Published in final edited form as:
Nat Genet. 2014 November ; 46(11): 1187–1196. doi:10.1038/ng.3118.

Discovery of new risk loci for IgA nephropathy implicates genes involved in immunity against intestinal pathogens

A full list of authors and affiliations appears at the end of the article.

Abstract

We performed a genome-wide association study (GWAS) of IgA nephropathy (IgAN), the most common form of glomerulonephritis, with discovery and follow-up in 20,612 individuals of European and East Asian ancestry. We identified six novel genome-wide significant associations, four in *ITGAM-ITGAX*, *VAV3* and *CARD9* and two new independent signals at *HLA-DQB1* and *DEFA*. We replicated the nine previously reported signals, including known SNPs in the *HLA-DQB1* and *DEFA* loci. The cumulative burden of risk alleles is strongly associated with age at disease onset. Most loci are either directly associated with risk of inflammatory bowel disease (IBD) or maintenance of the intestinal epithelial barrier and response to mucosal pathogens. The geo-spatial distribution of risk alleles is highly suggestive of multi-locus adaptation and the genetic risk correlates strongly with variation in local pathogens, particularly helminth diversity, suggesting a possible role for host-intestinal pathogen interactions in shaping the genetic landscape of IgAN.

IgA nephropathy (IgAN) is the most common form of primary glomerulonephritis and the leading cause of end-stage kidney failure in China[1]. The diagnosis is made by kidney biopsy, which shows predominant deposition of IgA-containing immune complexes in the glomerular mesangium, leading to glomerulonephritis, glomerular sclerosis, and progressive loss of kidney function. The etiology of IgAN is poorly understood and the genetic

Corresponding Authors: Ali Gharavi, MD, Department of Medicine, Division of Nephrology, Columbia University, 1150 St Nicholas Ave, Russ Berrie Pavilion #413, New York, NY 10032, USA, Tel: 212-851-5556, Fax: 212-851-5461, 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, Tel: 203-737-4420, Fax: 203-785-7560, richard.lifton@yale.edu.

Author contributions:
K.K., R.P.L., and A.G.G. conceptualized and designed the study; S.S.C., F.S., H.J.S., G.A., C.I., B.F.V., N.D., L.D.V., C.B., E.S., F.E.B., A.A., S.S., M.R., A. Amore, L.P., R.C., M.S., P.R., R.M., G.M.G., G.C., M.B., F.L., L.A., M.D., M.M., A.M., G.F., E.B., G.B., C.P., R.M., C.M., D.D.L, D.S., A.P., R.P., S.F., S.C., M.G., M. Gigante, L.G., P.Z., D.M., V.T., F.E., T.R., J.F., T.K., J.N., K.M., L.P., M.Z., M.M.W., M.R.B., K.P., D.G., J.B., L.T., F.B., G. Canaud., A.B., M.M., U.P., H.S., S.G., I.N., Y.C., J.X., P.H., N.C., H.Z., R.J.W., J.N., B.A.J., J.F., B.S., and D.C recruited study participants, contributed DNA samples, and performed clinical characterization of subjects; D.G., J.B., J.F., A.B., B.S., and D.C contributed genotype data; Y.L., S.P., S.S., C.F., Y.C., J.X., and P.H. prepared DNA samples; Y.L., S.P., S.S., C.F., Y.C., J.X., and P.H. assisted in genotyping, sequencing and wet lab experiments; K.K. and Y.L. managed clinical and genetic data; K.K., M.V., D.F., S.L., and A.G.G. analyzed data; K.K., R.P.L., and A.G.G. wrote the manuscript.

Competing Financial Interests:
The authors declare no competing financial interests.

Accession Codes:
EGAS00000000031, phs000431.v1, phs000431.v2

URLs:
The 15 SNP IgA nephropathy genetic risk calculator: www.columbiamedicine.org/divisions/gharavi/calc_genetic.php
For a full list of URLs used in data analysis, please refer to the Supplemental Note.

architecture is complex. The disease is most prevalent in East Asians, less frequent in Europeans, and relatively rare in individuals of African ancestry. For example, Asian-Americans have a 4-fold higher incidence of end-stage renal disease due to IgAN compared to European-Americans, and nearly 7-fold higher compared to African-Americans[2]. IgAN affects individuals of all age groups, with a peak incidence in the 2nd or 3rd decade of life; the factors determining age of onset are unknown.

To date, there have been three GWAS for IgAN[3–5]. The results of these studies demonstrate a strong contribution of the major histocompatibility (MHC) locus to disease risk. The two largest studies, both based on Asian discovery cohorts, detected four additional non-HLA loci, including chromosome 1q32, comprising a common deletion of the complement factor H related *CFHR3* and *CFHR1* genes (*CFHR3,1-delta);* 8p23 comprising the α-defensin (*DEFA*) gene cluster; 17p13 (including *TNFSF13*), and 22q12 (including *HORMAD2* and several other genes)[3,4]. Cumulatively, these GWAS loci explain about 5% of the total disease risk. Additionally, variation in risk allele frequency explains a substantial fraction of the observed ethnic variation in disease prevalence, with risk alleles having substantially higher frequencies in Asians compared to Europeans[3]. These findings raise the possibility that additional disease loci might have been missed owing to fixation of risk alleles in Asian populations. To identify new disease loci, we performed a GWAS twice the size of the prior largest study and have analyzed a discovery cohort based predominantly on European subjects.

RESULTS

Study Design

In stage I (discovery) we performed a genome-wide analysis in 2,747 biopsy-confirmed cases and 3,952 controls, including three new cohorts comprising 1,553 cases and 3,050 controls of European ancestry and the previously published Han Chinese discovery cohort of 1,194 cases and 902 controls (Table 1, Supplementary Tables 1–3, Supplementary Note). For each cohort, we performed principal component analyses to assure adequate ancestry matching between cases and controls (Supplementary Figure 1). All individual samples were imputed to a common set of >1 million SNPs (Supplementary Table 4) using ancestry-matched HapMap-3 reference panels (Supplementary Figure 2). Primary association testing was performed after accounting for imputation uncertainty and significant principal components of ancestry. We detected minimal effect of population stratification within each cohort (λ 1.01–1.06, Supplementary Figure 3). The association results from individual cohorts were combined using genome-wide fixed effects meta-analysis. We identified multiple suggestive signals and several distinct peaks exceeding genome-wide significance in the joint analysis of the discovery cohorts (Supplementary Figure 4). Top signals, defined by $P < 5 \times 10^{-5}$, were genotyped in additional 4,911 cases and 9,002 controls (stage II), followed by meta-analysis to identify genome-wide significant signals across the combined cohorts of 20,612 individuals. This two-stage design was adequately powered to detect ORs as small as 1.15–1.25 (Supplementary Table 1).

In the combined analysis, we identified six new genome-wide significant signals (Figure 1, Supplementary Figure 5, Table 2, and Supplementary Tables 5, 6, and 7). These included

four signals in three novel loci, chr.1p13 (*VAV3* locus), chr.9q34 (*CARD9* locus), and chr. 16p11 (*ITGAM-ITGAX* locus), and two new independent signals within the previously known *HLA-DQ/DR* and *DEFA* regions. We also confirmed associations at all nine previously identified loci at chr.6p21 (*HLA-DQ/DR, TAP1/PSMB8,* and *HLA-DP loci*), chr. 1q32 (*CFHR3,1-delta* locus), chr.8p23 (*DEFA* locus), chr.17p13 (*TNFSF13* locus), and chr. 22q12 (*HORMAD2* locus).

New IgAN susceptibility loci

Chr.16p11: ITGAM-ITGAX locus—This locus represented the strongest novel non-HLA signal (Figure 1b). The top signal, rs11574637, is an intronic SNP in *ITGAX* encoding leukocyte-specific integrin αX, a component of complement receptor 4 (CR4) involved in leukocyte cell adhesion, migration, and phagocytosis of complement-coated particles by monocytes and macrophages[6]. This SNP was genome-wide significant in the discovery phase (OR 1.47, $P = 2.8 \times 10^{-10}$) and in the combined meta-analysis (OR 1.32, $P = 8.1 \times 10^{-13}$). It is noteworthy that the risk allele (T) at this locus represents an ancestral (chimp) allele with frequency of 0.82 in Europeans and 1.0 in Asians, explaining why this strong signal was not detected in prior GWAS based on Asian discovery cohorts. Prior studies have shown that rs11574637 is associated with risk of systemic lupus erythematosus (SLE)[7]. Interestingly, the IgAN risk allele (T) is protective against SLE, suggesting complex interplay between these two disorders causing nephritis.

In addition, we detected another genome-wide significant intergenic SNP in this region, rs11150612 ($P = 1.3 \times 10^{-11}$), which is poorly correlated with rs11574637 ($r^2 = 0$ for Asians and $r^2 = 0.12$ in Europeans). Stratified conditional analysis strongly suggests that rs11150612 represents an independent signal and will require confirmation in larger European cohorts (conditioned OR 1.13, $P = 1.6 \times 10^{-6}$, Supplementary Table 8). The risk allele at rs11150612 is a derived (non-chimp) allele with frequency of 0.36 in Europeans and 0.75 in Asians. This allele is also associated with increased expression of *ITGAX* in peripheral blood cells[8] (Supplementary Table 9). Moreover, examination of 1000 Genomes data revealed that this risk allele is in strong LD with an *ITGAX* missense variant predicted to be damaging (rs2230429, P517R, r^2=0.97, but not typed in our study, Supplementary Table 10).

Chr.9q34: CARD9 locus—We observed a genome-wide significant signal at rs4077515 (OR 1.16, $P = 1.2 \times 10^{-9}$, Figure 1c), which was supported by both Asian and European cohorts (Supplementary Table 6). The rs4077515-T risk allele results in p.Ser12Asn substitution in *CARD9* (encoding Caspase recruitment domain-containing protein 9, an adapter protein that promotes activation of NF-κB in macrophages). This substitution is associated with higher expression of *CARD9* in monocytes[9], lymphoblastoid cell lines[10], and peripheral blood cells[8] (Supplementary Table 9). This same allele also confers increased risk of ulcerative colitis and Crohn's disease[11,12] (Supplementary Table 11).

Chr.1p13: VAV3 locus—The top signal, rs17019602 (Figure 1d) is an intronic SNP in *VAV3,* a gene encoding a guanine nucleotide exchange factor for Rho GTPases that is important for B- and T-lymphocyte development and antigen presentation[13,14] (OR 1.17, P

= 6.8×10^{-9}). Both Asian and European cohorts support this association (Supplementary Table 6). A common variant in *VAV3* has previously been associated with hypothyroidism, likely secondary to autoimmune etiology[15]. However, the hypothyroidism risk allele shows no linkage disequilibrium with rs17019602 ($r^2 = 0$), indicating that the IgAN signal represents a distinct allele at this locus.

Identification of novel and ethnicity-specific signals at known loci

Chr.6p21: Novel signal at HLA-DQ/DR locus—The strongest signal in the present GWAS represents a novel association within the *HLA-DQ/DR* locus (rs7763262, OR 1.41, P = 1.8×10^{-38}; Supplementary Figure 6). This signal persisted after conditioning on the previously described SNPs in the region (conditioned OR 1.31, P = 6.2×10^{-14}, Supplementary Table 12); the three previously reported SNPs remained significant after conditioning on rs7763262. Notably, we detect a stronger effect of rs7763262 in Europeans (OR 1.49, P = 1.2×10^{-30}) compared to Asians (OR 1.30, P = 1.2×10^{-10}, Supplementary Table 6, OR difference P = 0.012). To identify specific HLA alleles that may underlie associations in this region, we imputed classical *HLA* alleles (Supplementary Table 13). Stepwise conditional analysis identified four independent genome-wide significant associations (Supplementary Table 14), including *DQA1*0101* (OR 1.53, P = 1.7×10^{-15}), *DQA1*0102* (OR 0.68, P = 1.7×10^{-14}), *DQB1*0201* (OR 0.71, P = 2.6×10^{-13}), and *DQB1*0301* (OR 1.33, P = 2.2×10^{-12}). On conditional analysis, these classical alleles account for most of the SNP associations at this interval (Supplementary Table 15).

Chr.6p21: Population-specific effects at TAP1/PSMB8 locus—The previously reported risk allele at this locus (rs2071543, a Q49K missense variant in *PSMB8*)[3] represents a strong cis-eQTL associated with increased peripheral blood expression of *TAP2, PSMB8,* and *PSMB9*[8], which encode proteins involved in antigen processing and presentation (Supplementary Table 9). In this study, rs2071543 displayed significant heterogeneity across different cohorts ($I^2 = 76\%$, Cochrane's P < 0.05) attributable to ethnicity-specific effects (Supplementary Table 6). This SNP was genome-wide significant in Asians (OR 1.41, P = 2.1×10^{-9}), but no association was observed in Europeans (OR 0.99, P = 0.85). This difference was not explained by differences in risk allele frequency in Asian and European controls (0.80 and 0.87 respectively), suggesting variation in LD structure between Europeans and Asians, or the presence of an Asian-specific risk allele at this locus.

Chr.8p23: DEFA locus—A GWAS in Asians previously implicated rs2738048 in this locus, which contains a cluster of related genes encoding the α-defensin anti-microbial peptides[4]. We detected a new genome-wide significant signal in this region represented by rs10086568 (OR 1.16, P = 1.0×10^{-9}, Figure 1e). All cohorts regardless of ethnicity supported this new association. In contrast, we observed only a weak association at rs2738048 (OR 1.10, P = 1.6×10^{-4}), with evidence of significant heterogeneity across different cohorts (Cochrane's P < 0.05). In the ethnicity-specific analyses, the association of rs2738048 was evident only in Asian cohorts (OR 1.23, P = 1.3×10^{-7} in Asians; OR 1.02, P = 0.58 in Europeans; Supplementary Table 6), and this finding was not explained by differences in risk allele frequency in Asian and European controls (0.68 and 0.69

respectively). Because rs2738048 and rs10086568 are not in linkage disequilibrium ($r^2 < 0.03$), mutual conditioning had little effect on these results (Supplementary Table 16). To date, variation at this locus has not been identified by GWAS of other phenotypes, suggesting that the *DEFA* association may be specific to IgAN.

Replication of four other known loci and total variance explained

Our GWAS provided genome-wide significant confirmation of three previously reported loci on *chr.1q32* (*CHFR3,1-delta)*, *chr.6p21 (HLA-DP*), and the *chr.22q12 (HORMAD2)* and confirmed one of the two previously reported SNPs on *chr.17p13 (TNFSF13,* rs3803800) (Table 2, Figure 1, and Supplementary Figure 6). We also confirmed the additive effect of the *TNFSF13* and *HORMAD2* risk alleles on serum IgA levels (Supplementary Figure 7). Cumulatively, the 15 new and replicated GWAS loci explained 6.2% of the risk in the European cohorts and 7.6% of the variation in disease risk in the Chinese cohorts.

The genetic risk score is associated with the age of disease onset

We hypothesized that a higher burden of genetic susceptibility alleles may also influence the severity or onset of kidney disease. To test this hypothesis, we computed a genetic risk score as the weighted sum of the number of the alleles multiplied by the log of the OR for each of the individual loci. We detected a highly significant association between the genetic risk score and age of diagnosis among the 3,409 cases with available data, with 14 of 15 risk alleles individually contributing to this association. Risk alleles promoted earlier disease onset (Figure 2b and c, Supplementary Table 17), with each quintile of the genetic risk score changing the age of onset by 1.2 years ($P = 2.8 \times 10^{-13}$). This effect was robust to adjustments for cohort or ethnicity. Nonetheless, these loci explained only about 1.4% of the total variance in age of disease onset. Additional analysis of single SNP-phenotype correlations pointed to rs7763262-C risk allele (*HLA-DQ/DR* locus) as most strongly associated with age of diagnosis ($P = 3.2 \times 10^{-4}$) and greater risk of progression to end-stage kidney disease (per allele HR 1.72, $P = 3.6 \times 10^{-3}$). Exploratory analyses of other parameters of disease severity and progression were generally not statistically significant (Supplementary Tables 17–19).

Geospatial pattern of genetic risk suggests polygenic adaptation

We previously demonstrated that the worldwide distribution of IgAN risk alleles was correlated with distance from Africa and paralleled the prevalence of IgAN[2,3]. The distribution for the 15-SNP risk score derived from the present study showed an even greater difference among worldwide populations and was more correlated with geography (52 HGDP populations, $r = 0.33$, $p < 1.0 \times 10^{-16}$, Supplementary Figure 8a). We observed no evidence of hard selective sweeps at any of the individual loci by haplotype-based selection tests in Asians and Europeans[16]. For several loci, ancestral alleles have lower frequencies in Africans, suggesting that local selective pressures could be operating in Africa. The observed correlation of risk score with distance from Africa is unlikely to be a chance event; based on 10,000 permutations of 15 randomly drawn SNPs matched for average allele frequency to each IgAN SNP, we found that the observed geo-spatial correlation was in the upper tail of the null distribution (empiric $P = 0.026$, Supplementary Figure 8b). The IgAN

risk allele frequencies were also highly differentiated across HapMap III populations (average Fst of 0.237, Supplementary Table 20). Notably, the risk alleles with larger effect size displayed greater differences in frequency among populations, further suggesting a non-random change in allele frequencies across populations (Supplementary Figures 8d and e). Taken together, these observations are best explained by polygenic adaptation to local environments (soft selective sweeps acting simultaneously on multiple existing loci) or more complex selective pressures not easily detectable by classical tests of selection[17,18].

Overlap with susceptibility loci for other phenotypes

We identified many overlaps with susceptibility loci for other phenotypes documented in the NHGRI GWAS catalogue, suggesting shared pathogenic pathways (Figure 2a and Supplementary Table 11). We found both concordant and opposing effects with other immune mediated diseases. The *HLA-DQ/DR* region had the largest number of overlapping associations; IgAN risk alleles within this locus conferred increased risk of rheumatoid arthritis[19], systemic sclerosis[20], alopecia areata[21], Graves' disease[22], follicular lymphoma[23], type I diabetes[19] and IgA deficiency[24]. However, these risk alleles for IgAN also reduced risk for SLE[25], multiple sclerosis[26], ulcerative colitis[27], and hepatocellular carcinoma[28]. At the same time, because of extensive LD within the HLA region, some of these associations may be reflective of signal inter-correlation rather than true pleiotropic effects. Among non-HLA loci, IgAN risk alleles also conferred increased risk for IBD (*CARD9* locus)[11,12], elevated serum non-albumin protein and IgA levels (*TNFSF13* locus)[29], AMD (*CFHR3,1-delta* locus)[30], and T1DM (*HORMAD2* locus)[31]. Opposing effects were detected for SLE (*ITGAM-ITGAX* and *CFHR3,1-delta*)[7,32] and IBD (*HORMAD2* locus)[12,33].

Notably, detailed annotations revealed that the majority of IgAN loci encode proteins implicated in maintenance of the intestinal barrier and regulation of mucosal immune response to pathogens (Table 3). Three IgAN risk loci are associated with Crohn's disease and/or ulcerative colitis (*CARD9*, *HORMAD2* and *HLA*-DQB1)[11,12,34]. *ITGAM* and *TNSF13* participate in regulation of IgA-producing cells in the intestine[35,36]; ITGAM is also required for interaction between FcαR (CD89) and secretory IgA, the main form of IgA at mucosal sites[37,38]. α-defensins are expressed by the intestinal Paneth cells and protect from food- and water-borne pathogens in the intestine; deficiencies in α-defensins-5 and -6 have been associated with Crohn's disease[39,40]. Finally, CARD9, VAV and PSMB8/9 are involved in NF-κB activation and are essential for maintenance of the intestinal epithelial barrier and control of the local inflammatory response to infection and CARD9 deficiency produces susceptibility to invasive fungal infections[41–43].

Enrichment of the GWAS for SNPs implicated in autoimmune or inflammatory traits

We hypothesized that additional associations with other autoimmune and inflammatory disorders may be present below our replication threshold. Therefore, we performed a gene-set analysis of 582 non-HLA SNPs previously associated with any autoimmune or inflammatory trait listed in the NHGRI GWAS catalogue. In total, 87/582 (15%) were associated with the risk of IgAN at a nominal $P < 0.05$ (Figure 3a, Supplementary Table 21). This distribution was never observed in 10,000 permutations of phenotype on genotype, indicating a highly significant excess of positive associations (empiric $P < 0.0001$,

Supplementary Figure 9). We also detected a consistent excess of direct protein-protein interactions among gene products encoded by the significant and suggestive loci (Supplementary Figure 10). Among the most prominent autoimmune signals was the *PADI4* locus, previously associated with risk of rheumatoid arthritis[44] (rs12568771, OR 1.12, $P = 1.8 \times 10^{-6}$, Supplementary Table 5). These data make clear that additional associations with other autoimmune and inflammatory disorders are present below our replication threshold and should be pursued in follow-up studies.

When the suggestive and significant loci were tested for enrichment in KEGG pathways, the top overrepresented pathways were “Intestinal Immune Network for IgA Production” (overlap coefficient of 25%, $P < 1.0 \times 10^{-16}$, Figure 3b) and “Leishmania Infection”, a protozoan infection involving the skin, viscera and mucosa (overlap coefficient of 15%, $P = 6.8 \times 10^{-15}$). Notably, the pathway enrichment scores and all network connectivity parameters were consistently increased with the addition of the top SNPs at varying FDR levels, providing additional support for the role of these loci in the pathogenesis of IgAN (Supplementary Figure 10).

Association of the IgAN genetic risk score with pathogen diversity

The enrichment for pathways involving intestinal immunity and mucosal pathogens strongly suggested that the distinctive geographic pattern of IgAN risk alleles might have been shaped by an adaption to local environment. To better define potential environmental factors that could account for such an adaptive process, we performed an association analysis of the IgAN genetic risk score for HGDP populations with 14 ecological variables previously defined for these populations reflecting local climate, pathogen load, and dietary factors[45] (Supplementary Table 22a). The genetic risk was nominally associated with climatic and dietary factors. However, there was a very strong positive association of the IgAN genetic risk score with local pathogen diversity (measured as the number of different pathogen species in the area, including viruses, bacteria, protozoa, and helminthes, $r = 0.61$, $P = 6.0 \times 10^{-7}$, Figure 4a). In the analysis of individual pathogen classes, the strongest association was for helminth diversity ($r = 0.68$, $P = 1.0 \times 10^{-8}$, Figure 4b), which accounted for nearly all the association with pathogen diversity on a stepwise regression analysis. In the final combined model, only helminth diversity and geography were independently associated with the IgAN genetic risk score (Supplementary Table 22b).

Discussion

In this study, we identify six novel signals that contribute to IgAN, including four in novel loci (*ITGAM-ITGAX*, *VAV3* and *CARD9*) and two in known regions (*HLA-DQB1*, *DEFA*), and replicate nine of the previously reported genome-wide significant signals. The loci discovered in this study reside at the intersection of multiple canonical pathways, and point to critical steps in the pathogenesis of IgAN (maintenance of the intestinal mucosal barrier, activation of mucosal IgA production, NF-κB signaling, defense against intracellular pathogens, and complement activation). Collectively, these 15 independent risk alleles significantly influence the age of disease onset. Moreover, we demonstrate significant

overlap of these loci with other autoimmune and inflammatory disorders, placing IgAN in this disease spectrum.

The striking association of risk allele frequencies with geography and local helminth diversity is most consistent with multi-locus adaptation to environment. While our analysis cannot exclude unmeasured environmental factors or other pathogens that are associated with helminth diversity, helminth infection itself is a potential source of selection pressure. Helminth infection has been a major source of morbidity and mortality in human history, and even today occurs in 25% of the world population[46], with the highest global burden of soil-transmitted helminthes infections occurring in Asia, significantly contributing to pediatric mortality[46,47]. Intriguingly, secondary forms of IgAN are known to develop in the setting schistosomiasis, a common helminth infection[48]. Recent data also indicate that schistosome infection specifically impairs the ability of ITGAM-positive ($CD11b^+$) dendritic cells to stimulate CD4+ T-cells[49]. These findings strongly suggest that the increased incidence of IgAN in some geographic areas may represent an untoward consequence of protective adaptation to mucosal invasion by local pathogens. The enhanced immune response conferred by risk alleles would simultaneously explain the known association of mucosal infections as a trigger for IgAN.

Host-pathogen interactions have similarly exerted a critical influence on the genetic architecture of IBD[12]. Consistent with this finding, IgAN loci are either directly associated with risk of IBD (*HLA-DQ/DR, CARD9, HORMAD2*) or encode proteins involved in maintenance of the intestinal mucosal barrier or regulation of mucosal immune response (*DEFA, TNFSF13, VAV3, ITGAM-ITGAX, PSMB8;* Table 3). Network and enrichment analyses further point to perturbations of the immune pathway of intestinal IgA production as a central defect in the disease pathogenesis (Figure 3, Supplementary Figure 10, and Supplementary Table 21). These results clearly link intestinal mucosal inflammatory disorders and IBD with risk of IgAN and may explain why these two diseases co-occur more often than expected by chance[50]. These data are also consistent with the clinical observation that mucosal infections frequently trigger episodes of glomerulonephritis in IgAN, and with the key role of IgA in defense at mucosal surfaces[51].

Finally, these results demonstrated that most IgAN risk loci are shared with other immune-mediated diseases and identified 87 suggestive associations with non-HLA autoimmune and inflammatory SNPs. These analyses predict that follow-up studies of autoimmune and inflammatory variants, particularly among patients with early onset of disease, will yield additional genome-wide significant associations and further clarify links to environmental risk factors.

Methods

Study Design and Power Analysis

The study was designed in two stages. Stage I (the discovery phase) involved a genome-wide meta-analysis of four discovery cohorts (2,747 cases and 3,952 controls) imputed to a common set of >1 million SNPs. Stage II (the replication phase) involved genotyping of the top signals from stage I in ten additional cohorts of European and Asian ancestry (4,911

cases and 9,002 controls). We carried out power calculations for this design under the following assumptions: a disease prevalence of 1%; a log-additive risk model; perfect LD between a marker and a disease allele; a follow-up significance threshold of 5×10^{-5}; and joint (stage I and II) significance level of 5×10^{-8}. The power of our study was calculated for a range of disease allele frequencies (0.10–0.50) and effect sizes (genotypic risk ratio 1.10–1.50). The effect sizes detectable at $\alpha = 5\times10^{-8}$ with a power of 80% were also estimated (Supplementary Table 1). The calculations were performed using CaTS software[75]. All subjects provided informed consent to participate in genetic studies and the Institutional Review Board of Columbia University as well as local ethic review committees for each of the individual cohorts approved our study protocol.

GWAS Discovery Study (Stage I)

The cohorts, genome-wide genotyping, genotype quality control, ancestry analysis, and imputations are described in detail in the Supplementary Note and Supplementary Tables 2–4. We implemented strict quality control filters for each of the cohorts, including elimination of samples with low call rates, duplicates, ancestry outliers, samples with cryptic relatedness or samples with detected gender mismatch (Supplementary Table 2). We applied principal component (PC) -based ancestry-matching algorithms to reduce any potential bias of population stratification (Supplementary Table 3). After implementation of ancestry matching, we dramatically reduced the number of significant PCs for each cohort and we demonstrated that cases and controls were evenly distributed along the PC axes without significant outliers (Supplementary Figure 1). To improve coverage across different platforms, we performed imputation to a common set of >1 million HapMap-III SNPs (Supplementary Table 4 and Supplementary Note). Only SNPs with high imputation quality ($r^2>0.8$) were included in association analyses. After ancestry matching, imputation, and quality control, there were four cohorts included in stage I: the Italian Discovery Cohort of 1,045 cases and 1,340 controls (1,132,157 imputed markers), the Chinese Discovery Cohort of 1,194 cases and 902 controls (1,027,812 imputed markers), the French Discovery Cohort of 205 cases and 159 controls (1,032,453 imputed markers) and the US Discovery Cohort of 303 cases and 1,551 controls (1,118,683 imputed markers). The primary association testing was performed within each cohort individually under a multiplicative (log-additive) model and after accounting for imputation uncertainty using an allelic dosage method. Significant principal components of ancestry were included as covariates in the association analysis of each individual cohort. Ancestry-adjusted effect estimates and standard errors were derived for each SNP and the results were combined genome-wide using fixed effects. The meta-analysis results were verified using two independent software packages (PLINK v.1.07[76] and METAL[77]). The genome-wide distributions of *P* values were examined visually using QQ-plots for each individual cohort as well as for the combined analysis. We also estimated genomic inflation factors for each genome-wide analysis[78] (Supplementary Figure 3). The final meta-analysis QQ-plot showed no global departures from the expected distribution of P values and the overall genomic inflation factor was estimated at 1.047 (Supplementary Figure 4).

Follow-up of Suggestive Signals (Stage II)

Based on the examination of QQ-plots from Stage I, we selected a *P*-value threshold of 5×10^{-5} to define signals for follow-up analyses. This threshold corresponds to the positive FDR of 13% (Q-value software)[79]. The threshold defined 435 top SNPs that were subsequently prioritized for replication. Of the 435 SNPs, 320 (74%) were localized within the known susceptibility loci, including 286 SNPs across the HLA loci, 30 SNPs on chr. 22q12.2 (*HORMAD2* locus) and additional 4 SNPs on chr.1q32 (*CFHR3/1-delta* locus). The remainder 115 SNPs were clustered into distinct loci on the basis of their physical location and regional patterns of LD. Conditional logistic regression analysis was carried out to confirm correct SNP grouping and to detect independent signals. For follow-up genotyping, we prioritized independent SNPs with the lowest P-value within each independent locus. We additionally required that each SNP is successfully typed or imputed in at least three of the four analyzed cohorts. We excluded loci supported only by a single SNP ("singleton signals" defined by absence of supporting signals with P<0.01 within the same block of LD). In case genotyping failed, we selected a back-up SNP based on strength of association, LD with the top SNP, quality of genotyping or imputation, and ability to design working primers. Additionally, we included representative SNPs for the two recently discovered GWAS loci in Chinese[4], the *TNFSF13* locus (rs3803800 and rs4227) and the *DEFA* locus (rs2738048). In total, we successfully acquired and analyzed genotype data for 50 carefully selected SNPs representative of the top 37 distinct genomic regions in 13,913 replication samples (4,911 cases and 9,002 controls). The composition of the replication cohorts, genotyping methods and genotype quality control are summarized in the Supplementary Note and Supplementary Table 2. The association analyses were first carried out individually within each of the 10 included cohorts. Similar to stage I, the results were next combined using a fixed effects model. For each SNP, we derived pooled effect estimates, their standard errors, and 95% confidence intervals. We also estimated the degree of heterogeneity using heterogeneity index (I^2) and Cochrane's Q test in the combined analysis[80]. The complete summary of association results for all 50 SNPs tested in replication cohorts is provided in Supplementary Tables 5, 6, and 7.

Imputation Analysis of Classical HLA Alleles

For each of the cohorts with available genome-wide genotype data, we imputed classical HLA alleles at *-A, -B, -C, -DQB1, -DQA1*, and *-DRB1* loci. We used HapMap Caucasian Utah (CEU) samples as reference for imputation of Caucasian cohorts and combined Han Chinese Beijing (HCB) and Japanese of Tokyo (JPT) samples for Asians. The reference panels were constructed by phasing combined SNP genotype and HLA typing data. The phasing and imputation were performed using two independent methods: MACH[81] and BEAGLE-3[82]. Any poorly imputed alleles (R-sq < 0.3) were eliminated from association testing at the level of individual cohorts. The imputed allelic concordance rate between the two methods was 98.1%. In addition, direct sequencing of the informative coding segments of *HLA-DQB1* gene in a random subset of 155 samples demonstrated that our imputation had 89.0% sensitivity and 91.5% specificity. The association testing in each cohort was performed using allelic dosage method with adjustment for significant principal components in PLINK[76]. The final results were combined across cohorts using fixed effects meta-analysis in METAL[77] (Supplementary Table 13). Conditional analyses were performed

using stepwise logistic regression with Bayesian Information Criterion (BIC) as a selection criterion (Supplementary Table 14, *Step* function, R version 3.0)

Pairwise Epistasis Screen

We screened all possible pairwise interaction terms for association with disease using 1-df LRT comparing two nested logistic models: one with main effects only and one with main effects and a multiplicative (log-additive) interaction term. We included cohort membership as a fixed covariate in both models. We excluded 7 pairwise interaction terms between SNPs in partial linkage disequilibrium ($r^2>0.1$) resulting in a total of 98 independent interactions tested (Supplementary Table 23). The results were ranked in the order of significance and positive false discovery rate (q-values) were calculated. Suggestive interaction terms were defined as exceeding a significance threshold that was Bonferroni-corrected for the number of independent tests ($p < 0.05/98$ or 5×10^{-4}).

Interrogation of Protein-Protein Interaction (PPI) Networks

We interrogated two comprehensive PPI network datasets using two independent methods. First, we used the Disease Association Protein-Protein Link Evaluator (DAPPLE)[83]. This is a network connectivity tool based on InWeb[84], an integrated database of known PPIs with 12,793 nodes and 169,810 high-confidence interactions based on MINT, IntAct, BIND, PPrel, ECrel, and Reactome. Statistical significance of network connectivity parameters for individual proteins and for the entire seed set was assessed using 1,000 within-degree node-label permutations (Supplementary Figure 10). As an independent confirmatory analysis, we downloaded the Protein Interaction Network Analysis (PINA) dataset[85], which combines annotated PPI data from 6 databases (MINT, IntAct, DIP, BioGRID, HPRD, and MIPS/MPact). This large network consisted of 14,784 nodes and 107,802 unique edges (last release December 10th, 2012). To integrate our GWAS results with PPI data, and to identify modules enriched in disease-associated genes, we used a dense module searching method (dmGWAS v.2.0)[86]. Briefly, we performed a global search for modules with maximum proportion of low P-values by designating the top-scoring GWAS genes as seeds and selecting neighboring nodes (with a shortest path to any node in the module ≤ 2) that optimize subgraph's overall significance. The extracted subnetworks were merged and visualized using R (igraph v.0.5.2).

Other Methods of Prioritizing Candidate Genes

To interrogate putative functional SNPs that were not typed or imputed in our dataset, we systematically identified all variants that were in high LD ($r^2 > 0.5$) with the 15 IgAN GWAS SNPs based on 1000 Genomes data. These variants were further annotated using ANNOVAR[87], SeattleSeq[88], and HaploReg2[89] (Supplementary Table 10). We also analyzed a subset of 1,073 SNPs that represented tags for the known common copy number polymorphisms[90]. Additionally, we identified all genes whose expression was correlated with the IgAN susceptibility SNPs in cis- or trans- and at $P < 10^{-5}$ (Supplementary Table 9). For this purpose, we used the following recently published eQTL datasets: (1) meta-analysis of transcriptional profiles from peripheral blood cells of 5,311 Europeans[8], (2) primary immune cells (B-cells and monocytes) from 288 healthy Europeans[9]; (3) 400 lymphoblastoid cell lines (LCL) derived from asthmatic children[10], and (4) eqtl.uchicago

browser with compiled data across several tissues. Finally, we utilized GRAIL (Gene Relationships Across Implicated Loci), an online tool that uses PubMed text mining results to assess network connectivity between genes residing in implicated GWAS loci[91]. To prioritize candidate genes, each individual gene was tested for significant enrichment in GRAIL connectivity to genes residing in other loci.

Genetic Risk Score

To assess cumulative effects of the newly detected loci, we built a logistic regression model based on the 15 SNP predictors with independent contribution to disease risk. The risk score was calculated as a weighted sum of the number of risk alleles at each locus multiplied by the log of the adjusted OR for each of the individual loci. The percentage of the total variance explained was estimated by Nagelkerke's pseudo R^2 from the logistic regression model with the risk score as a quantitative predictor and disease state as an outcome (SPSS Statistics v.21.0, IBM 2013).

Geospatial Risk Analysis

For this purpose, we used publicly available genotype data of HapMap III (1,184 individuals representative of 11 populations) and the Human Genome Diversity Panel (HGDP; 1,050 individuals representative of 52 worldwide populations). The HGDP individuals have been previously genotyped for 660,918 markers using Illumina 650Y arrays (Stanford University). High quality genotype data was available for 13 out of 15 IgAN SNPs, with missing genotypes for rs10086568 and rs7763262. We imputed rs7763262 with high confidence (imputation $r^2 > 0.99$) using all combined HapMap-III populations for reference. Instead of rs10086568, we used a near-perfect proxy rs9644778 (r^2=94%, D′=1.00), which was also genome-wide significant in our study ($P = 1.8 \times 10^{-9}$). Using these data, we calculated individual risk score profiles for all individuals in the HGDP dataset. The risk score was standardized across populations using a Z-score method: Standardized Risk Score = (Individual Risk Score – Worldwide Mean)/Worldwide Standard Deviation. The median standardized risk scores for each population were compared across continents. We correlated standardized risk profiles with the longitude, latitude, and geographic distance from Africa.

Testing for Genetic Drift

To evaluate if the observed allelic differentiation is due to genetic drift, we analyzed 10,000 sets of SNPs randomly drawn from the genome but matched to the IgAN SNPs based on average minor allelic frequency on a per-SNP basis. In each permutation round, we scored all 1,050 HGDP individuals with the risk score calculated from the set of randomly selected SNPs. The risk scores were correlated with the distance from Africa to generate distributions of null statistics against which we compared the observed geospatial correlation. Empirical P-value was defined as the number of permuted statistics more extreme than the observed statistic divided by the total number of permutations (Supplementary Figure 8). Empiric P-value < 0.05 was considered statistically significant. The permutation procedure was implemented using a custom script in PERL programming language.

Correlations with Environmental Variables

We investigated correlations between the newly defined genetic risk score and 14 environmental variables previously defined for each of the HGDP populations (Supplementary Table 22a). The environmental variables were downloaded directly from Fumagalli et al.[45], and included climatic factors (relative humidity, mean annual temperature, precipitation rate, net short wave radiation flux, and physical distance from the sea), subsistence strategies (relative amount of agriculture, animal husbandry, fishing, hunting, and gathering) and pathogen diversity (number of different species of viruses, bacteria, protozoa, and helminthes). We applied Pearson's correlation analysis, as well as partial correlation to test median standardized genetic risk before and after controlling for geographic distance from Africa (SPSS Statistics v.21.0). Because many of the ecological factors are inter-correlated, we also applied a stepwise feature selection algorithm (BIC selection criterion) to construct the best predictive regression model of genetic risk (*step* function, R v.3.0). At entry, we included each of the broad predictor categories separately (climate, subsistence, pathogens), followed by all 14 predictors combined, with additional adjustment for the distance from Africa (Supplementary Table 22b).

Clinical Phenotype-Genotype Correlations

We analyzed baseline demographic and clinical data from the time of renal biopsy, including age, gender, body mass index, serum creatinine (SCr), albumin (Alb), hemoglobin (Hgb), 24-hour protein excretion (P24), microscopic hematuria, systolic blood pressure (SBP), diastolic blood pressure (DBP), and history of gross hematuria. The diagnosis of hypertension was based on SBP ≥ 140 mmHg, or DBP ≥ 90 mmHg, or history of antihypertensive medication use. The level of protein excretion was measured by a 24-hour urine collection or estimated based on urinary protein-to-creatinine ratio; the proteinuria values were normalized using ln(P24+1) transformation. The degree of renal tissue injury was graded using the Haas[92] classification. Estimated glomerular filtration rate (eGFR) was evaluated using the Modification of Diet in Renal Disease (MDRD) equation for Europeans[93] and the modified MDRD version for Chinese[94]. Chronic kidney disease (CKD) was classified based on the eGFR intervals according to the Kidney Disease Outcomes Quality Initiative (K/DOQI) practice guidelines[95]. End stage renal disease (ESRD) was defined by eGFR < 15 ml/min/1.73m^2 or initiation of renal replacement therapy (dialysis or kidney transplantation). Longitudinal data after kidney biopsy were available for 1,607 patients with a mean follow-up time of 7.9 years. Out of 1,607 patients, 459 reached the endpoint of ESRD within the follow-up period. For screening genotype-phenotype correlations, we used linear regression for quantitative traits, logistic regression for binary traits, and Cox proportional hazards models for survival analysis with SNP predictors coded under additive genetic model. The associations for eGFR, P24, Alb, Hgb, histopathology scores, and serum levels of IgA and IgA1 were adjusted for age, gender, and cohort/ethnicity. Association testing for the age of diagnosis and onset of ESRD were performed before and after adjustment for sex and cohort/ethnicity. The analysis of kidney disease progression was adjusted for age, sex, cohort/ethnicity, baseline eGFR (minimally adjusted model) as well as P24 and Haas histopathology score (full model). Statistical analyses were implemented in R version 3.0 and SPSS Statistics version 21 (IBM).

Genetic Overlap with Other Phenotypes

To systematically cross-annotate IgAN susceptibility loci against all previously published GWAS findings, we downloaded the latest NHGRI GWAS catalogue (September 2013)[96]. We filtered all published SNPs that were (1) associated with any disease phenotype or trait at a genome-wide significance ($p < 5 \times 10^{-8}$) and (2) resided within the genomic regions of association with IgAN. For each SNP association, we manually verified the direction of effect for a reference allele based on original publications. Next, each selected SNP from the catalogue was queried against our GWAS results to extract the odds ratios and p-values for associations with IgAN. The directionality of allelic effects was assessed to identify pleiotropic alleles with concordant or opposed effects (Supplementary Table 11). We calculated a maximum r^2 between SNPs associated with each catalogued trait and the 15 SNPs from our study based on the data from HapMap-III and 1000 Genomes project. We defined overlapping susceptibility alleles if r^2 exceeded 0.50. Lastly, we constructed a susceptibility overlap map that connects each of the IgAN loci to the previously associated GWAS traits and highlights associations with SNPs in high LD with the top IgAN signals (Figure 2a).

Testing Inflammatory/Autoimmune Subset Hypothesis

We analyzed 582 unique SNPs representative of all non-HLA autoimmune and inflammatory disease-associated GWAS loci out of the 11,276 listed in the NHGRI GWAS catalogue (September, 2013)[96]. The association results for this set were visually examined for overrepresentation of significant signals using a QQ-plot (Figure 3a). Next, we tested the autoimmune hypothesis using a previously published GWAS-HD approach[97]. This involved testing 582 unique SNPs simultaneously for association with IgAN using the GWAS discovery cohorts. To preserve the LD pattern between SNPs, the IgAN phenotype was permuted 10,000 times within each cohort. In each round of permutation, corresponding association analysis was performed using logistic regression after adjustment for cohort membership, and a sum of the Wald (1-d.f.) association statistics of the 582 SNPs was calculated. The empirical *P* value was calculated as the proportion of the permutation samples whose sum statistic was larger than that in the observed sample (Supplementary Figure 9).

Gene Annotation and Network Analysis of Autoimmune/Inflammatory SNPs

Based on the observed distribution of *P*-values, we defined two arbitrary thresholds for inclusion of suggestive signals in downstream network analyses: positive FDR < 10% (corresponding to $P < 5.9 \times 10^{-3}$) and positive FDR < 25% (corresponding to $P < 0.05$). The SNPs meeting these criteria were clustered into distinct loci based on genomic location and pairwise linkage disequilibrium. The disease locus was defined by nearest recombination hotspots in the 3′ and 5′ direction of the top SNP and overlapping intervals were merged into a single locus. All genes that intersect this interval, including 100-kb upstream and 40-kb downstream of the largest isoform (to include regulatory DNA), were considered as contained within the disease locus. The candidate gene sets (union of all genes within the candidate loci), were used as seeds in the sequential GRAIL and DAPPLE analyses (Supplementary Figure 10). These gene sets were also used for pathway analysis using Gene

Set Enrichment Analysis (GSEA)[98]. The KEGG pathway enrichment map (Figure 3b) was constructed using the Enrichment Map (v.1.2)[99]. Network graphs were visualized in Cytoscape (v.2.8).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Krzysztof Kiryluk[1], Yifu Li[1], Francesco Scolari[2,3], Simone Sanna-Cherchi[1], Murim Choi[4,5], Miguel Verbitsky[1], David Fasel[1], Sneh Lata[1], Sindhuri Prakash[1], Samantha Shapiro[1], Clara Fischman[1], Holly J. Snyder[1], Gerald Appel[1], Claudia Izzi[2,6], Battista Fabio Viola[7], Nadia Dallera[2,3], Lucia Del Vecchio[8], Cristina Barlassina[8], Erika Salvi[8], Francesca Eleonora Bertinetto[9,10], Antonio Amoroso[9,10], Silvana Savoldi[11], Marcella Rocchietti[11], Alessandro Amore[12], Licia Peruzzi[12], Rosanna Coppo[12], Maurizio Salvadori[13], Pietro Ravani[14,15], Riccardo Magistroni[16], Gian Marco Ghiggeri[17], Gianluca Caridi[17], Monica Bodria[17], Francesca Lugani[17], Landino Allegri[18], Marco Delsante[18], Mariarosa Maiorana[18], Andrea Magnano[18], Giovanni Frasca[19], Emanuela Boer[20], Giuliano Boscutti[21], Claudio Ponticelli[22], Renzo Mignani[23], Carmelita Marcantoni[24], Domenico Di Landro[24], Domenico Santoro[25], Antonello Pani[26], Rosaria Polci[27], Sandro Feriozzi[27], Silvana Chicca[28], Marco Galliani[28], Maddalena Gigante[29], Loreto Gesualdo[30], Pasquale Zamboli[31], Dita Maixnerová[32], Vladimir Tesar[32], Frank Eitner[33,34], Thomas Rauen[33], Jürgen Floege[33], Tibor Kovacs[35,36], Judit Nagy[35,36], Krzysztof Mucha[37], Leszek Pączek[37], Marcin Zaniew[38], Małgorzata Mizerska-Wasiak[39], Maria Roszkowska-Blaim[39], Krzysztof Pawlaczyk[40], Daniel Gale[41], Jonathan Barratt[42,43], Lise Thibaudin[44], Francois Berthoux[44], Guillaume Canaud[45,46], Anne Boland[47], Marie Metzger[48], Ulf Panzer[49], Hitoshi Suzuki[50], Shin Goto[51], Ichiei Narita[51], Yasar Caliskan[52], Jingyuan Xie[53], Ping Hou[54], Nan Chen[53], Hong Zhang[54], Robert J. Wyatt[55,56], Jan Novak[57], Bruce A. Julian[58], John Feehally[42,43], Benedicte Stengel[48], Daniele Cusi[8], Richard P. Lifton[4,59], and Ali G. Gharavi[1]

Affiliations

[1]Dept. of Medicine, Div. of Nephrology, College of Physicians and Surgeons, Columbia University, New York, New York, USA [2]Div. of Nephrology, Azienda Ospedaliera Spedali Civili of Brescia, Montichiari Hospital, Univ of Brescia, Brescia, Italy [3]Dept. of Medical and Surgical Specialties, Radiological Sciences, University of Brescia, Brescia, Italy [4]Dept. of Genetics, Yale University School of Medicine, New Haven, Connecticut, USA [5]Dept. of Biomedical Sciences, Seoul National University College of Medicine, Seoul, Korea [6]Prenatal Diagnosis Unit, Department of Obstetrics and Gynecology, University of Brescia, Brescia, Italy [7]Div. of Nephrology, Azienda Ospedaliera Spedali Civili of Brescia, Spedali Civili Hospital, Univ of Brescia, Brescia, Italy [8]Renal Div., DMCO, San Paolo Hospital, School of Medicine, University of Milan, Milan, Italy [9]Immunogenetics and Biology of Transplantation, Città della Salute e della Scienza, University Hospital of Turin, Italy [10]Medical

Genetics, Dept. of Medical Sciences, University of Torino, Torino, Italy [11]Nephrology and Dialysis Unit, Ospedali di Cirié e Chivasso, Cirié, Torino, Italy [12]Nephrology, Dialysis, and Transplantation Unit, Regina Margheritra Hospital, Torino, Italy [13]Div. of Nephrology and Renal Transplantation, Carreggi Hospital, Florence, Italy [14]Dept. of Medicine, University of Calgary, Calgary, Canada [15]Dept. of Community Health Sciences, University of Calgary, Calgary, Canada [16]Div. of Nephrology Dialysis and Transplantation, Azienda Ospedaliero Universitaria Policlinico di Modena, Università di Modena e Reggio Emilia, Italy [17]Div. of Nephrology, Dialysis and Transplantation, Giannina Gaslini Institute, Genova, Italy [18]Div. of Nephrology, Azienda Ospedaliero-Universitaria and Chair of Nephrology, University of Parma, Parma, Italy [19]Div. of Nephrology, Dialysis and Renal Transpantation, Riuniti Hospital, Ancona, Italy [20]Div. of Nephrology and Dialysis, Gorizia Hospital, Gorizia, Italy [21]Div. of Nephrology, Azienda Ospedaliero-Universitaria Ospedali Riuniti di Trieste, Trieste, Italy [22]CP Humanitas Clinical and Research Center, Milan, Italy [23]Div. of Nephrology and Dialysis, Infermi Hospital, Rimini, Italy [24]Div. of Nephrology, Cannizzaro Hospital, Catania, Italy [25]Div. of Nephology and Dialysis, Chair of Nephrology, University of Messina, Azienda Ospedaliero-Universitaria Policlinico, Messina, Italy [26]Dept. of Nephrology and Dialysis, G. Brotzu Hospital, Cagliari, Italy [27]Nephrology and Dialysis, Hospital of Viterbo, Viterbo, Italy [28]Div. of Nephrology and Dialysis, Sandro Pertini Hospital, Rome, Italy [29]Dept. of Medical and Surgical Sciences, University of Foggia, Foggia, Italy [30]Section of Nephrology, Department of Emergency and Organ Transplantation, University of Bari, Bari, Italy [31]Dept. of Nephrology, Second University of Naples, Naples, Italy [32]Dept. of Nephrology, 1st Faculty of Medicine and General University Hospital, Charles University, Prague, Czech Republic [33]Dept. of Nephrology, RWTH University of Aachen, Aachen, Germany [34]Kidney Diseases Research, Bayer Pharma AG, Wuppertal, Germany [35]Nephrology Center, Medical Faculty, University of Pécs, Pécs, Hungary [36]Second Dept. of Internal Medicine, Medical Faculty, University of Pécs, Pécs, Hungary [37]Dept. of Immunology, Transplantology, and Internal Medicine, Medical University of Warsaw, Warsaw, Poland [38]Children's Hospital, Krysiewicza 7/8, Poznań, Poland [39]Dept. of Pediatrics and Nephrology, Medical University of Warsaw, Warsaw, Poland [40]Dept. of Nephrology, Transplantology, and Internal Medicine, Poznan Medical University, Poznan, Poland [41]University College London-Centre for Nephrology, Royal Free Hospital Pond Street, London [42]The John Walls Renal Unit, University Hospitals of Leicester, Leicester, United Kingdom [43]Dept. of Infection, Immunity and Inflammation, University of Leicester, Leicester, United Kingdom [44]Nephrology, Dialysis, and Renal Transplantation Dept., University North Hospital, Saint Etienne, France [45]Service de Néphrologie Transplantation Adultes, Hôpital Necker - Enfants Malades, Paris, France [46]INSERM, Institut Necker Enfants Malades, Université Paris Descartes, Sorbonne Paris Cité, Paris, France [47]Centre National de Génotypage, CEA, Institut de Génomique, Evry, France [48]INSERM, Centre for Research in Epidemiology and Population Health, Villejuif, France and University Paris-Sud, Villejuif, France [49]III Medizinische Klinik, Universitätsklinikum Hamburg-Eppendorf, Hamburg, Germany [50]Division of

Nephrology, Dept. of Internal Medicine, Juntendo University Faculty of Medicine, Tokyo, Japan [51]Division of Clinical Nephrology and Rheumatology, Niigata University, Niigata, Japan [52]Division of Nephrology, Istanbul Faculty of Medicine, Istanbul University, Istanbul, Turkey [53]Dept. of Nephrology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China [54]Renal Div., Peking University First Hospital, Peking University Institute of Nephrology, Beijing, China [55]Div. of Pediatric Nephrology, University of Tennessee Health Sciences Center, Memphis, Tennessee, USA [56]Children's Foundation Research Center, Le Bonheur Children's Hospital, Memphis, Tennessee, USA [57]Dept. of Microbiology, University of Alabama at Birmingham, Birmingham, Alabama, USA [58]Dept. of Medicine, University of Alabama at Birmingham, Birmingham, Alabama, USA [59]Howard Hughes Medical Institute, Yale University School of Medicine, New Haven, Connecticut, USA

Acknowledgments

We are grateful to all study participants for their contribution to this work. This study was supported by R01DK082753 (A.G.G., J.N.), R01DK095510 (A.G.G. and R.P.L.), and by the Center for Glomerular Diseases at Columbia University. R.P.L. is an investigator of the Howard Hughes Medical Institute. K.K. is supported by K23DK090207, R03DK099564, and Carl W. Gottschalk Research Scholar Grant from the American Society of Nephrology (ASN). S.S.C. is supported by R21DK098531 and the American Heart Association (AHA) grant 13GRNT14680075. G.M.G. is supported by the Joint Italian Ministry of Health and NIH "Ricerca Finalizzata" and by the "Fondazione Malattie Renali nel Bambino". D.C. and the HYPERGENES Consortium are supported by InterOmics (PB05 MIUR-CNR Italian Flagship Project). Additionally, we would like to acknowledge individuals from the following organizations: IgA Nephropathy Foundation of America for facilitating recruitment of individuals in the United States; Columbia University Glomerular Center (New York, USA), including: J. Radhakrishnan, D. Cohen, C. Kunis, A. Bomback and P. Canetta for referrals of IgAN cases; The Polish Registry Congenital Malformations (PRCM, Poznan, Poland) including A. Materna-Kiryluk and A. Latos-Bieleńska (supported by the Polish Ministry of Health), for facilitating the recruitment of the Polish IgAN cohort; Judit Nagy at the University of Pécs, Hungary (supported by SROP-4.2.2/B-10/1/2010-0029); GN-PROGRESS study, including F. Martinez, F. Vrtovsnik, and D. Droz for adjudicating all IgAN cases as well as individual center investigators including: X. Belenfant (Hôpital A Grégoire, Montreuil); B. Charpentier, A. Durrbach (AP-HP Hôpital Bicêtre, Kremlin-Bicêtre); G. Rostoker (Hôpital C Galien, Quincy/Senart); J. Rossert, C. Jacquot (AP-HP, Hôpital Européen G Pompidou, Paris); P. Lang, P. Remy (AP-HP Hôpital H. Mondor, Créteil); O. Kourilsky (Hôpital L. Michel, Evry); J-P. Grünfeld, D. Chauveau (AP-HP Hôpital Necker, Paris); G. Deray, H. Izzedine (AP-HP Hôpital Pitié Salpétrière, Paris); C. Legendre, F. Martinez (AP-HP Hôpital Saint-Louis, Paris); and P. Ronco, E. Rondeau (AP-HP Hôpital Tenon, Paris). We would also like to thank L. Sturg from the Biostatistics Division, Dalla Lana School of Public Health at the University of Toronto, Canada for contributing the R code for HD-GWAS analysis.

References

1. Liu ZH. Nephrology in china. Nat Rev Nephrol. 2013; 9:523–8. [PubMed: 23877587]
2. Kiryluk K, et al. Geographic differences in genetic susceptibility to IgA nephropathy: GWAS replication study and geospatial risk analysis. PLoS Genet. 2012; 8:e1002765. [PubMed: 22737082]
3. Gharavi AG, et al. Genome-wide association study identifies susceptibility loci for IgA nephropathy. Nat Genet. 2011; 43:321–7. [PubMed: 21399633]
4. Yu XQ, et al. A genome-wide association study in Han Chinese identifies multiple susceptibility loci for IgA nephropathy. Nat Genet. 2012; 44:178–82. [PubMed: 22197929]
5. Feehally J, et al. HLA has strongest association with IgA nephropathy in genome-wide analysis. J Am Soc Nephrol. 2010; 21:1791–7. [PubMed: 20595679]
6. Corbi AL, Larson RS, Kishimoto TK, Springer TA, Morton CC. Chromosomal location of the genes encoding the leukocyte adhesion receptors LFA-1, Mac-1 and p150,95. Identification of a gene cluster involved in cell adhesion. J Exp Med. 1988; 167:1597–607. [PubMed: 3284962]

7. Hom G, et al. Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. N Engl J Med. 2008; 358:900–9. [PubMed: 18204098]
8. Westra HJ, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet. 2013; 45:1238–43. [PubMed: 24013639]
9. Fairfax BP, et al. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. Nat Genet. 2012; 44:502–10. [PubMed: 22446964]
10. Dixon AL, et al. A genome-wide association study of global gene expression. Nat Genet. 2007; 39:1202–7. [PubMed: 17873877]
11. Franke A, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nat Genet. 2010; 42:1118–25. [PubMed: 21102463]
12. Jostins L, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. 2012; 491:119–24. [PubMed: 23128233]
13. Fujikawa K, et al. Vav1/2/3-null mice define an essential role for Vav family proteins in lymphocyte development and activation but a differential requirement in MAPK signaling in T and B cells. J Exp Med. 2003; 198:1595–608. [PubMed: 14623913]
14. Swat W, Fujikawa K. The Vav family: at the crossroads of signaling pathways. Immunol Res. 2005; 32:259–65. [PubMed: 16106078]
15. Eriksson N, et al. Novel associations for hypothyroidism include known autoimmune risk loci. PLoS One. 2012; 7:e34442. [PubMed: 22493691]
16. Pickrell JK, et al. Signals of recent positive selection in a worldwide sample of human populations. Genome Res. 2009; 19:826–37. [PubMed: 19307593]
17. Pritchard JK, Pickrell JK, Coop G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. Curr Biol. 2010; 20:R208–15. [PubMed: 20178769]
18. Teshima KM, Coop G, Przeworski M. How reliable are empirical genomic scans for selective sweeps? Genome Res. 2006; 16:702–12. [PubMed: 16687733]
19. Wellcome Trust Case Control C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007; 447:661–78. [PubMed: 17554300]
20. Radstake TR, et al. Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. Nat Genet. 2010; 42:426–9. [PubMed: 20383147]
21. Petukhova L, et al. Genome-wide association study in alopecia areata implicates both innate and adaptive immunity. Nature. 2010; 466:113–7. [PubMed: 20596022]
22. Chu X, et al. A genome-wide association study identifies two new risk loci for Graves' disease. Nat Genet. 2011; 43:897–901. [PubMed: 21841780]
23. Smedby KE, et al. GWAS of follicular lymphoma reveals allelic heterogeneity at 6p21.32 and suggests shared genetic susceptibility with diffuse large B-cell lymphoma. PLoS Genet. 2011; 7:e1001378. [PubMed: 21533074]
24. Ferreira RC, et al. Association of IFIH1 and other autoimmunity risk alleles with selective IgA deficiency. Nat Genet. 2010; 42:777–80. [PubMed: 20694011]
25. Yang W, et al. Meta-analysis followed by replication identifies loci in or near CDKN1B, TET3, CD80, DRAM1, and ARID5B as associated with systemic lupus erythematosus in Asians. Am J Hum Genet. 2013; 92:41–51. [PubMed: 23273568]
26. International Multiple Sclerosis Genetics C et al. Risk alleles for multiple sclerosis identified by a genomewide study. N Engl J Med. 2007; 357:851–62. [PubMed: 17660530]
27. Okada Y, et al. HLA-Cw*1202-B*5201-DRB1*1502 haplotype increases risk for ulcerative colitis but reduces risk for Crohn's disease. Gastroenterology. 2011; 141:864–871. e1–5. [PubMed: 21699788]
28. Kumar V, et al. Genome-wide association study identifies a susceptibility locus for HCV-induced hepatocellular carcinoma. Nat Genet. 2011; 43:455–8. [PubMed: 21499248]
29. Osman W, et al. Association of common variants in TNFRSF13B, TNFSF13, and ANXA3 with serum levels of non-albumin protein and immunoglobulin isotypes in Japanese. PLoS One. 2012; 7:e32683. [PubMed: 22558069]

30. Chen W, et al. Genetic variants near TIMP3 and high-density lipoprotein-associated loci influence susceptibility to age-related macular degeneration. Proc Natl Acad Sci U S A. 2010; 107:7401–6. [PubMed: 20385819]
31. Barrett JC, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat Genet. 2009; 41:703–7. [PubMed: 19430480]
32. Zhao J, et al. Association of genetic variants in complement factor H and factor H-related genes with systemic lupus erythematosus susceptibility. PLoS Genet. 2011; 7:e1002079. [PubMed: 21637784]
33. 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]
34. McGovern DP, et al. Genome-wide association identifies multiple ulcerative colitis susceptibility loci. Nat Genet. 2010; 42:332–7. [PubMed: 20228799]
35. Litinskiy MB, et al. DCs induce CD40-independent immunoglobulin class switching through BLyS and APRIL. Nat Immunol. 2002; 3:822–9. [PubMed: 12154359]
36. Kunisawa J, et al. Microbe-dependent CD11b+ IgA+ plasma cells mediate robust early-phase intestinal IgA responses in mice. Nat Commun. 2013; 4:1772. [PubMed: 23612313]
37. van Egmond M, et al. Human immunoglobulin A receptor (FcalphaRI, CD89) function in transgenic mice requires both FcR gamma chain and CR3 (CD11b/CD18). Blood. 1999; 93:4387–94. [PubMed: 10361137]
38. Van Spriel AB, Leusen JH, Vile H, Van De Winkel JG. Mac-1 (CD11b/CD18) as accessory molecule for Fc alpha R (CD89) binding of IgA. J Immunol. 2002; 169:3831–6. [PubMed: 12244179]
39. Bevins CL, Salzman NH. Paneth cells, antimicrobial peptides and maintenance of intestinal homeostasis. Nat Rev Microbiol. 2011; 9:356–68. [PubMed: 21423246]
40. Wehkamp J, et al. Reduced Paneth cell alpha-defensins in ileal Crohn's disease. Proc Natl Acad Sci U S A. 2005; 102:18129–34. [PubMed: 16330776]
41. Sokol H, et al. Card9 mediates intestinal epithelial cell restitution, T-helper 17 responses, and control of bacterial infection in mice. Gastroenterology. 2013; 145:591–601. e3. [PubMed: 23732773]
42. Liu JY, et al. Vav proteins are necessary for correct differentiation of mouse cecal and colonic enterocytes. J Cell Sci. 2009; 122:324–34. [PubMed: 19139088]
43. Vigorito E, Gambardella L, Colucci F, McAdam S, Turner M. Vav proteins regulate peripheral B-cell survival. Blood. 2005; 106:2391–8. [PubMed: 15941910]
44. Freudenberg J, et al. Genome-wide association study of rheumatoid arthritis in Koreans: population-specific loci as well as overlap with European susceptibility loci. Arthritis Rheum. 2011; 63:884–93. [PubMed: 21452313]
45. Fumagalli M, et al. Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. PLoS Genet. 2011; 7:e1002355. [PubMed: 22072984]
46. Pullan RL, Smith JL, Jasrasaria R, Brooker SJ. Global numbers of infection and disease burden of soil transmitted helminth infections in 2010. Parasit Vectors. 2014; 7:37. [PubMed: 24447578]
47. Barry MA, Simon GG, Mistry N, Hotez PJ. Global trends in neglected tropical disease control and elimination: impact on child health. Arch Dis Child. 2013; 98:635–41. [PubMed: 23793132]
48. Barsoum RS. Schistosomal glomerulopathy: selection factors. Nephrol Dial Transplant. 1987; 2:488–97. [PubMed: 3126449]
49. Ferragine CE, Walls CD, Davies SJ. Modulation of innate antigen-presenting cell function by pre-patent schistosome infection. PLoS Negl Trop Dis. 2013; 7:e2136. [PubMed: 23556020]
50. Ambruzs JM, Walker PD, Larsen CP. The histopathologic spectrum of kidney biopsies in patients with inflammatory bowel disease. Clin J Am Soc Nephrol. 2014; 9:265–70. [PubMed: 24262508]
51. Wyatt RJ, Julian BA. IgA nephropathy. N Engl J Med. 2013; 368:2402–14. [PubMed: 23782179]
52. Fujimoto K, et al. A new subset of CD103+CD8alpha+ dendritic cells in the small intestine expresses TLR3, TLR7, and TLR9 and induces Th1 response and CTL activity. J Immunol. 2011; 186:6287–95. [PubMed: 21525388]

53. Uematsu S, et al. Regulation of humoral and cellular gut immunity by lamina propria dendritic cells expressing Toll-like receptor 5. Nat Immunol. 2008; 9:769–76. [PubMed: 18516037]
54. Bertin J, et al. CARD9 is a novel caspase recruitment domain-containing protein that interacts with BCL10/CLAP and activates NF-kappa B. J Biol Chem. 2000; 275:41082–6. [PubMed: 11053425]
55. Prunotto M, et al. Autoimmunity in membranous nephropathy targets aldose reductase and SOD2. J Am Soc Nephrol. 2010; 21:507–19. [PubMed: 20150532]
56. Rivas MA, et al. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. Nat Genet. 2011; 43:1066–73. [PubMed: 21983784]
57. Beaudoin M, et al. Deep resequencing of GWAS loci identifies rare variants in CARD9, IL23R and RNF186 that are associated with ulcerative colitis. PLoS Genet. 2013; 9:e1003723. [PubMed: 24068945]
58. Lanternier F, et al. Deep dermatophytosis and inherited CARD9 deficiency. N Engl J Med. 2013; 369:1704–14. [PubMed: 24131138]
59. Levison SE, et al. Genetic analysis of the Trichuris muris-induced model of colitis reveals QTL overlap and a novel gene cluster for establishing colonic inflammation. BMC Genomics. 2013; 14:127. [PubMed: 23442222]
60. Chu H, et al. Human alpha-defensin 6 promotes mucosal innate immunity through self-assembled peptide nanonets. Science. 2012; 337:477–81. [PubMed: 22722251]
61. McCarthy DD, et al. Mice overexpressing BAFF develop a commensal flora-dependent, IgA-associated nephropathy. J Clin Invest. 2011; 121:3991–4002. [PubMed: 21881212]
62. Castigli E, et al. TACI is mutant in common variable immunodeficiency and IgA deficiency. Nat Genet. 2005; 37:829–34. [PubMed: 16007086]
63. 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]
64. Cella M, et al. A human natural killer cell subset provides an innate source of IL-22 for mucosal immunity. Nature. 2009; 457:722–5. [PubMed: 18978771]
65. Fritz DK, et al. A mouse model of airway disease: oncostatin M-induced pulmonary eosinophilia, goblet cell hyperplasia, and airway hyperresponsiveness are STAT6 dependent, and interstitial pulmonary fibrosis is STAT6 independent. J Immunol. 2011; 186:1107–18. [PubMed: 21160052]
66. Ernst M, et al. Defective gp130-mediated signal transducer and activator of transcription (STAT) signaling results in degenerative joint disease, gastrointestinal ulceration, and failure of uterine implantation. J Exp Med. 2001; 194:189–203. [PubMed: 11457894]
67. Rockman SP, et al. Expression of interleukin-6, leukemia inhibitory factor and their receptors by colonic epithelium and pericryptal fibroblasts. J Gastroenterol Hepatol. 2001; 16:991–1000. [PubMed: 11595063]
68. Kalabis J, et al. Stimulation of human colonic epithelial cells by leukemia inhibitory factor is dependent on collagen-embedded fibroblasts in organotypic culture. FASEB J. 2003; 17:1115–7. [PubMed: 12692084]
69. Visekruna A, et al. Proteasome-mediated degradation of IkappaBalpha and processing of p105 in Crohn disease and ulcerative colitis. J Clin Invest. 2006; 116:3195–203. [PubMed: 17124531]
70. Wu F, et al. Genome-wide gene expression differences in Crohn's disease and ulcerative colitis from endoscopic pinch biopsies: insights into distinctive pathogenesis. Inflamm Bowel Dis. 2007; 13:807–21. [PubMed: 17262812]
71. Schmidt N, et al. Targeting the proteasome: partial inhibition of the proteasome by bortezomib or deletion of the immunosubunit LMP7 attenuates experimental colitis. Gut. 2010; 59:896–906. [PubMed: 20581238]
72. van Heel DA, et al. A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. Nat Genet. 2007; 39:827–9. [PubMed: 17558408]
73. Dubois PC, et al. Multiple common variants for celiac disease influencing immune gene expression. Nat Genet. 2010; 42:295–302. [PubMed: 20190752]
74. Yang SK, et al. Genome-wide association study of ulcerative colitis in Koreans suggests extensive overlapping of genetic susceptibility with Caucasians. Inflamm Bowel Dis. 2013; 19:954–66. [PubMed: 23511034]

75. 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]
76. 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]
77. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010; 26:2190–1. [PubMed: 20616382]
78. Devlin B, Roeder K, Bacanu SA. Unbiased methods for population-based association studies. Genet Epidemiol. 2001; 21:273–84. [PubMed: 11754464]
79. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A. 2003; 100:9440–5. [PubMed: 12883005]
80. Huedo-Medina TB, Sanchez-Meca J, Marin-Martinez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? Psychol Methods. 2006; 11:193–206. [PubMed: 16784338]
81. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol. 2010; 34:816–34. [PubMed: 21058334]
82. 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]
83. Rossin EJ, et al. Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. PLoS Genet. 2011; 7:e1001273. [PubMed: 21249183]
84. Lage K, et al. A human phenome-interactome network of protein complexes implicated in genetic disorders. Nat Biotechnol. 2007; 25:309–16. [PubMed: 17344885]
85. Wu J, et al. Integrated network analysis platform for protein-protein interactions. Nat Methods. 2009; 6:75–7. [PubMed: 19079255]
86. Jia P, Zheng S, Long J, Zheng W, Zhao Z. dmGWAS: dense module searching for genome-wide association studies in protein-protein interaction networks. Bioinformatics. 2011; 27:95–102. [PubMed: 21045073]
87. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010; 38:e164. [PubMed: 20601685]
88. Ng SB, et al. Targeted capture and massively parallel sequencing of 12 human exomes. Nature. 2009; 461:272–6. [PubMed: 19684571]
89. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 2012; 40:D930–4. [PubMed: 22064851]
90. Conrad DF, et al. Origins and functional impact of copy number variation in the human genome. Nature. 2010; 464:704–12. [PubMed: 19812545]
91. Raychaudhuri S, et al. Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. PLoS Genet. 2009; 5:e1000534. [PubMed: 19557189]
92. Haas M. Histologic subclassification of IgA nephropathy: a clinicopathologic study of 244 cases. Am J Kidney Dis. 1997; 29:829–42. [PubMed: 9186068]
93. Levey AS, et al. Using standardized serum creatinine values in the modification of diet in renal disease study equation for estimating glomerular filtration rate. Ann Intern Med. 2006; 145:247–54. [PubMed: 16908915]
94. 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; 79:1130–1134. [PubMed: 17186473]
95. K/DOQI clinical practice guidelines for chronic kidney disease: evaluation classification stratification. American journal of kidney diseases : the official journal of the National Kidney Foundation. 2002; 39:S1–266. [PubMed: 11904577]
96. Hindorff LA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A. 2009; 106:9362–7. [PubMed: 19474294]

97. Sun L, et al. Multiple apical plasma membrane constituents are associated with susceptibility to meconium ileus in individuals with cystic fibrosis. Nat Genet. 2012; 44:562–9. [PubMed: 22466613]

98. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005; 102:15545–50. [PubMed: 16199517]

99. Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PLoS One. 2010; 5:e13984. [PubMed: 21085593]

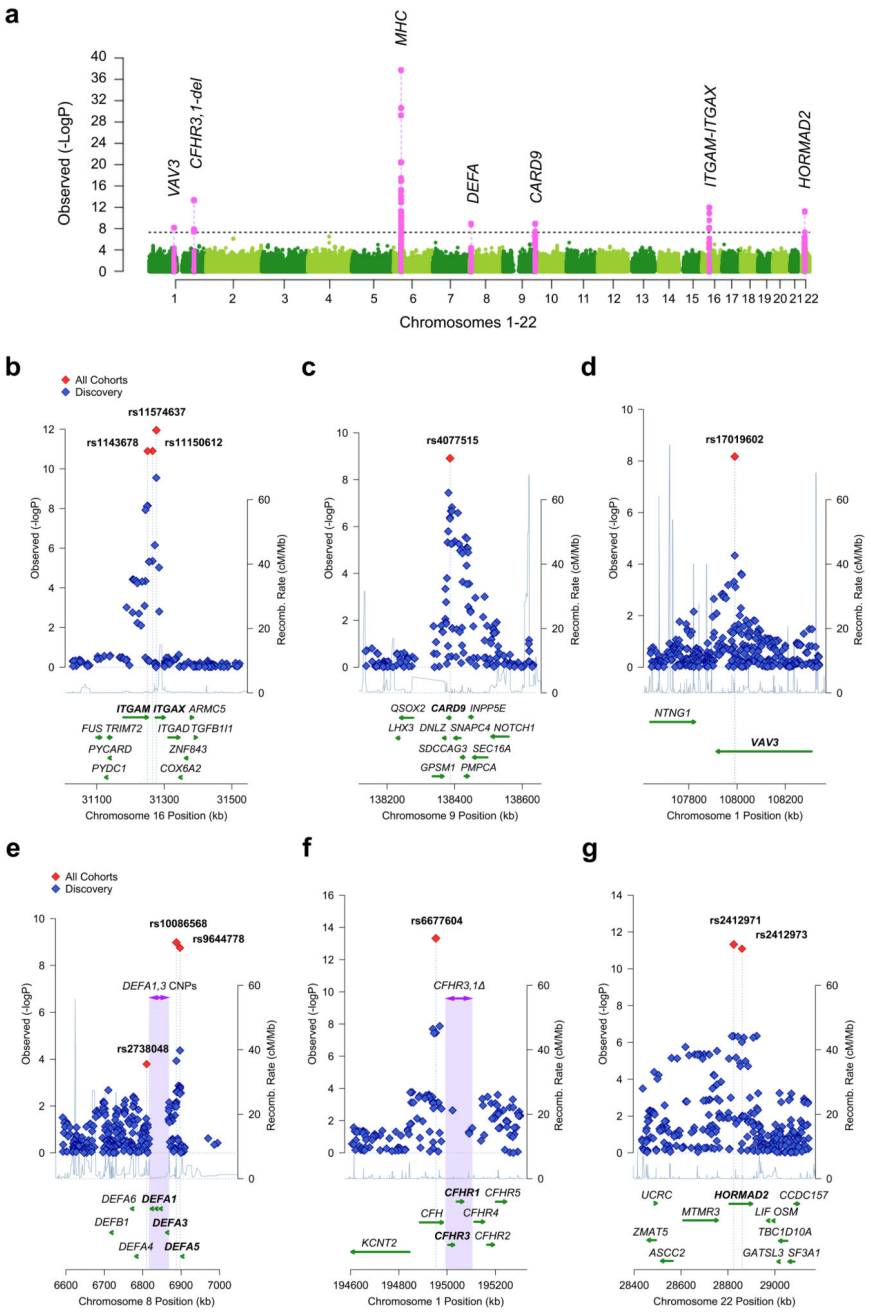

Figure 1. Results of the combined meta-analysis across all 20,612 individuals

Manhattan plot **(a)** and regional plots for genome-wide significant loci outside of the HLA region: **(b)** *ITGAM-ITGAX* locus, **(c)** *CARD9* locus, **(d)** *VAV3* locus, **(e)** *DEFA* locus (shaded area represents the region of common duplications involving D*EFA1* and *DEFA3* genes), **(f)** *CFHR3,1-delta* locus (shaded area represents the deletion of *CFHR3* and *CFHR1* genes), **(g)** *HORMAD2* locus. X-axis represents physical distance in kb (hg-18 coordinates); Y-axis represents -log *P* values for association statistics.

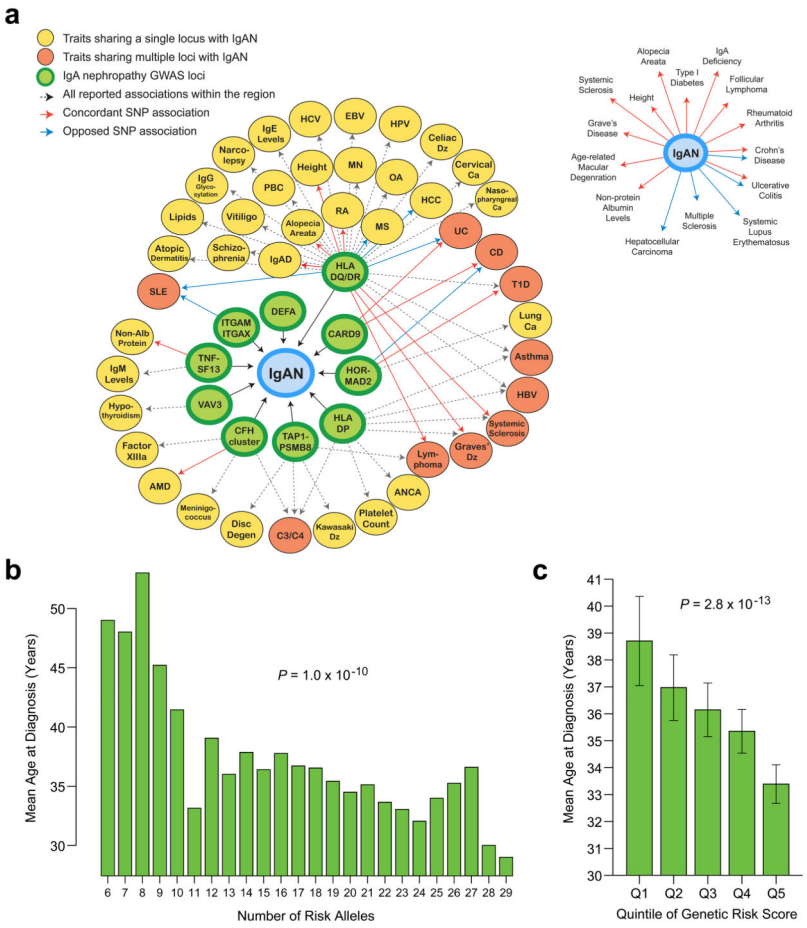

Figure 2. Pleiotropic effects of IgAN GWAS loci and their cumulative effect on the age at disease onset

(a) A genetic susceptibility map was constructed based on all overlapping genome-wide significant loci reported in the NHGRI GWAS catalogue: diseases sharing a single locus with IgAN are indicated in yellow; diseases sharing multiple loci with IgAN are indicated in orange; solid arrows represent allelic associations that are identical to, or in tight LD (r2 > 0.5) with the IgAN risk alleles: concordant effects are indicated in red and opposed effects in blue; dotted arrows represent all other phenotype associations in the region. Of note, candidate gene or regional association studies were not included in this analysis. **Inset:** collapsed representation of pleiotropic relationships between IgAN and other phenotypes (only shared allelic effects are included with concordant effects indicated in red and opposed effects in blue). **(b)** Average age at diagnosis as a function of an individual's risk allele burden (N=3,409 individuals with available data). **(c)** Average age at diagnosis by quintile of genetic risk (error bars represent 95% confidence interval for the mean). Abbreviations: IgAD: IgA Deficiency; RA: Rheumatoid Arthritis; PBC: Primary Biliary Cirrhosis; MN: Membranous Nephropathy; OA: Osteoarthritis; HCC: Hepatocellular Carcinoma; SLE: Systemic Lupus Erythematosus; UC: Ulcerative Colitis; CD: Crohn's Disease; T1D: Type I Diabetes; AMD: Age-related Macular Degeneration.

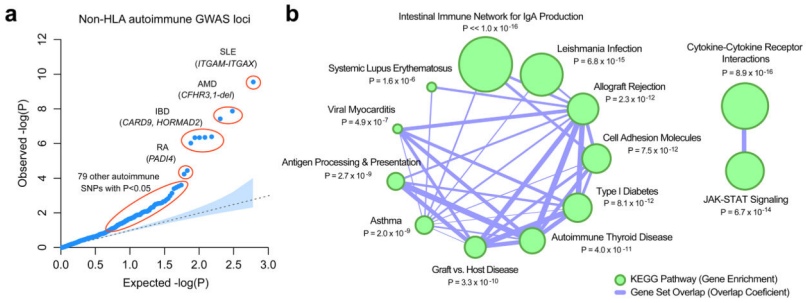

Figure 3. Autoimmunity/inflammatory loci and risk of IgAN

(a) A quantile-quantile plot of IgAN associations for 582 unique non-HLA SNPs previously associated with autoimmune or immune-mediated diseases at $p < 5 \times 10^{-8}$ in the NHGRI GWAS catalogue. When tested for association with IgAN, an unexpectedly large number of SNPs deviate from the null expectation (empiric $p < 1 \times 10^{-4}$, Supplementary Figure 9). **(b)** The KEGG enrichment map for the genes residing within autoimmunity loci associated with IgAN at $p < 0.05$ ($q < 0.25$). The size of nodes reflects −log10-transformed P-values of the adjusted hypergeometric enrichment test in GSEA. The edges represent pathway similarity as defined by an overlap coefficient. The top overrepresented KEGG pathway is the "Intestinal Immune Network for IgA Production" (gene set overlap coefficient = 25%, enrichment $p < 1.0 \times 10^{-16}$). Individual genes intersecting top-ranked KEGG pathways are provided in Supplementary Figure 10c.

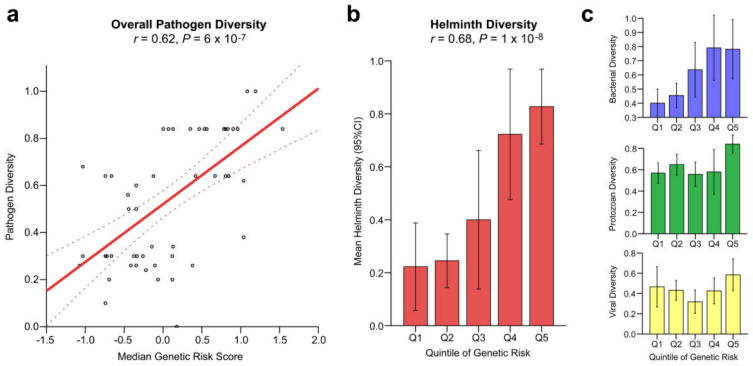

Figure 4. IgAN genetic risk is correlated with worldwide pathogen diversity

(a) Correlation between IgAN genetic risk score (X-axis) and the level of local pathogen diversity (Y-axis) among the HGDP populations (linear regression line and its 95% confidence intervals, Pearson's correlation coefficient = 0.62, $P = 6 \times 10^{-7}$); **(b)** Stepwise feature selection among all pathogen subgroups confirmed helminth diversity as the single best predictor of IgAN genetic risk (Pearson's correlation coefficient = 0.68, $P = 1 \times 10^{-8}$); **(c)** Weaker correlation was also evident for bacterial diversity (top panel, Pearson's correlation coefficient = 0.56, $P = 1 \times 10^{-5}$), but not for protozoan or viral diversity (middle and bottom panels). The pathogen diversity metrics were scaled and standardized across all populations; error bars represent 95% confidence interval for the mean; for detailed analysis refer to Supplementary Table 22.

Table 1

Summary of study cohorts

the final numbers of included cases and controls by cohort after implementation of all quality control filters.

GWAS Cohorts*	Ancestry	N Cases	N Controls	N Total	Genotyping Rate
Italian Discovery Cohort	European	1,045	1,340	2,385	99.9%
French Discovery Cohort	European	205	159	364	99.6%
US Discovery Cohort	European	303	1,551	1,854	99.7%
Chinese Discovery Cohort	East Asian	1,194	902	2,096	99.9%
Total Discovery:		**2,747**	**3,952**	**6,699**	--
Chinese Replication Cohort	East Asian	2,046	1,385	3,431	99.4%
UK Replication Cohort	European	464	4,783	5,247	99.9%
Japanese Replication Cohort	East Asian	445	395	840	99.3%
German Replication Cohort	European	393	371	764	99.6%
French Replication Cohort	European	432	436	868	99.5%
Czech Replication Cohort	European	247	230	477	99.7%
Polish Replication Cohort	European	123	200	323	99.6%
Hungarian Replication Cohort	European	220	237	457	98.7%
Italian Replication Cohort	European	413	780	1,193	99.1%
Turkish Replication Cohort	European	128	185	313	99.5%
Total Replication:		**4,911**	**9,002**	**13,913**	--
Total All Cohorts:		**7,658**	**12,954**	**20,612**	--

*The summary of quality control analyses and case/control exclusions is provided in Supplementary Table 2.

Table 2

Combined results for known and novel independent GWAS signals.

						Discovery Meta-analysis N=6,699 (2,747 cases & 3,952 controls)		Replication Meta-analysis N=13,913 (4,911 cases & 9,002 controls)		All Cohorts Combined N=20,612 (7,658 cases & 12,954 controls)					
Chr	**Location[*] (bp)**	**SNP [**]**	**Risk Allele**	**Freq. European Controls**	**Freq. Asian Controls**	**OR**	**P-value**	**OR**	**P-value**	**OR**	**P-value**	**Q-test**	**I^2**	**Locus Name**	**Novel**
1	107,990,381	rs17019602	G	0.19	0.19	1.20	4.7E-05	1.16	2.9E-05	1.17	6.8E-09	0.50	0	*VAV3*	**Novel**
1	194,953,541	rs6677604	G	0.80	0.93	1.36	3.5E-08	1.33	2.6E-07	1.35	4.8E-14	0.53	0	*CFHR3,1-del*	**Known**
6	32,532,860	rs7763262	C	0.69	0.72	1.51	1.7E-20	1.35	5.5E-20	1.41	1.8E-38	0.07	39	*HLA-DR/DQ*	**Novel**
6	32,767,856	rs9275224	G	0.51	0.59	1.33	1.2E-13	1.38	5.6E-18	1.36	5.9E-30	0.56	0	*HLA-DR/DQ*	**Known**
6	32,778,286	rs2856717	G	0.62	0.77	1.26	6.4E-08	1.27	3.1E-09	1.27	1.1E-15	0.27	19	*HLA-DR/DQ*	**Known**
6	32,789,609	rs9275596	T	0.65	0.80	1.43	7.7E-15	1.46	4.1E-18	1.44	2.5E-31	0.09	39	*HLA-DR/DQ*	**Known**
6	32,919,607	rs2071543	G	0.87	0.80	1.22	2.3E-04	1.09	8.8E-02	1.15	1.5E-04	<0.01	76	*TAP2/PSMB9*	**Known**
6	33,194,426	rs1883414	G	0.68	0.78	1.27	1.3E-08	1.17	1.1E-04	1.22	1.5E-11	0.79	0	*HLA-DP*	**Known**
8	6,810,195	rs2738048	T	0.69	0.68	1.05	2.1E-01	1.12	1.6E-04	1.10	1.6E-04	0.04	44	*DEFA*	**Known**
8	6,887,746	rs10086568	A	0.33	0.27	1.17	1.2E-04	1.16	2.1E-06	1.16	1.0E-09	0.78	0	*DEFA*	**Novel**
9	138,386,317	rs4077515	T	0.40	0.28	1.22	4.1E-07	1.12	1.5E-04	1.16	1.2E-09	0.55	0	*CARD9*	**Novel**
16	31,265,261	rs11150612	A	0.36	0.75	1.21	4.4E-06	1.17	5.1E-07	1.18	1.3E-11	0.57	0	*ITGAM-ITGAX*	**Novel**
16	31,276,375	rs11574637	T	0.82	1.00	1.47	2.8E-10	1.22	5.6E-05	1.32	8.1E-13	0.70	0	*ITGAM-ITGAX*	**Novel**
17	7,403,693	rs3803800	A	0.20	0.32	1.12	1.2E-02	1.13	2.5E-04	1.12	9.3E-06	0.38	7	*TNFSF13*	**Known**
22	28,824,371	rs2412971	G	0.54	0.67	1.21	4.6E-07	1.20	2.2E-06	1.20	4.8E-12	0.12	35	*HORMAD2*	**Known**

[*] Based on NCBI version 36 (hg-18) genome assembly

[**] Only non-redundant SNPs with mutually independent effects are included; the complete list of analyzed SNPs is provided in the Supplementary Table 5.

Table 3

IgAN GWAS loci and their role in the intestinal immunity and inflammation.

Locus (Genes)	Canonical Pathways *	Function and role in intestinal mucosal immunity
ITGAM, ITGAX	Granulocyte pathway, Monocyte pathway, Cell adhesion molecules (CAMs), Hematopoietic cell lineage, Leishmania infection, Leukocyte transendothelial migration, Regulation of actin cytoskeleton	• *ITGAM and ITGAX* encode integrins αM and αX that mark intestinal dendritic cells that maintain the balance between inflammation and tolerance. ITGAM and ITGAX also combine with integrin β2 chain to form leukocyte-specific complement receptors 3 and 4 (CR3 and CR4, respectively). • *ITGAM* is involved in the regulation of intestinal IgA-producing plasma cells in mice[36]. Integrin-αM-positive IgA plasma cells reside in Peyer's patches, require microbial stimulation for development, and exhibit more proliferation and more IgA production compared to integrin-αM-negative cells[36]. • In mice, intestinal dendritic cells that express high level of both αM and αX integrins are CD103+, express TLR5, produce retinoic acid, and induce T-cell-independent IgA class-switch recombination[52, 53]. • Schistosome infection specifically impairs the ability of ITGAM-positive (CD11b+) dendritic cells to stimulate CD4+ T-cells[49].
CARD9	NOD-like receptor signaling pathway, Innate immune system, Tuberculosis, Fungal infection	• *CARD9* encodes a molecular scaffold for the assembly of a BCL10 signaling complex that activates NF-κB, which is responsible for both innate and adaptive immune responses[54]. • The rs4077515 risk allele is associated with increased expression of *CARD9*, and has known association with increased risk of ulcerative colitis and Crohn's disease[11, 12, 34, 55, 56]. Conversely, a rare protein-truncating splice variant in *CARD9* confers additive protection from inflammatory bowel disease[56, 57]. • Familial CARD9 deficiency predisposes to invasive fungal infections[58]. • CARD9 mediates intestinal repair, T-helper 17 responses, and control of bacterial infection after intestinal epithelial injury in mice[41].
VAV3	Chemokine signaling pathway, Focal adhesion, Natural killer cell mediated cytotoxicity, T cell receptor signaling pathway, B cell receptor signaling pathway, Fc epsilon RI signaling pathway, Fc gamma R-mediated phagocytosis, Leukocyte transendothelial migration, Regulation of actin cytoskeleton	• VAV proteins (Vav1, 2, and 3) are guanine nucleotide exchange factors essential for adaptive immune function[13, 14] and NF-κB activation in B-cells, a process that stimulates IgA production[43]. • VAV proteins are also required for proper differentiation of colonic enterocytes and preventing spontaneous ulcerations of intestinal mucosa[42]. • VAV3 is a positional candidate for QTL for mouse intestinal inflammation in a parasite-induced (Trichuris muris) model-of infection[59].
DEFA1, DEFA3, DEFA4, DEFA5, DEFA6	Innate immune system	• α-defensins are antimicrobial peptides involved in mucosal defense. • *DEFA5* and *DEFA6* genes expressed by the intestinal Paneth cells. Deficiencies in α-defensins-5 and -6 have been associated with Crohn's disease[39, 40]. While α-defensin-5 is broadly antimicrobial, α-defensin-6 promotes mucosal innate immunity through self-assembled peptide nanonets[60].

Locus (Genes)	Canonical Pathways *	Function and role in intestinal mucosal immunity
TNFSF13	Cytokine-cytokine receptor interaction, Intestinal immune network for IgA production	• *TNFSF13* encodes APRIL, a powerful B-cell stimulating cytokine that promotes CD40-independent IgA class switching[35]. • The IgAN risk allele is associated with increased IgA levels[4]. • *TNFSF13* is induced by intestinal bacteria resulting in IgA class switching. APRIL levels are elevated in some patients with IgAN[61]. • Mutations in the TNFSF13 receptor (TACI) produce IgA deficiency or combined variable immunodeficiency, with increased propensity to mucosal infections[62].
LIF, OSM, HORMAD2, MTMR3	Cytokine-cytokine receptor interaction, Jak-STAT signaling pathway	• The IgAN risk allele at this locus is protective against Crohn's disease[11, 12, 63] and associated with increased serum IgA levels[3]. • LIF and OSM are IL-6 related cytokines that use gp130 for signal transduction, and have been previously implicated in mucosal immunity[64, 65]. Genetic disruption of gp130 signaling leads to gastrointestinal ulceration and inflammatory joint disease in mice[66]. LIF is secreted by pericrypt fibroblasts[67] and may be critical for proliferation and renewal of enterocytes[68].
PSMB8, PSMB9, TAP1, TAP2	Phagosome pathway, Antigen processing and presentation, Primary immunodeficiency, Proteosome, Activation of NFkB in B-cells	• PSMB8 and PSMB9 are interferon-induced subunits of the immunoproteosome that mediate intestinal NF-κB activation in IBD[69]. • PSMB8 is up-regulated in human intestinal tissue with active IBD lesions[70]. • Treatment with bortezomib (PSMB8 inhibitor) or *psmb8* deletion in mice attenuates experimental colitis[71].
HLA-DQA1, HLA-DQB1, HLA-DRB1	Antigen processing and presentation, Adaptive immune system, Intestinal immune network for IgA production, Allograft rejection, Graft versus host disease, Asthma, Autoimmune thyroid disease, Leishmania infection	• The IgAN risk allele is associated with increased risk of Celiac disease[72, 73] and increased risk of IgA deficiency[24]. • The IgAN risk allele has an opposed (protective) effect on the risk of ulcerative colitis[27, 74].

* Canonical pathways based on the Molecular Signature Database (KEGG, Biocarta, and Reactome).