Genome-wide association study identifies susceptibility loci for IgA nephropathy

(Article begins on next page)

22 September 2024

Genome-wide Association Study Identifies Susceptibility Loci for IgA Nephropathy

Ali G. Gharavi[1], **Krzysztof Kiryluk**[1], **Murim Choi**[2], **Yifu Li**[1], **Ping Hou**[1,3], **Jingyuan Xie**[1,4], **Simone Sanna-Cherchi**[1], **Clara J. Men**[2], **Bruce A. Julian**[5], **Robert J. Wyatt**[6], **Jan Novak**[5], **John C. He**[7], **Haiyan Wang**[3], **Jicheng Lv**[3], **Li Zhu**[3], **Weiming Wang**[4], **Zhaohui Wang**[4], **Kasuhito Yasuno**[2], **Murat Gunel**[2], **Shrikant Mane**[2,8], **Sheila Umlauf**[2,8], **Irina Tikhonova**[2,8], **Isabel Beerman**[2], **Silvana Savoldi**[9], **Riccardo Magistroni**[10], **Gian Marco Ghiggeri**[11], **Monica Bodria**[11], **Francesca Lugani**[1,11], **Pietro Ravani**[12], **Claudio Ponticelli**[13], **Landino Allegri**[14], **Giuliano Boscutti**[15], **Giovanni Frasca**[16], **Alessandro Amore**[17], **Licia Peruzzi**[17], **Rosanna Coppo**[17], **Claudia Izzi**[18], **Fabio Viola**[19], **Elisabetta Prati**[20], **Maurizio Salvadori**[21], **Renzo Mignani**[22], **Loreto Gesualdo**[23], **Francesca Bertinetto**[24], **Silvana Paola Mesiano**[24], **Antonio Amoroso**[24], **Francesco Scolari**[18], **Nan Chen**[4], **Hong Zhang**[3], and **Richard P. Lifton**[2]

[1]Department of Medicine, Columbia University College of Physicians and Surgeons, New York, New York, USA

[2]Department of Genetics, Howard Hughes Medical Institute, Yale University School of Medicine, New Haven, CT, USA

[3]Renal Division, Peking University First Hospital; Peking University, Institute of Nephrology; Key Laboratory of Renal Disease, Ministry of Health of China, Beijing, PR China

[4]Department of Nephrology, Ruijin Hospital, Shanghai Jiaotong University, School of Medicine, Shanghai 200025, P.R. China

[5]Departments of Microbiology and Medicine, University of Alabama at Birmingham, Birmingham, AL, USA

[6]Children'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

[7]Department of Medicine, Mount Sinai School of Medicine, New York, NY 10029

[8]Yale Center for Genome analysis, Yale University School of Medicine, New Haven, CT, USA

[9]Nephrology and Dialysis Unit, Ciriè Hospital, Torino, Italy

[10]Division of Nefrologia Dialisi e Trapianto, Azienda Ospedaliero-Universitaria Policlinico di Modena, Modena, Italy

[11]Laboratory on Pathophysiology of Uremia Istituto Giannina Gaslini, Genova, Italy

[12]University of Calgary, Alberta, CA

[13]Divisione di Nefrologia e Dialisi, Istituto Scientifico Humanitas, Milan, Italy

[14]Department of Clinical Medicine, Nephrology and Health Science, Section of Nephrology, University of Parma, Parma, Italy

[15]Department of Nephrology, Ospedale di Gorizia, Gorizia, Italy

[16]Nephrology and Dialysis Unit - Ospedali Riuniti, Ancona, Italy

[17]Nephrology, Dialysis and Transplantation, Regina Margherita University Hospital, Turin, Italy

[18]University of Brescia and Second Division of Nephrology, Montichiari Hospital, Montichiari (Brescia), Italy

[19]Division of Nephrology, Spedali Civili, Brescia; Brescia Italy

[20]Dialysis Center, Ospedale di Desenzano, Desenzano del Garda (Brescia), Italy

[21]Renal Unit Careggi University Hospital, Florence Italy

[22]Division of Nephrology, Ospedale degli Infermi, Rimini, Italy

[23]Department of Biomedical Sciences, University of Foggia, Foggia, Italy

[24]Department of Genetics, Biology and Biochemistry, University of Torino, Torino, Italy

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[1–3]. 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 Armitrage (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 (λ) 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 5×10^{-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 (5×10^{-6} to 1×10^{-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 1.3×10^{-5}, corresponding to a q-value 0.10

(Supplementary Figure S3)[20]. 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 ($p < 5 \times 10^{-8}$) 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 $\times$ 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[21] (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 (*CFHR1,3Δ*)[22,23]. We confirmed the association of rs6677604-A allele with *CFHR1,3*g=D 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.

DISCUSSION

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[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[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 *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[25], multiple sclerosis[31], narcolepsy[32] and hepatotoxicity from COX2 inhibitors[30] but is also highly protective for type I diabetes mellitus[26]. This haplotype is also protective in selective IgA deficiency[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 *TAP2, TAP1, PSMB8,* and *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[33,34]. *PSMB8* expression is increased in PBMCs from IgAN patients, motivating further investigation [35]. To our knowledge, this locus has not been identified in any prior GWAS.

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

The CFH protein plays a critical role in dampening the alternative complement cascade via inhibition of the C3 and C5 convertases[36]. The functions of the CFH-related proteins are less well understood[36,37]. Loss of function mutations in *CFH* produce uncontrolled C3 activation, leading to membranoproliferative glomerulonephritis type II, which is pathologically distinct from IgAN[36]. Other rare *CFH* mutations can produce hemolytic uremic syndrome, a thrombotic disorder[36], while distinct common haplotypes predispose to AMD and susceptibility to meningococcal infection[22–24]. Interestingly, the *CFH* haplotype bearing the *CFHR1,3Δ* 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[22,23]. Here, we found an unambiguous protective effect of the *CFHR1,3Δ*-containing haplotype in IgAN, strongly suggesting that *CFHR1,3Δ* 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[37], 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[38]. The functions of other genes such as *HORMAD2* and *MTMR3* have not been as well characterized[39]. 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[28]. 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 = 3.9×10^{-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[40] 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[22,27–31,41] and the increased power imparted by larger sample size[42], 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

References

1. Coresh J, et al. Prevalence of chronic kidney disease in the United States. JAMA. 2007; 298:2038–47. [PubMed: 17986697]

2. Tsukamoto Y, et al. Report of the Asian Forum of Chronic Kidney Disease Initiative (AFCKDI) 2007. "Current status and perspective of CKD in Asia": diversity and specificity among Asian countries. Clin Exp Nephrol. 2009; 13:249–56. [PubMed: 19288169]

3. Gesualdo L, Di Palma AM, Morrone LF, Strippoli GF, Schena FP. The Italian experience of the national registry of renal biopsies. Kidney Int. 2004; 66:890–4. [PubMed: 15327376]

4. D'Amico G. The commonest glomerulonephritis in the world: IgA nephropathy. Q J Med. 1987; 64:709–27. [PubMed: 3329736]

5. Nair R, Walker PD. Is IgA nephropathy the commonest primary glomerulopathy among young adults in the USA? Kidney Int. 2006; 69:1455–8. [PubMed: 16531983]

6. Varis J, et al. Immunoglobulin and complement deposition in glomeruli of 756 subjects who had committed suicide or met with a violent death. J Clin Pathol. 1993; 46:607–10. [PubMed: 8157744]

7. Suzuki K, et al. Incidence of latent mesangial IgA deposition in renal allograft donors in Japan. Kidney Int. 2003; 63:2286–94. [PubMed: 12753320]

8. Kiryluk K, et al. Genetic studies of IgA nephropathy: past, present, and future. Pediatr Nephrol. 2010

9. Barratt J, Feehally J. IgA nephropathy. J Am Soc Nephrol. 2005; 16:2088–97. [PubMed: 15930092]

10. Hastings MC, et al. Galactose-Deficient IgA1 in African Americans with IgA Nephropathy: Serum Levels and Heritability. Clin J Am Soc Nephrol. 2010

11. Gharavi AG, et al. Aberrant IgA1 glycosylation is inherited in familial and sporadic IgA nephropathy. J Am Soc Nephrol. 2008; 19:1008–14. [PubMed: 18272841]

12. Lin X, et al. Aberrant galactosylation of IgA1 is involved in the genetic susceptibility of Chinese patients with IgA nephropathy. Nephrol Dial Transplant. 2009; 24:3372–5. [PubMed: 19531670]

13. Moldoveanu Z, et al. Patients with IgA nephropathy have increased serum galactose-deficient IgA1 levels. Kidney Int. 2007; 71:1148–54. [PubMed: 17342176]

14. Mestecky J, et al. Defective galactosylation and clearance of IgA1 molecules as a possible etiopathogenic factor in IgA nephropathy. Contrib Nephrol. 1993; 104:172–82. [PubMed: 8325028]

15. Tomana M, et al. Circulating immune complexes in IgA nephropathy consist of IgA1 with galactose-deficient hinge region and antiglycan antibodies. J Clin Invest. 1999; 104:73–81. [PubMed: 10393701]

16. Gharavi AG, et al. IgA nephropathy, the most common cause of glomerulonephritis, is linked to 6q22-23. Nat Genet. 2000; 26:354–7. [PubMed: 11062479]

17. Bisceglia L, et al. Genetic heterogeneity in Italian families with IgA nephropathy: suggestive linkage for two novel IgA nephropathy loci. Am J Hum Genet. 2006; 79:1130–4. [PubMed: 17186473]

18. Paterson AD, et al. Genome-wide linkage scan of a large family with IgA nephropathy localizes a novel susceptibility locus to chromosome 2q36. J Am Soc Nephrol. 2007; 18:2408–15. [PubMed: 17634434]

19. Feehally J, et al. HLA Has Strongest Association with IgA Nephropathy in Genome-Wide Analysis. J Am Soc Nephrol. 2010

20. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A. 2003; 100:9440–5. [PubMed: 12883005]

21. de Bakker PI, et al. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. Nat Genet. 2006; 38:1166–72. [PubMed: 16998491]

22. Hughes AE, et al. A common CFH haplotype, with deletion of CFHR1 and CFHR3, is associated with lower risk of age-related macular degeneration. Nat Genet. 2006; 38:1173–7. [PubMed: 16998489]

23. Raychaudhuri S, et al. Associations of CFHR1-CFHR3 deletion and a CFH SNP to age-related macular degeneration are not independent. Nat Genet. 2010; 42:553–5. author reply 555–6. [PubMed: 20581873]

24. Davila S, et al. Genome-wide association study identifies variants in the CFH region associated with host susceptibility to meningococcal disease. Nat Genet. 2010; 42:772–6. [PubMed: 20694013]

25. Barcellos LF, et al. High-density SNP screening of the major histocompatibility complex in systemic lupus erythematosus demonstrates strong evidence for independent susceptibility regions. PLoS Genet. 2009; 5:e1000696. [PubMed: 19851445]

26. Erlich H, et al. HLA DR-DQ haplotypes and genotypes and type 1 diabetes risk: analysis of the type 1 diabetes genetics consortium families. Diabetes. 2008; 57:1084–92. [PubMed: 18252895]

27. Ferreira RC, et al. Association of IFIH1 and other autoimmunity risk alleles with selective IgA deficiency. Nat Genet. 2010; 42:777–80. [PubMed: 20694011]

28. Imielinski M, et al. Common variants at five new loci associated with early-onset inflammatory bowel disease. Nat Genet. 2009; 41:1335–40. [PubMed: 19915574]

29. Kamatani Y, et al. A genome-wide association study identifies variants in the HLA-DP locus associated with chronic hepatitis B in Asians. Nat Genet. 2009; 41:591–5. [PubMed: 19349983]

30. Singer JB, et al. A genome-wide study identifies HLA alleles associated with lumiracoxib-related liver injury. Nat Genet. 2010; 42:711–4. [PubMed: 20639878]

31. Zhou X, et al. HLA-DPB1 and DPB2 are genetic loci for systemic sclerosis: a genome-wide association study in Koreans with replication in North Americans. Arthritis Rheum. 2009; 60:3807–14. [PubMed: 19950302]

32. Mignot E, et al. Complex HLA-DR and -DQ interactions confer risk of narcolepsy-cataplexy in three ethnic groups. Am J Hum Genet. 2001; 68:686–99. [PubMed: 11179016]

33. Begley GS, Horvath AR, Taylor JC, Higgins CF. Cytoplasmic domains of the transporter associated with antigen processing and P-glycoprotein interact with subunits of the proteasome. Mol Immunol. 2005; 42:137–41. [PubMed: 15488952]

34. Muchamuel T, et al. A selective inhibitor of the immunoproteasome subunit LMP7 blocks cytokine production and attenuates progression of experimental arthritis. Nat Med. 2009; 15:781–7. [PubMed: 19525961]

35. Coppo R, et al. Upregulation of the immunoproteasome in peripheral blood mononuclear cells of patients with IgA nephropathy. Kidney Int. 2009; 75:536–41. [PubMed: 19037255]

36. Atkinson JP, Goodship TH. Complement factor H and the hemolytic uremic syndrome. J Exp Med. 2007; 204:1245–8. [PubMed: 17548524]

37. Heinen S, et al. Factor H-related protein 1 (CFHR-1) inhibits complement C5 convertase activity and terminal complex formation. Blood. 2009; 114:2439–47. [PubMed: 19528535]

38. Esashi E, et al. Oncostatin M deficiency leads to thymic hypoplasia, accumulation of apoptotic thymocytes and glomerulonephritis. Eur J Immunol. 2009; 39:1664–70. [PubMed: 19384873]

39. Wojtasz L, et al. Mouse HORMAD1 and HORMAD2, two conserved meiotic chromosomal proteins, are depleted from synapsed chromosome axes with the help of TRIP13 AAA-ATPase. PLoS Genet. 2009; 5:e1000702. [PubMed: 19851446]

40. Grossman SR, et al. A composite of multiple signals distinguishes causal variants in regions of positive selection. Science. 2010; 327:883–6. [PubMed: 20056855]

41. Maller J, et al. Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. Nat Genet. 2006; 38:1055–9. [PubMed: 16936732]

42. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007; 447:661–78. [PubMed: 17554300]

43. Zipfel PF, et al. Deletion of complement factor H-related genes CFHR1 and CFHR3 is associated with atypical hemolytic uremic syndrome. PLoS Genet. 2007; 3:e41. [PubMed: 17367211]

44. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–75. [PubMed: 17701901]

45. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet. 2006; 38:209–13. [PubMed: 16415888]

46. Clayton D, Leung HT. An R package for analysis of whole-genome association studies. Hum Hered. 2007; 64:45–51. [PubMed: 17483596]

47. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet. 2007; 81:1084–97. [PubMed: 17924348]

48. Gusev A, et al. Whole population, genome-wide mapping of hidden relatedness. Genome Res. 2009; 19:318–26. [PubMed: 18971310]

49. Conrad DF, et al. Origins and functional impact of copy number variation in the human genome. Nature. 2010; 464:704–12. [PubMed: 19812545]

50. Craddock N, et al. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. Nature. 2010; 464:713–20. [PubMed: 20360734]
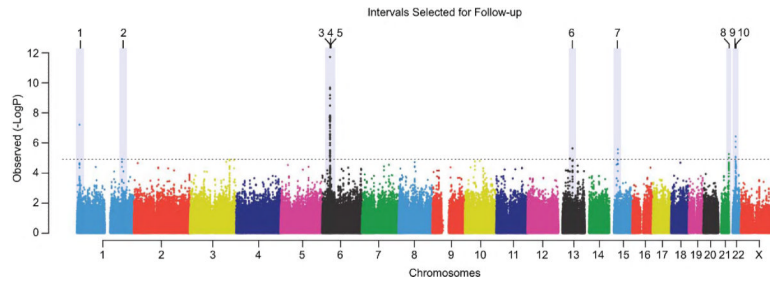
Figure 1. Manhattan plot of p-values for SNP associations to IgAN
Plot of the observed p-values versus chromosomal location; highlighted are the ten independent loci followed up in additional cohorts. The dashed line corresponds to the follow-up threshold.
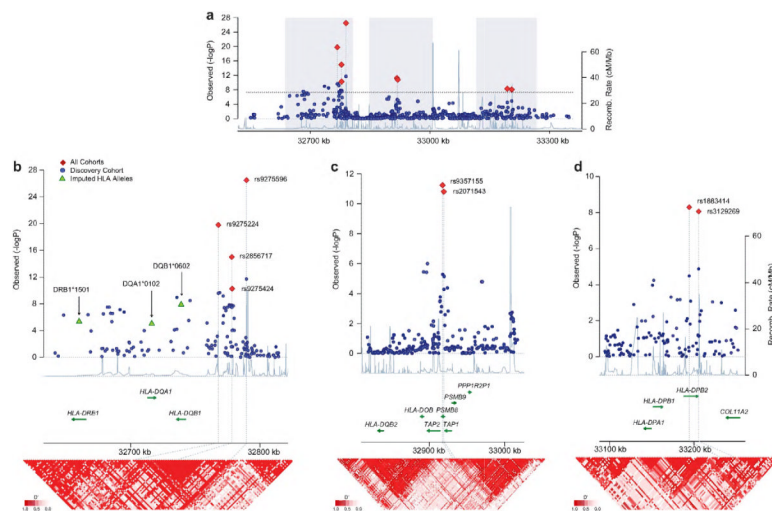
Figure 2. High resolution view of the MHC locus
The X-axis represents physical distance (kb). The left Y-axis represent the −log(p-values) for the association statistics. The −log(p-values) in the discovery and combined cohorts are shown as blue circles and red diamonds, respectively. The right Y-axis 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.
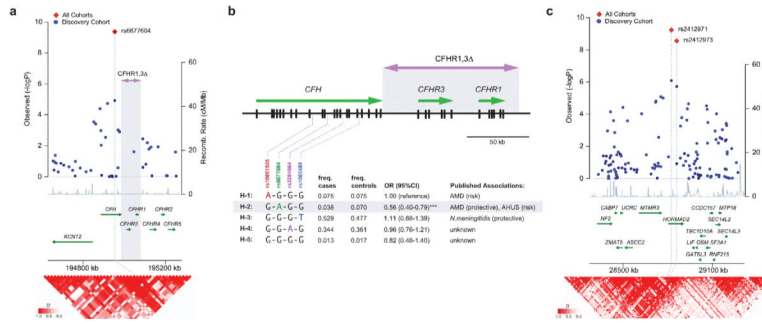
Figure 3. Analysis of the Chr. 1 and Chr 22. loci
(a) Regional association plot of the chromosome 1q32 locus; while the most strongly associated SNP resides within the *CFH* gene, it is a perfect proxy for *CFHR1,3Δ*. The lower panel represents the LD heatmap (D') calculated based on the genotype data 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 shown[22–24,36,37,41,43]. 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 × 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.
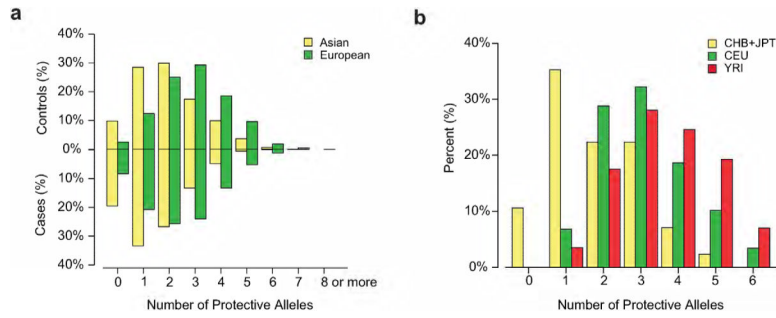
Figure 4. Differences in the distributions of protective alleles by ethnicity
(a) Distributions of protective alleles by ethnicity and case-control status. Numbers of protective alleles were scored for the combined Asian (N=3,556) and European (N=2,410) cohorts. Europeans harbor much greater numbers 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

Cohort	Ethnicity	Genotyped		After QC	
		Cases	Controls	Cases	Controls
Discovery Cohort	Han Chinese	1,228	966	1,194	902
Follow-up Cohort 1	Han Chinese	740	750	712	748
Follow-up Cohort 2	European	1,273	1,201	1,238	1,172
All Cohorts Combined:		**3,241**	**2,917**	**3,144**	**2,822**

NIH-PA Author Manuscript

Table 2

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

Chr	Location (kb)	SNP (minor allele)	Beijing Discovery Cohort[a] N = 2,096 (1,194 cases / 902 controls)			Shanghai Replication Cohort[a] N = 1,460 (712 cases / 748 controls)			European Replication Cohort[b] N = 2,410 (1,238 cases / 1,172 controls)			All Cohorts Combined[b] N = 5,966 (3,144 cases / 2,822 controls)				
			MAF (cases /controls)	OR	P-value	MAF (cases /controls)	OR	P-value	MAF (cases /controls)	OR	P-value	Per allele	OR Het.	Hom.	P-value	Q
1	194,918	rs3766404 (C)	0.052 / 0.086	0.59	1.84×10^{-5}	0.078 / 0.080	0.98	8.18×10^{-1}	0.12 / 0.14	0.82	1.46×10^{-2}	0.77	0.79	0.45	4.24×10^{-5}	0.01[*]
1	194,953	rs6677604 (A)	0.041 / 0.073	0.55	1.20×10^{-5}	0.052 / 0.070	0.73	3.22×10^{-2}	0.17 / 0.23	0.71	1.19×10^{-5}	0.68	0.69	0.41	2.96×10^{-10}	0.17
6	32,778	rs2856717 (T)	0.19 / 0.26	0.66	3.31×10^{-8}	0.14 / 0.20	0.69	1.51×10^{-4}	0.28 / 0.33	0.77	3.32×10^{-6}	0.73	0.69	0.59	8.44×10^{-16}	0.44
6	32,789	rs9275596 (C)	0.14 / 0.22	0.56	1.91×10^{-12}	0.09 / 0.16	0.54	6.29×10^{-8}	0.20 / 0.27	0.70	7.40×10^{-10}	0.63	0.62	0.43	1.59×10^{-26}	0.31
6	32,917	rs9357155 (A)	0.15 / 0.20	0.69	5.19×10^{-6}	0.12 / 0.18	0.64	1.79×10^{-5}	0.11 / 0.13	0.77	8.26×10^{-4}	0.71	0.66	0.62	2.11×10^{-12}	0.35
6	32,919	rs2071543 (A)	0.16 / 0.22	0.70	7.19×10^{-6}	0.14 / 0.20	0.65	1.59×10^{-5}	0.12 / 0.14	0.81	1.66×10^{-3}	0.73	0.67	0.64	5.77×10^{-12}	0.27
6	33,194	rs1883414 (T)	0.19 / 0.24	0.73	3.26×10^{-5}	0.17 / 0.20	0.82	3.55×10^{-2}	0.29 / 0.33	0.82	2.17×10^{-4}	0.78	0.77	0.61	4.84×10^{-9}	0.55
6	33,205	rs3129269 (T)	0.21 / 0.27	0.73	1.32×10^{-5}	0.20 / 0.23	0.83	3.48×10^{-2}	0.33 / 0.38	0.83	6.67×10^{-4}	0.79	0.79	0.61	8.54×10^{-9}	0.42
22	28,824	rs2412971 (A)	0.31 / 0.39	0.72	8.21×10^{-7}	0.24 / 0.28	0.83	2.79×10^{-2}	0.46 / 0.51	0.82	1.61×10^{-3}	0.80	0.75	0.66	1.86×10^{-9}	0.29
22	28,859	rs2412973 (A)	0.32 / 0.39	0.73	1.91×10^{-6}	0.26 / 0.30	0.83	2.68×10^{-2}	0.46 / 0.51	0.83	2.09×10^{-3}	0.80	0.76	0.66	4.46×10^{-9}	0.28

Q: p-value for the Cochrane's Q statistic;

The per-allele, heterozygote and homozygote OR's are indicated for the combined cohort.

[a] Cochran-Armitage trend test;

[b] Stratified analysis using Mantel's extension of Cochran-Armitage trend test;

[*] significant heterogeneity (P<0.05).

Table 3

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

Test SNP	Conditioning SNP(s)	Beijing Discovery Cohort N = 2,096 (1,194 cases / 902 controls)		Shanghai Follow-up Cohort N = 1,460 (712 cases / 748 controls)		European Follow-up Cohort N = 2,410 (1,238 cases / 1,172 controls)		All Cohorts Combined N = 5,966 (3,144 cases / 2,822 controls)	
		Unconditioned p-value	Conditioned p-value	Unconditioned p-value	Conditioned p-value	Unconditioned p-value	Conditioned p-value	Unconditioned p-value	Conditioned p-value
rs2856717		3.30×10^{-8}	0.280	1.51×10^{-4}	0.271	3.32×10^{-6}	0.354	8.44×10^{-16}	0.114
rs9275596	rs9275596	1.91×10^{-12}	NA	6.29×10^{-8}	NA	7.40×10^{-10}	NA	1.59×10^{-26}	NA
rs9357155		5.19×10^{-6}	2.29×10^{-3}	1.79×10^{-5}	3.12×10^{-4}	8.26×10^{-4}	8.83×10^{-4}	2.11×10^{-12}	6.87×10^{-9}
rs1883414		1.32×10^{-5}	2.16×10^{-4}	0.0348	0.164	6.67×10^{-4}	3.64×10^{-4}	8.54×10^{-9}	9.94×10^{-8}
rs2856717		3.30×10^{-8}	0.236	1.51×10^{-4}	0.225	3.32×10^{-6}	0.303	8.44×10^{-16}	0.0754
rs9275596	rs9275596,	1.91×10^{-12}	NA	6.29×10^{-8}	NA	7.40×10^{-10}	NA	1.59×10^{-26}	NA
rs9357155	rs9357155	5.19×10^{-6}	NA	1.79×10^{-5}	NA	8.26×10^{-4}	NA	2.11×10^{-12}	NA
rs1883414		1.32×10^{-5}	7.04×10^{-5}	0.0348	0.059	6.67×10^{-4}	7.18×10^{-4}	8.54×10^{-9}	3.13×10^{-8}
rs2856717		3.30×10^{-8}	0.278	1.51×10^{-4}	0.241	3.32×10^{-6}	0.272	8.44×10^{-16}	0.0760
rs9275596	rs9275596,	1.91×10^{-12}	NA	6.29×10^{-8}	NA	7.40×10^{-10}	NA	1.59×10^{-26}	NA
rs9357155	rs9357155,	5.19×10^{-6}	NA	1.79×10^{-5}	NA	8.26×10^{-4}	NA	2.11×10^{-12}	NA
rs1883414	rs1883414	1.32×10^{-5}	NA	0.0348	NA	6.67×10^{-4}	NA	8.54×10^{-9}	NA

rs9275596 and rs2856717 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.

No. of Protective Alleles	Beijing Discovery Cohort (N=2,074)* 1,176 cases/ 898 controls			Asian Replication Cohort (N=1,397)* 685 cases / 712 controls			European Replication Cohort (N=2,160)* 1,098 cases / 1,062 controls		
	Frequency (Cases / Controls)	Average Risk Score (+/− SD)	OR (95% CI)	Frequency (Cases / Controls)	Average Risk Score (+/− SD)	OR (95% CI)	Frequency (Cases / Controls)	Average Risk Score (+/− SD)	OR (95% CI)
0 (highest risk)	0.17 /0.07	0.00	1.00 (reference)	0.24 / 0.13	0.00	1.00 (reference)	0.07 / 0.03	0.00	1.00 (reference)
1	0.31 / 0.26	−0.37 (+/−0.09)	0.50 (0.36–0.69)	0.38 / 0.32	−0.30 (+/−0.15)	0.66 (0.48–0.90)	0.19 / 0.12	−0.11 (+/−0.04)	0.59 (0.36–0.97)
2	0.29 / 0.29	−0.77 (+/−0.14)	0.40 (0.29–0.56)	0.24 / 0.31	−0.65 (+/−0.23)	0.43 (0.31–0.60)	0.26 / 0.24	−0.23 (+/−0.05)	0.39 (0.25–0.63)
3	0.16 / 0.20	−1.17 (+/−0.15)	0.31 (0.22–0.44)	0.10 / 0.14	−1.06 (+/−0.26)	0.40 (0.27–0.60)	0.26 / 0.30	−0.35 (+/−0.06)	0.30 (0.19–0.48)
4	0.06 / 0.12	−1.61 (+/−0.17)	0.20 (0.13–0.31)	0.04 / 0.08	−1.44 (+/−0.28)	0.28 (0.16–0.47)	0.15 / 0.19	−0.47 (+/−0.06)	0.28 (0.17–0.45)
5 (lowest risk)	0.01 / 0.06	−2.11 (+/−0.25)	0.09 (0.05–0.16)	0.004 / 0.03	−1.86 (+/−0.36)	0.10 (0.03–0.33)	0.08 / 0.13	−0.65 (+/−0.10)	0.21 (0.12–0.35)
OR change highest vs. lowest risk [a]		11.1			10.0			4.8	
P-value [b]		6.76×10^{-27}			3.13×10^{-14}			6.24×10^{-17}	
C-stat (95%CI) [c]		0.63 (0.60–0.65)			0.61 (0.58–0.64)			0.60 (0.58–0.62)	
Nagelkerke R-sq [d]		0.072			0.054			0.042	

*
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.

[a] Fold-change in odds ratio between highest and lowest risk group.

[b] P-value for the risk score prediction model.

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

[d] Nagelkerke's pseudo R^2 indicates the fraction of the variance in risk explained by the risk score model.