Genome-wide association study identifies susceptibility loci for IgA nephropathy

This is the author's manuscript

Original Citation:

Availability:
This version is available  http://hdl.handle.net/2318/83966 since 2016-01-07T19:19:51Z

Published version:
DOI:10.1038/ng.787

Terms of use:
Open Access
Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)
This is an author version of the contribution published on:

The definitive version is available at:
La versione definitiva è disponibile alla URL:
[http://www.nature.com/ng/journal/v43/n4/full/ng.787.html]
Genome-wide Association Study Identifies Susceptibility Loci for IgA Nephropathy

Ali G. Gharavi1, Krzysztof Kiryluk1, Murim Choi2, Yifu Li11, Ping Hou1,3, Jingyuan Xie1,4, Simone Sanna-Cherchi1, Clara J. Men2, Bruce A. Julian5, Robert J. Wyatt5, Jan Novak5, John C. He7, Haiyan Wang3, Jicheng Ly3, Li Zhu3, Weiming Wang4, Zhaohui Wang4, Kasuhiyo Yano2, Murat Gune12, Shrikant Mane2,8, Sheila Umlauf2,8, Irina Tikhonova2,8, Isabel Beerman2, Silvana Savoldi9, Riccardo Magistroni10, Gian Marco Ghiggeri11, Monica Bodria11, Francesca Lugani1,11, Pietro Ravani12, Claudio Ponticelli13, Landino Allegri14, Giuliano Boscutti15, Giovanni Frasca16, Alessandro Amore17, Licia Peruzzi17, Rosanna Coppo17, Claudia Izzi18, Fabio Viola19, Elisabetta Prati20, Maurizio Salvadori21, Renzo Mignani22, Loreto Gesualdo22, Francesca Bertinetto24, Silvana Paola Mesiano24, Antonio Amoroso24, Francesco Scolari18, Nan Chen4, Hong Zhang3, and Richard P. Lifton2

1Department of Medicine, Columbia University College of Physicians and Surgeons, New York, New York, USA
2Department of Genetics, Howard Hughes Medical Institute, Yale University School of Medicine, New Haven, CT, USA
3Renal Division, Peking University First Hospital; Peking University, Institute of Nephrology; Key Laboratory of Renal Disease, Ministry of Health of China, Beijing, PR China
4Department of Nephrology, Ruijin Hospital, Shanghai Jiaotong University, School of Medicine, Shanghai 200025, P.R. China
5Departments of Microbiology and Medicine, University of Alabama at Birmingham, Birmingham, AL, USA
6Children’s Foundation Research Center at the Le Bonheur Children’s Hospital, and the Division of Pediatric Nephrology, University of Tennessee Health Sciences Center, Memphis, TN, USA
7Department of Medicine, Mount Sinai School of Medicine, New York, NY 10029
8Nephrology and Dialysis Unit, Ciriè Hospital, Torino, Italy

Correspondence: Ali G. Gharavi, MD, Department of Medicine, Columbia University College of Physicians and Surgeons, 1150 St Nicholas Ave, Russ Berrie Pavilion 412, New York, New York 10032, USA. Telephone 212-851-5556 Fax 212-305-5520 ag2239@columbia.edu. Richard P. Lifton, MD, PhD, Department of Genetics, Howard Hughes Medical Institute, Yale University School of Medicine, 300 Cedar Street, TAC S-341D, New Haven, CT 06520, USA Telephone: 203-737-4420 Fax: 203-785-7560 richard.lifton@yale.edu.


Genotyping and wet lab experiments: S.M., S.U., I.T., C.J.M., M.C., P.H., J.X. and Y.L.
Data management: K.K., Y.L., S.S.C., and M.C.
Data analysis: K.K., M.C., A.G.G., and R.P.L.
Analytical support and discussion: K.Y. and M.G.
Conception and overall supervision of project: A.G.G. and R.P.L.

The authors have no competing financial interests.
Abstract

We performed a genome-wide association study of IgA nephropathy (IgAN), a major cause of kidney failure worldwide. Discovery was in 1,194 cases and 902 controls of Chinese Han ancestry, with targeted follow-up in Chinese and European cohorts comprising 1,950 cases and 1,920 controls. We identified three independent loci in the major histocompatibility complex (MHC), a common deletion of CFHR1 and CFHR3 at Chr. 1q32 and a locus at Chr. 22q12 that each surpassed genome-wide significance (p-values for association between 1.59 × 10^{-26} and 4.84 × 10^{-9} and minor allele odds ratios of 0.63–0.80). These five loci explain 4–7% of the disease variance and up to a 10-fold variation in interindividual risk. Many of the IgAN–protective alleles impart increased risk of other autoimmune or infectious diseases, and IgAN risk allele frequencies closely parallel the variation in disease prevalence among Asian, European and African populations, suggesting complex selective pressures.

Chronic kidney disease is a major cause of morbidity and mortality affecting 10–20% of the world population, with glomerulonephritis accounting for a significant proportion of cases\(^1\). IgA nephropathy (IgAN) is the most common form of glomerulonephritis and the most common cause of kidney failure among Asian populations\(^2,4\). The diagnosis of IgAN requires documentation by kidney biopsy demonstrating proliferation of the glomerular mesangium with deposition of immune complexes predominantly composed of Immunoglobulin A (IgA) and complement C3 proteins\(^3,5,6\). Registry data as well as autopsy and kidney-donor biopsy series suggest significant variation in prevalence among different ethnicities: IgAN is most frequent among Asians, with a disease prevalence as high as 3.7% detected among Japanese kidney donors\(^7\), but is rare among individuals of African ancestry\(^5\) and of intermediate prevalence among Europeans (up to 1.3%)\(^6\).
The pathogenesis of IgAN is uncertain\(^8,9\). The finding of IgA1 glycosylation abnormalities among European, Asian, and African-American populations has suggested a shared pathogenesis among different groups\(^10\)-\(^15\). Moreover, familial aggregation of IgAN has been reported among all ethnicities, suggesting a genetic component to disease\(^8,16\). To date linkage studies have identified several loci predisposing to IgAN, but underlying genes are not known\(^8,16\)-\(^18\). A single, unreplicated genome-wide association study (GWAS) in a small European cohort (533 cases) has reported association of IgAN with the MHC complex\(^19\).

We report a GWAS for IgAN in a cohort of 3,144 IgAN cases of Chinese and European ancestry, leading to the identification of five loci for this disease.

**RESULTS**

**Study design and genotyping of discovery cohort**

To detect loci conferring susceptibility to IgAN, we performed a two-stage GWAS (Table 1). In the discovery phase, genome-wide genotyping was performed on the Illumina 610 quad platform in 1,228 biopsy-proven IgAN cases and 966 healthy controls of Chinese Han ancestry recruited from Beijing (Table 1 and Supplementary Table S1). The top signals in the discovery phase were further evaluated in an independent cohort of Han Chinese descent (Shanghai cohort, 740 cases and 750 controls) and a European cohort of Italian and North American origin (combined by stratified analysis, 1,273 cases and 1,201 controls). Subsequently, we analyzed the Beijing, Shanghai and European cohorts together to identify genome-wide significant loci.

**Genome-wide association analysis**

In analysis of genome-wide genotyping data we applied stringent quality control filters, resulting in elimination of 5% of samples due to low call rate, duplication, cryptic relatedness or gender mismatch and 16.8% of markers primarily due to low minor allelic frequency (<0.01, see supplementary notes and Supplementary Table S2). After quality control, the genotyping call rate was 0.9992. We next applied the standard 1-degree of freedom Cochran Armitage (CA) trend test to analyze 498,322 SNPs in the discovery cohort of 1,194 cases (650 males/544 females, average age 31.1 years) and 902 controls (608 males/294 females, average age 31.5 years). The quantile-quantile plot showed no global departure from the expected distribution of p-values and the inflation factor (\(\lambda\)) was 1.024, indicating negligible population stratification (Supplementary Figure S1 and Figure 1). Accordingly, principal component analysis (PCA) demonstrated that cases and controls were matched along the axes of significant principal components, and PCA correction did not substantially change the distribution of the association statistic or the genomic inflation factor (\(\lambda\)= 1.022, Supplementary Figure S2, Supplementary Table S3). We concluded that our association results were not biased by differences in ancestry or population structure between cases and controls.

The genome-wide association analysis revealed 27 SNPs exceeding genome-wide thresholds for significance (p \(\leq 5 \times 10^{-8}\), Figure 1). These 27 signals all resided in a 0.54 Mb interval within the major histocompatibility complex (MHC) on Chr. 6p21, with the top signal at rs9275596 (p = 1.9 \times 10^{-12}). Interestingly, fourteen MHC SNPs with suggestive p-values (5 \(\times 10^{-6}\) to 1 \(\times 10^{-4}\)) showed little or no linkage disequilibrium with rs9275596 (Figure 2a).

**Follow-up of top signals from discovery stage**

After removal of MHC SNPs, there remained additional loci showing departure from the expected p-value distribution. We ranked signals based on the false discovery rate and chose to follow-up loci with p-value \(\leq 1.3 \times 10^{-5}\), corresponding to a q-value \(\leq 0.10\).
Power calculations indicated that this strategy would provide 80% power to detect loci with allelic frequencies > 0.10 and relative risk > 1.5 with genome-wide significance (p < 5 × 10^{-8}) in the combined cohort (Supplementary Table S4). In total, 65 SNPs from 10 distinct loci met these criteria (including three potentially independent loci in MHC and two in the Chr. 22q12.2 interval). We genotyped the top-scoring SNP's and one additional SNP from each of these intervals in follow-up cohorts (total 20 SNPs in 3,870 individuals after quality control, table 1). Tests of association were performed within each cohort, followed by a combined analysis with the discovery cohort using Mantel’s extension of CA trend test (Table 2 and Supplementary Table S5).

Five of the ten loci selected for follow-up surpassed the threshold for significant genome-wide association - three loci within 6p21, one locus at 1q32, and one locus at 22q12.2 (Table 2, Supplementary Table S5, S6). Each signal demonstrated significant association with consistent effect size for the same risk allele in each individual cohort, with little evidence for heterogeneity.

The strongest association in the combined cohort was located within a ~170 kb interval that includes the HLA-DRB1, -DQA1, and -DQB1 genes (rs9275596, OR = 0.63, p=1.6 × 10^{-26}). This SNP achieves genome-wide significance with a consistent effect size in each cohort (Table 2, Figure 2b) and has strong supporting association from a nearby SNP in strong LD (rs2856717).

This locus, however, did not explain all of the signal at 6p21. Conditioning for the effect of rs9275596 eliminated evidence for association for the majority of SNPs in close proximity, however two distinct loci maintained genome-wide significance. The second independent locus is defined by rs9357155 (which has an r^2 = 0.01 with rs9275596 in the combined cohort) and shows an OR = 0.74 and a p-value of 6.9 × 10^{-9} for association with IgAN after conditional analysis (Table 3, Figure 2c). This SNP lies in a ~100 kb segment of LD and lies 128 kb centromeric to rs9275596. This LD segment contains the genes TAP2, TAP1, PSMB8, and PSMB9, and the supporting SNPs in this region (rs2071543) is a missense variant in PSMB8 (Q49K) that is at a position completely conserved among all orthologs (most distantly related ortholog is in platypus; Tables 2 and 3, Figure 2c and Supplementary Tables S7, S8).

After conditioning for the effects of both rs9275596 and rs9357155, a third locus within MHC, defined by rs1883414, which lies 400 kb centromeric to rs9275596 (and which shows r^2=0.005 and 0.002 with rs9275596 and rs9357155, respectively), shows a conditioned OR of 0.77 and p-value of 3.1 × 10^{-8} for association (Table 3). This signal in the HLA-DPA1 – DPB1 –DPB2 region is supported by a second SNP (rs3129269) and demonstrated consistent effect size across cohorts (Tables 2, 3, Figure 2d, and Supplementary Tables S7, S8).

To better delineate the risk associated with the MHC region and detect potential functional variants, we imputed classical HLA alleles in the discovery cohort (Supplementary Table S9). This demonstrated a genome-wide significant association with a protein-altering variant of known functional significance, the DQB1*0602 allele (OR = 0.47, p = 6.6 × 10^{-9}). DQB1*602 is in strong LD with another functional allele, DRB1*1501, but conditional analysis suggested that DQB1*602 best explains this association signal (Supplementary Table S10). The strength of the DQB1*602 association is probably underestimated due to the limitations of current imputation algorithms (sensitivity of 56.6% for detection of the DQB1*602 allele, Supplementary Table S11).
A major signal outside the MHC locus resided in a 100-kb segment on Chr. 1q31-q32.1 containing complement factor H (CFH) and the related CFHR3, CFHR1, CFHR4, CHFR2, CFHR5 genes (rs6677604, OR = 0.68, p = 3.0 × 10^{-10} in the combined cohort). This locus was also the top signal in our genome-wide CNP analysis (Supplementary Figure 4, Supplementary Table S12). The top SNP, rs6677604, is located in intron 12 of CFH and is supported by multiple highly correlated SNPs (Figure 3a, Table 2). After controlling for rs6677604, there were no other independent signals in the entire CFH region. The association results at rs6677604 were far less significant under a recessive model (p=5.6 × 10^{-5}), supporting an additive risk. The rs6677604-A allele is protective in all three cohorts but has a much higher allele frequency in Europeans (0.23 in European controls vs. 0.07 in Chinese controls, Table 2). This allele perfectly tags a common deletion spanning the CFHR1 and CFHR3 genes (CFHR1,3Δ22,23). We confirmed the association of rs6677604-A allele with CFHR1,3=Δ in our cohort: PCR of multiple amplicons within CFHR1 and CFHR3 failed and the CFHR1 protein could not be detected in serum from all A/A homozygotes tested (Supplementary Figure S5). We carefully evaluated evidence for association of IgAN with alleles in CFH that confer risk of macular degeneration (AMD) and found no contribution to risk (e.g., the Y402H variant, tagged by rs10801555, showed OR=1.0, p=0.99 in discovery cohort; Figure 3b). Haplotype-based analysis in the Beijing discovery cohort demonstrated protection by the haplotype containing the rs6677604-A allele (OR= 0.56, p=1×10^{-6} vs. all other haplotypes in the discovery cohort, Figure 3b, Supplementary Figure S6) but no significant effect of other haplotypes.

The fifth signal in the GWAS resided in an intronic SNP in HORMAD2 on Chr. 22.q12.2 (rs2412971, OR = 0.80, p = 1.9 × 10^{-9}) and was supported by a second SNP within 35kb of this signal (rs2412973, OR = 0.80, p = 4.5 × 10^{-9}). After controlling for rs2412971, there were no other independent signals in this region. The association extends across a large LD segment that encompasses genes including HORMAD2, MTMR3, LIF, and OSM (Figure 3c).

Cumulative effects on disease risk

To determine the cumulative risk conferred by these loci, we computed a genetic risk score, calculated as the weighted sum of the number of protective alleles multiplied by the log of the odds ratio for each of the individual loci (Table 4, Supplementary Table S13, S14). The disease risk varied up to 10-fold between individuals with no protective alleles compared those with five or more. The risk score model was similar in all cohorts and collectively explained 5–7% of the variation in disease risk in the Chinese cohorts and ~4% of the risk in the European cohort (Table 4). The risk score did not reproducibly correlate with any of the parameters of disease severity, such as estimated GFR, degree of proteinuria, or histologic severity grade.

Most interestingly, consistent with the known higher prevalence of IgAN in Asians, the frequency of protective alleles was significantly lower in the Chinese cohort compared to the European group. The differences in the distribution of protective alleles were highly significant between the Asian and European cohorts (Figure 4a, p = 4.8 × 10^{-72} and p = 6.4 × 10^{-60} for differences within cases and controls, respectively). To confirm this finding in independent populations, we examined three HapMap groups and similarly found that frequencies of risk alleles correlate with disease frequency among these populations: risk allele frequencies were highest in Asians, intermediate in Europeans, and lowest in Africans (Figure 4b, Supplementary Figure S7). For example, the protective allele at the chromosome 1 locus shows a frequency of 0.08 in Asians, 0.24 in Europeans and 0.49 in Africans.
In this GWAS, we identified five loci imparting significant and consistent effects on the risk of IgAN across three independent cohorts. These five loci explained up to a ten-fold variation in interindividual risk and cumulatively accounted for 4–7% of the disease variance. The effect sizes at these loci are relatively large and consistent across the European and Chinese cohorts, with four having inverse OR ≥ 1.4, which is comparable to those detected in previous studies of autoimmune or inflammatory diseases\(^\text{21,24–30}\). The risk allele frequencies also strongly paralleled the prevalence of IgAN among different populations.

We detected a major signal in the MHC region, which was identified but not localized in a recent GWAS with 533 affected subjects\(^\text{19}\). Close scrutiny in the markedly larger cohorts reported here revealed that this signal originated from three distinct loci within HLA and we also identified two additional non-HLA loci. Evidence supporting the presence of three independent risk loci on Chr. 6p21 includes their position within distinct LD segments, as well as genome-wide significance after conditioning for the other two loci, with consistent effects within each cohort.

The strongest HLA signal was in the \textit{HLA-DRB1/DQB1} region. Imputation of classical alleles suggested that this signal is fully or partially conveyed by a strong protective effect of the DRB1*1501-DQB1*0602 haplotype; the strength of this association was likely underestimated by limitations of imputation. This haplotype is relatively common in the European and Asian populations (frequency ~ 0.1–0.2) and in contrast to its protective effect for IgAN has been associated with increased risk of SLE\(^\text{31}\), multiple sclerosis\(^\text{31}\), narcolepsy\(^\text{32}\) and hepatotoxicity from COX2 inhibitors\(^\text{30}\) but is also highly protective for type I diabetes mellitus\(^\text{26}\). This haplotype is also protective in selective IgA deficiency\(^\text{27}\), yet we found no association with IgA levels at this locus among cases (Supplementary Table S15). This region has a complex LD structure, and our conditional analysis suggests the possibility of an independent signal within this region (at rs9275424, Supplementary Table S7, S8). High-resolution mapping and direct genotyping of classical alleles will be required to further dissect this interval and identify the functional variant(s).

The second independent interval at 6p21 contained \textit{TAP2, TAP1, PSMB8,} and \textit{PSMB9}, interferon-regulated genes that have been implicated in antigen generation and processing for presentation by MHC I molecules; they also play an important role in modulation of cytokine production and cytotoxic T-cell response\(^\text{33,34}\). \textit{PSMB8} expression is increased in PBMCs from IgAN patients, motivating further investigation\(^\text{35}\). To our knowledge, this locus has not been identified in any prior GWAS.

The third signal at 6p21 comprised the \textit{HLA-DPA1, -DPB1,} and \textit{-DPB2} genes. This locus is associated with risk of chronic hepatitis B infection\(^\text{29}\) (a major clinical problem in China) and systemic sclerosis stratified for anti-DNA topoisomerase I or anticientromere autoantibodies\(^\text{31}\), but the risk alleles associated with these phenotypes are not in LD with any of the IgAN risk alleles.

The \textit{CFH} protein plays a critical role in dampening the alternative complement cascade via inhibition of the C3 and C5 convertases\(^\text{36}\). The functions of the \textit{CFH}-related proteins are less well understood\(^\text{36,37}\). Loss of function mutations in \textit{CFH} produce uncontrolled C3 activation, leading to membranoproliferative glomerulonephritis type II, which is pathologically distinct from IgAN\(^\text{36}\). Other rare \textit{CFH} mutations can produce hemolytic uremic syndrome, a thrombotic disorder\(^\text{36}\), while distinct common haplotypes predispose to AMD and susceptibility to meningococcal infection\(^\text{22–24}\). Interestingly, the \textit{CFH} haplotype bearing the \textit{CFHR1,3} variant may be protective in AMD, but detection of an independent effect has been complicated owing to the presence of additional haplotypes imparting both
high and low risk. Here, we found an unambiguous protective effect of the CFHR1,3-Δ-containing haplotype in IgAN, strongly suggesting that CFHR1,3-Δ is the functional variant. Nevertheless, it is not clear how loss of CFHR1 and/or CFHR3 may confer protection for IgAN. The protective effects may be due to the competing roles of CFH and CFHR1 proteins, such that loss of CFHR1 enhances CFH effects, reducing inflammation at tissue surfaces.

The Chr. 22q12.2 locus spans a large interval that contains OSM and LIF, encoding cytokines implicated in mucosal immunity and inflammation. Of particular interest, inactivation of Osm results in autoimmune glomerulonephritis in the mouse. The functions of other genes such as HORMAD2 and MTMR3 have not been as well characterized. Interestingly, the rs2412973-A allele, which is protective for IgAN, has also been associated with increased risk of early-onset inflammatory bowel disease (IBD) and altered expression of MTMR3 expression in individuals with ulcerative colitis. This finding is of interest given the known clinical association between IBD and secondary forms of IgAN, but the underlying signal within this locus remains to be clarified. Lastly, the protective allele at this locus is also associated with lower serum IgA levels among cases (p = 3.9 x 10^-3, Supplementary Table S15, Supplementary Figure S8).

It is noteworthy that many of the protective alleles for IgAN have been implicated as risk factors other immune-mediated and infectious disorders, demonstrating that complex selection pressures (potentially balancing selection) influence the frequencies of these alleles among world populations. Statistical proof of balancing selection on allele frequencies or genotypes may be particularly challenging if alleles have been maintained in the population over very long evolutionary periods. Interestingly, a recent genome-wide survey detected a signal of selection in the vicinity of the CFH gene cluster and there is a large difference in the frequency of the rs6677604-A allele among world populations (Supplementary Table S16).

The loci identified in this study provide significant insight into the genetic architecture of sporadic IgAN, identifying novel pathogenic pathways and connections to other immune-mediated disorders. Based on our power calculations, we identified virtually all loci imparting an OR ≥1.5 in the Chinese discovery cohort but additional loci with large effects may be present among Europeans. Considering the effectiveness of GWAS for studies of immunologic disorders and the increased power imparted by larger sample size, genome-wide examination of larger cohorts will likely define additional genetic components of IgA nephropathy.

**Supplementary Material**

Refer to Web version on PubMed Central for supplementary material.

**Acknowledgments**

We are grateful to all study participants for their contribution to this work. We also thank the staff of the Yale West Campus Center for Genome Analysis for their excellent support. The authors also appreciate the assistance of Catherine V. Barker and Susan Y. Woodford with sample collection. This study was supported by RC1DK087445 (AGG, RPL), R01DK082753 (AGG, JN, BAJ, RJW), and KL2 RR24157 (KK), the Center for Glomerular Diseases at Columbia University, the Yale Center Translational Science Award, and the Yale Center for Human Genetics and Genomics. RPL is an investigator of the Howard Hughes Medical Institute.

**References**


**Figure 1. Manhattan plot of p-values for SNP associations to IgAN**
Plot of the observed p-values versus chromosomal location; highlighted are the ten independent loci followed up in additional cohorts. The dashed line corresponds to the follow-up threshold.
Figure 2. High resolution view of the MHC locus
The X-axis represents physical distance (kb). The left Y-axis represent the −log(p-values) for the association statistics. The −log(p-values) in the discovery and combined cohorts are shown as blue circles and red diamonds, respectively. The right Y-axis represents the average recombination rates based on the phased HapMap haplotypes. The recombination rates are shown by the light blue line. (a) The three intervals associated with IgA nephropathy reside within a 0.54 Mb segment on chromosome 6. The shaded areas correspond to regional plots in lower panels; (b) Regional plot for the interval containing HLA-DQB1, DQA1, and DRB1. The classical HLA alleles imputed in the discovery cohort (green triangles) formed a protective haplotype DQB1*0602-DQA1*0102-DRB1*1501. (c) Regional plot for the second MHC interval: SNPs typed in the combined cohorts reside within the PSMB8 gene. (d) Regional plot for the HLA-DPB2, DPB1, and DPA1 interval. The lower panels for (b–d) represent LD heatmaps (D') calculated based on the actual genotype data of the Beijing cohort.
Figure 3. Analysis of the Chr. 1 and Chr 22. loci

(a) Regional association plot of the chromosome 1q32 locus; while the most strongly associated SNP resides within the CFH gene, it is a perfect proxy for CFHR1,3Δ. The lower panel represents the LD heatmap (D') calculated based on the genotype data of the Beijing cohort. (b) Haplotype analysis revealed five common haplotypes (H-1 to H-5) in the Beijing discovery cohort (freq. > 0.01). The haplotype frequencies, corresponding tag-SNPs and reported disease associations are shown22–24,36,37,41,43. The H2 haplotype perfectly tags CFHR1,3Δ. The odds ratios (ORs) and 95% confidence intervals (95% CIs) are calculated in reference to H-1, which has an identical frequency among cases and controls. *** p=7.7 x 10^-6 for comparison of H-2 versus all other haplotypes. (c) Regional association plot of the chromosome 22 locus: the strongest association stems from the SNPs residing within HORMAD2, but the area of association spans over ~ 0.7 Mb region containing multiple genes.
Figure 4. Differences in the distributions of protective alleles by ethnicity
(a) Distributions of protective alleles by ethnicity and case-control status. Numbers of protective alleles were scored for the combined Asian (N=3,556) and European (N=2,410) cohorts. Europeans harbor much greater numbers of protective alleles. The differences in the distribution of protective alleles between Asians and Europeans are highly significant within both case and control groups (Chi-square p=4.9 × 10^{-72} and p=6.4 × 10^{-60} for cases and controls, respectively). (b) Distributions of protective alleles among the three HapMap populations: there were highly significant differences between Asian (CHB+JPT) vs. Europeans (CEU, p=1.3×10^{-3}) and Asian vs. Yorubans (YRI, p=7.1×10^{-6}) populations.
### Table 1

Summary of Study Cohorts

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Ethnicity</th>
<th>Genotyped Cases</th>
<th>Genotyped Controls</th>
<th>After QC Cases</th>
<th>After QC Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discovery Cohort</td>
<td>Han Chinese</td>
<td>1,228</td>
<td>966</td>
<td>1,194</td>
<td>902</td>
</tr>
<tr>
<td>Follow-up Cohort 1</td>
<td>Han Chinese</td>
<td>740</td>
<td>750</td>
<td>712</td>
<td>748</td>
</tr>
<tr>
<td>Follow-up Cohort 2</td>
<td>European</td>
<td>1,273</td>
<td>1,201</td>
<td>1,238</td>
<td>1,172</td>
</tr>
<tr>
<td><strong>All Cohorts Combined:</strong></td>
<td></td>
<td><strong>3,241</strong></td>
<td><strong>2,917</strong></td>
<td><strong>3,144</strong></td>
<td><strong>2,822</strong></td>
</tr>
</tbody>
</table>
### Table 2

**Association results for 10 SNPs representing 5 independent regions that reach genome-wide significance in combined analyses.**

<table>
<thead>
<tr>
<th>Chr</th>
<th>Location (kb)</th>
<th>SNP (minor allele)</th>
<th>Beijing Discovery Cohort[^a] N = 2,096 (1,194 cases / 902 controls)</th>
<th>Shanghai Replication Cohort[^b] N = 1,460 (712 cases / 748 controls)</th>
<th>European Replication Cohort[^b] N = 2,410 (1,238 cases / 1,172 controls)</th>
<th>All Cohorts Combined[^b] N = 5,966 (3,144 cases / 2,822 controls)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>194,918</td>
<td>rs3766404 (C)</td>
<td>MAF (cases / controls): 0.052 / 0.086</td>
<td>OR: 0.59</td>
<td>P-value: 1.84 × 10^-5</td>
<td>MAF (cases / controls): 0.078 / 0.080</td>
</tr>
<tr>
<td>1</td>
<td>194,953</td>
<td>rs6677604 (A)</td>
<td>MAF (cases / controls): 0.061 / 0.073</td>
<td>OR: 0.49</td>
<td>P-value: 8.18 × 10^-2</td>
<td>MAF (cases / controls): 0.052 / 0.070</td>
</tr>
<tr>
<td>6</td>
<td>32,778</td>
<td>rs2856717 (T)</td>
<td>MAF (cases / controls): 0.078 / 0.086</td>
<td>OR: 0.66</td>
<td>P-value: 3.31 × 10^-8</td>
<td>MAF (cases / controls): 0.078 / 0.080</td>
</tr>
<tr>
<td>6</td>
<td>32,789</td>
<td>rs9275596 (C)</td>
<td>MAF (cases / controls): 0.063 / 0.070</td>
<td>OR: 0.56</td>
<td>P-value: 1.51 × 10^-4</td>
<td>MAF (cases / controls): 0.078 / 0.080</td>
</tr>
<tr>
<td>6</td>
<td>32,917</td>
<td>rs9357155 (A)</td>
<td>MAF (cases / controls): 0.078 / 0.080</td>
<td>OR: 0.69</td>
<td>P-value: 1.91 × 10^-12</td>
<td>MAF (cases / controls): 0.12 / 0.14</td>
</tr>
<tr>
<td>6</td>
<td>32,919</td>
<td>rs2071543 (A)</td>
<td>MAF (cases / controls): 0.078 / 0.080</td>
<td>OR: 0.64</td>
<td>P-value: 1.79 × 10^-5</td>
<td>MAF (cases / controls): 0.14 / 0.14</td>
</tr>
<tr>
<td>6</td>
<td>32,922</td>
<td>rs1883414 (T)</td>
<td>MAF (cases / controls): 0.080 / 0.080</td>
<td>OR: 0.66</td>
<td>P-value: 3.55 × 10^-2</td>
<td>MAF (cases / controls): 0.12 / 0.14</td>
</tr>
<tr>
<td>6</td>
<td>32,923</td>
<td>rs3129269 (T)</td>
<td>MAF (cases / controls): 0.080 / 0.080</td>
<td>OR: 0.64</td>
<td>P-value: 1.59 × 10^-5</td>
<td>MAF (cases / controls): 0.14 / 0.14</td>
</tr>
<tr>
<td>22</td>
<td>28,824</td>
<td>rs2412971 (A)</td>
<td>MAF (cases / controls): 0.078 / 0.080</td>
<td>OR: 0.66</td>
<td>P-value: 2.17 × 10^-4</td>
<td>MAF (cases / controls): 0.12 / 0.14</td>
</tr>
<tr>
<td>22</td>
<td>28,859</td>
<td>rs2412973 (A)</td>
<td>MAF (cases / controls): 0.078 / 0.080</td>
<td>OR: 0.66</td>
<td>P-value: 8.26 × 10^-4</td>
<td>MAF (cases / controls): 0.12 / 0.14</td>
</tr>
</tbody>
</table>

Q: p-value for the Cochran's Q statistic;  
The per-allele, heterozygote and homozygote ORs are indicated for the combined cohort.  
[^a] Cochran-Armitage trend test;  
[^b] Stratified analysis using Mantel's extension of Cochran-Armitage trend test;  
[^c] Significant heterogeneity (P<0.05).
### Table 3

Stepwise conditional analysis of association among the signals in the HLA region.

<table>
<thead>
<tr>
<th>Test SNP</th>
<th>Conditioning SNP(s)</th>
<th>Beijing Discovery Cohort N = 2,096 (1,194 cases / 902 controls)</th>
<th>Shanghai Follow-up Cohort N = 1,460 (712 cases / 748 controls)</th>
<th>European Follow-up Cohort N = 2,410 (1,238 cases / 1,172 controls)</th>
<th>All Cohorts Combined N = 5,966 (3,144 cases / 2,822 controls)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2856717</td>
<td>rs9275596</td>
<td>Unconditioned p-value: 3.30 × 10⁻⁸</td>
<td>Conditioned p-value: 0.236</td>
<td>1.51 × 10⁻⁴</td>
<td>0.225</td>
</tr>
<tr>
<td>rs9275596</td>
<td>rs2856717</td>
<td>Unconditioned p-value: 1.91 × 10⁻¹²</td>
<td>Conditioned p-value: NA</td>
<td>6.29 × 10⁻⁸</td>
<td>NA</td>
</tr>
<tr>
<td>rs9357155</td>
<td>rs1883414</td>
<td>Unconditioned p-value: 5.19 × 10⁻⁶</td>
<td>Conditioned p-value: 2.29 × 10⁻³</td>
<td>1.79 × 10⁻⁵</td>
<td>3.12 × 10⁻⁴</td>
</tr>
<tr>
<td>rs1883414</td>
<td>rs2856717</td>
<td>Unconditioned p-value: 1.32 × 10⁻⁵</td>
<td>Conditioned p-value: 2.16 × 10⁻⁴</td>
<td>0.0348</td>
<td>0.164</td>
</tr>
<tr>
<td>rs2856717</td>
<td>rs9275596</td>
<td>Unconditioned p-value: 3.30 × 10⁻⁸</td>
<td>Conditioned p-value: 0.236</td>
<td>1.51 × 10⁻⁴</td>
<td>0.225</td>
</tr>
<tr>
<td>rs9275596</td>
<td>rs2856717</td>
<td>Unconditioned p-value: 1.91 × 10⁻¹²</td>
<td>Conditioned p-value: NA</td>
<td>6.29 × 10⁻⁸</td>
<td>NA</td>
</tr>
<tr>
<td>rs9357155</td>
<td>rs1883414</td>
<td>Unconditioned p-value: 5.19 × 10⁻⁶</td>
<td>Conditioned p-value: 2.29 × 10⁻³</td>
<td>1.79 × 10⁻⁵</td>
<td>3.12 × 10⁻⁴</td>
</tr>
<tr>
<td>rs1883414</td>
<td>rs2856717</td>
<td>Unconditioned p-value: 1.32 × 10⁻⁵</td>
<td>Conditioned p-value: 2.16 × 10⁻⁴</td>
<td>0.0348</td>
<td>0.164</td>
</tr>
</tbody>
</table>

rs2856717 and rs9275596 represent the major HLA signal near DQB1. rs9357155 and rs1883414 represent the other two independent signals in the HLA region.
## Table 4

Cumulative effect of replicated loci stratified by the number of protective alleles.

<table>
<thead>
<tr>
<th>No. of Protective Alleles</th>
<th>Frequency (Cases / Controls)</th>
<th>Average Risk Score (+/- SD)</th>
<th>OR (95% CI)</th>
<th>Average Risk Score (+/- SD)</th>
<th>OR (95% CI)</th>
<th>Average Risk Score (+/- SD)</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (highest risk)</td>
<td>0.17 / 0.07</td>
<td>0.00</td>
<td>1.00 (reference)</td>
<td>0.24 / 0.13</td>
<td>0.00</td>
<td>1.00 (reference)</td>
<td>0.07 / 0.03</td>
</tr>
<tr>
<td>1</td>
<td>0.31 / 0.26</td>
<td>-0.37 (+/-0.09)</td>
<td>0.50 (0.36–0.69)</td>
<td>0.38 / 0.32</td>
<td>-0.30 (+/-0.15)</td>
<td>0.66 (0.48–0.90)</td>
<td>0.19 / 0.12</td>
</tr>
<tr>
<td>2</td>
<td>0.29 / 0.29</td>
<td>-0.77 (+/-0.14)</td>
<td>0.40 (0.29–0.56)</td>
<td>0.24 / 0.31</td>
<td>-0.65 (+/-0.23)</td>
<td>0.43 (0.31–0.60)</td>
<td>0.26 / 0.24</td>
</tr>
<tr>
<td>3</td>
<td>0.16 / 0.20</td>
<td>-1.17 (+/-0.15)</td>
<td>0.31 (0.22–0.44)</td>
<td>0.10 / 0.14</td>
<td>-1.06 (+/-0.26)</td>
<td>0.40 (0.27–0.60)</td>
<td>0.26 / 0.30</td>
</tr>
<tr>
<td>4</td>
<td>0.06 / 0.12</td>
<td>-1.61 (+/-0.17)</td>
<td>0.20 (0.13–0.31)</td>
<td>0.04 / 0.08</td>
<td>-1.44 (+/-0.28)</td>
<td>0.28 (0.16–0.47)</td>
<td>0.15 / 0.19</td>
</tr>
<tr>
<td>≥5 (lowest risk)</td>
<td>0.01 / 0.06</td>
<td>-2.11 (+/-0.25)</td>
<td>0.09 (0.05–0.16)</td>
<td>0.004 / 0.03</td>
<td>-1.86 (+/-0.36)</td>
<td>0.10 (0.08–0.33)</td>
<td>0.08 / 0.13</td>
</tr>
<tr>
<td>OR change highest vs. lowest risk</td>
<td>11.1</td>
<td>10.0</td>
<td>4.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**P-value**

- **a** the risk scores were calculated based on the odds ratios and allele frequencies for each specific cohort. Only individuals with non-missing genotypes for all 10 alleles were included in this analysis.

- **b** Fold-change in odds ratio between highest and lowest risk group.

- **c** P-value for the risk score prediction model.

- **d** The C-statistic indicates the area under the receiver operating characteristic (ROC) curve for the risk score prediction model.

- **e** Nagelkerke's pseudo $R^2$ indicates the fraction of the variance in risk explained by the risk score model.