Susceptibility to primary biliary cirrhosis (PBC) is strongly associated with human leukocyte antigen (HLA)-region polymorphisms. To determine if associations can be explained by classical HLA determinants, we studied Italian, 676 cases and 1440 controls, genotyped with dense single-nucleotide polymorphisms (SNPs) for which classical HLA alleles and amino acids were imputed. Although previous genome-wide association studies and our results show stronger SNP associations near DQB1, we demonstrate that the HLA signals can be attributed to classical DRB1 and DPB1 genes. Strong support for the predominant role of DRB1 is provided by our conditional analyses. We also demonstrate an independent association of DPB1. Specific HLA-DRB1 genotypes (*08, *11 and *14) account for most of the DRB1 association signal. Consistent with previous studies, DRB1*08 (P = 1.59 × 10^{-11}) was the strongest predisposing allele, whereas DRB1*11 (P = 1.42 × 10^{-10}) was protective. Additionally, DRB1*14 and the DPB1 association (DPB1*03:01; P = 9.18 × 10^{-7}) were predisposing risk alleles. No signal was observed in the HLA class 1 or class 3 regions. These findings better define the association of PBC with HLA and specifically support the role of classical HLA-DRB1 and DPB1 genes and alleles in susceptibility to PBC.

**Genes and Immunity** (2012) 13, 461–468; doi:10.1038/gene.2012.17; published online 10 May 2012

**Keywords:** genetic risk; risk allele; imputation; antigen-binding pocket; autoimmune disease

**INTRODUCTION**

The human major histocompatibility complex, human leukocyte antigen (HLA), has been implicated in the etiopathogenesis of primary biliary cirrhosis (PBC), similar to many other autoimmune diseases. Genome-wide association studies (GWAS) of PBC, including our own, find the strongest association with single-nucleotide polymorphisms (SNPs) within the HLA region.1–3 In these studies, the peak association signal is between HLA-DQA1 and HLA-DQB1. Multiple studies of PBC also show association with particular classical HLA alleles in PBC (reviewed in Invernizzi). These studies have variably implicated different DRB1 alleles in the European populations with most studies, including all larger cohorts showing association of DRB1*08.*5,6 Our previous studies in an Italian cohort with PBC showed the association of DRB1*08 as predisposing, and DRB1*11 and DRB1*13 as protective alleles.6 A study using a small cohort (32 German PBC cases and 47 controls) suggest that DPB1 associations may also be present in Europeans.7 However, a comprehensive study of the HLA region associations has not been performed, and like other autoimmune diseases, it is unclear which determinants are actually causally related to pathogenesis.

To further study the HLA associations in PBC, in the current study, we used the most recent advances in imputation algorithms and sequence information resources, including the 1000 genome database to accurately impute missing SNPs, and importantly, HLA classical alleles. Specifically, our investigation rests on recent development and resources for imputing HLA classical alleles, including a reference set of European subjects.8 For our study, we used an inference set of SNP genotypes from both GWAS and a designed chip array, the Immunochip,9 which contains a set of SNPs that have been used in multiple studies of HLA.10 We perform a series of conditioning analyses that clarify which HLA genes and alleles underlie the major component of the genetic associations of PBC.

**RESULTS**

Analyses show strong association of imputed SNPs and HLA determinants

To further define PBC–HLA region associations, we analyzed association using imputed genotypes with high probabilities and information scores (see Materials and methods). These studies...
utilized genotypes from both GWAS and Immunochip arrays that contained large numbers of SNPs in the major histocompatibility complex region (Table 1, Supplementary Table S1, and see Materials and methods). Strong association was observed with the peak association \( (P = 9.83 \times 10^{-17}) \), with rs115721871 at position 32653792, distal to DQB1 (Figure 1a, Table 2 and Supplementary Table S2). Although the strongest associations were with non-coding SNPs, multiple classical genes in HLA show strong association with PBC (Table 2). For the classical HLA genes, the strongest association was with DRB*08 \( \left( P = 1.59 \times 10^{-1}} \right) \). The DQB1*04:02 and DQA1*04:01 in tight linkage disequilibrium (LD) with DRB*08 \( \left( r^2 = 0.84 \right. \text{and} 0.89, \text{respectively} \) showed nearly equivalent signal \( \left( 1.38 \times 10^{-1} 1.90 \times 10^{-10}, \text{respectively} \right) \). Very strong association was also observed for DRB1*11 \( \left( P = 1.42 \times 10^{-10} \right) \) with a weaker association with the DQB1 allele (DQB1*03:01, \( P = 6.10 \times 10^{-10} \)) that is in LD \( \left( r^2 = 0.75 \right) \) with DRB1*11. Less strong associations were observed with DRB1*14, DQB1*05:03 \( \left( 6.89 \times 10^{-7} \right. \text{and} 6.21 \times 10^{-7}, \text{respectively} \) and DRB1*03:01 \( \left( P = 9.18 \times 10^{-7} \right) \). DQB1*05:03 is in nearly complete LD with DRB1*14 \( \left( r^2 = 0.97 \right) \). DPB1*03:01 is not in LD with any of the DRB1, DQB1 or DQA1 classical alleles or amino acids (AAs; \( r^2 < 0.01 \)). There was no association \( (P > 10^{-10}) \) observed for classical alleles in HLA A, B, C or DPA1.

As expected from the analysis of classical HLA alleles, PBC also showed strong association with specific AAs in these genes. Most of the HLA AA showing association signals corresponded to the key residues that distinguish the specific classical alleles, which for DRB1, included lysine (L) at AA74 in DRB1 (DRB*08), glutamate (E) at AA58 (DRB*11), alanine (A) at AA57 (DRB1*14) and histidine (H) at AA60 (DRB1*14). Similar results were observed for specific DQB1 and DQA1 AAs that are in strong LD with specific DRB1 alleles and AAs (Supplementary Table S2).

Conditioning studies using classical HLA genes

To examine whether these associations could be explained for by known coding differences in genes, we next performed a series of conditional analyses. These studies were done by conditioning on a combination of the alleles from an HLA gene (for example, DRB1) to control for the association that might be attributable to each gene, albeit some of the effect may not be directly attributed to that gene because of extensive LD across this region. The residual signals after controlling for the effect of various combinations of classical alleles and AA residues in these HLA genes show that both DRB1 and DQB1 could account for most of the association signal (Figure 1b–e, Table 2 and Supplementary Table S3). In addition, the signal in the DPB1 region was only marginally decreased, conditioning on DRB1, DQB1 or DQA1. Conditioning on DPB1 eliminated the signal in the DPB1 region and showed a modest increase in the signal in the across DRB1, DQA1 and DQB1.

To further assess these conditional analyses, we also examined the relative difference of the conditioning by different HLA genes by examining beta estimates and their differences. The beta estimate is the measure of the increase in log-odds that can be attributed to each copy of a given minor allele. The largest effect is from the composite of DRB1 alleles as shown by the residual beta estimates (and odds ratios (ORs)) after conditioning and the mean change in the beta estimates (Table 3). This is most evident examining the SNPs with the strongest signals from association (original signal \( < 5 \times 10^{-10} \)). For example, the DRB1 conditioning had a much larger mean change in beta estimate \( (–0.42) \) compared with DQB1 \( (–0.236) \left( P = 0.5 \right) \). Additional conditional analyses using combinations with DQA1 demonstrated that DQA1 could not substitute for DRB1 or DPB1 in any of the combinations tested (data not shown).

Conditioning on specific HLA alleles

We next examined the effect of conditioning on specific DRB1, DQA1, DQB1 and DPB1 classical genes and AAs. A clear pattern emerged showing that the association of groups of SNPs was specifically controlled by different alleles. These results are highlighted in Table 4, and in a more complete version Supplementary Table S3. Individually, the specific SNPs conditioned a part of the association signal largely corresponding to those SNPs in moderate or strong LD \( (r^2 > 0.5) \) with the particular classical specificity or AA. Similar effects were observed for specific alleles in one gene in strong LD with another gene. This is particularly evident for the DRB1*14 and DQB1*05:03, in which the effects of controlling for either of these alleles was virtually indistinguishable (Supplementary Table S3). Other pairs (for example, DRB1*08 and DQB1*04:02) showed small but consistent differences in which the DRB1 allele diminished more signal than when the DQB1 allele was used in the conditioning analyses (Supplementary Table S3). Notably, the strongest AA association was at position 74 that is in the antigen-binding pocket of DRB1.11,12

Similar to when we conditioned on genes (‘all’ alleles at a particular gene), the DPB1 region SNP signal was only substantially reduced when DPB1*03:01 (Table 4 and Supplementary Table S3) or when specific DPB1 AAs were used in conditioning. The strongest effects were observed for the lysine at AA position 11 and the methionine at AA position 76 that are both members of the 16 AA in the putative antigen-binding pocket of this gene.11 Thus, the vast majority of the HLA region association signal can be accounted individually by conditioning on one of four specific

### Table 1. Summary of subject genotyping information

|                  | GWAS\( ^a \) | IC\( ^b \) | GWAS-only | IC-only | Both\( ^b \) | Total |
|------------------|-------------|---------|-----------|--------|------------|-------|
| Genotyping\( ^a \) |             |         |           |        |            |       |
| Italian PBC cases| 453         | 622     | 54        | 223    | 399        | 676   |
| Italian controls | 1042        | 597     | 843       | 398    | 199        | 1440  |
| MHC region SNP genotypes |         |         |           |        |            |       |
| Number of SNPs\( ^a \) | 1548        | 4885    | 604       | 3941   | 944        | 5489  |
| SNPs overlapping 1000 genome reference\( ^a \) | 1435       | 4386    | 595       | 3546   | 840        | 4981  |
| SNPs overlapping HLA reference\( ^a \) | 648        | 1444    | 166       | 962    | 482        | 1610  |

Abbreviations: GWAS, genome-wide association study; IC, Immunochip; HLA, human leukocyte antigen; PBC, primary biliary cirrhosis; SNP, single-nucleotide polymorphism. *Additional details on subject set genotyping is in Supplementary Table S1. **GWAS: genotyped using Illumina platforms with > 550 K GWAS SNPs. The samples included in previous GWAS study. *IC: subjects genotyped using IC array. **Subjects genotyped using both GWAS platform and IC array. *Number of SNPs within the region (chromosome 6 bps 28911802–33813043 (HG19 map)) used for analyses. **Number of SNPs genotyped in region that overlap the reference sets used for imputation.
alleles, three in DRB1 (*08, *11, and *14) and one in DPB1 (*03:01). Combinations of these specific alleles accounted for most of the remaining signal (Supplementary Table S3 and Figure 1f) and are also reflected in the strong reduction in beta estimates (Table 3). However, there are signals from several SNPs in the DRB1–DQB1 region that are not accounted for by these conditioning studies. None of these SNPs with signals P < 10^{-5} after conditioning on DRB*08, *11 and *14, and DPB1*03:01 were among the stronger associated SNPs before conditioning (all with original association P-values > 10^{-6}). In particular, the strongest associated SNP after conditioning (rs9268668, P = 1.67 x 10^{-7}) showed no signal before conditioning (P = 0.40). Whether these residual or new signals are also because of other specific classical HLA genes is not clear; however, conditioning on all DRB1 and DPB1 alleles ablated all signals with resulting P-values > 10^{-5} (Figure 1f), suggesting that additional sequence differences (for example, putative regulatory SNPs) do not have to be postulated.

Most of the signal observed for specific AAs was also specifically eliminated when conditioning on the DRB1 or DPB1 classical alleles. However, there were several exceptions in which the association signal was not readily decreased by controlling for single classical HLA alleles. These AAs included DRB1-AA47F, DRB1 AA74A, DQB1-AA26G and DQB1-AA74S. For these AAs, the signal was ablated when conditioning on two DRB1 alleles (DRB1*08 and DRB1*11) (Supplementary Table S3). Conversely, conditioning on these AA could not account for the association of the most of the other SNPs that were not ablated by single classical HLA alleles (Supplementary Table S3). Therefore, it may be less likely that these particular AAs are critical to explaining the association patterns we observed. However, we cannot exclude a specific functional role for these AAs, and it is notable that DRB1-AA47 and DRB1-AA74 are both in the antigen-binding pocket of DRB111,12 and that conditioning on DRB1-AA47F did ablate several of the association signals that were not controlled by individual DRB1 classical alleles.

Genotypic associations
We also examined genotypic associations, including combinations of susceptibility alleles, and combinations of risk susceptibility and protective alleles. Examining individuals with combinations of risk alleles, DRB1*08 combined with DPB1*03:01 or DRB1*14 combined with DPB1*03:01, we found higher ORs for disease association than when examining only individuals with single susceptibility alleles (Supplementary Table S4). There were insufficient numbers of DRB1*08/DRB1*14 (frequency < 1%) to evaluate these hetero-
results were compatible with an additive model of action between protective alleles had OR of 0.5 or 0.38, respectively (Table 5). All conversely, individuals with an excess of one or two, or more showed an OR of 3.05 and 5.25 between cases and controls, and Individuals with an excess of one or two, or more risk alleles (Table S4). The count of predisposing alleles minus protective alleles (Table 5 and Supplementary).

Our results are consistent with a confounding factors. Our results are consistent with a previous investigation by providing the most comprehensive analysis of the entire HLA region while correcting for multiple confounding factors. Our results are consistent with a predominant role for class II genes and, we believe, exclude any substantial effect from either HLA class I or class III genes (there were no residual signals for these genes with \( P < 0.0005 \) after accounting for class II genes). This contrasts other autoimmune diseases in which HLA class I or class III has a predominant role (for example, myasthenia gravis \( ^{13} \)) or strong class I gene effects are observed independent of class II associations for example, type 1 diabetes \( ^{14} \) and multiple sclerosis \( ^{15,16} \). Our study strongly suggests that the major gene in HLA that underlies susceptibility to PBC is DRB1. Overall, DRB1 alleles show the strongest associations and conditioning studies show that DRB1 could account for almost all (except the DPB1 region) of the association signal. HLA-DQB1 shows association that is only marginally less than that observed for DRB1. However, several points suggest that these associations are secondary to the strong LD between DRB1 and DQB1: (1) the overall strength of association of particular DRB1 alleles is stronger than the corresponding DQB1 allele; (2) conditioning on DRB1 could account for all DQB1 associations; and (3) residual beta-estimates after conditioning showed a substantially stronger DRB1 than DQB1 effect.
genes. These results are also consistent with findings in some, but not other, autoimmune diseases in which an independent effect of DPB1 has been reported. These include juvenile idiopathic arthritis,\textsuperscript{17} type 1 diabetes,\textsuperscript{18} multiple sclerosis\textsuperscript{16,19} and particular autoantibodies in systemic lupus erythematosus.\textsuperscript{20}

Table 3. Comparison of residual association beta estimates and ORs after conditioning studies

| All markers\(^a\) | Association\(^b\) | Controlled DRB1\(^c\) | Controlled DQB1 | Controlled DQA1 | Controlled DPB1 | Controlled DRB1 and DPB1 | Controlled DQB1 and DPB1 | Controlled DRB1*8, 11, 14 and DPB1*03:01 |
|-------------------|-------------------|----------------------|-----------------|-----------------|----------------------|------------------------|--------------------------|-----------------------------------|
| Mean B estimate   | 0.242             | 0.188                | 0.205           | 0.211           | 0.237                | 0.180                  | 0.197                    | 0.190                             |
| Mean change in beta estimate | Reference | -0.055             | -0.037          | -0.026          | -0.005              | -0.062                 | -0.045                   | -0.052                            |
| Mean OR           | 1.274             | 1.207                | 1.228           | 1.234           | 1.268                | 1.197                  | 1.218                    | 1.209                             |

Table 4. Conditioning on specific HLA DRB1 and DPB1 alleles

| Group\(^d\) | Marker/amino acid/HLA determinant\(^b\) | Base pair position (HG19)\(^c\) | P-value\(^d\) | Cond. P-value DRB1*08\(^e\) | Cond. P-value DRB1*11\(^f\) | Cond. P-value DRB1*14\(^g\) | Cond. P-value DPB1*03:01\(^h\) |
|-------------|---------------------------------------|----------------------------------|---------------|----------------------------|-----------------------------|----------------------------|-------------------------------|
| 1           | rs114991247                           | 32583504                         | 5.45E-13      | 2.41E-03                  | 3.96E-13                    | 2.15E-18                   | 3.75E-14                     |
| 1           | rs116240177                           | 32561662                         | 8.20E-13      | 4.82E-03                  | 1.17E-12                    | 3.66E-18                   | 4.29E-14                     |
| 1           | DRB1-AA74L                            | 32551948                         | 1.33E-11      | 4.72E-01                  | 2.26E-11                    | 3.67E-16                   | 4.18E-12                     |
| 1           | HLA-DRB1*08                           | 32552064                         | 1.59E-11      | 1.00E+00                  | 2.73E-11                    | 4.55E-16                   | 1.77E-12                     |
| 1           | DQA1-AA69T                            | 32602979                         | 2.75E-11      | 8.93E-01                  | 3.55E-11                    | 5.99E-16                   | 2.35E-12                     |
| 1           | HLA-DQB1*04:02                        | 32631061                         | 1.38E-10      | 5.79E-01                  | 1.86E-10                    | 4.02E-15                   | 1.63E-11                     |
| 1           | HLA-DQA1*04:01                        | 32608306                         | 1.90E-10      | 9.58E-01                  | 1.73E-10                    | 5.55E-15                   | 2.01E-11                     |
| 2           | rs9461777                             | 32575735                         | 3.02E-12      | 2.39E-14                  | 3.35E-03                    | 9.62E-15                   | 1.35E-13                     |
| 2           | rs115671039                           | 32603212                         | 1.98E-11      | 9.29E-13                  | 4.62E-02                    | 5.22E-13                   | 2.91E-14                     |
| 2           | HLA-DRB1*11                           | 32552064                         | 1.42E-10      | 1.34E-12                  | 9.99E-01                    | 3.43E-13                   | 9.25E-14                     |
| 2           | DRB1-AA58E                            | 32551996                         | 1.45E-10      | 1.34E-12                  | 9.98E-01                    | 4.38E-13                   | 9.30E-14                     |
| 2           | DQB1-AA71T                            | 32632664                         | 5.25E-10      | 3.65E-09                  | 2.62E-04                    | 4.68E-10                   | 1.69E-12                     |
| 2           | HLA-DQB1*03:01                        | 32631061                         | 6.10E-09      | 3.29E-09                  | 9.76E-01                    | 3.95E-10                   | 3.25E-11                     |
| 3           | rs380911                              | 32209451                         | 8.94E-09      | 1.36E-10                  | 4.52E-06                    | 3.71E-04                   | 5.26E-09                     |
| 3           | rs115240978                           | 32672383                         | 1.28E-08      | 1.30E-10                  | 2.24E-06                    | 6.75E-03                   | 8.16E-10                     |
| 3           | HLA-DQB1*05:03                        | 3261061                          | 6.21E-07      | 4.79E-09                  | 5.13E-05                    | 6.15E-01                   | 2.31E-08                     |
| 3           | HLA-DRB1*11                           | 32552064                         | 6.89E-07      | 5.29E-09                  | 4.83E-05                    | 9.99E-01                   | 2.46E-08                     |
| 3           | DRB1-AA57A                            | 32551999                         | 6.90E-07      | 5.29E-09                  | 4.82E-05                    | 9.99E-01                   | 2.46E-08                     |
| 3           | DRB1-AA60H                            | 32551990                         | 6.90E-07      | 5.29E-09                  | 4.82E-05                    | 9.99E-01                   | 2.46E-08                     |
| 3           | rs116751915                           | 33052502                         | 8.11E-09      | 9.21E-09                  | 3.09E-06                    | 1.76E-10                   | 1.69E-04                     |
| 3           | rs114761768                           | 33050442                         | 1.79E-08      | 3.09E-08                  | 4.73E-06                    | 4.90E-10                   | 3.98E-04                     |
| 4           | DPB1-AA76M                            | 33048662                         | 6.13E-08      | 3.78E-07                  | 2.11E-05                    | 1.07E-08                   | 8.71E-03                     |
| 4           | DPB1-AA1701                           | 33052958                         | 1.06E-07      | 2.61E-07                  | 1.61E-05                    | 3.21E-09                   | 8.41E-03                     |
| 4           | HLA-DPB1*03:01                        | 33049368                         | 9.18E-07      | 2.05E-06                  | 1.01E-06                    | 3.70E-09                   | 9.98E-01                     |

Abbreviations: AA, amino acid; Cond., conditioning; HLA, human leukocyte antigen; OR, odds ratio; SNP, single-nucleotide polymorphism. \(*\)All markers including SNPs (MHC region), classical HLA determinants and HLA AA. \(\textsuperscript{1}\)The beta estimate, change in beta estimate and odds ratios from SNPTST2V score test (see methods). \(\textsuperscript{2}\)Results after conditioning on the indicated HLA genes or alleles. \(\textsuperscript{3}\)Results using only markers with \(P\)-values < 5 x 10\(^{-6}\) in the initial association tests.
We note that our study does not directly address whether DRB3, DRB4, DRB5 or structural variations might have additional independent associations. At present, such studies are challenging because of absence of reference sets for imputation and/or difficulty in assessing these polymorphisms, including whether missing genotypes (excluded SNPs with call rates < 0.95) may have excluded analysis or inclusion of SNPs within these genes in available arrays.

This study has also addressed the association of specific HLA-gene alleles. Most of the HLA association with PBC can be attributed to specific associations with DRB1*08, DRB1*11, DRB1*14 and DPB1*0301. DRB1*08 has the strongest association, followed by DRB1*11, consistent with several previous studies.\(^6,7\) PBC associations with DRB1*14 has not been previously demonstrated; however, this weaker effect is supported by our conditioning studies that show that this classical allele can control for a set of associated SNPs and AAs that are not strongly influenced by other classical alleles (Table 3 and Supplementary Table S3). The DPB1*0301 association is consistent with a previous study of a small German cohort.\(^7\)

In a previous study, we observed that DRB1*13 was a protective allele.\(^5\) In the current study, the association of DRB1*13 was weak (P = 4.9 × 10^-3, OR = 0.69, 95% confidence limits = 0.53–0.89) compared with the previous study (P = 3.6 × 10^-5). This may be because of several factors: (1) the previous study did not explicitly control for population substructure; (2) the overlap of subjects with the previous study is < 25% and the difference may reflect statistical noise; and (3) the previous study used DNA typing rather than the imputation used in the current study. It may be worth noting that DRB1*13 like DRB1*11 has an alanine at AA position 74 and thus contributes to the protective effect observed for this AA (P = 1.33 × 10^-11, Supplementary Table S3). Similarly, our previous study\(^6\) showed only a marginal association of DRB1*14 (uncorrected P-value = 0.004) compared with a strong association (P = 6.9 × 10^-7) observed in our current study. Here, the conditioning study results, including the effect of controlling SNPs with very strong associations (see group 3, Table 4), provide additional support for the role of DRB1*14.

Finally, we have also considered specific HLA AAs. Most of the associated AAs are both nearly unique to the specific HLA classical alleles discussed above and also correspond to critical residues for the antigen-binding pocket. Thus, associated AAs in DRB1 at AA positions 37, 47, 57, 60, 67, 70 and 74; and DPB1 at AA positions 9, 11, 76, 84, 87 are antigen-pocket AAs.\(^11,12\) Consistent with our results, strong associations have been recently observed with serine at position 57 and leucine at position 74 in a Japanese PBC cohort.\(^7\) We also note that many of the associated DQB1 AAs are also in critical residues for antigen binding (DQB1- AA13, 26, 70, 71, 74).\(^11,12\) Of the DRB1 associated, only AA58 is not among the AAs in this functional class, whereas for DQB1, several are not in this functional class (DQB1 DRB1*08, 56, 75, 167, 185).

In conclusion, the most parsimonious explanation consistent with the current study is that classical HLA genes and the coding variations within these genes are responsible for the HLA associations with PBC. Although we cannot exclude the possibility that other sequence variations affecting, for example, gene regulation could be important, our data indicates that a limited set of classical DRB1 and DPB1 alleles are sufficient to explain the HLA associations with this disease. We believe the current data provides cogent information for understanding HLA-associations in PBC. Studies in other ethnic groups both within Europe and in other continental groups will also be important in further definition of the role of particular HLA genes and alleles. Lastly, our results provide additional rational for functional studies examining specific HLA genes and their relative binding to the putative disease-associated epitopes of the PDC-E2, the immunodominant autoantigen epitopes of PBC.\(^2,22\)

### MATERIALS AND METHODS

**Study population and design**

The Italian PBC cases were obtained through a multi-center study and met internationally accepted criteria for the diagnosis of PBC as detailed in a previous study.\(^5\) Each of the included cases also met ancestry criteria as defined below (see Ancestry). Controls were derived from several sources and this sample set information is detailed in Supplementary Table S1. After data filtering and ancestry, analyses contained 676 Italian PBC cases and 1440 Italian controls. All subjects enrolled in the study provided written informed consent and the study followed ethical guidelines of the most recent revision of the Declaration of Helsinki (Edinburgh, 2000).

All samples were genotyped with either Illumina (San Diego, CA, USA) genome-wide and/or Immunochip SNP platforms, and the participants included the data set from our previous GWAS as well as new samples (see Supplementary Table S1). With the exception of ancestry information and assessment of relatedness, the current study was restricted to genotypes in an approximately 4 Mb segment of human chromosome 6 (bp 289111802–33813043, HG19 map). This data set comprised a minimum of 1548 and a maximum of 5489 genotyped SNPs in each individual (Supplementary Table S1) and was used for the SNP and HLA imputations (see Imputation).

**Data filtering**

We used stringent quality-control criteria to ensure that high-quality data were included in the analyses. We excluded individuals who had > 5% missing data, and all individuals with cryptic relatedness and duplicate samples based on identity-by-descent status for genome-wide SNPs (P < 0.15) using PLINK.\(^24\)

We included only SNPs with < 5% missing data, Hard–Weinberg equilibrium P-values > 10^-4 in controls and > 10^-5 in combined cases and controls (to exclude most genotyping errors), applying these procedures in a stepwise approach separately for each data set. For each of the separately derived control genotyping sets (Supplementary Table S1), SNPs were excluded if they failed the above criteria within the individual control set or in combination with any of the other control groups, or in the complete data set. The Hard–Weinberg criteria were applied after exclusion of non-European individuals (see Ancestry). Finally, SNPs were excluded if allele frequency differed by > 10% between different control subject groups.

---

**Table 5. Cumulative effect of risk and protective classical HLA alleles for PBC susceptibility**

| Category\(^a\) | Controls number (%)\(^b\) | Cases number (%) | OR | P-value\(^c\) |
|--------------|---------------------|-----------------|----|------------|
| RISK— PROTECTIVE \(\leq -2\) | 177 (12.3) | 34 (5.0) | 0.38 | 5.87E-08 |
| RISK— PROTECTIVE \(\leq -1\) | 682 (48.1) | 185 (27.4) | 0.50 | 2.03E-12 |
| RISK— PROTECTIVE = 0 | 522 (36.3) | 246 (36.4) | 1.00 | NS |
| RISK— PROTECTIVE \(>1\) | 226 (15.7) | 245 (36.2) | 3.05 | 6.94E-25 |
| RISK— PROTECTIVE \(\geq 2\) | 32 (2.2) | 72 (10.7) | 5.25 | 1.37E-15 |

Abbreviations: AA, amino acid; HLA, human leukocyte antigen; OR, odds ratio; PBC, primary biliary cirrhosis.\(^1\) This table shows the results of categorizing each participant based on the sum of the each risk allele (positive number) and each protective allele (negative number). The risk alleles are DRB1*08, DRB1*14 and DPB1*0301. The protective alleles are DRB1*11 and DRB1*13. The alleles were determined from the most probable allele after imputation and haplotype analyses (see Materials and methods).\(^8\) The number of subjects that are in each category are shown together with the % of total controls (1440) or percent of cases (676).\(^9\) The two tailed Fisher exact P-value from contingency table analyses comparing cases and controls in each category.

---

© 2012 Macmillan Publishers Limited
Ancestry
European ancestry was determined using 883 genome-wide SNPs with minimal or no LD ($r^2<0.1$). SNPs analyzed using the STRUCTURE v2.1 program\textsuperscript{25} and subjects of known European, Amerindian, East Asian and West African origin as previously described.\textsuperscript{26} We used STRUCTURE to exclude non-European and admixed study participants, as this method allows exclusion/inclusion criteria to be set using reference populations. Subjects with $>15\%$ non-European ancestry were excluded from further analysis.

Italian ancestry was defined using principal components (PCs) analyses. For subjects with GWAS data, we used the same methods and criteria applied in a previous study with largely the same data set. Briefly, PCs analyses were performed using the EIGENSOFT statistical package,\textsuperscript{27} utilizing 54,000 SNPs distributed throughout the genome ($r^2<0.1$) that we have previously used to define population genetic substructure.\textsuperscript{2} These analyses used an independent set of Italian subjects for establishing membership ($\pm2\text{ s.d.}$) in first four PCs. In the current study, a substantial portion of the samples did not have GWAS data (Supplementary Table S1). For these samples, we used a set of 12,579 SNPs from the Immunochip for which our empirical analyses demonstrated the ability of this set to discern Italian ancestry and exclude both the other European ethnicities, including Sardinian Italians (Supplementary Figure S1). Using the GWAS-defined individuals, the Immunochip only genotyped samples were included using $2\text{ s.d.}$ in the first four PCs. In addition to the subject selection, we used the eigenvectors from the first four PCs (only the first four PCs were significant based on Tracy–Widom statistics) as covariates in our association analyses.

Imputation
We imputed SNPs, HLA classical alleles and HLA gene AAs using phased reference genotypes from both the 1000 genome-imputation project (interim release June 2011) (http://www.1000genomes.org/) and an HLA-defined reference set.\textsuperscript{3} For the 1000 genome imputation, we used IMPUTE version 2\textsuperscript{28} under default parameters. The reference haplotypes for this imputation were from 1094 subjects, including 381 European subjects and 98 Tuscan Italians. The number of genotyped (inference) SNPs that overlapped with the 1000 genome reference set ranged from 1435 SNPs (samples typed by GWAS), 4386 (samples typed by Immunochip) to 4981 SNPs (samples typed by GWAS plus Immunochip) (Supplementary Table S1). For subsequent data analyses, we utilized only imputed genotypes with maximum posterior probability scores of $>0.90$. Using this parameter, our empirical testing (leave one-out analyses) indicated that the maximum error rate for genotype assignment was $<0.05$, and the mean error rate was $<0.01$. To impute classical HLA alleles and corresponding AAs determinants, we utilized a reference separate data set of collected by the Type 1 Diabetes Genetics Consortium. This reference data contains genotype data for 2537 SNPs, selected to tag the major histocompatibility complex, and classical types for HLA-A, B, C, DRB1, DQA1, DQB1, DPB1 and DQA1 at 4-digit resolution in 2767 unrelated individuals of European descent.\textsuperscript{39} The Beagle software package\textsuperscript{40} was used for this imputation under default parameters. The number of inference SNPs that overlapped with this reference data set ranged from 648 (samples typed by GWAS), 1444 SNPs (samples typed by Immunochip) to 1610 SNPs (GWAS plus Immunochip) (Supplementary Table S1). Similar to the imputation using 1000 Genome data, only SNPs with posterior probabilities of $>0.90$ were included in our final analyses. For imputed SNPs that overlapped between the two imputation sets and algorithms used (ImpuTe V2.0 and Beagle), there was a nearly complete concordance of the association-testing results, indicating similar performance of these algorithms for this data set. After imputation and selecting only those markers meeting posterior probability criterion, this region contained a total of 49,885 markers including the genotyped SNPs that were included in association test analyses.

Association and conditional association tests of imputed SNPs and HLA determinants
SNPTEST V2.20\textsuperscript{36} (web) was used for the primary association analyses for the imputed markers. This software uses the genotype probabilities for the imputed SNPs or determinants and accounts for genotype uncertainty. The first four PC eigenvalue scores were used as continuous variables in the association test together with the gender covariate. Analyses were performed using the SNPTEST V2 Score test algorithm that enabled both the inclusion of the covariates and conditioning tests, and all of our reported results used an additive model. To minimize potential spurious results, we limited our main and conditioning analyses to markers with information scores (Inf) $>0.85$. This parameter is a measure of the observed statistical information for the estimate of SNP allele frequency (for additional information see https://mathgen.stats.ox.ac.uk/genetics_software/snpset/snpsetv2.pdf).

Conditioning on multiple markers either separately or together was performed using an additive model. For the HLA region, over 150 conditional analyses were performed using the SNPs and HLA determinants, including all HLA determinants with $P$-values $<10^{-6}$. Nominal $P$-values after correction for covariates and conditioning are provided throughout the manuscript. The $P$-values $<10^{-6}$ would remain significant after conservative (Bonferroni) correction for the number of markers ($<50,000$) tested after imputation.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

ACKNOWLEDGEMENTS
This study was supported by NIH grants: R01 DK056839, R01DK091823, and KO8AR055688 and HYPERGENES (European Network for Genetic-Epidemiological Studies HEALTH-F4-2007-201550).

REFERENCES
1 Hirschfeld GM, Liu X, Xu C, Lu Y, Xie G, Lu Y et al. Primary biliary cirrhosis associated with HLA, IL12A, and IL12RB2 variants. N Engl J Med 2009; 360: 2544–2555.
2 Liu X, Invernozzi P, Lu Y, Kosoy R, Lu Y, Bianchi I et al. Genome-wide meta-analyses identify three loci associated with primary biliary cirrhosis. Nat Genet 2010; 42: 658–660.
3 Mells GF, Floyd JA, Morley KL, 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.
4 Inverniy P. Human leukocyte antigen in primary biliary cirrhosis: an old story now reviving. Hepatology 2011; 54: 714–723.
5 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: 667–674.
6 Inverniy 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; 49: 1906–1912.
7 Mella SG, Raschmann E, Maier KP, Volk BA. Association of primary biliary cirrhosis with the allele HLA-DRB1*0301 in a German population. Hepatology 1995; 21: 398–402.
8 Pereya F, Jia X, McLaren PJ, Telenti A, de Bakker PI, Walker BD et al. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science 2010; 330: 1551–1557.
9 Connell J, Brown MA. Promise and pitfalls of the Immunoscan. Arthritis Res Ther 2011; 13: 101.
10 de Bakker PI, McVean G, Sabeti PC, Miretti MM, Green T, Marchini J et al. High-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. Nat Genet 2008; 38: 1166–1172.
11 Salamon H, Klitz W, Eastal S, Gao X, Erlich HA, Fernandez-Vina M et al. Evolution of HLA class II molecules: Allelic and amino acid site variability across populations. Genetics 1999; 152: 393–400.
12 Karp DR, Marthandan N, Marsh SG, Achn C, Arnett FC, Deluca DS et al. Novel sequence feature variant type analysis of the HLA genetic association in systemic sclerosis. Hum Mol Genet 2010; 19: 707–719.
13 Vandiedonck C, Beaurain G, Giraud M, Hue-Beauvais C, Eymard B, Tranchant C et al. Pleiotropic effects of the 8.1 HLA haplotype in patients with autoimmune myasthenia gravis and thymus hyperplasia. Proc Natl Acad Sci USA 2004; 101: 15464–15469.
14 Noble JA, Valdes AM, Varney MD, Carlson JA, Moonsamy P, Fear AL et al. HLA class I and genetic susceptibility to type 1 diabetes: results from the Type 1 Diabetes Genetics Consortium. Diabetes 2010; 59: 2972–2979.
15 Wu XM, Wang C, Zhang KN, Lin AY, Xia J, Hu GZ et al. Association of susceptibility to multiple sclerosis in Southern Han Chinese with HLA-DRB1, -DRB1 alleles and DRB1-DRB1 haplotypes: distinct from other populations. Mult Sci 2009; 15: 1422–1430.
16 Field J, Browning SR, Johnson LJ, Danoy P, Varney MD, Tait BD et al. A polymorphism in the HLA-DRB1 gene is associated with susceptibility to multiple sclerosis. PLoS One 2010; 5: e13454.
APPENDIX

The Italian PBC Genetic Study Group
Piero L. Almasio (Gastroenterology and Hepatology Unit, D.L.B.M.I.S., University of Palermo, Palermo), Domenico Alvaro (Department of Medical-Surgical Sciences and Biotechnologies, Fondazione Elettronomia Lorillard Spencer Cenci, University of Rome, Rome), Pietro Andreone (Dipartimento di Medicina Clinica, Università di Bologna, Bologna), Angela Andriulli (IRCCS Casa Sollievo della Sofferenza Hospital, San Giovanni Rotondo), Cristina Barlassina (Department of Medicine, Surgery and Dentistry, Università degli Studi di Milano, Milan), Antonio Benedetti (Università Politecnica delle Marche, Ancona), Francesca Bernuzzi (Center for Autoimmune Liver Diseases, IRCCS Istituto Clinico Humanitas, Rozzano), Ilaria Bianchi (Center for Autoimmune Liver Diseases, IRCCS Istituto Clinico Humanitas, Rozzano), Maria Consiglia Bragazzi (Department of Medical-Surgical Sciences and Biotechnologies, Fondazione Elettronomia Lorillard Spencer Cenci, University Sapienza of Rome, Rome), Maurizia Brunetto (Azienda Ospedaliera Universitaria Pisana, Pisa), Savino Bruno (Department of Internal Medicine, Ospedale Fatebene Fratelli e Oftalmico, Milan), Lisa Gallari (Center for Autoimmune Liver Diseases, IRCCS Istituto Clinico Humanitas, Rozzano), Giovanni Casella (Medical Department, Desio Hospital, Desio), Fabiola Civardi (Center for Autoimmune Liver Diseases, IRCCS Istituto Clinico Humanitas, Rozzano), Gianni Ciselli (Medical Department, Desio Hospital, Desio), Barbara Cocco (Azienda Ospedaliera Universitaria Pisana, Pisa), Agostino Colli (Department of Internal Medicine, AO Provincia di Lecco, Lecco), Massimo Colombo (Fondazione IRCCS Ca’ Granda, Ospedale Maggiore Policlinico, Milan), Silvia Colombo (Treviso Hospital, Treviso), Carmela Cursaro (Dipartimento di Medicina Clinica, Università di Bologna, Bologna), Lory Saveria Croce (University of Trieste and Fondazione Italiana Fegato (FIF), Trieste), Andrea Cisognani (San Paolo Hospital Medical School, Università di Milano, Milan), Francesca Donato (Fondazione IRCCS Ca’ Granda, Ospedale Maggiore Policlinico, Milan), Luca Fabris (University of Padova, Padova), Carlo Ferrari (Azienda Ospedaliero-Universitaria di Parma, Parma), Annarosa Floreani (Dept. of Surgical, Oncological and Gastroenterological Sciences, University of Padova, Padova), Andrea Galli (University of Florence, Florence), Ignazio Grattaglione (Italian College of General Practitioners, ASL Bari), Roberta Lazzari (Department of Surgical, Oncological and Gastroenterological Sciences, University of Padova, Padova), Fabio Macaluso (Gastroenterology & Hepatology Unit, D.L.B.M.I.S., University of Palermo, Palermo), Fabio Marra (University of Florence, Florence), Marco Marzioni (Università Politecnica delle Marche, Ancona), Alberto Mattalìa (Santa Croce Carle Hospital, Cuneo), Renzo Montanari (Ospedale di Negrar, Verona), Lorenzo Morini (Magenta Hospital, Magenta), Filomena Morisco (University of Naples, Federico II, Naples), Luca Moroni (Center for Autoimmune Liver Diseases, IRCCS Istituto Clinico Humanitas, Rozzano), Luigi Muratori (Department of Clinical Medicine, University of Bologna, Bologna), Paolo Muratori (Department of Clinical Medicine, University of Bologna, Bologna), Grazia Niro (IRCCS Casa Sollievo della Sofferenza Hospital, San Giovanni Rotondo), Antonio Picciotto (University of Genoa, Genoa), Piero Portincasa (Department of Interdisciplinary Medicine, University Medical School, Bari), Daniele Prati (Ospedale Alessandro Manzoni, Lecco, Fondazione IRCCS Ca’ Granda, Ospedale Maggiore Policlinico, Milan), Cleofe Prisco (Ospedale Niguarda Milan), Florian Rosina (Division of Gastroenterology & Hepatology, Center for Predictive Medicine, Gradenigo Hospital, Turin), Sonia Rossi (Department of Internal Medicine, Ospedale Fatebene Fratelli e Oftalmico, Milan), Carlo Selmi (IRCCS Istituto Clinico Humanitas, Rozzano), Giancarlo Spinti (Azienda Ospedaliera Valduce, Como), Mario Strazzabosco (Yale University, New Haven, Connecticut 06511, USA and University of Milan-Bicocca, Monza), Sonia Taralio (Division of Gastroenterology and Hepatology, Center for Predictive Medicine, Gradenigo Hospital, Turin), Claudio Tirielli (University of Trieste and Fondazione Italiana Fegato (FIF), Trieste), Pierluigi Toniutti (University of Udine, Udine), Maria Vinci (Ospedale Niguarda Milan), Simona Zini (San Paolo Hospital Medical School, Università di Milano, Milan).

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