic architecture of PBC, a gap remains between genetic information and clinical utility. Moreover, the associated variants have relatively modest effects on disease risk and may not be the disease-causing variants, and may instead be the proxies for them. A genetic risk score (GRS), which combines effects of multiple loci, has been demonstrated to be able to predict disease risk better than single loci in several autoimmune diseases[14–18]. Therefore, the GRS might become a prospective means of translating GWAS findings into clinical practice.

In this study, we utilized a weighted genetic risk score (wGRS) of twenty-six known risk loci identified in GWAS of PBC. We applied our model to two independent cohorts to evaluate the cumulative effects of known variants to the disease susceptibility, and to investigate how well it can discriminate between PBC cases and controls. We also demonstrate that inclusion of non-genetic factors in the model can markedly improve the discriminatory capacity.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Results

Association between selected SNPs and PBC

Twenty-six known SNPs identified at genome-wide level of significance were investigated in the discovery cohort. All the loci were significantly associated with PBC and shared the same risk allele as in the original studies. The SNP rs7774434 in the HLA region showed the strongest association (*P*=1.57E-026), whereas rs34536443 had the largest effect (OR=0.52, 95% CI: 0.39–0.70), albeit with very low minor allele frequency (1.9% in cases and 3.4% in controls). Table 1 lists the details of allele frequencies and statistical analysis results for each SNP. The ORs were used in the subsequent wGRS analysis.

Evaluation of the weight genetic risk score

We created a weighted genetic risk score utilizing 26 disease susceptibility loci, to assess the polygenic effect of PBC. For each individual, a wGRS was calculated to evaluate the risk of the disease. We compared the distribution of wGRS between PBC cases and controls and found PBC patients had a significantly greater risk score than normal controls (median 7.85 versus 7.16, Wilcoxon test *P* =1.61E-142, Figure 1).

To assess the ability of wGRS to discriminate correctly between cases and controls, we used receiving operating characteristic (ROC) curves and calculated area under curve (AUC). The wGRS composed of 26 SNPs showed good ability to identify individuals who are at the increased risk for developing PBC (AUC: 0.72, 95% CI: 0.706–0.735, Table 2). Comparing the wGRS model including all 26 SNPs and the model without the HLA tag SNP rs7774434, we found that removal of HLA SNP from the wGRS only slightly decreased the AUC to 0.70 (95% CI: 0.689–0.719). However, the AUC for rs7774434 alone decreased to 0.58 (95% CI: 0.566–0.597, Table 2). When taking gender into account, the AUC for 26-locus wGRS markedly increased to 0.82 (95% CI: 0.813–0.833, Figure 3).

To explore the effect of wGRS in more detail, we divided the wGRS into quartiles. The odds ratio increased with the increasing wGRS groups, using the first quartile as reference. Individuals in the top quartile have a 9.31 times higher risk than those in the reference group (95% CI: 7.44–11.66, P = 1.91E-84, Table 3).

Validation of wGRS in independent samples

To evaluate whether the genetic risk model is applicable to an independent population, we next examined the wGRS in an Italian cohort, which was generated by weighting the same 26 SNPs using the natural log of ORs from the discovery data. In the replication data set, the wGRS index was also significantly higher in patients with PBC, with median score of 7.83, compared to controls with a median score of 7.15 (Wilcoxon test *P* =9.88E-29, Figure 2). There was little difference between the wGRS model containing all the SNPs and the model with only non-HLA SNPs (AUC: 0.72 versus 0.69, Table 2). The AUC including gender showed better discriminatory ability than the one with genetic information alone (AUC 0.81 versus 0.72, Figure 3). As before, the wGRS were divided into quartiles. The individuals in the fourth quartiles had significantly increased risk compared to the individuals in the reference group (OR: 8.14, 95% CI: 4.97–13.31, P=6.85E-017, Table 3).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Measurement of the genetic variance

To estimate the cumulative fraction of genetic variance explained by the SNPs included in the wGRS, we used the liability threshold model. We found that no single locus accounted for more than 1.0% of the genetic variance in either the discovery or validation set. The percentage of genetic variance explained by each locus in the discovery set is listed in Table 1. It was estimated that together the twenty-six SNPs could explain 5.3% of genetic susceptibility to PBC in the discovery set. The validation set (Supplemental Table 2) showed similar results with the same set of SNPs explaining 6.2% of genetic susceptibility.

Discussion

Here, we present the first genetic risk score for PBC in two large cohorts, using recently established genetic variants from GWAS and Immunochip studies. We found that the 26-locus wGRS was significantly associated with the increased risk of PBC, with the odds ratio of ~9.3 in the highest risk group when compared to the lowest risk group. The wGRS displayed good discrimination between individuals with PBC and healthy controls. In addition, taking gender into consideration allows better prediction of the risk than genetic factors alone. Notably, our findings in the discovery data set were replicated in an independent sample. Like many previous studies, we developed the genetic risk score model in the same population used to test the score, an approach that risks over-fitting the model and yielding inflated results[14, 19, 20]. Here, we addressed this issue by applying the model in an independent population. Using the ORs derived from the discovery data, we managed to minimize the potential influence of over-fitting in the replication sample. Thus, our wGRS model is likely to accurately reflect the discriminatory ability of the twenty-six genetic loci.

To our knowledge, this is the first study exploring the combined effects of the genetic variants for predicting PBC risk. However, several limitations need to be considered. First, some SNPs used in the model were imputed rather than directly typed, which may lead to less accurate results. However, we only employed the SNPs imputed with high confidence. As a result, two SNPs were excluded in the analysis. Given that these two SNPs were estimated to explain a very small proportion of the heritability, exclusion of these two SNPs is expected to have little effect on the overall performance of the model. Second, addition of gender in our model enhanced the discriminatory power, and other factors such as age, age at disease onset and family risk may also affect the risk of PBC. However, these clinical data were not available in our study and therefore were not incorporated. Third, our analysis was restricted to individuals of European ancestry. Since PBC is a genetically heterogeneous disorder, caution must be taken when the model is used in other ethnic populations. Fourth, these studies have presumably not used the true disease risk variants but rather SNP surrogates that are associated with disease risk. It is likely that when the actual disease risk variants are defined the model will perform substantially better. Fifth, we note that the identification of the 26 SNPs that influence disease risk occurred using the same data sets that have been previously used to build genetic models and this could inflate the AUC for both the initial and replication data sets, a potential weakness of the current analysis. Lastly,, calibration of the model was not assessed in the current study. A well-calibrated genetic risk

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

score, when applied to a new population, should be able to predict the risk close to the observed. However, such a population was not available.

The HLA region SNP rs7774434 showed the strongest association with PBC among the 26 SNPs evaluated in the current study. Nevertheless, the SNP rs7774434 only conferred relatively small increments in predictive capability of the wGRS model. It is likely that our model underestimated the HLA contribution since more than one independent risk loci have been demonstrated by further fine-mapping studies[12, 21] and hierarchical effects of specific HLA D region determinants have been implicated[5, 21]. However, additional studies using additional HLA SNPs as well as imputed HLA-DRB1* determinants only had marginal effects on the replication model (data not shown).

Although the genetic variants for PBC have been identified at an unprecedented pace with the application of high-throughput genotyping approaches, the genetic variance of PBC explained by the risk loci identified so far is only ~5%. Hence, there remain additional genetic risk factors in PBC yet to be identified. It has been speculated that rare variants with larger effect and copy number variations may account for the missing heritability[22]. Moreover, complementary mechanisms, in particular epigenetic modifications have been suggested to be involved in the etiology of PBC[23, 24]. Accordingly, we assume that the performance of the model might be improved as novel genetic factors are included. We also note that other information including biomarkers, not available in the current study, could potentially be added to the model.

We also note that it is not clear if the risk score developed for PBC is specific for PBC. Future studies will address whether the risk score for PBC is specific or whether this risk score overlaps with several different autoimmune diseases including gluten sensitive enteropathy, systemic lupus erythematosus and rheumatoid arthritis. We should also note that it is premature for a clinician to consider genotyping a patient who may show signs suggestive of PBC. PBC, although clearly well-defined at the level of immunopathology and clinical presentation, has still remained elusive with respect to etiology and new therapies despite intensive genetic, molecular, immunological analysis and use of murine models[25–30]. Clearly any future model must take into account environmental risks, further definition of whether the risk score is specific for PBC or a generic autoimmune risk, and ultimately what a risk analysis would mean for a family member with PBC[31].

We demonstrated that the wGRS comprising 26 PBC risk SNPs were strongly associated with the susceptibility to PBC and displayed good ability to discriminate between cases and controls in two independent samples. A better understanding of the genetic complexity of PBC, together with the environmental factors is warranted to improve the risk assessment of this disease.

Materials and Methods

Study populations

Our study included two independent genome-wide association study data sets: (1) a discovery data set of 1840 PBC cases and 5163 population controls drawn from Wellcome

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Trust Case Control Consortium (WTCCC) PBC GWAS[10]; (2) a validation data set of 453 cases and 936 controls from Italian PBC GWAS[9]. The discovery data was obtained from the WTCCC official website (http://www.wtccc.org.uk/). The Italian dataset was obtained from the database of Genotypes and Phenotypes (dbGaP, http://www.ncbi.nlm.nih.gov/gap) through accession number phs000444.v1.p1. Details about sample characteristics, genotyping and quality control can be found in the original papers.

SNP selection and quality control

We selected SNPs based on previous genome-wide association studies and Immunochip studies of PBC[32]. Twenty-eight SNPs including one SNP in the HLA region achieving the genome wide level of significance were selected. Of the twenty-eight SNPs, fifteen SNPs were genotyped directly and thirteen SNPs were imputed. All the genotyped SNPs had call rate>95%, Hardy Weinberg Equilibrium $P > 1.0E-06$ and minor allele frequency (MAF) >1%. The SNPs that were not directly genotyped were pre-phased by the program SHAPEIT and then imputed by IMPUTE2, using 1000 Genomes release 20101123 reference panel[33, 34]. Imputed genotypes with posterior probability <0.9 and SNPs with info score <0.8, MAF<1% were excluded. Two SNPs, rs3024921 (*STAT4*) and rs4979462 (*TNFSF15*), which showed low imputation confidence and for which no proxies were available, were excluded from the wGRS model. Therefore, twenty-six SNPs in total were taken forward for analysis. The genotyping and imputation rate for individual SNPs included in the analysis is summarized in Supplemental Table 1.

The subjects who had one or more missing genotypes for the 26 SNPs were removed from further analysis. Thus, 1512 cases and 4168 controls remained in the discovery dataset, and 325 cases and 662 controls remained in the validation set. Principal components analysis (PCA) implemented in EIGENSTRAT was used to correct for the population stratification in the samples[35]. 89,081 and 92,802 SNPs with low linkage disequilibrium ($r^2<0.2$) were used for PCA in the discovery data and the replication data, respectively. The principal components were selected based on leveling off the PCA screen plots. The first four principal components were used as covariates in the discovery cohort, whereas the first principal component was included as covariate in the validation dataset.

Genetic variance explained

A liability threshold model was used to estimate the genetic variance explained by the 26 risk SNPs[36], assuming a disease prevalence of 0.04% in the general population[37]. Risk allele frequencies in the control population and an effect size corresponding to OR values derived from our analysis were used to calculate thresholds for each SNP.

Statistical Analysis

We calculated the wGRS as the weighted sum of risk allele counts, where the weight for each SNP was the natural log of the OR. The OR for each SNP was derived from the discovery dataset. We generated a wGRS_all that included all 26 SNPs, a wGRS_noHLA that included 25 SNPs without the HLA region SNP rs7774434, and a wGRS_HLA that only included rs7774434. The difference in the distribution of wGRS between PBC cases and controls was tested using non-parametric Mann-Whitney test. To examine the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

discriminant ability of the wGRS, we plotted receiving operating characteristics (ROC) curves and calculated the area under the curve (AUC). An AUC = 1.0 represents a perfect model, while an AUC of 0.5 represents a random model. In addition, we divided the combined cases and control subjects into quartiles based on the wGRS with 26 SNPs. Logistic regression test was carried out to compare the individuals in the second to fourth quartiles to the individuals in the first quartile. Logistic regression, OR estimation, Mann-Whitney test and AUC calculation of the wGRS were conducted in R software v2.15 (http://www.r-project.org/)[38]. Single SNP association tests in discovery dataset were performed using logistic regression analysis implemented in PLINK v1.07 (http://pngu.mgh.harvard.edu/~purcell/plink/)[39]. All the association tests were adjusted by the principal components and sex.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We are grateful to all the subjects of this study for their participation. Funding for the project was provided by the Wellcome Trust under award 076113. The study was sponsored by Foundation for the Author of National Excellent Doctoral Dissertation of PR China (201325 to R.T), and awards from the National Natural Science Foundation of China (81170380 and 81325002 to X.M). This work was supported in part by grants from the National Institutes of Health (R01AR065174 and K08AR057763 to W.L. and R01DK091823 to M.E.G and M.F.S).

References

1. Hirschfield GM, Gershwin ME. The immunobiology and pathophysiology of primary biliary cirrhosis. Annu Rev Pathol. 2013; 8:303–330. [PubMed: 23347352]
2. Hirschfield GM, Invernizzi P. Progress in the genetics of primary biliary cirrhosis. Semin Liver Dis. 2011; 31(2):147–156. [PubMed: 21538281]
3. Jones DE, Watt FE, Metcalf JV, Bassendine MF, James OF. Familial primary biliary cirrhosis reassessed: a geographically-based population study. J Hepatol. 1999; 30(3):402–407. [PubMed: 10190721]
4. Umemura T, Joshita S, Ichijo T, Yoshizawa K, Katsuyama Y, Tanaka E, et al. Human leukocyte antigen class II molecules confer both susceptibility and progression in Japanese patients with primary biliary cirrhosis. Hepatology. 2012; 55(2):506–511. [PubMed: 21953406]
5. Invernizzi P, Selmi C, Poli F, Frison S, Floreani A, Alvaro D, et al. Human leukocyte antigen polymorphisms in Italian primary biliary cirrhosis: a multicenter study of 664 patients and 1992 healthy controls. Hepatology. 2008; 48(6):1906–1912. [PubMed: 19003916]
6. Donaldson PT, Baragiotta A, Heneghan MA, Floreani A, Venturi C, Underhill JA, et al. HLA class II alleles, genotypes, haplotypes, and amino acids in primary biliary cirrhosis: a large-scale study. Hepatology. 2006; 44(3):667–674. [PubMed: 16941709]
7. Invernizzi P. Human leukocyte antigen in primary biliary cirrhosis: an old story now reviving. Hepatology. 2011; 54(2):714–723. [PubMed: 21563204]
8. 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(24):2544–2555. [PubMed: 19458352]
9. 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(8):658–660. [PubMed: 20639880]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

10. 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(4): 329–332. [PubMed: 21399635]

11. Nakamura M, Nishida N, Kawashima M, Aiba Y, Tanaka A, Yasunami M, et al. Genome-wide association study identifies TNFSF15 and POU2AF1 as susceptibility loci for primary biliary cirrhosis in the Japanese population. Am J Hum Genet. 2012; 91(4):721–728. [PubMed: 23000144]

12. 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. 2012; 44(10):1137–1141. [PubMed: 22961000]

13. Juran BD, Hirschfield GM, Invernizzi P, Atkinson EJ, Li Y, Xie G, et al. Immunochip analyses identify a novel risk locus for primary biliary cirrhosis at 13q14, multiple independent associations at four established risk loci and epistasis between 1p31 and 7q32 risk variants. Hum Mol Genet. 2012; 21(23):5209–5221. [PubMed: 22936693]

14. Chen H, Poon A, Yeung C, Helms C, Pons J, Bowcock AM, et al. A genetic risk score combining ten psoriasis risk loci improves disease prediction. PLoS One. 2011; 6(4):e19454. [PubMed: 21559375]

15. Yarwood A, Han B, Raychaudhuri S, Bowes J, Lunt M, Pappas DA, et al. A weighted genetic risk score using all known susceptibility variants to estimate rheumatoid arthritis risk. Ann Rheum Dis. 2013

16. De Jager PL, Chibnik LB, Cui J, Reischl J, Lehr S, Simon KC, et al. Integration of genetic risk factors into a clinical algorithm for multiple sclerosis susceptibility: a weighted genetic risk score. Lancet Neurol. 2009; 8(12):1111–1119. [PubMed: 19879194]

17. Romanos J, Rosen A, Kumar V, Trynka G, Franke L, Szperl A, et al. Improving coeliac disease risk prediction by testing non-HLA variants additional to HLA variants. Gut. 2014; 63(3):415–422. [PubMed: 23704318]

18. Scott IC, Seegobin SD, Steer S, Tan R, Forabosco P, Hinks A, et al. Predicting the risk of rheumatoid arthritis and its age of onset through modelling genetic risk variants with smoking. PLoS Genet. 2013; 9(9):e1003808. [PubMed: 24068971]

19. Zhou XJ, Qi YY, Hou P, Lv JC, Shi SF, Liu LJ, et al. Cumulative Effects of Variants Identified by Genome-wide Association Studies in IgA Nephropathy. Sci Rep. 2014; 4:4904. [PubMed: 24811838]

20. Tam CH, Ho JS, Wang Y, Lam VK, Lee HM, Jiang G, et al. Use of net reclassification improvement (NRI) method confirms the utility of combined genetic risk score to predict type 2 diabetes. PLoS One. 2013; 8(12):e83093. [PubMed: 24376643]

21. Invernizzi P, Ransom M, Raychaudhuri S, Kosoy R, Lleo A, Shigeta R, et al. Classical HLA-DRB1 and DPB1 alleles account for HLA associations with primary biliary cirrhosis. Genes Immun. 2012; 13(6):461–468. [PubMed: 22573116]

22. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009; 461(7265):747–753. [PubMed: 19812666]

23. Lu Q. The critical importance of epigenetics in autoimmunity. Journal of autoimmunity. 2013; 41:1–5. [PubMed: 23375849]

24. Selmi C, Cavaciocchi F, Lleo A, Cheroni C, De Francesco R, Lombardi SA, et al. Genome-wide analysis of DNA methylation, copy number variation, and gene expression in monozygotic twins discordant for primary biliary cirrhosis. Front Immunol. 2014; 5:128. [PubMed: 24734033]

25. Chen RC, Naiyanetr P, Shu SA, Wang J, Yang GX, Kenny TP, et al. Antimitochondrial antibody heterogeneity and the xenobiotic etiology of primary biliary cirrhosis. Hepatology. 2013; 57(4): 1498–1508. [PubMed: 23184636]

26. Leung PS, Wang J, Naiyanetr P, Kenny TP, Lam KS, Kurth MJ, et al. Environment and primary biliary cirrhosis: electrophilic drugs and the induction of AMA. Journal of autoimmunity. 2013; 41:79–86. [PubMed: 23352659]

27. Huang W, Kachapati K, Adams D, Wu Y, Leung PS, Yang GX, et al. Murine autoimmune cholangitis requires two hits: cytotoxic KLRG1(+) CD8 effector cells and defective T regulatory cells. Journal of autoimmunity. 2014; 50:123–134. [PubMed: 24556277]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

28. Lleo A, Zhang W, McDonald WH, Seeley EH, Leung PS, Coppel RL, et al. Shotgun proteomics: Identification of unique protein profiles of apoptotic bodies from biliary epithelial cells. Hepatology. 2014
29. Liaskou E, Hirschfield GM, Gershwin ME. Mechanisms of tissue injury in autoimmune liver diseases. Seminars in immunopathology. 2014
30. Dhirapong A, Yang GX, Nadler S, Zhang W, Tsuneyama K, Leung P, et al. Therapeutic effect of cytotoxic T lymphocyte antigen 4/immunoglobulin on a murine model of primary biliary cirrhosis. Hepatology. 2013; 57(2):708–715. [PubMed: 22996325]
31. Bogdanos DP, Smyk DS, Rigopoulou EI, Mytilinaiou MG, Heneghan MA, Selmi C, et al. Twin studies in autoimmune disease: genetics, gender and environment. Journal of autoimmunity. 2012; 38(2–3):J156–J169. [PubMed: 22177232]
32. Mells GF, Kaser A, Karlsen TH. Novel insights into autoimmune liver diseases provided by genome-wide association studies. Journal of autoimmunity. 2013; 46:41–54. [PubMed: 23931959]
33. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods. 2013; 10(1):5–6. [PubMed: 23269371]
34. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009; 5(6):e1000529. [PubMed: 19543373]
35. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38(8): 904–909. [PubMed: 16862161]
36. So HC, Gui AH, Cherny SS, Sham PC. Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. Genet Epidemiol. 2011; 35(5):310–317. [PubMed: 21374718]
37. Boonstra K, Beuers U, Ponsioen CY. Epidemiology of primary sclerosing cholangitis and primary biliary cirrhosis: a systematic review. J Hepatol. 2012; 56(5):1181–1188. [PubMed: 22245904]
38. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011; 12:77. [PubMed: 21414208]
39. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81(3): 559–575. [PubMed: 17701901]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

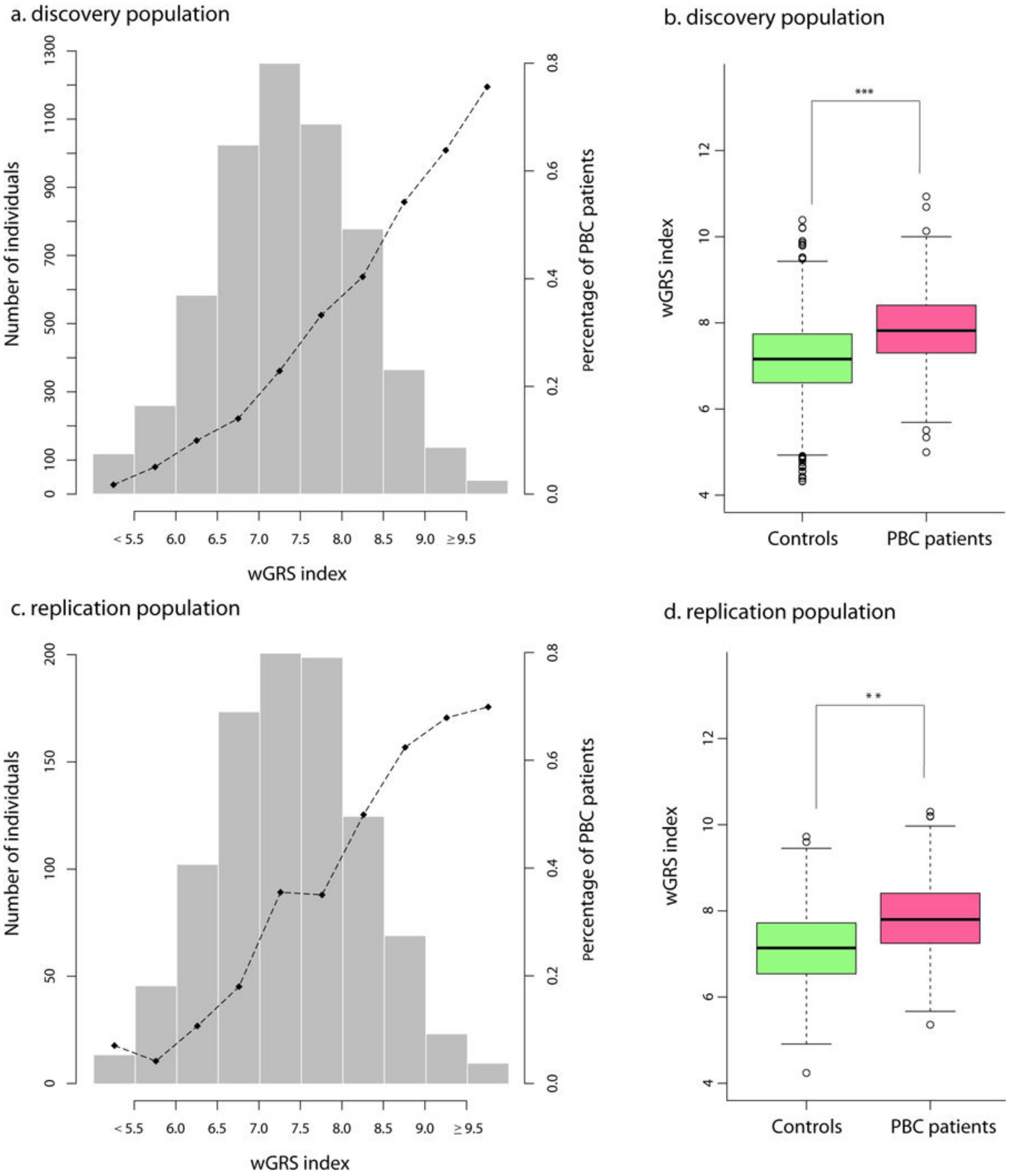

Figure 1.

Distributions of wGRS in discovery cohort (a&b); and in replication cohort (c&d). a, c: The histogram shows the distribution of wGRS for all individuals including PBC cases and healthy controls. Values smaller than 5.5 and greater than 9.5 were grouped. The black dots represent the percentage of PBC patients in the population of that bin (y axis on the right). b, d: PBC patients (red box) had a significantly higher wGRS than controls (green box) with p value < 1.0E-8. Boxes represent the 25th to 75th percentile across the wGRS; the median is shown as a thick line in the middle of the box; whiskers extend to values with 1.5 times the

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

difference between the 25th to 75th percentlies; and outliers are marked with circles. *** $P <$ 1E-32; ** $P <$ 1E-16

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

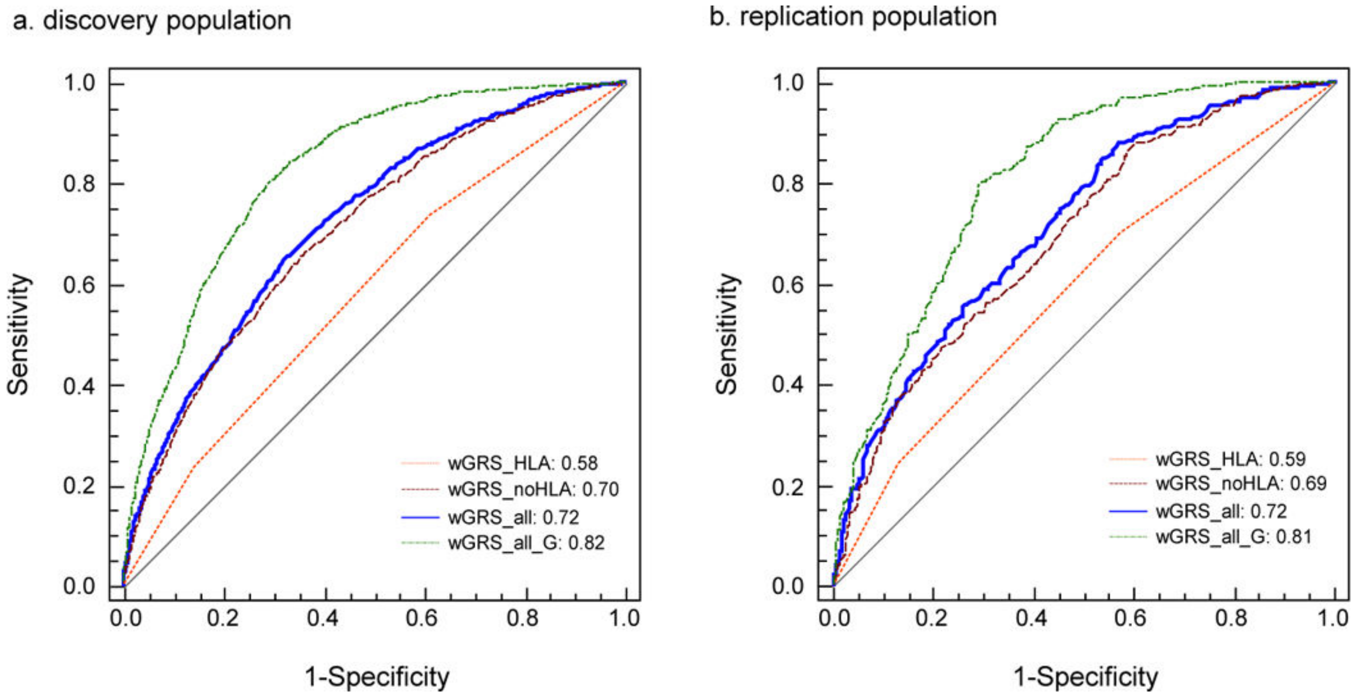

Figure 2.

ROC curves comparing wGRS in different models. wGRS_all includes 26 SNPs; wGRS_noHLA includes 25 non-HLA SNPs; wGRS_HLA includes the HLA tag SNP rs7774434; and wGRS_all_G includes 26 SNPs and gender in the model.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

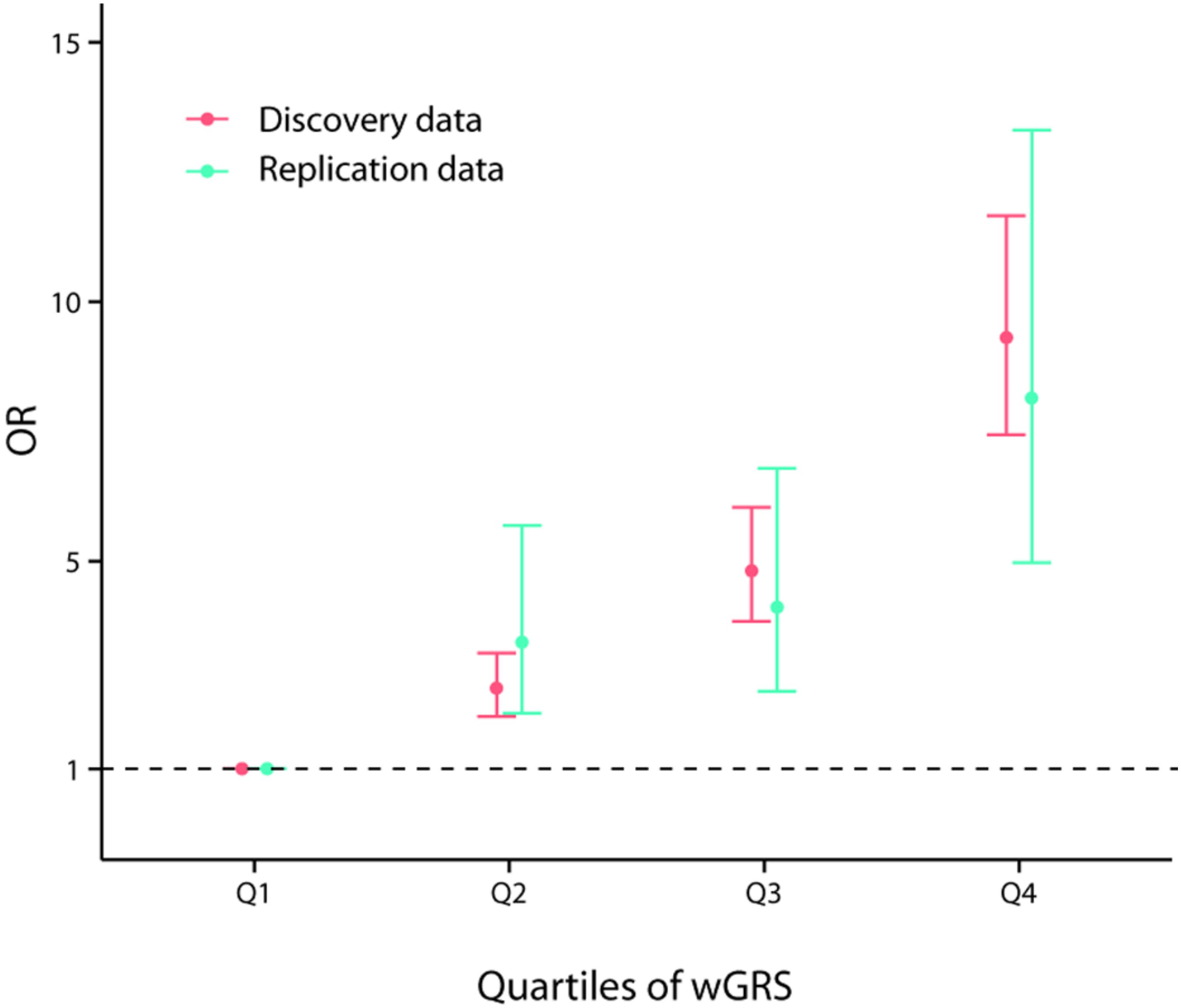

Figure 3.

PBC odds ratios of wGRS quartiles compared to the first quartile (reference). Red represents discovery data and green represents replication data. Vertical bars are 95% confidence intervals.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Associations of published risk SNPs for PBC in discovery sample

Chr	Candidate Gene	SNP	Allele (risk/non-risk)	Case frequency	Control frequency	OR (95%CI)	P-value	Genetic variance
1	MMEL1	rs3748816	G/A	0.39	0.36	1.15(1.05–1.25)	0.0017	0.07%
1	IL12RB2	rs72678531	C/T	0.24	0.17	1.63(1.47–1.81)	2.12E-020	0.54%
1	DENND1B	rs2488393	T/C	0.25	0.20	1.28(1.16–1.41)	9.22E-007	0.15%
3	PLCL2	rs1372072	A/G	0.40	0.37	1.13(1.04–1.23)	0.003773	0.05%
3	TIMMDC1	rs2293370	G/A	0.85	0.80	1.4(1.25–1.56)	5.45E-009	0.27%
7	TNPO3	rs35188261	A/G	0.16	0.11	1.55(1.37–1.74)	1.61E-012	0.28%
3	IL12A	rs2366643	T/C	0.64	0.57	1.35(1.24–1.47)	1.25E-011	0.33%
4	MANBA/NFKB1	rs7665090	G/A	0.57	0.52	1.2(1.11–1.3)	1.53E-005	0.13%
5	IL7R	rs6871748	T/C	0.78	0.72	1.34(1.22–1.48)	3.88E-009	0.26%
6	HLA region	rs7774434	G/A	0.49	0.38	1.58(1.46–1.72)	1.57E-026	0.76%
7	ELMO1	rs6974491	A/G	0.20	0.17	1.27(1.14–1.41)	7.86E-006	0.12%
11	RPS6KA	rs538147	G/A	0.65	0.61	1.19(1.09–1.29)	8.30E-005	0.11%
11	POU2AF	rs4938534	A/G	0.66	0.64	1.11(1.02–1.21)	0.01617	0.04%
11	CXCR5/ DDX6	rs80065107	T/C	0.84	0.79	1.33(1.19–1.49)	3.15E-007	0.20%
12	TNFRSF1A	rs1800693	G/A	0.45	0.40	1.25(1.15–1.36)	1.30E-007	0.19%
12	ATXN2/BRAP	rs11065979	T/C	0.48	0.44	1.17(1.08–1.27)	0.0002554	0.09%
13	TNFSF11	rs3862738	G/A	0.76	0.74	1.18(1.07–1.3)	0.0005927	0.08%
14	RAD51B	rs911263	A/G	0.76	0.71	1.29(1.18–1.42)	1.10E-007	0.21%
14	TNFAIP2	rs8017161	A/G	0.44	0.40	1.18(1.09–1.28)	9.24E-005	0.10%
16	CLEC16A	rs12708715	C/T	0.74	0.69	1.32(1.2–1.45)	5.05E-009	0.25%
16	IRF8	rs11117433	G/C	0.81	0.78	1.26(1.13–1.4)	2.07E-005	0.14%
17	ZPBP2 /ORMDL3	rs8067378	G/A	0.57	0.52	1.2(1.11–1.31)	9.00E-006	0.13%
17	MAPT	rs117220953	T/C	0.79	0.76	1.25(1.13–1.38)	9.07E-006	0.14%
19	TYK2	rs34536443	G/C	0.98	0.97	1.91(1.44–2.54)	9.04E-006	0.20%
19	SPIB	rs3745516	A/G	0.29	0.23	1.37(1.25–1.51)	5.06E-011	0.27%
22	SYNGR1	rs2267407	A/G	0.27	0.23	1.31(1.19–1.45)	4.34E-008	0.20%

The association test was performed using logistic regression adjusted for the first four principal components and gender.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

The association of each wGRS model with PBC

	Susceptible SNPs	HLA tag SNP	OR (95% CI)	AUC (95% CI)	Cases	Controls
Discovery dataset					1512	4168
wGRS_all	+	+	2.71(2.48–2.96)	0.72(0.706–0.735)		
wGRS_noHLA	+	−	2.70(2.46–2.97)	0.70(0.689–0.719)		
wGRS_HLA	−	+	2.54(2.07–3.11)	0.58(0.566–0.597)		
Replication dataset					325	662
wGRS_all	+	+	2.51(2.06–3.05)	0.72(0.685–0.751)		
wGRS_noHLA	+	−	2.38(1.95–2.90)	0.69(0.661–0.728)		
wGRS_HLA	−	+	2.19(1.40–3.40)	0.59(0.552–0.622)		

The ORs shown were calculated using the logistic regression testing the association of each wGRS with PBC. wGRS_all includes 26 SNPs; wGRS_noHLA includes 25 SNPs (without the HLA tag SNP) and wGRS_HLA only includes the HLA tag SNP.