nature genetics

Article

https://doi.org/10.1038/s41588-023-01422-x

Genome-wide association analyses define pathogenic signaling pathways and prioritize drug targets for IgA nephropathy

Received: 6 December 2021

Accepted: 5 May 2023

Published online: 19 June 2023

Check for updates

A list of authors and their affiliations appears at the end of the paper

IgA nephropathy (IgAN) is a progressive form of kidney disease defined by glomerular deposition of IgA. Here we performed a genome-wide association study of 10,146 kidney-biopsy-diagnosed IgAN cases and 28,751 controls across 17 international cohorts. We defined 30 genome-wide significant risk loci explaining 11% of disease risk. A total of 16 loci were new, including *TNFSF4/TNFSF18*, *REL*, *CD28*, *PF4V1*, *LY86*, *LYN*, *ANXA3*, *TNFSF8/TNFSF15*, *REEP3*, *ZMIZ1*, *OVOL1/RELA*, *ETS1*, *IGH*, *IRF8*, *TNFRSF13B* and *FCAR*. The risk loci were enriched in gene orthologs causing abnormal IgA levels when genetically manipulated in mice. We also observed a positive genetic correlation between IgAN and serum IgA levels. High polygenic score for IgAN was associated with earlier onset of kidney failure. In a comprehensive functional annotation analysis of candidate causal genes, we observed convergence of biological candidates on a common set of inflammatory signaling pathways and cytokine ligand–receptor pairs, prioritizing potential new drug targets.

IgA nephropathy (IgAN) is a common form of immune-mediated glomerulonephritis characterized by glomerular deposition of IgA-containing immune complexes. Examination of kidney-biopsy tissue and demonstration of glomerular IgA deposits are required to establish the diagnosis. No approved targeted therapies presently exist for IgAN, and a large fraction of cases progress to kidney failure requiring kidney transplantation or dialysis.

Because the diagnosis requires a kidney biopsy, genetic discoveries have been hindered by small patient cohorts. Approximately 15 independent risk loci have been previously identified for IgAN, implicating defects in the complement pathway, intestinal network of IgA production and innate immunity against mucosal pathogens[1–7]. These findings have led to the reformulation of the disease pathogenesis model, with most candidate mechanisms mapping to the immune system rather than the kidney[8,9]. Nevertheless, prior genome-wide association studies (GWASs) had a two-stage design, with the sample size of the discovery stage limiting the power to identify new loci.

Herein we report a large GWAS discovery study for IgAN, involving 38,897 individuals (10,146 kidney biopsy-defined cases and 28,751 controls) recruited across 17 international cohorts. With the discovery of 16 new non-HLA (human leukocyte antigen) risk loci, we provide strong support for a highly polygenic architecture of IgAN. We assess the functional consequences of the risk alleles, define causal cell types and pathways and explore genetic correlations and pleiotropic associations of the risk loci. Notably, we report the convergence of multiple risk loci on a set of common signaling pathways and ligand–receptor pairs involved in the regulation of IgA production, prioritizing plausible new molecular drug targets.

Results

Study design

We performed GWAS and meta-analysis of 17 independent international case–control cohorts (12 newly genotyped and five published cohorts) comprising 38,897 individuals (10,146 cases and 28,751 controls). All cases were defined by dominant mesangial IgA staining by immunofluorescence of kidney biopsy tissue. We excluded any cases with suspected secondary causes (for example, liver, autoimmune or malignant disease). The patient characteristics are summarized in Supplementary Table 1. The controls were unrelated to the cases, were derived from the same populations or geographic regions as the cases and had no known kidney disease. Of the 17 cohorts, 11 were of European ancestry, with participants from Italy, Poland, Germany, France,

e-mail: kk473@columbia.edu; ag2239@cumc.columbia.edu

https://doi.org/10.1038/s41588-023-01422-x

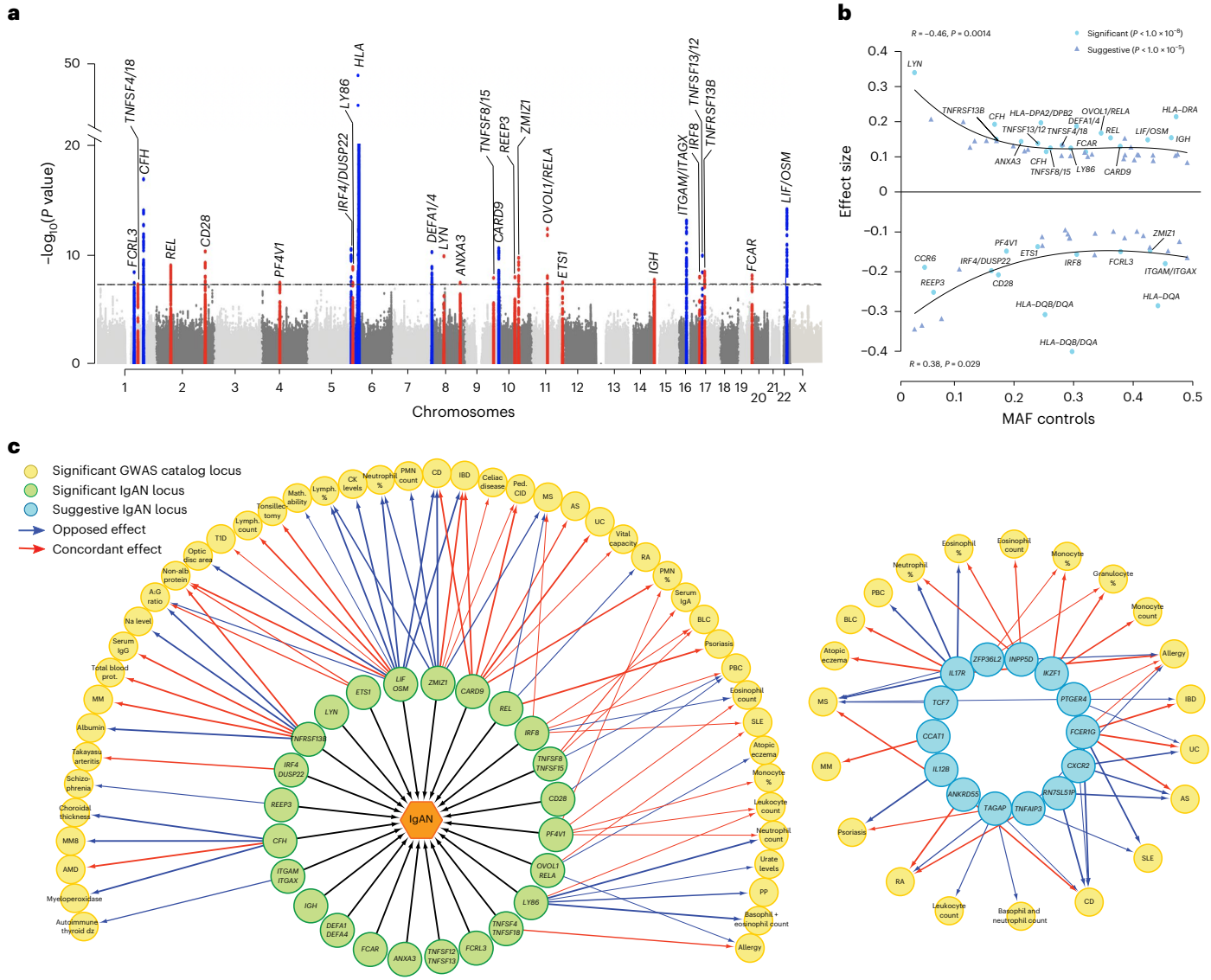


Fig. 1 | Cross-ancestry GWAS for IgAN. a, Manhattan plot for the combined meta-analysis across 38,897 individuals. The dashed horizontal line indicates a genome-wide significant *P* value = 5.0×10^{-8}. The *y* axis shows $-\log_{10}$ of two-sided *P* values (fixed-effects meta-analysis of all cohorts without correcting for multiple testing) and is truncated to accommodate the HLA signal. The *x* axis shows genomic position for each chromosome (1–22 and X). Red, new genome-wide significant loci associated with IgAN; dark blue, previously known loci reaching genome-wide significance in this study. Locus name based on the top candidate gene from our biological prioritization strategy. **b**, Effect size (β, *y* axis) as a function of MAF (*x* axis) for suggestive and significant GWAS loci. Minor alleles with positive effect sizes (risk alleles) are represented at the top, and negative effect sizes (protective alleles) are represented at the bottom. A three-degree polynomial regression curve was fitted to illustrate positive and negative correlations. Correlation coefficients (*R*) and their corresponding *P* values (*P*) are also provided. Light-blue circles represent genome-wide significant loci and are labeled using the most likely candidate gene per locus. Blue triangles represent suggestive loci. **c**, Pleiotropic effects of non-HLA GWAS loci for IgAN based on the NHGRI GWAS catalog. Only genome-wide significant associations in LD ($r^2 > 0.5$) with IgAN top SNPs are included as edges. Yellow are diseases and traits sharing at least one locus with IgAN. Edge thickness is proportional to the LD between the IgAN top SNP and the lead association SNP for GWAS catalog traits. Concordant effects are indicated in red, and opposed effects are indicated in blue. Green nodes represent IgAN GWAS loci, and light-blue nodes are IgAN-suggestive loci. Only 14 suggestive loci sharing at least one pleiotropic association with a genome-wide significant IgAN locus are depicted.

Belgium, Czech Republic, Hungary, Croatia, Turkey, Spain, Sweden, the United Kingdom, the United States, Canada and Argentina, and six were of East Asian ancestry, with participants from China, Japan and the Republic of Korea. A total of 14 cohorts (8,139 cases and 17,713 controls) were genotyped with high-density SNP arrays, ancestrally matched to controls and imputed using ancestry-specific reference panels, and three additional cohorts (2,007 cases and 11,038 controls) were genotyped with the Immunochip (Supplementary Table 2).

Detailed descriptions of each cohort are provided in the Supplementary Note. Our primary discovery involved the combined analysis of all 17 cohorts under an additive genetic model. Additional exploratory analyses were conducted to identify potential ancestry-specific and sex-specific loci, including under alternative (dominant and recessive) genetic models.

Genome-wide significant loci

Our primary analysis consisted of a combined meta-analysis across all cohorts (Fig. 1, Table 1 and Supplementary Tables 3–5). We confirmed multiple independently genome-wide significant ($P < 5 \times 10^{-8}$) signals in the HLA region, with an overall $\lambda = 1.048$ and 1.042 before and after excluding the HLA region (Extended Data Fig. 1). In addition, we detected 24 independent non-HLA loci at genome-wide

https://doi.org/10.1038/s41588-023-01422-x

Table 1 | New and known genome-wide significant loci based on meta-analysis

Chr.	Genomic location (bp[a])	SNP[b]	Locus	Risk allele	Freq. European controls	Freq. Asian controls	Q test	I^2	OR (95% CI)	Fixed effects P value[c]	TransMeta random effects P value[c]	Supporting cohorts[d]	New
1	157,542,162	rs849815	*FCRL3*	A	0.66	0.50	0.57	0	1.14 (1.09–1.19)	3.9×10^{-9}	3.6×10^{-9}	GWAS+IC	Known
1	173,146,357	rs4916312	*TNFSF4/18*	A	0.35	0.07	0.68	0	1.14 (1.09–1.20)	5.0×10^{-8}	6.3×10^{-8}	GWAS+IC	New
1	196,686,918	rs6677604	*CFH*	G	0.80	0.93	0.97	0	1.21 (1.20–1.34)	1.5×10^{-17}	5.9×10^{-17}	GWAS+IC	Known
1	196,603,302	rs12029571	*CFH*	A	0.22	0.35	0.88	0	1.12 (1.07–1.18)	2.5×10^{-6}	3.3×10^{-6}	GWAS	Known
2	61,092,678	rs842638	*REL*	T	0.44	0.15	0.002[e]	64	1.17 (1.11–1.23)	9.6×10^{-10}	1.1×10^{-9}	GWAS	New
2	204,584,759	rs3769684	*CD28*	T	0.95	0.49	0.93	0	1.19 (1.13–1.25)	5.1×10^{-11}	2.1×10^{-10}	GWAS+IC	New
4	74,725,320	rs6828610	*PF4V1*	G	0.16	0.27	0.98	0	1.14 (1.10–1.22)	3.5×10^{-8}	3.5×10^{-8}	GWAS	New
6	249,571	rs12201499	*IRF4/DUSP22*	C	0.12	0.28	0.39	5	1.18 (1.15–1.29)	3.1×10^{-11}	4.0×10^{-11}	GWAS	Known
6	7,214,676	rs12530084	*LY86*	C	0.77	0.51	0.49	0	1.13 (1.10–1.21)	1.3×10^{-9}	1.7×10^{-9}	GWAS	New
6	32,389,305	rs9268557	*HLA-DRA*	C	0.51	0.57	<0.001[e]	72	1.24 (1.27–1.36)	4.5×10^{-47}	1.5×10^{-51}	GWAS+IC	Known
6	32,667,829	rs9275355	*HLA-DQB/DQA*	C	0.23	0.33	0.21	25	1.26 (1.30–1.43)	1.7×10^{-34}	2.1×10^{-37}	GWAS	Known
6	32,599,999	rs9272105	*HLA-DQA*	A	0.60	0.45	0.001[e]	73	1.25 (1.20–1.30)	1.2×10^{-28}	5.2×10^{-31}	GWAS+IC	Known
6	32,681,631	rs9275596	*HLA-DQB/DQA*	T	0.66	0.81	0.05	46	1.33 (1.27–1.39)	3.2×10^{-36}	1.4×10^{-35}	GWAS+IC	Known
6	33,074,288	rs3128927	*HLA-DPA/DPB*	C	0.73	0.83	0.38	6	1.22 (1.22–1.34)	1.5×10^{-25}	5.4×10^{-25}	GWAS+IC	Known
8	6,808,722	rs2075836	*DEFA1/4*	T	0.31	0.30	0.90	0	1.21 (1.14–1.28)	5.8×10^{-11}	2.2×10^{-10}	GWAS	Known
8	56,852,496	rs75413466	*LYN*	A	0.02	0.06	0.80	0	1.40 (1.27–1.56)	1.4×10^{-10}	2.4×10^{-10}	GWAS+IC	New
8	124,765,474	rs34354351	*ANXA3*	T	0.17	0.32	0.84	0	1.15 (1.10–1.21)	3.5×10^{-8}	3.9×10^{-8}	GWAS	New
9	117,643,362	rs13300483	*TNFSF8/15*	T	0.24	0.31	0.88	0	1.13 (1.09–1.18)	1.3×10^{-8}	2.5×10^{-8}	GWAS+IC	New
9	139,266,496	rs4077515	*CARD9*	T	0.41	0.29	0.41	2	1.14 (1.10–1.18)	2.6×10^{-11}	1.1×10^{-11}	GWAS+IC	Known
10	65,363,048	rs57917667	*REEP3*	G	0.02	0.19	0.91	0	1.22 (1.18–1.40)	1.1×10^{-8}	1.3×10^{-8}	GWAS	New
10	81,043,743	rs1108618	*ZMIZ1*	A	0.60	0.49	0.43	0	1.14 (1.09–1.18)	1.9×10^{-10}	3.9×10^{-10}	GWAS+IC	New
11	65,555,524	rs10896045	*OVOL1/RELA*	A	0.30	0.48	0.03[e]	51	1.18 (1.13–1.24)	4.7×10^{-13}	8.5×10^{-14}	GWAS	New
11	128,487,069	rs7121743	*ETS1*	C	0.16	0.47	0.79	0	1.13 (1.09–1.20)	3.4×10^{-8}	3.2×10^{-8}	GWAS	New
14	107,222,014	rs751081288	*IGH*	A	0.43	0.53	0.74	0	1.17 (1.11–1.23)	1.9×10^{-8}	2.1×10^{-8}	GWAS	New
16	31,357,760	rs11150612	*ITGAM/ITGAX*	A	0.64	0.27	0.10	37	1.16 (1.12–1.21)	8.4×10^{-14}	3.4×10^{-13}	GWAS+IC	Known
16	86,017,715	rs1879210	*IRF8*	T	0.64	0.86	0.96	0	1.14 (1.09–1.20)	9.9×10^{-9}	1.4×10^{-8}	GWAS+IC	New
17	7,462,969	rs3803800	*TNFSF12/13*	A	0.21	0.32	0.19	27	1.15 (1.10–1.20)	1.2×10^{-10}	5.2×10^{-11}	GWAS+IC	Known
17	16,851,450	rs57382045	*TNFRSF13B*	A	0.11	0.33	0.84	0	1.16 (1.11–1.22)	3.4×10^{-9}	3.6×10^{-9}	GWAS	New
19	55,397,217	rs1865097	*FCAR*	A	0.30	0.38	0.49	0	1.12 (1.08–1.16)	7.7×10^{-9}	1.3×10^{-8}	GWAS+IC	New
22	30,512,478	rs4823074	*LIF/OSM*	G	0.54	0.67	0.51	0	1.16 (1.14–1.24)	7.8×10^{-15}	9.2×10^{-16}	GWAS	Known

[a]Genome Reference Consortium Human Build 37 (hg19). [b]Only independent SNPs in each locus are included. [c]Two-sided *P* values for variant association without multiple testing correction. [d]GWAS cohorts: $n=8{,}139$ cases and 17,713 controls; GWAS+IC cohorts: $n=10{,}146$ cases and 28,751 controls. [e]Heterogeneity $P<0.05$. Chr., chromosome; freq., frequency; *Q* test, two-sided *P* value for Cochrane's *Q* statistic; I^2, heterogeneity index; IC, Immunochip.

significance, including eight known loci (*CFH*, *FCRL3*, *IRF4/DUSP22*, *DEFA1/4*, *CARD9*, *ITGAM/ITGAX*, *TNFSF13/12*, *LIF/OSM*) and 16 new loci (*TNFSF4/18*, *CD28*, *REL*, *PF4V1*, *LY86*, *LYN*, *ANXA3*, *TNFSF8/15*, *ZMIZ1*, *REEP3*, *OVOL1/RELA*, *ETS1*, *IGH*, *IRF8*, *TNFRSF13B* and *FCAR*; Extended Data Fig. 2), and 48 suggestive loci at $P<1\times10^{-5}$ (Supplementary Tables 3 and 4).

Secondary analyses by ancestry revealed another locus (*CCR6*, $P=3.9\times10^{-8}$) evident only in the East Asian cohorts (Extended Data Fig. 3a) and 11 additional suggestive signals in this ancestral group (Supplementary Table 5a). One of the suggestive signals in the combined meta-analysis (*PADI3/PADI4* locus) reached genome-wide significance in the East Asian meta-analysis under a recessive model (Extended Data Fig. 3b). The European-only meta-analysis showed a total of 19 suggestive signals, but no new genome-wide significant loci (Supplementary Table 5b). No sex-specific loci were found in a sex-stratified meta-analysis or in the analysis of sex chromosomal markers.

We next performed stepwise conditional analyses of the 24 significant non-HLA loci, but only the *CFH* locus showed evidence of two independent genome-wide significant variants (Supplementary Table 6 and Extended Data Fig. 4a). Stepwise conditional analyses of the *HLA* region revealed a complex pattern of association, with at least five independently significant SNPs (Extended Data Fig. 4b and Supplementary Table 7). In addition, the patterns of association across the HLA region

https://doi.org/10.1038/s41588-023-01422-x

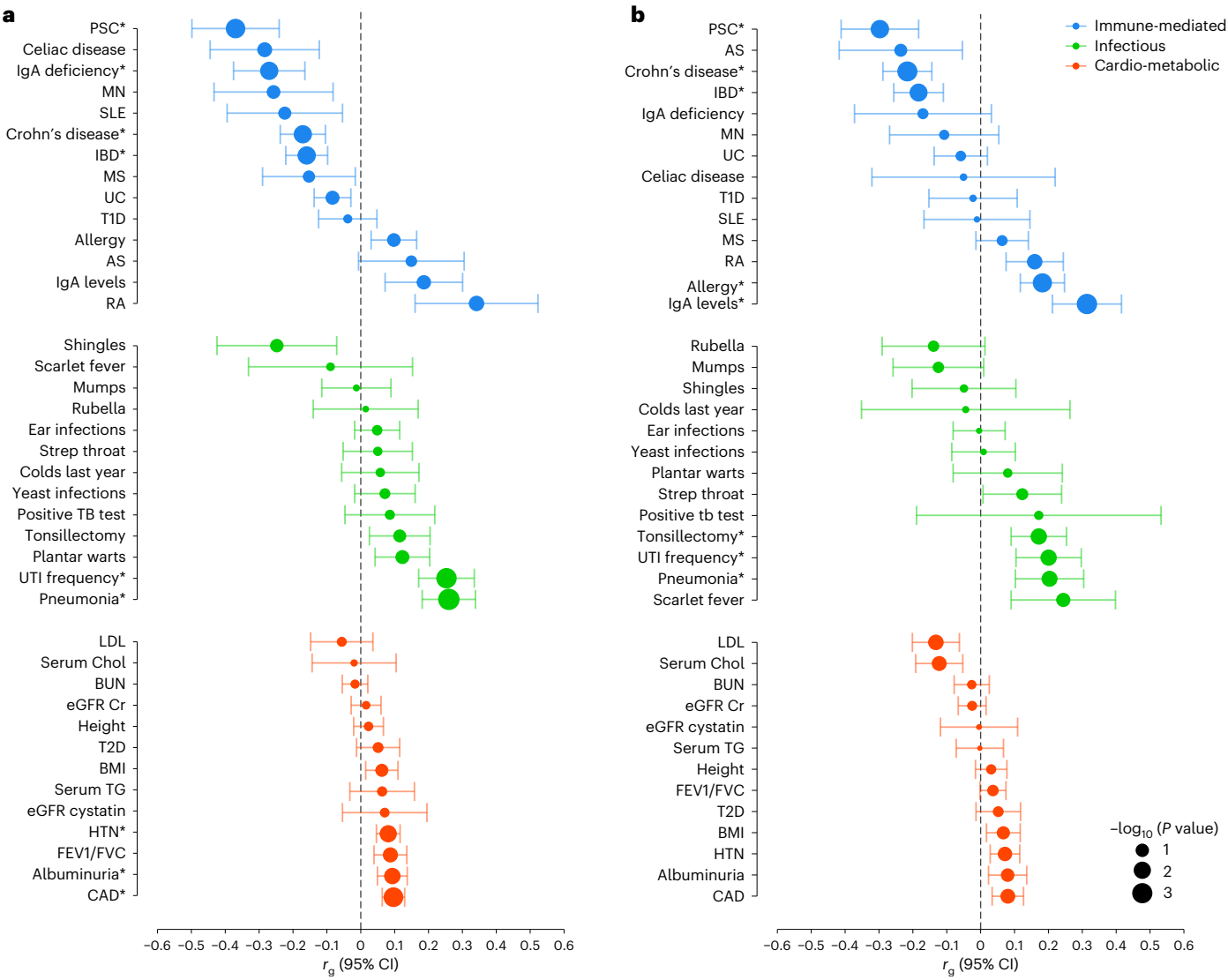


Fig. 2 | Genome-wide genetic correlation analysis between IgAN and other complex traits. a, Including HLA region. **b**, Excluding HLA region. The traits are organized by immune-mediated (blue), infectious (green) and cardiometabolic (orange) categories and sorted based on the genetic correlation coefficient (r_g). The data are presented as r_g point estimates (center) with bars corresponding to 95% CIs. PSC, primary sclerosing cholangitis; MN, membranous nephropathy; SLE, systemic lupus erythematosus; IBD, inflammatory bowel disease; MS, multiple sclerosis; UC, ulcerative colitis; T1D, type 1 diabetes; AS, ankylosing spondylitis; RA, rheumatoid arthritis; Chol, total serum cholesterol levels; TG, total serum triglycerides levels; LDL, low-density lipoprotein levels; BUN, blood urea nitrogen; eGFR Cr, estimated glomerular filtration rate using serum creatinine levels; T2D, type 2 diabetes; BMI, body mass index; eGFR cystatin, estimated glomerular filtration rate using serum cystatin levels; HTN, essential hypertension; FEV1/FVC, forced expiration volume at 1 s over forced vital capacity; CAD, coronary artery disease. The asterisk indicates nominal two-sided $P < 0.05$ for a test of genetic correlation without multiple testing correction. Supplementary Table 13 provides references to all GWAS used in this analysis along with the statistics for each correlation test.

differed when stepwise conditioning was performed separately in East Asian and European cohorts—a total of five independent signals were detected in Europeans and four in East Asians. In the meta-analysis, we observed an inverse relationship between minor allelic frequency (MAF) of the top variants and their effect sizes (Fig. 1b). We estimated the overall SNP-based heritability of IgAN at 0.23 (95% confidence interval (CI): 0.15–0.30), and the heritability estimates were comparable between East Asian and European subgroups at 0.27 (95% CI: 0.10–0.43) and 0.24 (95% CI: 0.09–0.40), respectively. Excluding the HLA region reduced the SNP-based heritability estimate to 0.12 (95% CI: 0.10–0.13). The genetic risk score (GRS) based on 30 independently significant SNPs (30-SNP GRS) explained 11% of overall disease variance, a notable improvement compared to the 6% explained by the previous 15-SNP GRS[4].

Classical HLA alleles

To better understand the signal at the HLA region, we imputed amino acid sequences and classical HLA alleles at four-digit resolution using ancestry-specific reference panels (Methods). The analysis of imputed amino acid sequences in class II genes pointed to *DRB1* as the gene with most strongly associated polymorphic positions in both ancestral groups (Extended Data Fig. 5). In East Asian cohorts, stepwise conditioning demonstrated independently significant associations at DRβ1 positions 11 and 71, and the same positions were also associated with the disease in Europeans. Specifically, proline at position 11 (in linkage disequilibrium (LD) with alanine at position 71) conferred significant protection, while arginine at position 71 (in LD with valine at position 11) was associated with increased risk in both ancestral groups (Supplementary Table 8). In Europeans, we additionally observed significant

effects of glycine (protective) and leucine (risk) at position 11, but these substitutions are infrequent in East Asians.

The association patterns of classical HLA alleles were complex but generally consistent with the analysis of amino acid sequences. In East Asians, we observed a protective effect of the *DRB1*1501–DQA1*0102–DQB1*0602* haplotype (DR15 serotype), and an independent risk effect of *DRB1*0405*, with no significant associations after conditioning for both *DRB1*1501* and *DRB1*0405* (Supplementary Table 9). In Europeans, we confirmed a strong protective association of the *DRB1*1501–DQA1*0102–DQB1*0602* haplotype. *DRB1*0405* had low frequency in Europeans; thus, this association was not observed. Instead, we observed three additional independent European haplotypes (rare in East Asians), including two protective haplotypes, *DRB1*0301–DQA1*0501–DQB1*0201* (DR3 serotype) and *DRB1*0701–DQA1*0201–DQB1*0202/0203* (DR7 serotype), and one risk haplotype *DRB1*0101–DQA1*0101–DQB1*0501* (DR1 serotype; Supplementary Table 10). After conditioning on the four independently significant *DRB1* alleles residing on these haplotypes, we observed additional independent protective associations of *DQA1*0102* and *DPA1*0103* in Europeans. There were no independently significant associations for the class I genes.

Pleiotropic associations of individual IgAN loci

To describe the full spectrum of pleiotropic associations of individual risk variants, we cross-annotated our non-HLA signals against all studies listed in the National Human Genome Research Institute (NHGRI) GWAS catalog (Supplementary Tables 11 and 12). We identified concordant and opposed associations for multiple autoimmune and inflammatory diseases, suggesting shared pathogenic pathways with IgAN. Among the loci with the highest level of pleiotropy were *LIF/OSM* and *ZMIZ1*, but autoimmune pleiotropy was also evident for *CARD9*, *TNFSF8/15*, *REL*, *OVOL1/RELA*, *IRF4/DUSP22* and *IRF8* loci. Some new loci, including *TNFRSF13B*, *PF4V1*, *LY86* and *ETS1*, showed concordant effects on blood levels of distinct immune cell types or immunoglobulins, suggesting that these loci alter immune cell proliferation and immunoglobulin production. When we expanded this analysis to all suggestive loci, we found that 14 of the 47 suggestive loci were associated with the same autoimmune or blood immune cell traits as the genome-wide significant loci, prioritizing these 14 loci for future follow-up studies (Fig. 1c).

Shared genetic architecture with IgA levels and related traits

To interrogate shared susceptibility between IgAN and other diseases, we explored genome-wide genetic correlations with immune, infectious and cardiometabolic traits using bivariate LD score regression (Fig. 2 and Supplementary Table 13)[10]. We found negative genetic correlations with primary sclerosing cholangitis ($r_g = -0.37$, $P = 4.1 \times 10^{-3}$) and inflammatory bowel disease ($r_g = -0.16$, $P = 9.9 \times 10^{-3}$), and positive correlations with pneumonia ($r_g = 0.26$, $P = 9.0 \times 10^{-4}$) and urinary tract infection ($r_g = 0.25$, $P = 2.1 \times 10^{-3}$). After excluding the HLA region, we observed a positive genetic correlation with serum IgA levels ($r_g = 0.31$, $P = 2.1 \times 10^{-3}$), allergy ($r_g = 0.18$, $P = 5.2 \times 10^{-3}$) and tonsillectomy ($r_g = 0.17$, $P = 0.036$), a procedure performed for recurrent pharyngeal infections and sometimes used to treat relapsing IgAN[11]. We next interrogated all independent IgAN risk alleles in our recent GWAS for serum IgA levels[12] (Supplementary Table 14). Of 25 non-HLA loci, nine were nominally ($P < 0.05$) associated with increased serum IgA levels, all with concordant effects. Conversely, of 31 significant loci for IgA levels, 12 were nominally associated with the risk of IgAN, also with concordant effects. The intersection includes the following four genome-wide significant loci in both GWAS: *TNFSF12/13*, *TNFSF8/15*, *OVOL1/RELA* and *LIF/OSM*. At the same time, the effects at the HLA region were either opposed or not associated with serum IgA levels, consistent with our genetic correlation analyses in which positive correlation became significant only after excluding the HLA region.

Mouse phenotypes support IgA dysregulation in IgAN

We tested the candidate gene set defined by our significant GWAS loci for overlap with human ortholog gene sets producing 27 phenotype categories when genetically manipulated in mice. We observed the top-most significant enrichments in 'immune system' ($P = 1.3 \times 10^{-12}$) and 'hematopoietic system' ($P = 3.2 \times 10^{-9}$) phenotypes (Supplementary Table 15). Within these categories, we observed significant enrichments in genes whose disruption in mice was associated with 'abnormal IgA levels' ($P = 6.4 \times 10^{-6}$; Extended Data Fig. 6), including *TNFSF13*, *TNFSF13B*, *ITGAM*, *RELA*, *REL*, *CD28* and *LYN* genes. These observations corroborate our findings of overlapping loci between serum IgA levels and IgAN and further highlight the role of dysregulated IgA production in the disease pathogenesis. Moreover, this analysis also supports the named genes as causal at the corresponding loci and nominates appropriate animal models for experimental follow-up.

Global pathway and tissue/cell-type enrichment analyses

We next used several unbiased strategies to explore biological pathway and tissue enrichments using genome-wide approaches. Pathway-enrichment analysis using multimarker analysis of genomic annotation (MAGMA)[13] revealed 24 enriched gene sets (Extended Data Fig. 7). The most strongly enriched Gene Ontology (GO) terms after excluding the HLA region were 'immune system processes' (enrichment $P = 1.4 \times 10^{-9}$) and 'immune response' (enrichment $P = 2.6 \times 10^{-9}$). Examination of significant non-HLA loci revealed enrichments in pathways involved in innate and adaptive immunity, with the most significant enrichment in the 'cytokine–cytokine receptor interactions' (enrichment $P = 4.0 \times 10^{-11}$; Fig. 3a).

To map the most likely causal tissues and cell types, we partitioned SNP-based heritability by tissue and cell-type-specific FUN-LDA scores[14]. We found the most statistically significant heritability enrichments in blood, immune and gastrointestinal mucosa cells (Fig. 3b and Supplementary Table 16). The top enriched cell types were primary neutrophils from peripheral blood ($P = 5.9 \times 10^{-10}$), PMA (phorbol-12-myristate-13-acetate)-I-stimulated primary T helper cells ($P = 2.1 \times 10^{-9}$) and primary B cells from peripheral blood ($P = 2.0 \times 10^{-8}$). Analogous analysis performed using experimental mouse datasets pointed to small intestine inflammatory cells under basal conditions and after *Salmonella* infection as the top tissue (Extended Data Fig. 7). Additional independent analytical methods (DEPICT[15] and GARFIELD[16]) similarly prioritized extrarenal tissues as likely causal in IgAN, converging on hematopoietic, immune, and gastrointestinal tissues as the most likely tissues to harbor causal cell types (Fig. 3c,d and Supplementary Tables 17 and 18).

Transcription factor (TF) enrichment analysis

We tested for potential intersection of GWAS loci with a comprehensive database of TF chromatin immunoprecipitation (ChIP)–seq datasets using the regulatory element locus intersection (RELI) algorithm[17]. In the analysis of genome-wide significant and suggestive loci, we detected significant intersection with binding sites for up to 32 TFs in 52 immune cell types, with the most significant enrichments for RELA (corrected $P = 5.3 \times 10^{-13}$) and NFKB1 (corrected $P = 1.9 \times 10^{-12}$; Fig. 4d and Supplementary Table 19). Nearly half of these TFs interact with Epstein–Barr virus super-enhancers, which control B-cell proliferation and have previously been found to intersect multiple autoimmune loci[17,18]. Moreover, some of the prioritized TFs, such as RUNX (runt-related transcription factor)[19] and SMAD (sma- and mad-related protein)[20] family, are well known to regulate IgA levels, and *RUNX3*, *RUNX2* and *OVOL1/RELA* loci are substantially associated with serum IgA levels[12], further supporting perturbations in IgA homeostasis as a primary pathogenetic factor in IgAN.

Protein–protein interactions (PPIs) and ligand–receptor pairs

We next tested whether candidate genes within our significant loci encode proteins that are likely to have physical interactions. Using

https://doi.org/10.1038/s41588-023-01422-x

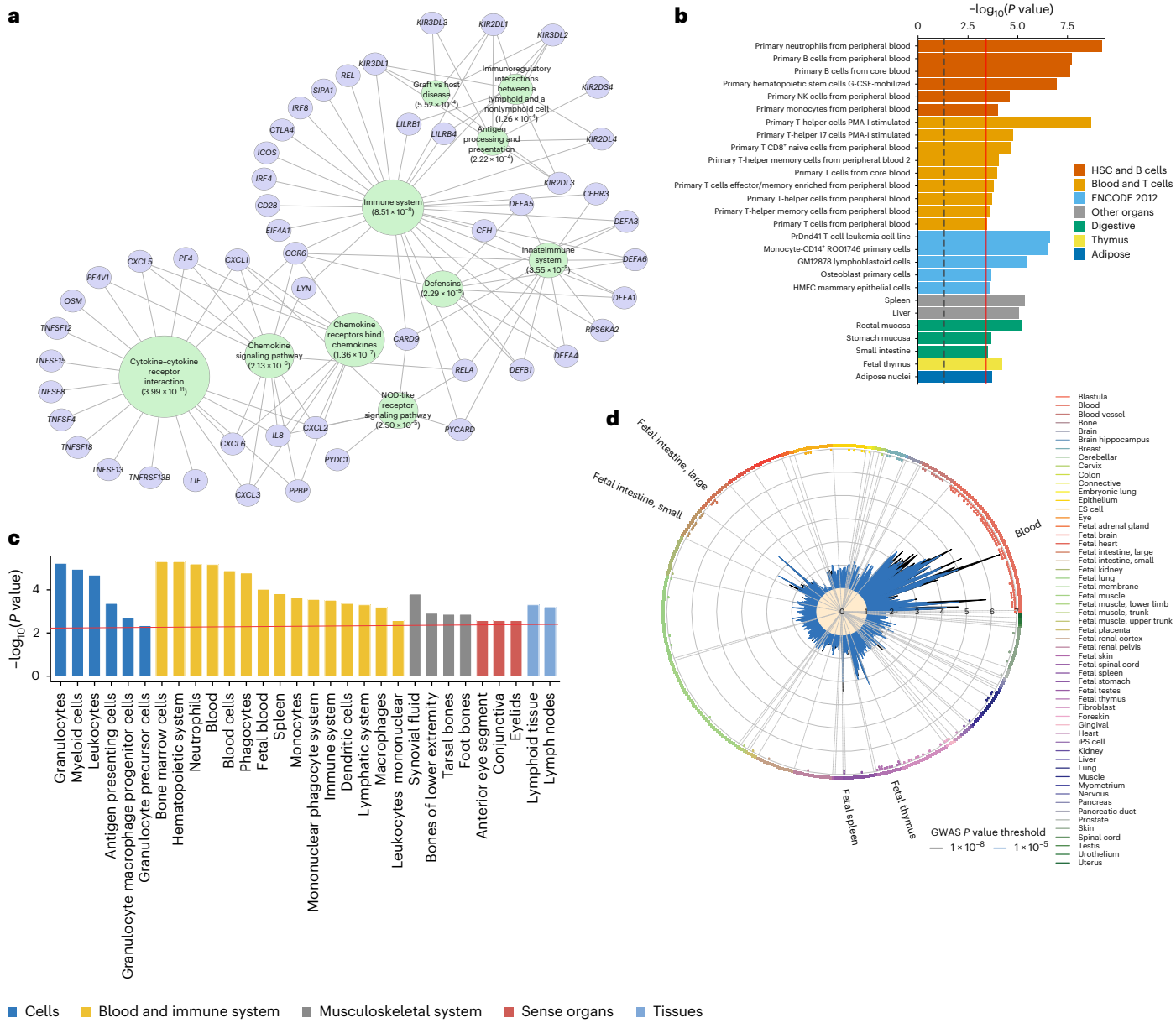


Fig. 3 | Global pathway, cell-type and tissue enrichment analyses. a, KEGG, REACTOME and BIOCARTA pathway-enrichment map based on the gene set defined by genome-wide significant risk loci excluding HLA. The top ten most substantially enriched pathways and their intersecting GWAS genes are shown. Node size reflects $-\log_{10}$-transformed two-sided *P* values of the multiple testing-adjusted hypergeometric enrichment test in GSEA. **b**, Cell-type-specific heritability enrichment for functional annotations based on FUN-LDA scoring system for all ENCODE and Roadmap Epigenomics cell types and tissues. The *x* axis depicts $-\log_{10}$ of two-sided *P* values for heritability enrichment without multiple testing correction with only significant results grouped by the tissue type depicted. Solid red line represents the Bonferroni-corrected significance threshold ($P = 3.9 \times 10^{-4}$). Dashed black line represents the $-\log_{10}$ of the nominal uncorrected *P* value ($P = 0.05$). The most significant heritability enrichments were found in blood immune cells and gastrointestinal tissues. **c**, Tissue and cell-type enrichment analysis with DEPICT; only cells and tissues with a false discovery rate (FDR) < 0.05 are shown. The *y* axis represents the $-\log_{10}$ of the two-sided empirical *P* value without multiple testing correction. The *x* axis shows the first-level Medical Subject Headings (MeSH) annotations. The strongest enrichment is observed for blood and immune cells. The red horizontal line corresponds to FDR = 0.05. **d**, Global GWAS enrichment in DNase I-hypersensitive sites (DHS) using GARFIELD. Radial lines show odds ratios at two genome-wide significance thresholds (*T*) for all DHS cells and tissues on the outer circle. Dots in the inner ring of the outer circle denote significant GARFIELD enrichments for $T < 1.0 \times 10^{-5}$ (outermost) and $T < 1.0 \times 10^{-8}$ (innermost) loci by a two-sided enrichment test after multiple-testing correction for the number of effective annotations. Similar to FUN-LDA, GWAS results are most enriched in DHS sites in blood and immune cells, and intestinal mucosal tissue (labeled). ES cell, embryonic stem cell; iPS cell, induced pluripotent stem cell.

a refined database of high-confidence PPIs, we constructed a network with 76 candidate proteins defined by GWAS using InWeb_IM[21] and GeneMANIA[22]. The final network composed of 53 nodes and 63 edges exhibited an excess of direct physical interactions compared to null expectation ($P < 1.0 \times 10^{-16}$; Fig. 4c). Gene set enrichment analyses (GSEA) of individual modules in this network (Supplementary Table 20) identified strong enrichments in stress and defense responses (module 1), chemokine signaling pathways (module 4), immune responses (module 5), cytokine-mediated signaling (module 6) and regulation of nuclear factor kappa B (NF-κB) signaling (module 7). Consistent with the observed enrichments in chemokine and cytokine pathways and global cytokine–receptor interactions, we

https://doi.org/10.1038/s41588-023-01422-x

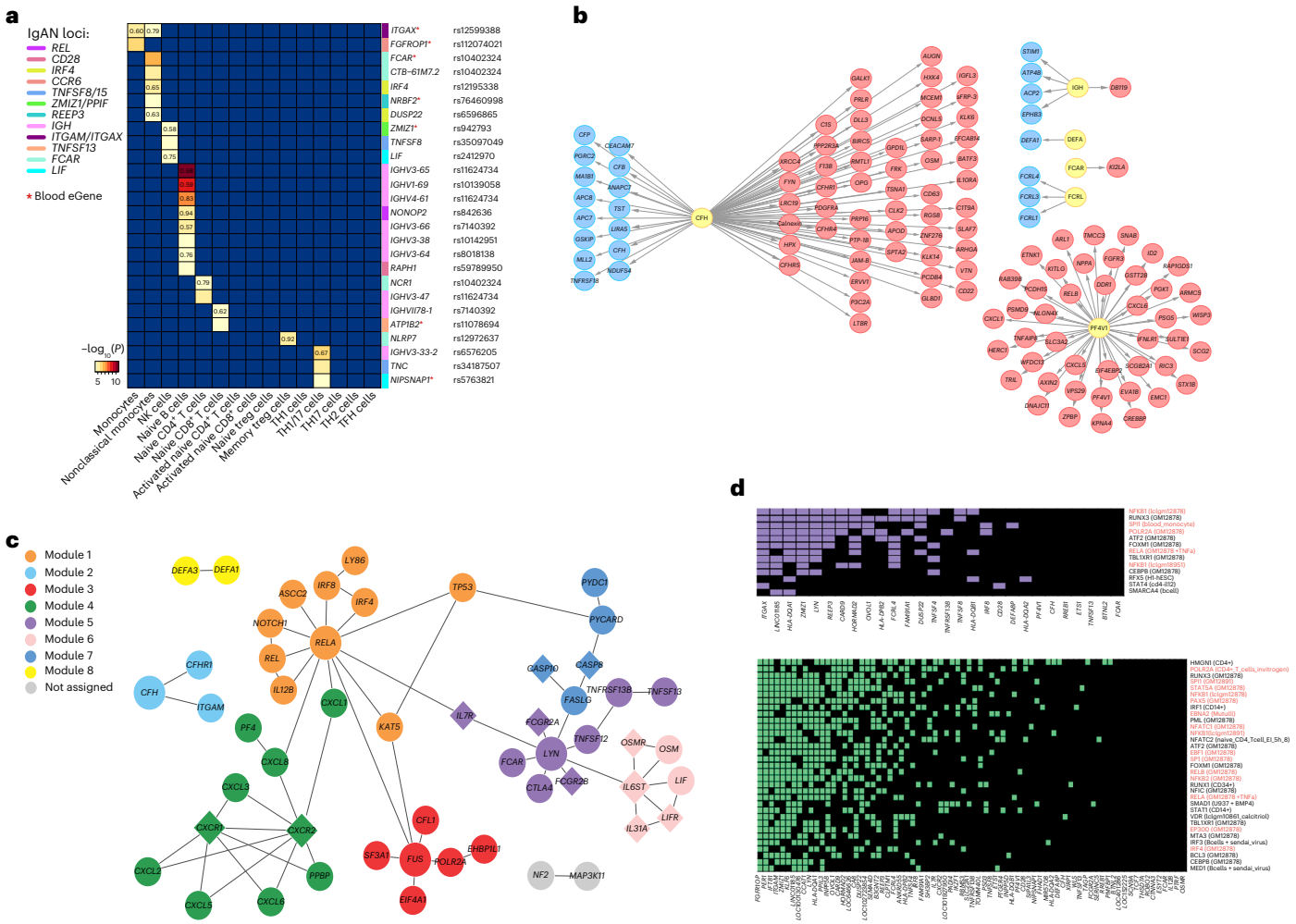


Fig. 4 | Functional annotations of non-HLA loci. a, *Cis*-eQTL effects in primary immune cells. The *x* axis shows 15 immune cell types (DICE project). The *y* axis shows significant eGene–eSNP pairs with shared loci depicted by a color bar. Two-sided eQTL *P* values without multiple testing correction were used for a color scale. Posterior probability for a shared causal variant (PP4) is shown for the eQTL effects that colocalize at PP4 > 0.50. An asterisk indicates eGenes with blood *cis*-eQTLs. **b**, pQTL (protein quantitative trait loci) effects in blood. IgAN risk alleles or their proxies (yellow nodes) with significant blood pQTLs depicted by blue (reduced protein levels) and red (increased protein levels) nodes. Edge thickness corresponds to the LD between the lead GWAS and pQTL SNPs at a given locus. **c**, Protein–protein interaction network for candidate genes at GWAS loci. Modules represent genes that are more connected to one another than they are to other genes. Each of the following modules exhibits a functional enrichment network based on GO biological processes: module 1 (orange) represents response to stress and defense response networks; module 2 (light blue) represents regulation of inflammatory response network; module 3 (red) represents mRNA splicing via spliceosome network; module 4 (green) represents chemokine-mediated signaling pathway; module 5 (purple) represents an immune response network; module 6 (pink) represents cytokine-mediated signaling pathway; module 7 (dark blue) represents regulation of I-κB kinase/NF-κB signaling and apoptotic signaling pathway; and module 8 (yellow) represents innate immune response in mucosa and antibacterial humoral response. The gray module has no functional enrichment. **d**, Intersection with TF ChIP–seq peaks with the significant (top) and suggestive (bottom) IgAN risk loci. The *x* axis shows IgAN risk loci. The *y* axis shows top significant TFs ranked by the number of intersecting loci. A colored box at the intersection indicates that a given locus has at least one IgAN-associated variant located within a ChIP–seq peak for the given TF. Datasets were considered significant if their RELI empirical *P* values corrected for multiple testing were $<1 \times 10^{-4}$. TFs binding to EBNA2 super-enhancers are colored in red; ChIP–seq dataset cell types are indicated in parentheses. Related cell lines for a given TF (for example, GM12878 and GM12891) were merged for clarity. I-κB, inhibitor of nuclear factor kappa B.

identified enrichment in soluble ligand–receptor pairs, attributable to 16 ligand–receptor pairs spanning 12 independent significant or suggestive loci (enrichment $P = 0.01$; Supplementary Table 21). This included APRIL and its receptor TACI (transmembrane activator and CAML interactor) encoded by two independent genome-wide significant loci (*TNFSF12/13* and *TNFRSF13B*, respectively), both implicated in IgA homeostasis. Several IL6-related cytokine–receptor pairs were also identified (IL6–IL6ST, LIF (leukemia inhibitory factor)–LIFR/IL6ST and OSM (oncostatin M)–OSMR/LIFR/IL6ST), with *OSM/LIF* being encoded by a single genome-wide significant locus and related receptors being encoded by two independent suggestive loci, *OSMR/LIFR* and *IL6ST*. Notably, APRIL is known to alter the glycosylation of IgA1 (ref. 23), IL6, LIF and OSM are involved in mucosal immunity, and IL6 and LIF lead to enhanced production of galactose-deficient IgA1 (refs. 24–26). These ligand–receptor pairs nominate candidate genes within corresponding loci and delineate potentially targetable pathways in IgAN.

Functional annotations of individual GWAS loci

We performed genomic annotations of our significant loci, including intersection with tissue and cell-type-specific functional scores, colocalization with expression quantitative trait loci (eQTLs) in primary immune cells, whole blood and other tissues and cross-annotation with blood proteome and metabolome data (Methods)[27]. Most signals mapped to noncoding regions, and there were only two missense

variants (rs4077515 *CARD9* p.(Ser12Asn) and rs3803800 *TNFSF13* p.(Asn96Ser)) among the top SNPs (Supplementary Table 22). The *CARD9* risk allele (rs4077515-T), a nonsynonymous S12N substitution in exon 2 of *CARD9*, is associated with increased blood transcript level of *CARD9* and a significant splice QTL in Genotype-Tissue Expression (GTEx). The protective allele is associated with a truncation of the functional CARD (caspase recruitment domain), while the risk allele is associated with higher levels of the intact, active isoform, affecting both expression and splicing of *CARD9* (Extended Data Fig. 8).

For top signals mapping to noncoding regions, we found 79 significant *cis*-eQTL effects with 17 IgAN colocalizations at 20 independent non-HLA risk loci (Fig. 4a and Supplementary Tables 23 and 24). Twelve loci had 27 significant *cis*-eQTL effects across 15 primary immune cell types, and 17 of the 27 *cis*-eQTLs colocalized with IgAN with PP4 > 0.5 (Supplementary Table 23). In GTEx, we further found 19 *cis*-eQTL effects for eight IgAN loci across the 28 available tissues and cell types. As an example, two loci (*ITGAM/ITGAX* and *IRF4/DUSP22*) mapped specifically to monocytes, an understudied cell type in IgAN. The top signals at these loci intersect monocyte-specific functional elements by FUN-LDA and colocalize with monocyte-specific eQTLs, with the risk alleles associated with upregulation and downregulation of *ITGAX* and *IRF4/DUSP22*, respectively. As another example of cell-type specificity, the *ZMIZ1* locus colocalized with an eQTL in natural killer (NK) cells, with the risk allele associated with lower expression of *ZMIZ1*, which encodes an inhibitor of JAK/STAT (Janus kinase/signal transducer and activator of transcription) signaling and is also involved in transforming growth factor-beta (TGF-β) signaling and intestinal inflammation[28,29]. In whole blood, notable eQTL colocalizations included the *FCRL3* risk locus, where the risk allele was associated with reduced transcript levels of *FCRL3* and *FCRL5*, and with lower levels of circulating FCRL3 protein (Supplementary Table 25). As FCRL3 is a specific receptor for secretory IgA[30,31], we prioritized *FCRL3* as the most likely causal gene at this locus.

Three independent IgAN risk loci with colocalizing *cis*-eQTLs also exhibited *trans*-eQTL effects, suggesting that these loci induce a more global transcriptional perturbation in blood cells (Supplementary Table 26). For example, the *CARD9* locus was associated with 12 *trans*-eQTL effects, nine of which involve genes in the 'type I interferon signaling pathway' (enrichment $P = 9.5 \times 10^{-18}$). The *TNFSF8/15* locus was associated with eight *trans*-eQTL effects with three representing 'cytokines involved in lymphocyte differentiation' (enrichment $P = 4.3 \times 10^{-3}$). Interestingly, the *ITGAM/ITGAX* locus had only one *trans*-eQTL association, lowering mRNA level of *IGHG4*, encoded by an independent IgAN risk locus on chromosome 14.

Several loci were also associated with perturbations in blood proteome or metabolome. The *PF4V1* locus colocalized with a *PF4V1* *cis*-eQTL and exhibited multiple protein QTL (pQTL) associations with blood protein levels (Supplementary Table 25), including four *cis*- and 40 *trans*-pQTL proteins. These proteins were most enriched in the GO process of 'positive regulation of neutrophil chemotaxis' (enrichment $P = 1.3 \times 10^{-3}$), providing additional support for *PF4V1* as the likely causal gene[32]. Similarly, the *CFH* locus, where a protective allele tags a common deletion of *CFHR1* and *CFHR3* (ref. 2), was associated with reduced expression of *CFHR1* and *CFHR3* in the liver, kidney and other tissues (Supplementary Tables 27 and 28). This allele was also associated with reduced levels of circulating FHR1 (encoded by *CFHR1*) and higher blood levels of factor H (Supplementary Table 25). Moreover, this locus exhibited a widespread proteomic and metabolomic signature in blood, with 64 additional *trans*-pQTL associations including seven proteins involved in the 'regulation of complement cascade' (enrichment $P = 2.1 \times 10^{-10}$; Fig. 4b), and altered blood levels of multiple inflammation-related metabolites (Supplementary Table 29)[33,34].

Integrative prioritization of biological candidate genes

To systematically prioritize the 308 candidate genes encoded within the 24 non-HLA risk loci, we scored for convergence of in silico annotation methods by assigning one point for each of the following criteria: (1) genes most proximal to the top SNP at the locus; (2) genes with a nonsynonymous coding variant tagged ($r^2 \geq 0.8$) by the top SNP; (3) genes with a 3D chromatin interaction predicted by the activity-by-contact (ABC) model[35] or (4) GeneHancer[36], with enhancers that are intersected by variants tagged ($r^2 \geq 0.8$) by the top SNP or contained within a 95% credible set for the locus; (5) eGenes controlled by at least one eQTL (any GTEx tissue) tagged by the top SNP; (6) eGenes colocalizing with the risk locus in peripheral blood or (7) primary immune cells at PP4 > 0.5; (8) pGenes encoding blood proteins controlled by at least one *cis*-pQTL tagged by the top SNP; (9) genes prioritized by PPI network connectivity at $P < 0.05$; (10) genes with shared mouse knockout phenotypes; (11) genes within shared MAGMA pathways; (12) genes prioritized by DEPICT; and (13) genes prioritized by manual literature review. Using this approach, we prioritized 26 'biological candidate genes', 19 (73%) of which were also most proximal genes to the top SNP (Fig. 5). This approach had 79% ranking concordance with the variant-to-gene (V2G) scoring method[37].

Prioritization of plausible drug targets

To facilitate drug repurposing and to prioritize new targets with GWAS support, we evaluated whether any of the 308 genes contained within significant loci encoded a protein or directly interacted with a protein that was a pharmacologically active drug target either approved or in development for human disease. In total, 13 GWAS loci (54%) encoded 17 proteins that were already targeted by existing drugs, and 11 loci (46%) encoded 14 proteins with a direct PPI target (Fig. 6 and Supplementary Table 30). Among the top 26 high-priority 'biological candidates', 11 (42%) were targeted directly or indirectly by the existing drugs. This included the following: (1) inhibitors of the alternative complement pathway that are currently in clinical trials for glomerulopathies[38]; (2) drugs targeting B cells by inhibiting APRIL or TACI interactions that are already in clinical trials for IgAN; (3) drugs that inhibit T-cell activation by targeting ligands of CD28 protein, such as Belatacept (approved for allograft rejection) or Abatacept (approved for rheumatoid arthritis); (4) drugs that inhibit IL8 (ABX-IL8) or IL8 receptor (Clotrimazole) and (5) drugs that inhibit NF-κB pathway, such as Bardoxolone that is already in clinical trials for glomerular disorders. We also note that some of our top prioritized causal genes with expression increased by the risk alleles, such as *CARD9*, *ITGAX*, *PF4V1*, *CFHR1* or *FCAR*, do not yet have effective drug inhibitors. Other loci encode secreted proteins that appear protective, such as FCRL3 and TNFSF4, suggesting that targeting their upregulation may present a rational therapeutic strategy. Our data additionally imply that activation of transcriptional programs controlled by ZMIZ1 and IRF4, but reduced activation of NF-κB, may confer a protective effect.

Genome-wide polygenic risk score (GPS) and clinical correlations

Based on GWAS statistics after excluding Immunochip cohorts, we designed and optimized a GPS for IgAN. The best-performing

Fig. 5 | Prioritization of candidate genes at non-HLA loci. Blue boxes indicate prioritization criteria based on genomic coordinates (the nearest gene to the index SNP, exonic variant in LD with the top SNP or top signal intersecting chromatin interaction site with the gene promoter). Red boxes indicate the presence of additional functional criteria (any GTEx eQTL effect, blood and immune cell eQTL colocalization, pQTL effects, PPI network connectivity, shared mouse knockout phenotype, shared pathways by MAGMA, prioritized by DEPICT and prioritized by manual PubMed review). The priority score represents a sum of the 13 scoring criteria depicted in blue and red. The genes with the maximum score at each locus (light green) were defined as 'biological candidate genes'. V2G scores are also provided for comparison. Additional annotation indicates drug-target genes (orange). Only 58 of 311 positional candidate genes with a priority score >3 (or top V2G score per locus) are depicted.

https://doi.org/10.1038/s41588-023-01422-x

GPS was based on LDPred method and assumed 1% causal variants genome-wide. When tested in the independent German Chronic Kidney Disease (GCKD) study[39,40], the GPS explained 7.3% of disease risk ($P = 3.1 \times 10^{-12}$; *C*-statistic, 0.65; 95% CI, 0.61–0.68). We then performed a comprehensive analysis of clinical disease features associated with the GPS (Supplementary Table 31a). Consistent with previous observations for the 15-SNP GRS[4], the GPS was inversely associated with the age at diagnosis. The GPS was also significantly associated

Locus	Top SNP	Candidate gene	Nearest gene	Nonsynonymous variant	ABC promoter interaction	GeneHancer interaction	eQTL in GTEx (any tissue)	Blood eQTL colocalization	Immune cell colocalization	pQTL	PPI	Mouse KO phenotype	MAGMA pathways	DEPICT prioritized	PubMed review	Priority score	V2G score	Drug target	PPI with drug target
1q31	rs6677604	*CFH*	■	■	■	■	■			■	■	■	■	■		10	0.20	■	
		CFHR1				■	■			■	■				■	5	0.19		■
		CFHR3				■	■			■			■		■	4	0.17		
1q23	rs849815	*FCRL3*			■	■	■	■		■				■	■	7	0.36		
		FCRL4	■		■		■			■				■		5	0.22		
		FCRL5			■	■	■	■								4	0.31		
1q25	rs4916312	*TNFSF4*	■		■	■	■	■				■	■		■	8	0.21	■	
		TNFSF18			■		■					■	■			4	0.21		
2q33	rs3769684	*CD28*	■		■	■				■		■		■	■	7	0.11	■	
		CTLA4			■	■					■	■			■	5	0.07	■	■
		RAPH1			■	■	■					■				4	0.05		
2p16	rs842638	*REL*	■		■	■	■				■	■			■	7	0.18		■
		PUS10			■	■	■									!	0.21		
4q13	rs6828610	*PF4V1*	■		■	■	■	■		■	■		■		■	9	0.33		■
		CXCL1			■	■	■			■	■		■	■	■	8	0.27		
		CXCL8			■	■	■				■		■			5	0.17	■	
		CXCL6			■					■	■		■			4	0.23		
		CXCL5			■		■			■	■					4	0.14		
6p25	rs12201499	*IRF4*			■		■		■		■	■	■		■	7	0.11		
		DUSP22	■		■	■	■		■						■	6	0.08		
6p24	rs12530084	*LY86*			■						■	■	■		■	5	NA		
		RREB1	■		■	■						■				4	0.14		
8p23	rs2075836	*DEFA1*								■	■		■		■	4	0.07		
		DEFA4	■		■								■	■		4	0.10		
		DEFA3					■				■				■	!	0.14		
8q12	rs75413466	*LYN*	■		■	■					■	■			■	6	0.22	■	■
8q24	rs34354351	ANXA13	■		■	■							■			4	0.10		
		FAM91A1			■	■	■									!	0.18		
9q34	rs4077515	*CARD9*	■	■	■	■	■	■				■	■	■	■	10	0.42		
		INPP5E				■	■	■				■		■		5	0.24		
		SNAPC4		■	■	■	■									4	0.19		
9q32	rs13300483	*TNFSF15*			■	■	■					■	■	■	■	7	0.12	■	
		TNFSF8	■		■	■	■						■		■	6	0.22		■
10q22	rs1108618	*ZMIZ1*	■		■	■	■	■	■			■			■	8	0.21		
		PPIF			■	■	■	■			■				■	6	0.19	■	
10q21	rs57917667	*REEP3*	■		■	■	■				■					5	0.12		
		NRBF2			■	■	■									!	0.14		
11q13	rs10896045	*OVOL1*	■		■	■	■					■				5	0.28		
		RELA				■					■	■	■		■	5	0.06	■	■
		EHBP1L1				■	■				■	■				4	NA		
		CFL1				■	■				■	■				4	0.22		
11q24	rs7121743	*ETS1*	■		■	■						■			■	5	0.15		
14q32	rs751081288	*IGH*	■			■	■		■			■			■	6	0.24		
16p11	rs11150612	*ITGAX*	■	■	■	■	■	■	■			■		■	■	10	0.25		
		ITGAM			■	■	■				■	■		■	■	7	0.18		■
		PYCARD			■	■					■	■	■			5	0.14		
		ITGAD			■		■					■	■			4	0.22		
		FUS			■	■					■	■				4	0.17	■	
		KAT8			■	■	■					■				4	0.17		■
16q24	rs1879210	*IRF8*	■		■	■					■	■	■	■	■	8	0.14	■	
17p13	rs3803800	*TNFSF13*	■	■		■	■				■	■	■		■	8	0.38	■	
		TNFSF12				■	■				■	■	■	■		6	0.38	■	
17p11	rs57382045	*TNFRSF13B*	■	■		■					■	■	■		■	7	0.11		■
19q13	rs1865097	*FCAR*	■		■		■				■			■	■	6	0.26	■	
22q12	rs4823074	*LIF*			■	■	■		■		■	■	■		■	8	0.26		■
		OSM			■	■				■	■				■	5	NA		
		ASCC2			■	■	■					■				4	0.13		
		HORMAD2	■		■	■					■					4	0.17		

https://doi.org/10.1038/s41588-023-01422-x

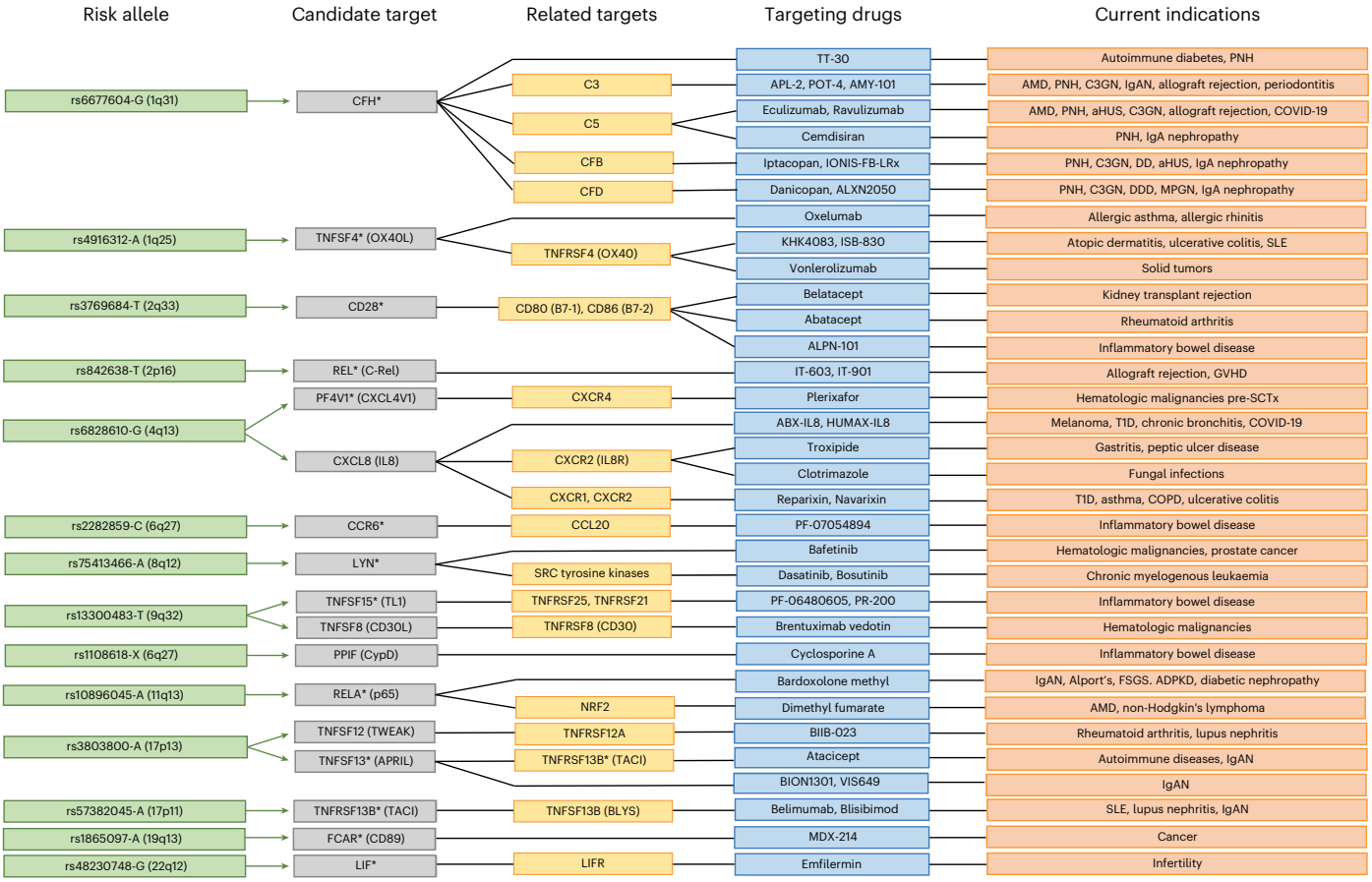


Fig. 6 | Drug targets among candidate causal genes. IgAN risk alleles (green), prioritized positional candidate genes (gray), related genes in PPI (for example, ligands/receptors) or same pathway (yellow), targeting drugs approved or currently in clinical trials including agonists and antagonists (blue) and diseases targeted by these drugs (orange). High-priority targets defined in Fig. 5 are indicated by an asterisk. GWAS loci with candidate causal genes not targeted by existing drugs are not depicted.

with higher lifetime risk of kidney failure among 2,879 cases with long-term follow-up (hazard ratio (HR) = 1.17 per s.d.; 95% CI, 1.09–1.24; $P = 3.3 \times 10^{-6}$). For example, individuals in the top 20% tail of the GPS distribution had 34% increased risk of kidney failure (HR = 1.34; 95% CI, 1.15–1.56; $P = 2.0 \times 10^{-4}$), while individuals in the top 10% tail had 48% increased risk of kidney failure (HR = 1.48; 95% CI, 1.22–1.79; $P = 6.6 \times 10^{-5}$) compared to the rest of the cohort (Fig. 7a). However, the GPS was not a significant predictor of ESKD at the time of kidney biopsy after accounting for age, sex and other clinical risk factors (Supplementary Table 31b).

To explore additional clinical associations of the GPS, we performed a meta-analysis of phenome-wide association study (meta-PheWAS) across 590,515 participants of the UK Biobank (UKBB) and Electronic Medical Records and Genomics-III (eMERGE-III) datasets (Fig. 7b). We detected positive correlation of the GPS with hematuria, the most common manifestation of IgAN ($P = 7.3 \times 10^{-21}$). Other notable associations included a protective association with celiac disease ($P = 4.2 \times 10^{-148}$) and several risk associations, including with rheumatoid arthritis ($P = 1.1 \times 10^{-39}$), hypothyroidism ($P = 2.0 \times 10^{-15}$), epistaxis or throat hemorrhage ($P = 2.6 \times 10^{-9}$) and asthma ($P = 1.5 \times 10^{-6}$). These associations remained significant after removing the HLA region from the GPS (Fig. 7c and Supplementary Table 32). Notably, the directions of effect were generally consistent with our genome-wide genetic correlation analyses of IgAN with related traits, providing a validation of the shared polygenic architecture for these traits.

Discussion

Our GWAS of 10,146 cases and 28,751 controls defined 30 independently significant risk loci and provided support for a highly polygenic architecture of IgAN. The SNP-based heritability of IgAN was estimated at ~23%, and high polygenic risk was associated with earlier disease onset and greater lifetime risk of kidney failure, suggesting that polygenic background is predictive of a more aggressive disease. Future studies are needed to test whether our polygenic stratification is useful in the diagnosis, clinical risk assessment or prediction of treatment responsiveness.

Our results reinforce the hypothesis that the genetic regulation of IgA production represents the key pathogenic pathway in IgAN. Significant risk loci were enriched in human orthologs of mouse genes that, when genetically modified, cause abnormal IgA levels. Moreover, 21 of 25 independent genome-wide significant non-HLA risk loci for IgAN have a concordant effect on serum IgA levels, and four of these are also genome-wide significant in our recent GWAS for serum IgA levels[12].

We observed positive genetic correlations with IgA levels, infections and tonsillectomy, indicating a genetic link between the IgA system, mucosal infections and IgAN. The association with tonsillectomy is especially intriguing, because IgAN is often triggered by pharyngitis, and tonsillectomy has been employed as a treatment for IgAN[11]. In contrast, the observed negative genetic correlations with inflammatory bowel disease may be due to genetically increased production of secretory IgA that has known homeostatic anti-inflammatory and immunosuppressive effects at the level of the gut mucosa[41]. Moreover,

https://doi.org/10.1038/s41588-023-01422-x

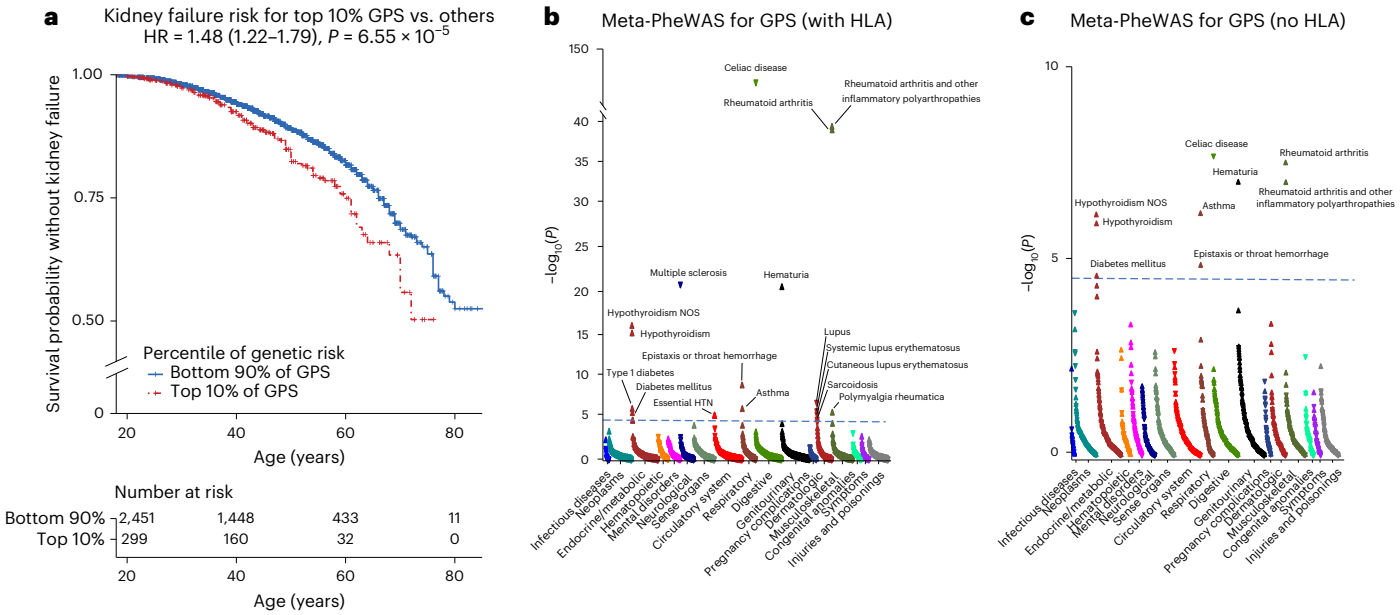


Fig. 7 | Clinical associations of the GPS for IgAN. a, Survival analysis of lifetime risk of kidney failure for IgAN cases in the top 90th percentile of the GPS distribution (n = 2,879 cases with follow-up data). The x axis shows age starting from 18 years, and the y axis shows survival probability without kidney failure with the number of participants at risk at each age cut-off of 20, 40, 60 and 80 years depicted below. HR (95% CI) of kidney failure adjusted for sex, site and ancestry; two-sided P value from the adjusted Cox proportional hazards model is also provided. **b**,**c**, PheWAS for the GPS with (**b**) or without (**c**) the HLA region, based on joint meta-analysis of eMERGE-III (n = 102,138) and UKBB (n = 488,377) datasets. The x axis indicates electronic health record phenotypes (phecodes) grouped by system and sorted by significance. The y axis indicates the level of statistical significance expressed as $-\log_{10}$(P value). All P values are two-sided and correspond to a fixed effects meta-analysis of both datasets without correction for multiple testing. Dashed horizontal line represents the significance threshold after Bonferroni correction for the number of phenotypes; significant associations are labeled. An upward triangle indicates a positive association (increased risk) and a downward triangle indicates a negative association (decreased risk) with increasing GPS.

our analyses of partitioned heritability clearly support a causal role of extrarenal tissues, prioritizing immune, hematopoietic and intestinal mucosal cells. This extrarenal mapping of causal tissues is consistent with the established observation that IgAN commonly recurs after kidney transplantation[42].

Our GWAS loci encoded proteins that were more likely to interact physically despite being encoded by distant genomic regions. This included several ligand–receptor pairs that are amenable to therapeutic targeting. IgAN currently lacks effective targeted therapies, and recent pharmaceutical database analyses indicate that drug targets with genetic support are more likely to advance in the development pipeline[43]. Similar to recent strategies for type 1 diabetes[44] and other autoimmune conditions[45], we prioritized several candidate genes whose products are targeted by drugs that are presently approved or in clinical development for another condition, and which could be repurposed for IgAN. Mechanistic studies are still needed to confirm the candidate target genes prioritized by our in silico annotations.

Our study has several limitations. First, we pooled data across cohorts recruited across diverse timeframes, clinical settings, ancestries and nationalities. To reduce heterogeneity and bias, we used stringent biopsy-based diagnostic criteria, standardized covariates, genetic matching by platform and ancestry and a uniform statistical analysis for each cohort. Second, we note that our meta-analysis included Immunochip and other lower-resolution platforms, resulting in uneven coverage (and power) across the genome. Third, our case cohorts included only biopsy-diagnosed individuals; thus, we are missing patients presenting with mild symptoms that are not routinely biopsied. Our clinical correlation analyses are limited by the retrospective nature of data and the lack of uniform histopathology grading. Fourth, the use of population controls might have led to some control misclassification, although this problem is likely minimized by the low IgAN prevalence. Fifth, our GWAS was performed in European (68.7%) and East Asian (31.3%) ancestries; thus, our results may not be generalizable to other populations. Our functional annotations are also limited by the fact that some of the functional genomic datasets may not be well matched based on ancestry to our study. Our meta-analysis clearly favors loci supported by all cohorts, and we were less powered for the discovery of ancestry-specific or sex-specific effects. Finally, we were not able to evaluate the contribution of rare variants in this study, and sequencing studies are still needed to evaluate relative contributions of rare and common variants to the overall disease risk.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-023-01422-x.

References

1. Feehally, J. et al. HLA has strongest association with IgA nephropathy in genome-wide analysis. *J. Am. Soc. Nephrol.* **21**, 1791–1797 (2010).
2. Gharavi, A. G. et al. Genome-wide association study identifies susceptibility loci for IgA nephropathy. *Nat. Genet.* **43**, 321–327 (2011).
3. Kiryluk, K. et al. Geographic differences in genetic susceptibility to IgA nephropathy: GWAS replication study and geospatial risk analysis. *PLoS Genet.* **8**, e1002765 (2012).
4. Kiryluk, K. et al. Discovery of new risk loci for IgA nephropathy implicates genes involved in immunity against intestinal pathogens. *Nat. Genet.* **46**, 1187–1196 (2014).

5. Yu, X. Q. et al. A genome-wide association study in Han Chinese identifies multiple susceptibility loci for IgA nephropathy. *Nat. Genet.* **44**, 178–182 (2012).
6. Li, M. et al. Identification of new susceptibility loci for IgA nephropathy in Han Chinese. *Nat. Commun.* **6**, 7270 (2015).
7. Li, M. et al. Genome-wide meta-analysis identifies three novel susceptibility loci and reveals ethnic heterogeneity of genetic susceptibility for IgA nephropathy. *J. Am. Soc. Nephrol.* **31**, 2949–2963 (2020).
8. Kiryluk, K., Novak, J. & Gharavi, A. G. Pathogenesis of immunoglobulin A nephropathy: recent insight from genetic studies. *Annu Rev. Med.* **64**, 339–356 (2013).
9. Kiryluk, K. & Novak, J. The genetics and immunobiology of IgA nephropathy. *J. Clin. Invest.* **124**, 2325–2332 (2014).
10. Bulik-Sullivan, B. K. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
11. Feriozzi, S. & Polci, R. The role of tonsillectomy in IgA nephropathy. *J. Nephrol.* **29**, 13–19 (2016).
12. Liu, L. et al. Genetic regulation of serum IgA levels and susceptibility to common immune, infectious, kidney, and cardio-metabolic traits. *Nat. Commun.* **13**, 6859 (2022).
13. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
14. Backenroth, D. et al. FUN-LDA: a latent dirichlet allocation model for predicting tissue-specific functional effects of noncoding variation: methods and applications. *Am. J. Hum. Genet.* **102**, 920–942 (2018).
15. Pers, T. H. et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
16. Iotchkova, V. et al. GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. *Nat. Genet.* **51**, 343–353 (2019).
17. Harley, J. B. et al. Transcription factors operate across disease loci, with EBNA2 implicated in autoimmunity. *Nat. Genet.* **50**, 699–707 (2018).
18. Zhou, H. et al. Epstein–Barr virus oncoprotein super-enhancers control B cell growth. *Cell Host Microbe* **17**, 205–216 (2015).
19. Sugai, M., Watanabe, K., Nambu, Y., Hayashi, T. & Shimizu, A. Functions of Runx in IgA class switch recombination. *J. Cell. Biochem.* **112**, 409–414 (2011).
20. Malhotra, N. & Kang, J. SMAD regulatory networks construct a balanced immune system. *Immunology* **139**, 1–10 (2013).
21. Lage, K. et al. A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat. Biotechnol.* **25**, 309–316 (2007).
22. Warde-Farley, D. et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* **38**, W214–W220 (2010).
23. Han, S. S. et al. The role of TNF superfamily member 13 in the progression of IgA nephropathy. *J. Am. Soc. Nephrol.* **27**, 3430–3439 (2016).
24. Yamada, K. et al. Leukemia inhibitory factor signaling enhances production of galactose-deficient IgA1 in IgA nephropathy. *Kidney Dis. (Basel)* **6**, 168–180 (2020).
25. Suzuki, H. et al. Cytokines alter IgA1 O-glycosylation by dysregulating C1GalT1 and ST6GalNAc-II enzymes. *J. Biol. Chem.* **289**, 5330–5339 (2014).
26. Yamada, K. et al. Inhibition of STAT3 signaling reduces IgA1 autoantigen production in IgA nephropathy. *Kidney Int. Rep.* **2**, 1194–1207 (2017).
27. Vosa, U. et al. Large-scale *cis*- and *trans*-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat. Genet.* **53**, 1300–1310 (2021).
28. Owen, C. R., Yuan, L. & Basson, M. D. Smad3 knockout mice exhibit impaired intestinal mucosal healing. *Lab. Invest.* **88**, 1101–1109 (2008).
29. Yang, X. et al. Targeted disruption of SMAD3 results in impaired mucosal immunity and diminished T cell responsiveness to TGF-β. *EMBO J.* **18**, 1280–1291 (1999).
30. Agarwal, S. et al. Human Fc receptor-like 3 inhibits regulatory T cell function and binds secretory IgA. *Cell Rep.* **30**, 1292–1299 (2020).
31. Tolnay, M. Lymphocytes sense antibodies through human FCRL proteins: emerging roles in mucosal immunity. *J. Leukoc. Biol.* **111**, 477–487 (2021).
32. Choobdar, S. et al. Assessment of network module identification across complex diseases. *Nat. Methods* **16**, 843–852 (2019).
33. Zhang, X. et al. An integrative investigation of the therapeutic mechanism of *Ainsliaea fragrans* Champ. in cervicitis using liquid chromatography tandem mass spectrometry based on a rat plasma metabolomics strategy. *J. Pharm. Biomed. Anal.* **156**, 221–231 (2018).
34. Zhao, L. et al. Stachydrine ameliorates isoproterenol-induced cardiac hypertrophy and fibrosis by suppressing inflammation and oxidative stress through inhibiting NF-kappaB and JAK/STAT signaling pathways in rats. *Int. Immunopharmacol.* **48**, 102–109 (2017).
35. Fulco, C. P. et al. Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 (2019).
36. Fishilevich, S. et al. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford)* **2017**, bax028 (2017).
37. Mountjoy, E. et al. An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. *Nat. Genet.* **53**, 1527–1533 (2021).
38. Zipfel, P. F. et al. Complement inhibitors in clinical trials for glomerular diseases. *Front. Immunol.* **10**, 2166 (2019).
39. Titze, S. et al. Disease burden and risk profile in referred patients with moderate chronic kidney disease: composition of the German Chronic Kidney Disease (GCKD) cohort. *Nephrol. Dial. Transpl.* **30**, 441–451 (2015).
40. Wunnenburger, S. et al. Associations between genetic risk variants for kidney diseases and kidney disease etiology. *Sci. Rep.* **7**, 13944 (2017).
41. Hansen, I. S., Baeten, D. L. P. & den Dunnen, J. The inflammatory function of human IgA. *Cell. Mol. Life Sci.* **76**, 1041–1055 (2019).
42. Infante, B. et al. Recurrence of immunoglobulin A nephropathy after kidney transplantation: a narrative review of the incidence, risk factors, pathophysiology and management of immunosuppressive therapy. *Clin. Kidney J.* **13**, 758–767 (2020).
43. Nelson, M. R. et al. The support of human genetic evidence for approved drug indications. *Nat. Genet.* **47**, 856–860 (2015).
44. Robertson, C. C. et al. Fine-mapping, trans-ancestral and genomic analyses identify causal variants, cells, genes and drug targets for type 1 diabetes. *Nat. Genet.* **53**, 962–971 (2021).
45. Fang, H. et al. A genetics-led approach defines the drug target landscape of 30 immune-related traits. *Nat. Genet.* **51**, 1082–1091 (2019).

Krzysztof Kiryluk[1,2,134] ✉, Elena Sanchez-Rodriguez[1,134], Xu-Jie Zhou[3,134], Francesca Zanoni[1], Lili Liu[1], Nikol Mladkova[1], Atlas Khan[1], Maddalena Marasa[1], Jun Y. Zhang[1], Olivia Balderes[1], Simone Sanna-Cherchi[1,2], Andrew S. Bomback[1], Pietro A. Canetta[1], Gerald B. Appel[1], Jai Radhakrishnan[1], Hernan Trimarchi[4], Ben Sprangers[5,6], Daniel C. Cattran[7], Heather Reich[7], York Pei[7], Pietro Ravani[8], Kresimir Galesic[9], Dita Maixnerova[10], Vladimir Tesar[10], Benedicte Stengel[11], Marie Metzger[11], Guillaume Canaud[12], Nicolas Maillard[13], Francois Berthoux[13], Laureline Berthelot[14], Evangeline Pillebout[15], Renato Monteiro[15], Raoul Nelson[16], Robert J. Wyatt[17,18], William Smoyer[19], John Mahan[19], Al-Akash Samhar[20], Guillermo Hidalgo[21], Alejandro Quiroga[22], Patricia Weng[23], Raji Sreedharan[24], David Selewski[25], Keefe Davis[26], Mahmoud Kallash[27], Tetyana L. Vasylyeva[28], Michelle Rheault[29], Aftab Chishti[30], Daniel Ranch[31], Scott E. Wenderfer[32], Dmitry Samsonov[33], Donna J. Claes[34], Oleh Akchurin[35], Dimitrios Goumenos[36], Maria Stangou[36], Judit Nagy[37], Tibor Kovacs[37], Enrico Fiaccadori[38], Antonio Amoroso[39], Cristina Barlassina[40], Daniele Cusi[40], Lucia Del Vecchio[41], Giovanni Giorgio Battaglia[42], Monica Bodria[43], Emanuela Boer[44], Luisa Bono[45], Giuliano Boscutti[46], Gianluca Caridi[47], Francesca Lugani[47], GianMarco Ghiggeri[47], Rosanna Coppo[48], Licia Peruzzi[48], Vittoria Esposito[49], Ciro Esposito[49], Sandro Feriozzi[50], Rosaria Polci[50], Giovanni Frasca[51], Marco Galliani[52], Maurizio Garozzo[42], Adele Mitrotti[53], Loreto Gesualdo[53], Simona Granata[54], Gianluigi Zaza[54], Francesco Londrino[55], Riccardo Magistroni[56], Isabella Pisani[38], Andrea Magnano[38], Carmelita Marcantoni[57], Piergiorgio Messa[58], Renzo Mignani[59], Antonello Pani[60], Claudio Ponticelli[61], Dario Roccatello[62], Maurizio Salvadori[63], Erica Salvi[64], Domenico Santoro[65], Guido Gembillo[65], Silvana Savoldi[66], Donatella Spotti[67], Pasquale Zamboli[68], Claudia Izzi[69], Federico Alberici[69], Elisa Delbarba[69], Michał Florczak[70], Natalia Krata[70], Krzysztof Mucha[70,71], Leszek Pączek[70,71], Stanisław Niemczyk[72], Barbara Moszczuk[70,73], Malgorzata Pańczyk-Tomaszewska[74], Malgorzata Mizerska-Wasiak[74], Agnieszka Perkowska-Ptasińska[75], Teresa Bączkowska[76], Magdalena Durlik[76], Krzysztof Pawlaczyk[77], Przemyslaw Sikora[78], Marcin Zaniew[79], Dorota Kaminska[80], Magdalena Krajewska[80], Izabella Kuzmiuk-Glembin[81], Zbigniew Heleniak[81], Barbara Bullo-Piontecka[81], Tomasz Liberek[81], Alicja Dębska-Slizien[81], Tomasz Hryszko[82], Anna Materna-Kiryluk[83], Monika Miklaszewska[84], Maria Szczepańska[85], Katarzyna Dyga[85], Edyta Machura[85], Katarzyna Siniewicz-Luzeńczyk[86], Monika Pawlak-Bratkowska[86], Marcin Tkaczyk[86], Dariusz Runowski[87], Norbert Kwella[88], Dorota Drożdż[84], Ireneusz Habura[89], Florian Kronenberg[90], Larisa Prikhodina[91], David van Heel[92], Bertrand Fontaine[93,94], Chris Cotsapas[95], Cisca Wijmenga[96], Andre Franke[97], Vito Annese[98], Peter K. Gregersen[99], Sreeja Parameswaran[100], Matthew Weirauch[100,101], Leah Kottyan[100,101], John B. Harley[102], Hitoshi Suzuki[103], Ichiei Narita[104], Shin Goto[104], Hajeong Lee[105], Dong Ki Kim[105], Yon Su Kim[106], Jin-Ho Park[107], BeLong Cho[107,108], Murim Choi[106], Ans Van Wijk[109], Ana Huerta[110], Elisabet Ars[111], Jose Ballarin[111], Sigrid Lundberg[112], Bruno Vogt[113], Laila-Yasmin Mani[113], Yasar Caliskan[114], Jonathan Barratt[115], Thilini Abeygunaratne[116], Philip A. Kalra[116], Daniel P. Gale[117], Ulf Panzer[118], Thomas Rauen[119], Jürgen Floege[119], Pascal Schlosser[120], Arif B. Ekici[121], Kai-Uwe Eckardt[122,123], Nan Chen[124], Jingyuan Xie[124], Richard P. Lifton[125,126], Ruth J. F. Loos[127,128], Eimear E. Kenny[127,129,130], Iuliana Ionita-Laza[131], Anna Köttgen[120], Bruce A. Julian[132], Jan Novak[132], Francesco Scolari[69], Hong Zhang[3] & Ali G. Gharavi[1,2,134] ✉

[1]Department of Medicine, Vagelos College of Physicians and Surgeons, Columbia University, New York City, NY, USA. [2]Institute for Genomic Medicine, Columbia University, New York City, NY, USA. [3]Renal Division, Peking University First Hospital, Peking University Institute of Nephrology, Beijing, China. [4]Nephrology Service, Hospital Británico de Buenos Aires, Buenos Aires, Argentina. [5]Department of Microbiology and Immunology, Laboratory of Molecular Immunology, KU Leuven, Leuven, Belgium. [6]Division of Nephrology, University Hospitals Leuven, Leuven, Belgium. [7]Department of Nephrology, University of Toronto, Toronto General Hospital, Toronto, Ontario, Canada. [8]Division of Nephrology, Department of Internal Medicine, University of Calgary, Calgary, Alberta, Canada. [9]Zagreb Medical School, Zagreb, Croatia. [10]1st Faculty of Medicine and General University Hospital, Charles University, Prague, Czech Republic. [11]Centre for Research in Epidemiology and Population Health (CESP), Paris-Saclay University, Versailles Saint Quentin University, INSERM Clinical Epidemiology Team, Villejuif, France. [12]Université de Paris, Hôpital Necker–Enfants Malades, Paris, France. [13]Nephrology, Dialysis, and Renal Transplantation Department, University North Hospital, Saint Etienne, France. [14]CRTI, Nantes University, Nantes, France. [15]Center for Research on Inflammation, University of Paris, INSERM and CNRS, Paris, France. [16]Division of Pediatric Nephrology, Department of Pediatrics, University of Utah, Salt Lake City, UT, USA. [17]Division of Pediatric Nephrology, University of Tennessee Health Sciences Center, Memphis, TN, USA. [18]Children's Foundation Research Center, Le Bonheur Children's Hospital, Memphis, TN, USA. [19]Division of Pediatric Nephrology, Nationwide Children's Hospital, Columbus, OH, USA.

[20]Division of Pediatric Nephrology, Driscoll Children's Hospital, Corpus Christi, TX, USA. [21]Division of Pediatric Nephrology, Department of Pediatrics, HMH Hackensack University Medical Center, Hackensack, NJ, USA. [22]Division of Pediatric Nephrology, Helen DeVos Children's Hospital, Grand Rapids, MI, USA. [23]Division of Pediatric Nephrology, Mattel Children's Hospital, Los Angeles, CA, USA. [24]Division of Pediatric Nephrology, Medical College of Wisconsin, Milwaukee, WI, USA. [25]Division of Pediatric Nephrology, Mott Children's Hospital, Ann Arbor, MI, USA. [26]Division of Pediatric Nephrology, Department of Pediatrics, The Medical University of South Carolina (MUSC), Charleston, SC, USA. [27]Division of Pediatric Nephrology, SUNY Buffalo, Buffalo, NY, USA. [28]Division of Pediatric Nephrology, Department of Pediatrics, Nationwide Children's Hospital, Columbus, OH, USA. [29]Division of Pediatric Nephrology, University of Minnesota, Minneapolis, MN, USA. [30]Division of Pediatric Nephrology, University of Kentucky, Lexington, KY, USA. [31]Division of Pediatric Nephrology, Department of Pediatrics, University of Kentucky, Lexington, KY, USA. [32]Division of Pediatric Nephrology, Baylor College of Medicine/Texas Children's Hospital, Houston, TX, USA. [33]Division of Pediatric Nephrology, Boston Children's Hospital, Boston, MA, USA. [34]Division of Pediatric Nephrology, Department of Pediatrics, New York Medical College, New York City, NY, USA. [35]Division of Pediatric Nephrology, Department of Pediatrics, Weill Cornell Medical College, New York City, NY, USA. [36]The Aristotle University of Thessaloniki, Thessaloniki, Greece. [37]2nd Department of Internal Medicine, Nephrological and Diabetological Center, University of Pécs, Pécs, Hungary. [38]Department of Medicine and Surgery, University of Parma, Parma, Italy. [39]Department of Medical Sciences, University of Turin, Turin, Italy. [40]Renal Division, Dipartimento di Medicina, Chirurgia e Odontoiatria, San Paolo Hospital, School of Medicine, University of Milan, Milan, Italy. [41]Department of Nephrology and Dialysis, ASST Lariana, Como, Italy. [42]Unità Operativa di Nefrologia e Dialisi, Ospedale di Acireale, Acireale, Italy. [43]University of Messina, Messina, Italy. [44]Division of Nephrology and Dialysis, Gorizia Hospital, Gorizia, Italy. [45]Nephrology and Dialysis, A.R.N.A.S. Civico and Benfratelli, Palermo, Italy. [46]Nephrology, Dialysis and Renal Transplant Unit, S. Maria della Misericordia Hospital, ASUFC, Udine, Italy. [47]Division of Nephrology, Dialysis and Transplantation, IRCCS Giannina Gaslini Institute, Genova, Italy. [48]Regina Margherita Children's Hospital, Torino, Italy. [49]ICS Maugeri SpA, University of Pavia, Pavia, Italy. [50]Belcolle Hospital, Viterbo, Italy. [51]Division of Nephrology, Dialysis and Renal Transplantation, Riuniti Hospital, Ancona, Italy. [52]Sandro Pertini Hospital, Rome, Italy. [53]Nephrology, Dialysis and Transplantation Unit, Department of Emergency and Organ Transplantation, University of Bari Aldo Moro, Bari, Italy. [54]Renal Unit, Department of Medicine, University of Verona, Verona, Italy. [55]S. Andrea Hospital, La Spezia, Italy. [56]Department of Surgical, Medical, Dental, Oncologic and Regenerative Medicine, University of Modena and Reggio Emilia, Modena, Italy. [57]Nephrology Division, San Marco Hospital, Catania, Italy. [58]Nephrology Dialysis and Kidney Transplant Unit, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Università degli Studi di Milano, Milan, Italy. [59]Azienda Unità Sanitaria Locale Rimini, Rimini, Italy. [60]Department of Nephrology and Dialysis, G. Brotzu Hospital, Cagliari, Italy. [61]Nephrology and Dialysis, Ospedale Maggiore, Milan, Italy. [62]Nephrology and Dialysis Unit, G. Bosco Hub Hospital (ERK-net Member) and University of Torino, Torino, Italy. [63]Division of Nephrology and Renal Transplantation, Carreggi Hospital, Florence, Italy. [64]Renal Division, DMCO (Dipartimento di Medicina, Chirurgia e Odontoiatria), San Paolo Hospital, School of Medicine, University of Milan, Milan, Italy. [65]Unit of Nephrology and Dialysis, AOU G Martino, University of Messina, Messina, Italy. [66]Unit of Nephrology and Dialysis, ASL TO4–Consultorio Cirié, Turin, Italy. [67]San Raffaele Hospital, Milan, Italy. [68]NephroCare Italy, Naples, Italy. [69]Department of Medical and Surgical Specialties and Nephrology Unit, University of Brescia–ASST Spedali Civili, Brescia, Italy. [70]Department of Immunology, Transplantology and Internal Diseases, Medical University of Warsaw, Warsaw, Poland. [71]Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw, Poland. [72]Department of Internal Disease, Nephrology and Dialysotherapy, Military Institute of Medicine, Warsaw, Poland. [73]Department of Clinical Immunology, Medical University of Warsaw, Warsaw, Poland. [74]Department of Pediatrics and Nephrology, Medical University of Warsaw, Warsaw, Poland. [75]Department of Pathology, Medical University of Warsaw, Warsaw, Poland. [76]Department of Transplantation Medicine, Nephrology and Internal Diseases, Medical University of Warsaw, Warsaw, Poland. [77]Department of Nephrology, Transplantology and Internal Medicine, Poznan Medical University, Poznan, Poland. [78]Department of Pediatric Nephrology, Medical University of Lublin, Lublin, Poland. [79]Department of Pediatrics, University of Zielona Góra, Zielona Góra, Poland. [80]Clinical Department of Nephrology and Transplantation Medicine, Wroclaw Medical University, Wroclaw, Poland. [81]Department of Nephrology, Transplantology and Internal Diseases, Medical University of Gdansk, Gdansk, Poland. [82]2nd Department of Nephrology and Hypertension with Dialysis Unit, Medical University of Bialystok, Bialystok, Poland. [83]Department of Genetics, Medical University of Poznan, Poznan, Poland. [84]Department of Pediatric Nephrology and Hypertension, Jagiellonian University Medical College, Krakow, Poland. [85]Department of Pediatrics, Faculty of Medical Sciences in Zabrze, Medical University of Silesia in Katowice, Katowice, Poland. [86]Department of Pediatrics, Immunology and Nephrology, Polish Mother's Memorial Hospital Research Institute, Lodz, Poland. [87]Department of Nephrology, Kidney Transplantation and Hypertension, Children's Memorial Health Institute, Warsaw, Poland. [88]Department of Nephrology, Hypertension and Internal Medicine, University of Warmia and Mazury in Olsztyn, Olsztyn, Poland. [89]Department of Nephrology, Karol Marcinkowski Hospital, Zielona Góra, Poland. [90]Institute of Genetic Epidemiology, Department of Genetics and Pharmacology, Medical University of Innsbruck, Innsbruck, Austria. [91]Division of Inherited and Acquired Kidney Diseases, Veltischev Research and Clinical Institute for Pediatrics of the Pirogov Russian National Research Medical University, Moscow, Russia. [92]Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK. [93]Sorbonne University, INSERM, Center of Research in Myology, Institute of Myology, University Hospital Pitie-Salpetriere, Paris, France. [94]Assistance Publique-Hôpitaux de Paris (AP-HP), Service of Neuro-Myology, University Hospital Pitie-Salpetriere, Paris, France. [95]Departments of Neurology and Genetics, Yale University, New Haven, CT, USA. [96]University of Groningen, Groningen, the Netherlands. [97]Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany. [98]CBP American Hospital, Dubai, United Arab Emirates. [99]Robert S. Boas Center for Genomics and Human Genetics, Feinstein Institutes for Medical Research, North Shore LIJ Health System, New York City, NY, USA. [100]Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. [101]University of Cincinnati College of Medicine, Cincinnati, OH, USA. [102]US Department of Veterans Affairs Medical Center and Cincinnati Education and Research for Veterans Foundation, Cincinnati, OH, USA. [103]Department of Nephrology, Juntendo University Faculty of Medicine, Tokyo, Japan. [104]Division of Clinical Nephrology and Rheumatology, Kidney Research Center, Niigata University Graduate School of Medical and Dental Sciences, Niigata, Japan. [105]Internal Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea. [106]Biomedical Science, Seoul National University College of Medicine, Seoul, Republic of Korea. [107]Department of Family Medicine, Seoul National University College of Medicine and Seoul National University Hospital, Seoul, Republic of Korea. [108]Institute on Aging, Seoul National University College of Medicine, Seoul, Republic of Korea. [109]Amsterdam University Medical Centre, VU University Medical Center (VUMC), Amsterdam, the Netherlands. [110]Hospital Universitario Puerta del Hierro Majadahonda, REDINREN, IISCIII, Madrid, Spain. [111]Molecular Biology Laboratory and Nephrology Department, Fundació Puigvert, Instituto de Investigaciones Biomédicas Sant Pau, Universitat Autònoma de Barcelona, REDINREN, IISCIII, Barcelona, Spain. [112]Department of Nephrology, Danderyd University Hospital, and Department of Clinical Sciences, Karolinska Institutet, Stockholm, Sweden. [113]Department of Nephrology and Hypertension, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland. [114]Division of Nephrology, Saint Louis University, Saint Louis, MO, USA. [115]John Walls Renal Unit, University Hospitals of Leicester, Leicester, UK. [116]Salford Royal Hospital, Salford, UK. [117]Department of Renal Medicine, University College London, London, UK. [118]University of Hamburg, Hamburg, Germany. [119]Division of Nephrology and Clinical Immunology, RWTH Aachen University, Aachen, Germany. [120]Institute of Genetic Epidemiology, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany. [121]Institute of Human Genetics, Friedrich-Alexander-Universität

Erlangen-Nürnberg, Erlangen, Germany. [122]Department of Nephrology and Medical Intensive Care, Charité-Universitätsmedizin Berlin, Berlin, Germany. [123]Department of Nephrology and Hypertension, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany. [124]Department of Nephrology, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China. [125]Department of Genetics, Yale School of Medicine, New Haven, CT, USA. [126]Laboratory of Human Genetics and Genomics, The Rockefeller University, New York City, NY, USA. [127]The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York City, NY, USA. [128]Novo Nordisk Foundation Center for Basic Metabolic Research, Department of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [129]Department of Genetics and Genomic Sciences, Mount Sinai Health System, New York City, NY, USA. [130]Center for Population Genomic Health, Icahn School of Medicine at Mount Sinai, New York City, NY, USA. [131]Department of Biostatistics, Mailman School of Public Health, Columbia University, New York City, NY, USA. [132]Departments of Microbiology and Medicine, University of Alabama at Birmingham, Birmingham, AL, USA. [133]These authors contributed equally: Elena Sanchez-Rodriguez, Xu-Jie Zhou. [134]These authors jointly supervised this work: Krzysztof Kiryluk, Ali G. Gharavi. ✉e-mail: kk473@columbia.edu; ag2239@cumc.columbia.edu

Methods

Association analyses

Individual study cohorts are described in the Supplemental Note, including ethics oversight, recruitment, genotyping, imputation, and quality control analyses. We conducted genome-wide association analysis in each of the 17 cohorts using imputed dosage data under an additive model with adjustment for significant PCs in PLINK v1.9 (ref. 46). Only high-quality ($R^2 > 0.8$) common (MAF > 0.01) SNPs were included in GWAS. Subsequently, we performed meta-analyses using the fixed effects inverse-variance-weighted method (METAL v.2011-03-25)[47] and TransMeta random effects model (TransMeta software)[48]. Genome-wide distributions of *P* values were examined visually using quantile–quantile (QQ) plots for each cohort and for the combined analysis. The meta-analysis QQ plot showed no departure from the expected distribution of *P* values with the genomic inflation factor (λ) estimated at 1.04 (Extended Data Fig. 1). Similarly, our analyses under dominant and recessive models had no evidence of genomic inflation ($\lambda = 1.03$ for dominant and $\lambda = 0.94$ for recessive model; Extended Data Fig. 3). Sex-specific analyses were conducted by cohort separately for males and females and subsequently meta-analyzed. A total of 21,236 males and 17,661 females were used in the meta-analysis with overall $\lambda = 1.01$ for males and $\lambda = 0.99$ for females. A total of 1,990,322 imputed chromosome X markers ($R^2 > 0.8$ and MAF > 0.01) were analyzed separately by sex-encoding genotypes as (0, 2) in males and (0, 1, 2) in females. Significant PCs for each cohort were included as covariates in each model. Genome-wide significant loci were defined by at least one SNP with $P < 5.0 \times 10^{-8}$ that is successfully typed or imputed in ≥50% of cohorts. Signals with $P < 1.0 \times 10^{-5}$ were considered as suggestive.

Conditional analyses

To detect independent associations at individual loci, we conducted stepwise conditional analyses using the conditional and joint (COJO) association analysis method[49] implemented in GCTA v.1.92.0beta[49,50]. Using the summary statistics, we conducted conditional analyses with a threshold of $P \leq 5.0 \times 10^{-8}$ and the LD reference composed of 1000 Genomes Phase 3 (European and East Asian populations). Subsequent conditional analyses were performed for makers with a conditioned $P \leq 5.0 \times 10^{-8}$ until no residual genome-wide significant associations were observed (Supplementary Table 6).

HLA imputation

We used the SNP2HLA v.1.0 software to impute classical HLA alleles[51]. The Type 1 Diabetes Genetics Consortium reference panel of 5,225 Europeans and 8,961 markers was used for our European cohorts[51], and the Pan-Asian reference panel of 530 individuals and 8,245 markers was used for East Asian cohorts[52]. Only common and high-quality variants (MAF > 0.01, $R^2 > 0.8$) were used for association testing. For validation, we used exome sequence data (average depth 60× or above) available for a subset of 500 cases of European ancestry. Classical HLA alleles were called from exome sequence using HLAscan software[53], and the sequence-based results were compared to SNP2HLA. Using this approach, the imputation accuracies for *DRB1*1501*, *DRB1*0405*, *DRB1*0301*, *DRB1*0701* and *DRB1*0101* alleles were estimated at 98.3%, 99.8%, 98.8%, 96.7% and 98.0%, respectively.

HLA classical alleles analysis

We analyzed each imputed variant using logistic regression, assuming additive dosage effects and controlling for significant PCs of ancestry. For testing multi-allelic loci, we used the following logistic regression model:

$$\log(\text{odds}) = \beta_0 + \sum_{j=1}^{m-1} \beta_j X_{j,i} + \sum_{k=1}^{n} \beta_k \text{PC}_{k,i}$$

where *m* indicates a total number of alleles at a multi-allelic locus, *j* indicates a specific allele and $X_{j,i}$ is the imputed dosage for allele *j* in individual *i*; β_0 represents the intercept and β_j represents the additive effect of an allele *j*; $\text{PC}_{k,i}$ denotes the value for *k*th ancestry PC of individual *i*, *n* is the total number of significant PCs in the dataset; β_k is the effect size of principal component *k*. For statistical testing, we compared the log-likelihoods of the following two nested models: the full model containing the test locus and relevant covariates with the reduced model (null model) without the test locus, but with the same set of covariates.

HLA peptide sequence analysis

To test the effects of individual amino acid substitution sites, we applied a conditional haplotype analysis using fully phased haplotypes across the HLA region. We tested each single amino acid position by first identifying the *m* possible amino acid residues occurring at that position and then using $m - 1$ degrees of freedom test to derive *P* values with a single amino acid residue arbitrarily selected as a reference. For conditioning on individual amino acid sites, we used the following procedure: by adding a new amino acid position to the model, a total of κ additional unique haplotypes were generated and tested over the null model using the likelihood ratio test with κ degrees of freedom. If the new position was independently significant, we further updated the null model to include all unique haplotypes created by all amino acid residues at both positions to identify another independent position. The procedure was repeated until no significant (conditioned $P \leq 5.0 \times 10^{-8}$) site was observed. To resolve relationships between HLA alleles and individual amino acid substitutions, we performed joint haplotype phasing of amino acid residues and classical HLA alleles using PLINK v1.07 (ref. 54). HLA protein structure was visualized using UCSF Chimera v1.16 (ref. 55).

Heritability and genetic correlations

SNP-based heritability was estimated using LD score regression (LDSC software)[10] using 1000 Genomes phase 3 European and East Asian populations combined as reference[56]. We also estimated SNP heritability after excluding the HLA region (Chr.6: 28,000,000–33,000,000 bp). To investigate evidence for possible shared genetic effects between IgAN and other traits, we estimated genetic correlations using bivariate LD score regression[10]. For each phenotype, we used summary statistics from the largest available GWAS with a minimum coverage of 2 million SNPs. We excluded traits with estimated SNP-based heritability <1%. Genetic correlations were calculated with and without the HLA region. Summary statistics for immune and cardiometabolic traits were downloaded from the LD Hub or GWAS catalog or provided by the corresponding consortia. Summary statistics for infection-related phenotypes were provided by 23andMe[57].

Pleiotropy maps

GWAS loci were cross-annotated against GWAS catalog (last update: January 31, 2019). For each locus, we selected all variants in strong LD ($r^2 \geq 0.8$) with the top SNP. We then queried the GWAS Catalog for genome-wide significant ($P < 5.0 \times 10^{-8}$) associations of the selected SNPs with other traits. We manually confirmed the direction of allelic effects by reviewing original publications. In cases where there were multiple GWAS for the same trait, we selected studies with the largest sample size. To evaluate the overlap of pleiotropic effects between significant and suggestive IgAN loci, the traits associated with significant loci were queried against GWAS catalog for associations with any of the suggestive SNPs or their proxies ($r^2 \geq 0.8$). The results were visualized using Cytoscape v3.7.0 software.

Polygenic risk models

To assess the cumulative effect of risk loci, we performed a GRS analysis. We first created the following two new GRS models based on the new meta-analysis: the 30-SNP model that comprises 30 independent genome-wide significant SNPs, and the 77-SNP GRS model that includes

the same 30 SNPs plus 47 independent suggestive loci ($P < 1.0 \times 10^{-5}$). Each GRS was standard normalized against the control distribution. We evaluated the performance of each GRS by estimating Nagelkerke's pseudo R^2 and the area under the receiver operating characteristics curve. GPS was calculated using the following two methods: LDpred[58] and LD-pruning and *P* value thresholding ($P + T$)[59,60]. We used the combined meta-analysis including 2,408,512 high-quality SNPs that overlapped across all cohorts but excluding the Immunochip cohorts. For LDPred, the fraction of causal variants was used as a tuning parameter (ρ) across the range of 1, 0.3, 0.1, 0.03, 0.01, 0.003, 0.001, 3.0×10^{-4}, 1.0×10^{-4}, 1.0×10^{-5} and 1.0×10^{-6}; a GPS was calculated for each value of ρ and the best-performing score was selected. For $P + T$ method, we used a range of r^2 (0.2, 0.4, 0.6 and 0.8) and *P* value thresholds (1, 0.3, 0.1, 0.03, 0.01, 0.003, 0.001, 3.0×10^{-4}, 1.0×10^{-4}, 3.0×10^{-5}, 1.0×10^{-5}, 1.0×10^{-6}, 1.0×10^{-7}, 5.0×10^{-8} and 1.0×10^{-8}), and we again selected the best-performing model. The performance of 30-SNP GRS, 77-SNP GRS and best GPS were compared to the previously published 15-SNP GRS[4]. We additionally tested these models in the GCKD study[39,40], including 314 histologically confirmed IgAN cases versus 663 disease controls with biopsy-diagnosed kidney disease of another cause (Supplementary Note). The analyses were implemented in R v.3.5.2 software.

Gene set and pathway-enrichment analyses

We defined each IgAN-associated locus by first selecting all proxy SNPs in LD ($r^2 \geq 0.5$) with the lead SNP, and then extending the genomic region 250 kb upstream and downstream of the first and last proxy. Each region was annotated using the BiomaRt package (Bioconductor release 3.16), which retrieves Ensembl human gene annotations. Gene sets were created for all significant and suggestive loci, excluding the HLA region. For GSEA, we used the Molecular Signatures Database (MSigDB), including GO, Kyoto Encyclopedia of Genes and Genomes (KEGG), BioCarta, REACTOME, chemical and genetic perturbations and transcription factor targets. Statistical significance for enrichment was set at a false discovery rate <0.05. We additionally applied genome-wide gene set enrichment testing (excluding the HLA) using MAGMA (v.1.09) with default parameters[13]. We used DEPICT v1 release 194 (ref. 15) to perform pathway/gene set enrichment and tissue/cell-type analyses. For this analysis, we first used PLINK to identify independently associated SNPs setting $P < 5.0 \times 10^{-5}$ and $r^2 < 0.05$ in a physical window of 500 kb. We then used DEPICT to prioritize genes and identify tissue and cell-type annotations in which genes from the associated regions are expressed. Specifically, for each tissue, the DEPICT method performs a *t*-test comparing the tissue-specific expression of trait-associated genes versus all other genes. Next, for each tissue, empirical enrichment *P* values are computed by repeatedly sampling random sets of loci across the genome to estimate the null distribution for the enrichment statistic as previously described[61,62].

Prioritization of causal tissues and cell types

We estimated SNP-based heritability enrichment for functional categories in tissue/cell-type-specific regulatory elements using stratified LD score regression. This method regresses the chi-squared statistics of SNPs from summary statistics on their LD scores[10] and partitions heritability by functional annotation[63]. For this analysis, we used the meta-analysis statistics without the HLA region and excluding the Immunochip cohorts. Heritability enrichment was defined as the proportion of SNP heritability in a specific category, divided by the proportion of SNPs that belong to that category. We first calculated heritability enrichment for a control model of 96 noncell-type-specific functional categories and compared it to the enrichment in cell-type-specific annotations from the ENCODE and Roadmap Epigenomics[64], as well as mouse immune cell-specific functional categories from the Immunological Genome Project (ImmGen)[65]. We also evaluated tissue/cell-type-specific heritability enrichments based on the FUN-LDA scores[14]. As an alternative, we used the GARFIELD v2 (ref. 16) method to assess enrichment within the ENCODE and Roadmap-derived regulatory regions.

Analysis of relevant phenotypes in mice

We used the Mouse Genome Informatics (MGI) database to identify all genes, the disruption of which causes relevant phenotypes in mice[66]. The mouse phenotypes in MGI are categorized based on the mammalian phenotype ontology and emerge as a result of different genetic models, including targeted knockouts and chemically induced (ENU) and spontaneous mutations. MGI includes a total of 17,101 mouse genes with human orthologs[67]. There were 62 genes with mouse orthologs across the 24 non-HLA risk loci used for testing against MGI phenotypes to define substantially enriched categories.

Functional annotations of individual loci

For the purpose of detailed functional annotation, we calculated 95% credible sets for each of the significant loci using CAVIAR software[68]. We added variants that were neither typed nor imputed in our data, but in strong LD with the top SNP based on external reference ($r^2 \geq 0.8$ in 1000G European and East Asian populations). These SNP sets were annotated using ANNOVAR to first define any coding variants and their predicted effects. Using the FUN-LDA method, we next calculated the posterior probability of a functional effect for each of the selected variants, as described previously[14]. These SNPs were interrogated against the following datasets: (1) eQTLs for 13 human immune cell types from the Database of Immune Cell eQTLs (DICE) project[69]; (2) blood eQTLs from the eQTLGen consortium[27] (31,684 individuals of European ancestry); (3) tissue eQTLs and (4) splicing QTLs from GTEx[70]; (5) glomerular and tubular eQTLs[71]; (6) blood mQTLs from KORA ($n = 2,820$) and TwinsUK ($n = 7,824$) studies[72,73] and (7) blood pQTLs from three recent well-powered multi-ancestry studies[74–76]. We additionally performed colocalization analysis between IgAN and eQTL summary statistics for each GWAS locus using COLOC software[77]. We considered PP4 > 0.5 as supportive of a shared causal variant.

PPIs

PPIs were predicted using InWeb_InBioMap (InWeb_IM)[78]. InWeb_IM is a curated and computationally derived protein–protein network of 420,000 PPIs that has 2.8 times more interactions than other comparable resources. We used only high-confidence PPIs with confidence score ≥0.1. All annotated genes within the 76 significant and suggestive loci were used to probe the PPI database; the final network contained 53 nodes connected by 63 edges. Enrichment *P* value was computed using a hypergeometric test and corrected for multiple testing using the Benjamini–Hochberg method. The GLay community clustering algorithm was implemented for module detection and modules were visualized in GeNets[79]. Subsequently, the clustering with overlapping neighborhood expansion algorithm[80] was used to extract protein clusters using the default parameters with confidence scores as edge weights. Functional and pathway enrichments were identified using STRING[81] based on GO, KEGG and Reactome databases. We used ToppGene Suite[82] to calculate PPI enrichment *P* values. A Bonferroni-corrected $P < 0.05$ was interpreted as significant.

TF–DNA binding interactions

To identify TF binding sites enriched across IgAN risk loci, we used the RELI algorithm[17]. RELI uses a set of genetic variants as input, expands the set using LD blocks ($r^2 > 0.8$) and calculates the statistical intersection of the resulting loci with ChIP–seq datasets by counting the number of loci with one or more variants intersecting the TF ChIP–seq peaks. The LD blocks were calculated using 1000 Genomes Project East Asian and European populations combined. The null distribution was generated using 2,000 random repeats of the procedure and was used to calculate *z*-scores and empirical *P* values for the observed intersections. The final reported *P* values were Bonferroni-corrected (P_c) for the 1,544 TF

datasets tested, as previously published[17]. As input, we used a set of 28 independent significant loci ($P \leq 5.0 \times 10^{-8}$) and a set of 76 loci including significant and suggestive variants ($P < 1.0 \times 10^{-5}$). $P_c < 1.0 \times 10^{-4}$ was used as a significance cut-off for each set.

Ligand–receptor pairs

To identify the number of potential ligand–receptor pairs associated with IgAN risk, we queried the Database of Ligand–Receptor Partners (DLRP)[83]. This database includes 175 protein ligands, 131 protein receptors and 451 experimentally determined ligand–receptor pairings. To test for ligand–receptor enrichment in our dataset, we used a hypergeometric test for overlap between this dataset and the gene set defined by our significant and suggestive GWAS loci.

Analysis of drug targets

We obtained drug-target genes and corresponding drug information from DrugBank[84], the Therapeutic Targets Database[85], the Open Targets Platform[86,87] and GlobalData combined with manual literature searches. To search for potential drug targets, we extracted all genes in direct PPIs with IgAN risk genes by using the In_Web_IM database. We selected drug-target genes that had pharmacological activities and human orthologs, and that were targeted by any of the drugs that are approved or currently in development (experimental or in clinical trials).

Prioritization of biological candidate genes

Each of the positional candidate genes was scored adopting the following criteria and calculating the number of the satisfied criteria, including: (1) genes most proximal to the top SNP; (2) genes with coding variants in 95% credible sets and/or high LD ($r^2 > 0.8$) with the index SNP; (3) genes with promoter chromatin interaction by ABC model[35] or (4) GeneHancer[36] involving regions intersected by top SNP and its 95% credible sets/high LD proxies; (5) eGenes controlled by at least one eQTL (any tissue) tagged by the top SNP in any tissues (primary immune cells, whole blood, kidney, GTEx); (6) eGenes colocalized with the risk locus in blood or (7) primary immune cells with PP4 > 0.5; (8) pGenes controlled by at least one blood pQTL tagged by the top SNP; (9) genes prioritized by PPI network connectivity analysis at $P < 0.05$; (10) genes that when knocked out in mice produce at least two phenotype labels—'immune system', 'haematopoietic system' or 'cellular phenotype'; (11) genes prioritized by MAGMA, (12) DEPICT with gene-based $P < 0.05$ or (13) manual review of the literature as related to IgAN, IgA production or mucosal immunity. For each locus, the gene ranking using our scoring method was then compared to the recently proposed V2G score ranking[37].

Genotype–phenotype correlations

GRSs (15-SNP, 30-SNP, 77-SNP and GPS) were tested for clinical correlations in the subset of cases with available clinical data. We tested each risk score predictor for association with clinical disease features (outcomes) at the time of diagnosis, including age at biopsy, estimated glomerular filtration rate (eGFR), proteinuria, microhematuria, hypertension and gross hematuria. The GFR was estimated using the CKD–EPI formula in adults[88] and the Schwartz formula in children[89]. The GFR was normalized using logarithmic transformation, proteinuria was normalized with a ln(P24 + 1) transformation, microhematuria was defined as positive if 1+ or greater on a urine dipstick test; gross hematuria was defined by self-report; hypertension was defined as systolic pressure ≥140 mmHg or diastolic pressure ≥90 mmHg, or anti-hypertensive medication use. The outcome of kidney failure was defined as eGFR < 15 ml min^{-1} 1.73 m^{-2} or initiation of dialysis or kidney transplantation. All analyses were adjusted for age, gender, site and ancestry. The analyses were implemented in R v3.5.2.

PheWAS

We performed meta-PheWAS across the eMERGE-III and the UKBB datasets. The eMERGE-III dataset contains EHR information linked to GWAS data for 102,138 individuals[90,91]. The UKBB is comprised of 488,377 individuals with EHR information linked to GWAS data[92]. For meta-analysis, we harmonized phenotype data by converting all available ICD-10 codes to the ICD-9-CM system; the eMERGE participants had 20,783 unique ICD-9 codes and the UKBB participants had 10,221 unique ICD-9 codes. These codes were next mapped to 1,817 unique phecodes and tested using logistic regression adjusted for age, sex, study site, imputation batch and threee PCs of ancestry using PheWAS R package[93]. Meta-PheWAS was performed across both datasets using metagen under fixed effects[93]. For phenome-wide significance, we used the Bonferroni-corrected $P < 2.75 \times 10^{-5}$ to account for 1,817 independent phecodes tested.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Primary genotype data for previously published cohorts are available through dbGAP under accession number phs000431.v2.p1, and the new genotype data are available under dbGAP accession number phs000431.v3.p1. Our IRB determined that the use of the primary genotype data is restricted to genetic studies of kidney disease. GWAS summary statistics are available from the Kiryluk Lab website: https://www.columbiamedicine.org/divisions/kiryluk/study_gwas_stat_IgA_summary.php. The PAGE consortium control genotype data are available on dbGAP under accession number phs000356.v2.p1. The eMERGE-III imputed genotype and phenotype data are available through dbGAP, accession number phs001584.v2.p2. The UKBB genotype and phenotype data are available through the UKBB web portal https://www.ukbiobank.ac.uk/. The DICE dataset is available at https://dice-database.org/; the Blood eQTL meta-analysis by eQTLGen is available at https://www.eqtlgen.org/; the Kidney eQTL Atlas is available at http://susztaklab.com/eqtl; NEPTUNE eQTL Browser is available at http://nephqtl.org/; GTEx is available at https://gtexportal.org/home/; GWAS catalog is available at https://www.ebi.ac.uk/gwas; LD Hub is available at http://www.nealelab.is/tools-and-software; Open Targets platform is available at https://www.targetvalidation.org; DLRP can be accessed at https://dip.doe-mbi.ucla.edu/dip/DLRP.cgi; DrugBank can be accessed at https://www.drugbank.ca; GlobalData database is available at https://www.globaldata.com/industries-we-cover/pharmaceutical; FUN-LDA scores are available at http://www.columbia.edu/~ii2135/funlda.html; MSigDB and GSEA are available at http://software.broadinstitute.org/gsea/msigdb/; STRING is available at https://string-db.org; InWeb is available at http://www.lagelab.org/resources/; the Metabolomics GWAS Server is available at http://metabolomics.helmholtz-muenchen.de/gwas/; and the MGI database is available at http://www.informatics.jax.org.

Code availability

Only publicly available open-source software was used in the analyses; there was no custom software.

References

46. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
47. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
48. Shi, J. & Lee, S. A novel random effect model for GWAS meta-analysis and its application to trans-ethnic meta-analysis. *Biometrics* **72**, 945–954 (2016).
49. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).

50. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
51. Jia, X. et al. Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS ONE* **8**, e64683 (2013).
52. Pillai, N. E. et al. Predicting HLA alleles from high-resolution SNP data in three Southeast Asian populations. *Hum. Mol. Genet.* **23**, 4443–4451 (2014).
53. Ka, S. et al. HLAscan: genotyping of the HLA region using next-generation sequencing data. *BMC Bioinformatics* **18**, 258 (2017).
54. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
55. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
56. 1000 Genomes Project Consortium et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
57. Tian, C. et al. Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. *Nat. Commun.* **8**, 599 (2017).
58. Vilhjalmsson, B. J. et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
59. Khera, A. V. et al. Polygenic prediction of weight and obesity trajectories from birth to adulthood. *Cell* **177**, 587–596 (2019).
60. Khera, A. V. et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50**, 1219–1224 (2018).
61. Gorski, M. et al. 1000 Genomes-based meta-analysis identifies 10 novel loci for kidney function. *Sci. Rep.* **7**, 45040 (2017).
62. Pattaro, C. et al. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nat. Commun.* **7**, 10023 (2016).
63. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
64. Roadmap Epigenomics Consortium et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
65. Heng, T. S. & Painter, M. W. & Immunological Genome Project Consortium. The Immunological Genome Project: networks of gene expression in immune cells. *Nat. Immunol.* **9**, 1091–1094 (2008).
66. Eppig, J. T. et al. The Mouse Genome Database (MGD): facilitating mouse as a model for human biology and disease. *Nucleic Acids Res.* **43**, D726–D736 (2015).
67. Bult, C. J. et al. Mouse genome database 2016. *Nucleic Acids Res.* **44**, D840–D847 (2016).
68. Hormozdiari, F., Kostem, E., Kang, E. Y., Pasaniuc, B. & Eskin, E. Identifying causal variants at loci with multiple signals of association. *Genetics* **198**, 497–508 (2014).
69. Schmiedel, B. J. et al. Impact of genetic polymorphisms on human immune. *Cell* **175**, 1701–1715 (2018).
70. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
71. Qiu, C. et al. Renal compartment-specific genetic variation analyses identify new pathways in chronic kidney disease. *Nat. Med.* **24**, 1721–1731 (2018).
72. Suhre, K. et al. Human metabolic individuality in biomedical and pharmaceutical research. *Nature* **477**, 54–60 (2011).
73. Shin, S. Y. et al. An atlas of genetic influences on human blood metabolites. *Nat. Genet.* **46**, 543–550 (2014).
74. Sun, B. B. et al. Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).
75. Suhre, K. et al. Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat. Commun.* **8**, 14357 (2017).
76. Emilsson, V. et al. Co-regulatory networks of human serum proteins link genetics to disease. *Science* **361**, 769–773 (2018).
77. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
78. Li, T. et al. A scored human protein-protein interaction network to catalyze genomic interpretation. *Nat. Methods* **14**, 61–64 (2017).
79. Su, G., Kuchinsky, A., Morris, J. H., States, D. J. & Meng, F. GLay: community structure analysis of biological networks. *Bioinformatics* **26**, 3135–3137 (2010).
80. Nepusz, T., Yu, H. & Paccanaro, A. Detecting overlapping protein complexes in protein-protein interaction networks. *Nat. Methods* **9**, 471–472 (2012).
81. Szklarczyk, D. et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
82. Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. ToppGene suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* **37**, W305–W311 (2009).
83. Graeber, T. G. & Eisenberg, D. Bioinformatic identification of potential autocrine signaling loops in cancers from gene expression profiles. *Nat. Genet.* **29**, 295–300 (2001).
84. Knox, C. et al. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.* **39**, D1035–D1041 (2011).
85. Zhu, F. et al. Therapeutic target database update 2012: a resource for facilitating target-oriented drug discovery. *Nucleic Acids Res.* **40**, D1128–D1136 (2012).
86. Carvalho-Silva, D. et al. Open targets platform: new developments and updates two years on. *Nucleic Acids Res.* **47**, D1056–D1065 (2019).
87. Ochoa, D. et al. Open targets platform: supporting systematic drug-target identification and prioritisation. *Nucleic Acids Res.* **49**, D1302–D1310 (2021).
88. Levey, A. S. et al. A new equation to estimate glomerular filtration rate. *Ann. Intern. Med.* **150**, 604–612 (2009).
89. Schwartz, G. J. et al. New equations to estimate GFR in children with CKD. *J. Am. Soc. Nephrol.* **20**, 629–637 (2009).
90. Stanaway, I. B. et al. The eMERGE genotype set of 83,717 subjects imputed to ~40 million variants genome wide and association with the herpes zoster medical record phenotype. *Genet. Epidemiol.* **43**, 63–81 (2019).
91. Khan, A. et al. Medical records-based genetic studies of the complement system. *J. Am. Soc. Nephrol.* **32**, 2031–2047 (2021).
92. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
93. Denny, J. C. et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* **26**, 1205–1210 (2010).

Acknowledgements

We are grateful to all study participants across multiple nephrology centers worldwide for their contributions to this manuscript. This work was supported by the following institutions, grants and funding agencies: Columbia University, Columbia Glomerular Center, IGA Nephropathy Foundation of America and National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) grants R01-DK105124 (to K.K., J. Novak, B.A.J.), RC2-DK116690 (to K.K.) and R01-DK082753 (to A.G.G., J. Novak, K.K., F.S., B.A.J., R.J.W.). Additional support was provided by R01LM013061 (to K.K.), U01HG008680 (to K.K.), U01AI152960 (to K.K.), R01-DK078244 (to J. Novak, B.A.J.), R01-AI149431 (to J. Novak, B.A.J.), National Science Foundation of

China (82022010 to X.-J.Z.; 82070733 to H.Z.), Beijing Natural Science Foundation (Z190023 to X.-J.Z.), Deutsche Forschungsgemeinschaft (DFG) Project-ID 192904750—CRC 992 Medical Epigenetics (to A. Köttgen, P. Schlosser), DFG Project-ID 431984000—CRC 1453 (to A. Köttgen) and EQUIP Program for Medical Scientists, Faculty of Medicine, University of Freiburg (to P. Schlosser). J.B.H., S.P., L.K. and M.W. were supported by R01HG010730, U01AI130830, R01NS099068, R01AI024717, R01AR073228, U01AI150748, R01AI148276, P01AI150585 and the US Department of Veterans Affairs Merit Award I01-BX001834 to J.B.H. D.P.G. was supported by the St Peter's Trust for Kidney, Bladder and Prostate Research. The UK cohort data were generated as a result of a grant from Kidney Research UK and the Medical Research Council (to J. Barratt and D.P.G.). The German STOP-IgAN study was supported by the DFG (German Research Foundation)—CRU 5011—Project-ID 445703531 (to J.F., T.R.). The GCKD study was funded by grants from the German Ministry of Education and Research (BMBF, 01ER0804) and the KfH Foundation for Preventive Medicine. Unregistered grants to support the study were provided by Bayer, Fresenius Medical Care and Amgen. Genotyping was supported by Bayer Pharma AG. The recruitment of Polish cases with IgAN was sponsored by the Polish Kidney Genetics Network (POLYGENES), a collaborative effort between Columbia University and Poznań University of Medical Sciences, Poland. The recruitment of Czech patients with IgAN was supported by the research project of General University Hospital in Prague (RVO-VFN64165). The recruitment of the Russian cohort was supported by the Government Assignment of the Russian Ministry of Health, Assignment 200080056 of the Veltischev Research and Clinical Institute for Pediatrics of the Pirogov Russian National Research Medical University. The recruitment of the Swedish cohort was supported by the Nephrology and Rheumatology Departments at Karolinska University Hospital and Karolinska Institutet, Stockholm, and grants from the Swedish Society of Medicine. We thank the Immunopathology Working Group of the Italian Society of Nephrology (ISN) for inviting their member sites to contribute to this study. We are also grateful to the Pediatric Nephrology Research Consortium (PNRC) for hosting and cosponsoring the GIGA-kids study that recruited pediatric patients with IgAN across the PNRC sites. We are additionally grateful to C. Bowers for coordinating GIGA-kids and PMRC-based recruitment, J. Narus for coordinating recruitment at the University of Utah, E. Elenberg and S. Shah for helping enroll participants from Texas Children's Hospital and A. Suwanichkul for sample shipping. The recruitment of the Korean cohort was supported by the Seoul National University Hospital Human Biobank, a member of the National Biobank of Korea, financed by the Ministry of Health and Welfare, Republic of Korea. We are grateful to the International Multiple Sclerosis Genetics Consortium (IMSGC), the Inflammatory Bowel Disease Genetics Consortium (IBDGC), the International Myositis Consortium (IMC), the Feinstein Institute for Medical Research, the French Biological Ressource Center for MS Genetics (REFGENSEP), Genethon and INSERM for contributing Immunochip controls for the purpose of this study. We also thank the Population Architecture Using Genomics and Epidemiology (PAGE) consortium, funded by the NHGRI with cofunding from the NIMHD, for providing population controls genotyped with MEGA chip for this study. Molecular graphics shown in Extended Data Fig. 5c were generated using UCSF Chimera, developed by the Resource for Biocomputing, Visualization and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311. The funding sources were not involved in the study design, collection, analysis and interpretation of data, writing of the report or the decision to submit the paper for publication.

Author contributions

K.K. and A.G.G. conceived the study, provided overall supervision of the project and made the decision to publish the findings. E.S.-R. and N. Mladkova performed quality control, imputation and association analyses for GWAS and Immunochip discovery cohorts. E.S.-R. performed final statistical analyses, meta-analyses, fine-mapping studies, functional annotations and drug-target analyses. F.Z. and A. Khan designed polygenic risk scores, tested for genome-wide genetic correlations and performed clinical correlation analyses. L.L. performed analysis of the HLA locus. J.B.H., M.W., S.P. and L.K. performed RELI transcription factor analysis. I.I.-L. consulted on the statistics and functional annotation of GWAS loci. K.K., A.G.G. and J. Novak provided biological interpretation of GWAS loci. X.-J.Z. and H.Z. coordinated recruitment, genotyping and analysis of the Beijing cohorts. F.S. coordinated recruitment activities across the Italian network of clinical recruitment sites. Y.C. coordinated recruitment activities across the Turkish recruitment sites. H.T. coordinated recruitment across the Argentinian network of clinical recruitment sites. V.T. and D.M. coordinated the recruitment of the Czech cohort. J.X. and N.C. coordinated recruitment, genotyping and analysis of the Shanghai cohorts. H.S. recruited and clinically characterized the Japanese discovery cohort. H.L., J.-H.P., B.C., Y.S.K. and D.K.K. recruited and clinically characterized the Korean discovery cohort. J.F. and T.R. contributed DNA and clinical data for the German STOP-IgAN cohort. K.K., R.N., R.J.W. and W.S. led the GIGA-kids study in collaboration with the Pediatric Nephrology Research Consortium (PNRC). M. Marasa, O.B. and J.Y.Z. coordinated recruitment at Columbia University and managed clinical data and DNA samples. J. Barratt and D.P.G. provided summary statistics for the UK GWAS cohort. R.P.L. contributed previously published European and Beijing GWAS cohorts. A.B.E., K.-U.E. and A. Köttgen contributed German cases to the Immunochip discovery cohorts, and P. Schlosser performed polygenic risk score validation studies in the GCKD cohort. D.v.H., C.C., C.W., A.F., V.A. and P.K.G. contributed Illumina Immunochip idat files for population controls. R.J.F.L. and E.E.K. contributed Illumina MEGA chip idat files for population controls. S.S.-C., A.S.B., P.A.C., G.B.A., J.R., B. Sprangers, D.C.C., H.R., Y.P., P.R., K.G., B. Stengel, M. Metzger, G. Canaud, N. Maillard, F.B., L. Berthelot, E.P., R. Monteiro, J.M., A.-A.S., G.H., A.Q., P.W., R.S., D. Selewski, K. Davis, M. Kallash, T.L.V., M.R., A.C., D. Ranch, S.E.W., D. Samsonov, D.J.C., O.A., D.G., M. Stangou, J. Nagy, T.K., E.F., A.A., C.B., D.C., L.D.V., G.G.B., M.B., E.B., L. Bono, G.B., G. Caridi, F. Lugani, G. Ghiggeri, R.C., L. Peruzzi, V.E., C.E., S.F., R.P., G.F., M. Galliani, M. Garozzo, A. Mitrotti, L.G., S. Granata, G.Z., F. Londrino, R. Magistroni, I.P., A. Magnano, C.M., P.M., R. Mignani, A.P., C.P., D. Roccatello, M. Salvadori, E.S., D. Santoro, G. Gembillo, S.S., D. Spotti, P.Z., C.I., F.A., E.D., M.F., N. Krata, K.M., L. Pączek, S.N., B.M., M.P.-T., M.M.-W., A.P.-P., T.B., M.D., K.P., P.S., M.Z., D.K., M. Krajewska, I.K.-G., Z.H., B.B.-P., T.L., A.D.-S., T.H., A.M.-K., M. Miklaszewska, M. Szczepańska, K. Dyga, E.M., K.S.-L., M.P.-B., M.T., D. Runowski, N. Kwella, D.D., I.H., F.K., L. Prikhodina, B.F., I.N., S. Goto, M.C., A.V.W., A.H., E.A., J. Ballarin, S.L., B.V., L.-Y.M., T.A., P.A.K., U.P. and B.A.J. contributed to the recruitment and clinical characterization of patients with IgAN and controls recruited from their respective clinical centers. All authors have read and approved the final version of the manuscript.

Competing interests

B.A.J. and J. Novak are cofounders, co-owners of, and consultants for Reliant Glycosciences, LLC, and are co-inventors on US patent application 14/318,082 (assigned to UAB Research Foundation). A.G.G. has served on an advisory board for Novartis, Travere and Natera and receives research grant funding from the Renal Research Institute and Natera. K.K. has served on an advisory board for Goldfinch Bio and Gilead. The other authors report no competing interests.

Additional information

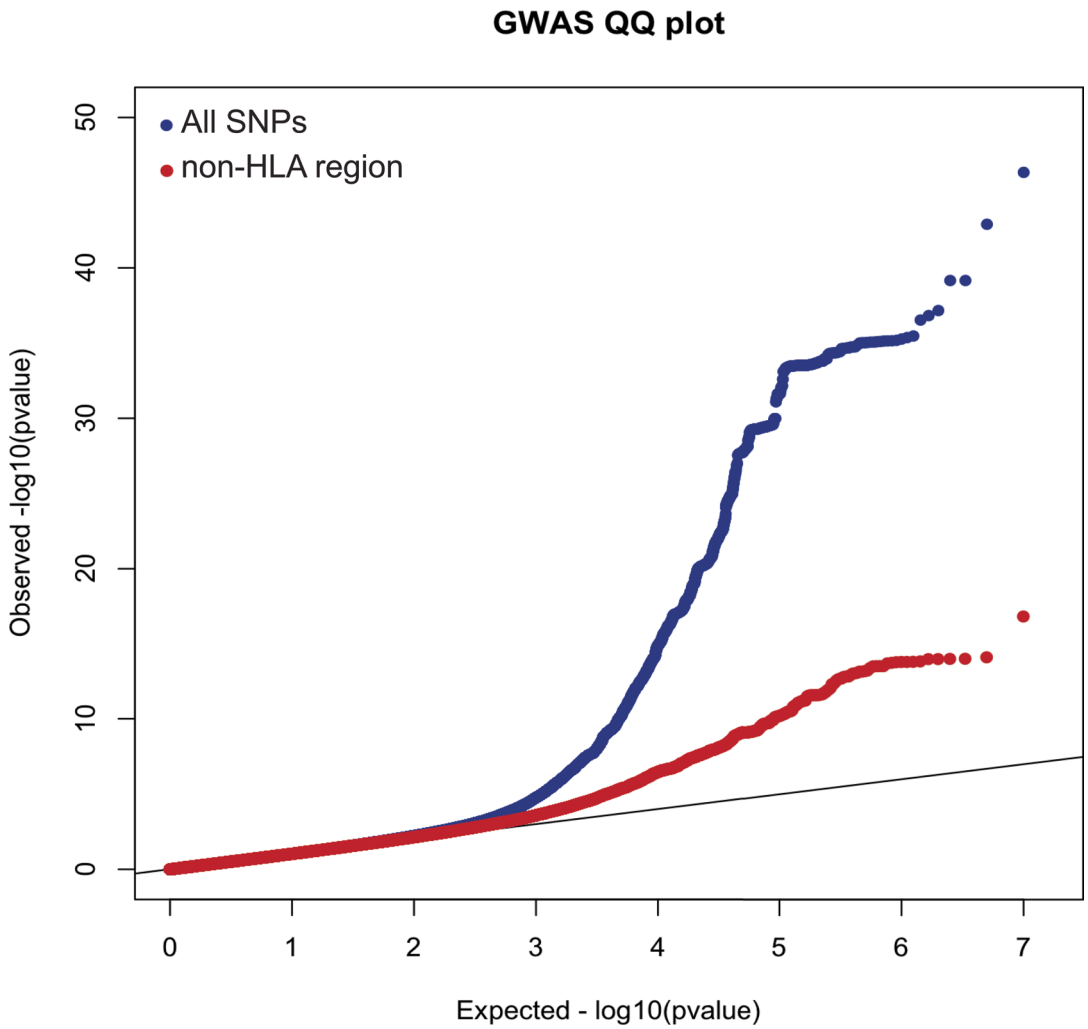
Extended data is available for this paper at https://doi.org/10.1038/s41588-023-01422-x.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-023-01422-x.

Correspondence and requests for materials should be addressed to Krzysztof Kiryluk or Ali G. Gharavi.

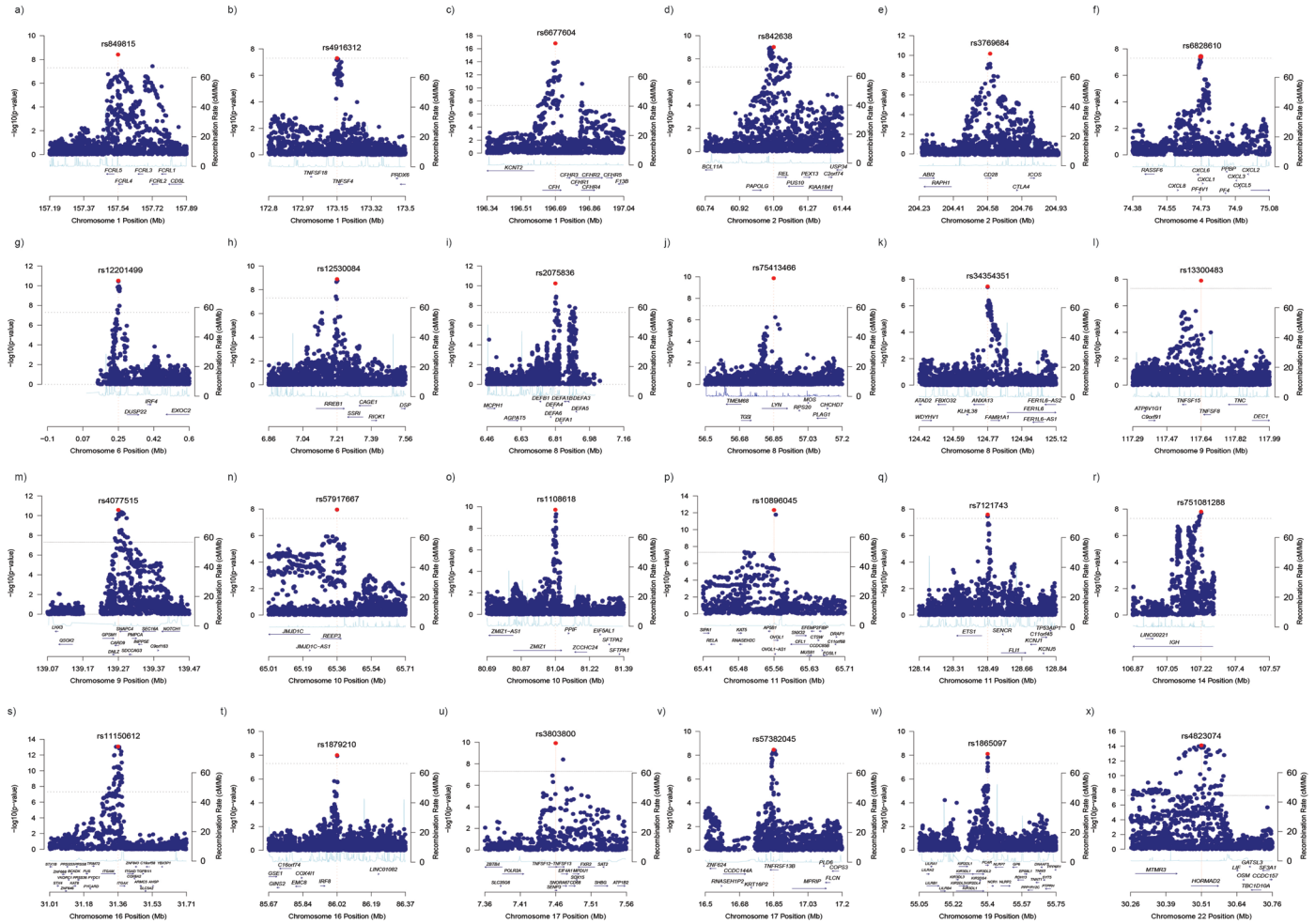
Peer review information *Nature Genetics* thanks Xueqing Yu, Richard Oram and Amy McKnight for their contribution to the peer review of this work.

https://doi.org/10.1038/s41588-023-01422-x



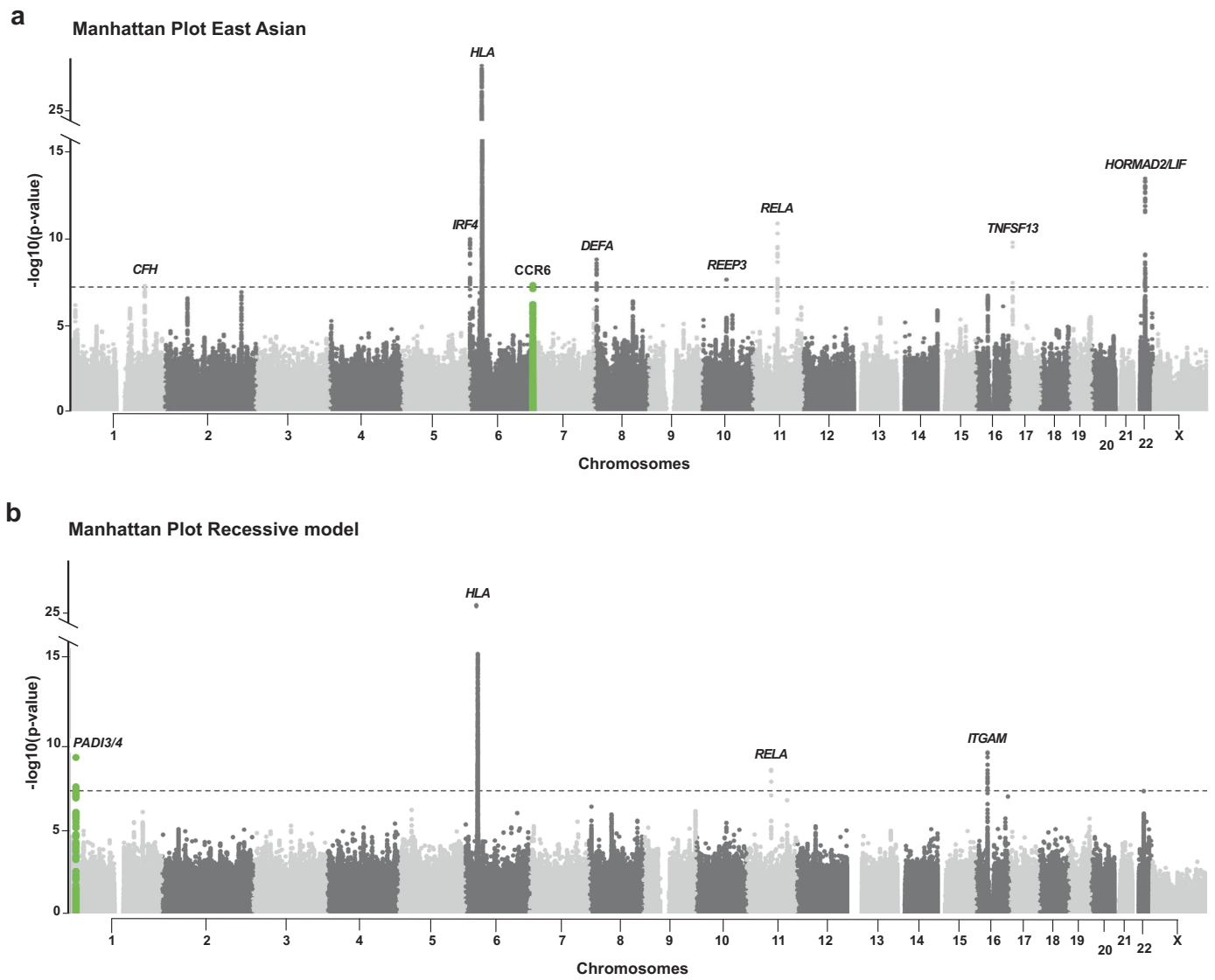
Extended Data Fig. 1 | Quantile-quantile (QQ) plot of the combined meta-analysis across 38,897 individuals. Blue dots represent the QQ plot based on all SNPs in the meta-analysis, and red dots represent the QQ plot after exclusion of SNPs within the MHC region. The overall genomic inflation factor (λ) was 1.048 with the MHC region, and 1.042 without the MHC region. The *y*-axis depicts $-\log_{10}$ of the observed two-sided *P*-values for a fixed effects meta-analysis of all cohorts without correction for multiple testing. The *x*-axis depicts $-\log_{10}$ of the expected two-sided *P*-values under the null hypothesis.

https://doi.org/10.1038/s41588-023-01422-x



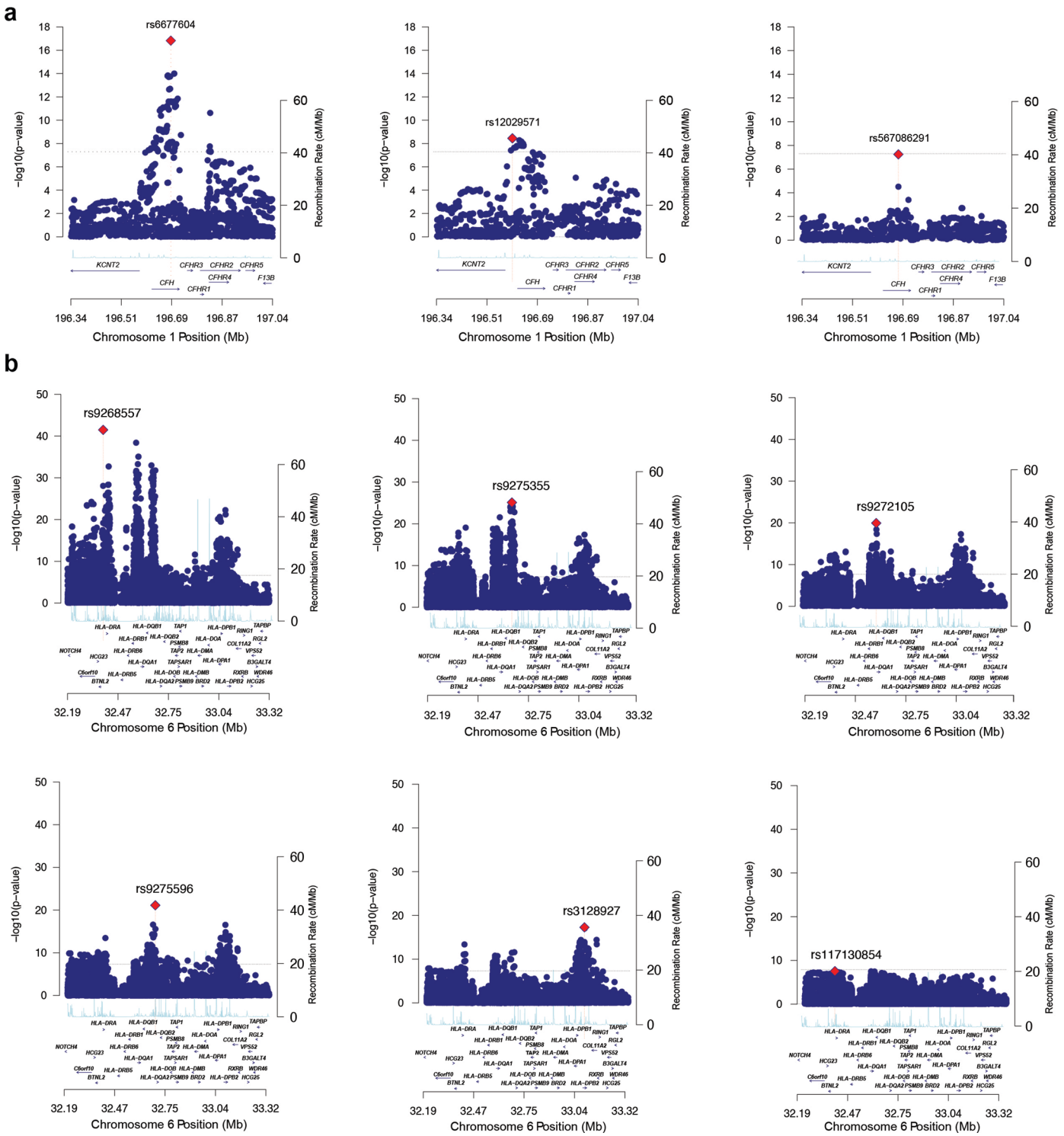
Extended Data Fig. 2 | Regional plots for non-HLA genome-wide significant loci. a, The *FCRL* locus. **b**, The *TNFSF4* locus. **c**, The *CFH* locus. **d**, The *REL* locus. **e**, The *CD28* locus. **f**, The *PF4V1/CXCL8* locus. **g**, The *IRF4* locus. **h**, The *RREB1* locus. **i**, The *DEFA* locus. **j**, The *LYN* locus. **k**, The *ANXA3* locus. **l**, The *TNFSF8* locus. **m**, The *CARD9* locus. **n**, The *REEP3* locus. **o**, The *ZMIZ1* locus. **p**, The *RELA* locus. **q**, The *ETS1* locus. **r**, The *IGH* locus. **s**, The *ITGAM* locus. **t**, The *IRF8* locus. **u**, The *TNFSF13* locus. **v**, The *TNFRSF13B* locus. **w**, The *FCAR* locus. **x**, The *HORMAD2/LIF* locus. The *x*-axis shows the physical position in Mb (hg19 coordinates) and known genes. The left *y*-axis presents $-\log_{10}$ of two-sided *P*-values for variant association statistics (fixed effects meta-analysis under an additive model without correction for multiple testing), and the right *y*-axis shows the recombination rate across the region. The dotted horizontal line indicates a genome-wide significance threshold of 5.0×10^{-8}.

https://doi.org/10.1038/s41588-023-01422-x



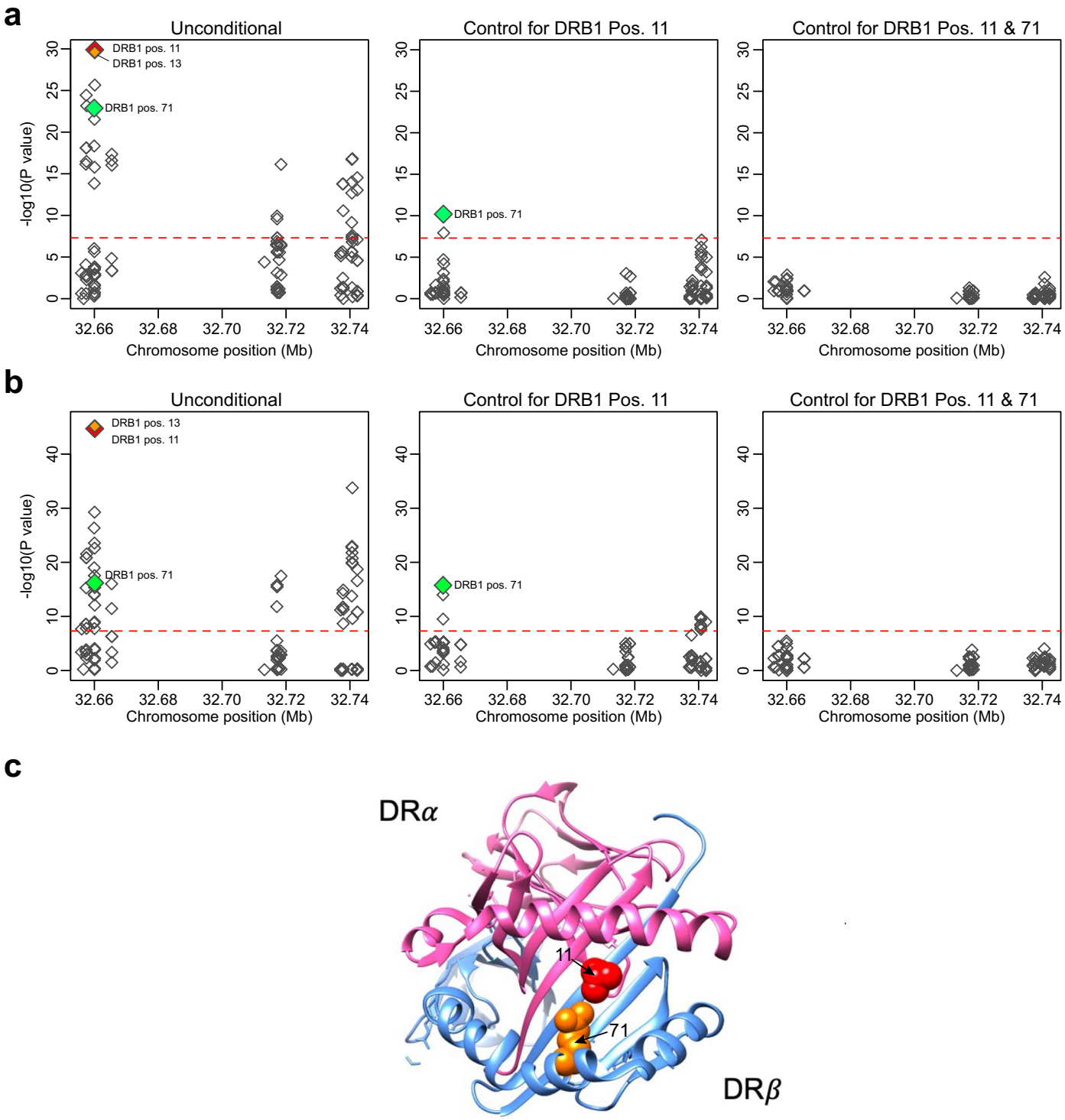
Extended Data Fig. 3 | Manhattan plots for East Asian GWAS subgroup analyses. a, East Asian meta-analysis under an additive genetic model (λ = 1.040) revealed a new genome-wide significant locus on chr. 6 (*CCR6*, green). **b**, East Asian meta-analysis under a recessive model (λ = 0.940) revealed a new genome-wide significant locus on chr. 1 (encoding *PADI3* and *PADI4*, green). The *y*-axis shows -$\log_{10}$ of two-sided *P*-values for a fixed effects meta-analysis under an additive or recessive genetic model without correction for multiple testing (note that *y*-axis is truncated to accommodate the HLA peak). The *x*-axis shows genomic position along each chromosome (1-22 and X). The dotted horizontal line indicates $P = 5.0 \times 10^{-8}$.

https://doi.org/10.1038/s41588-023-01422-x



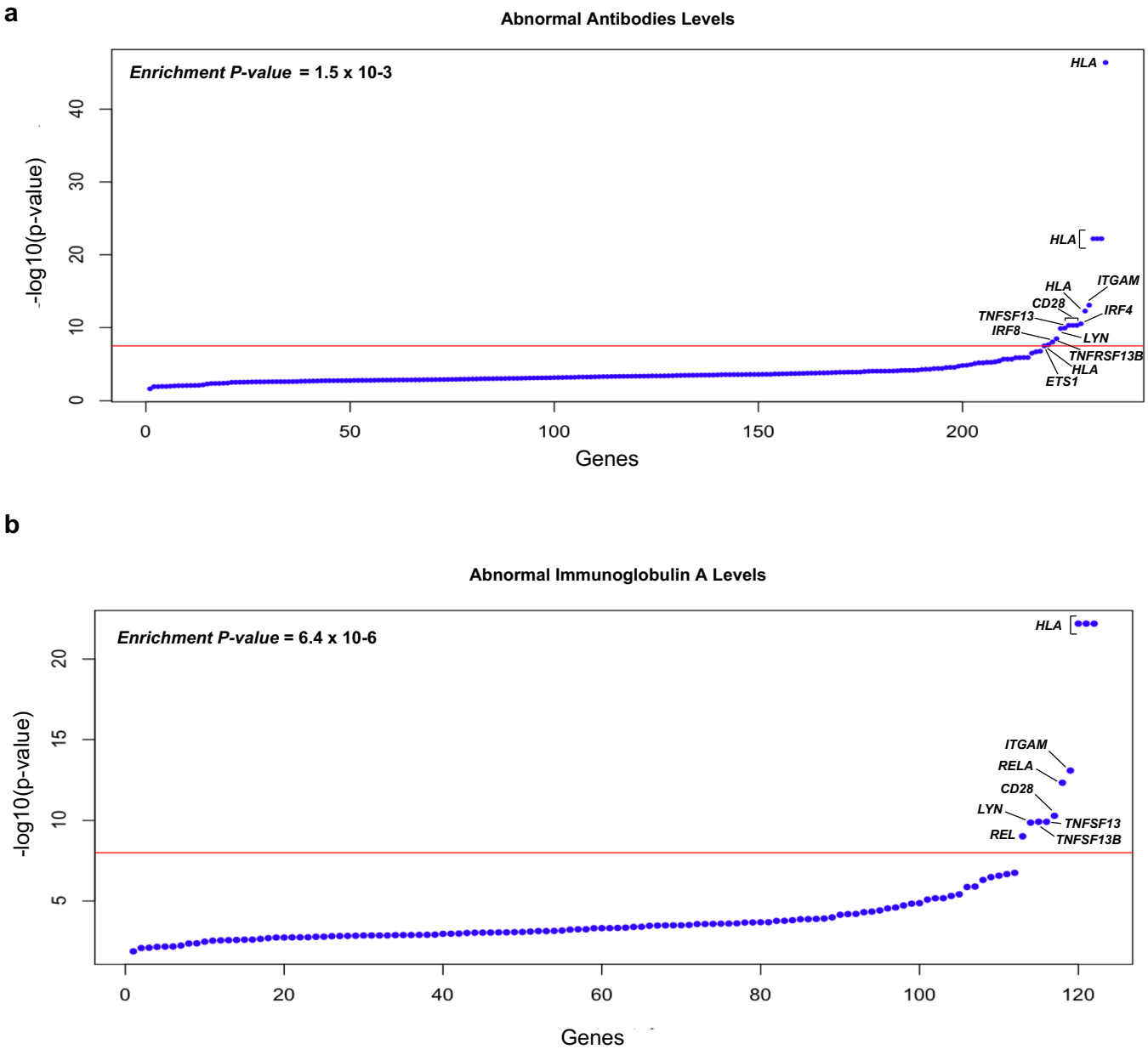
Extended Data Fig. 4 | Stepwise conditional analyses of the *CFH* and *HLA* loci. a, *CFH* locus. Initial meta-analysis results without conditioning (top left); after conditioning for the top SNP rs6677604 (top middle); and after controlling for the two significant SNPs rs6677604 and rs12029571 (top right). **b**, *HLA* locus. Initial meta-analysis results without conditioning (middle left); after conditioning for the top SNP rs9268557 (middle), after controlling for rs9268557 and rs9275355 (middle right), after controlling for rs9268557, rs9275355 and rs9272105 (bottom left), after controlling for rs9268557, rs9275355, rs9272105 and rs9275596 (bottom middle), and after controlling for rs9268557, rs9275355 rs9272105, rs9275596 and rs3128927 (bottom right) with no additional significant signals. The *x*-axis shows genomic position in Mb (hg19 coordinates) and known genes. The left *y*-axis presents $-\log_{10}$ *P*-values for association statistics (two-sided *P*-values for a fixed effects meta-analysis under an additive genetic model without correction for multiple testing). The right *y*-axis (light-blue line) shows the average recombination rate across the region. The dotted horizontal line indicates a genome-wide significance threshold of 5.0×10^{-8}. The top SNP in each panel is marked by a red diamond.

https://doi.org/10.1038/s41588-023-01422-x



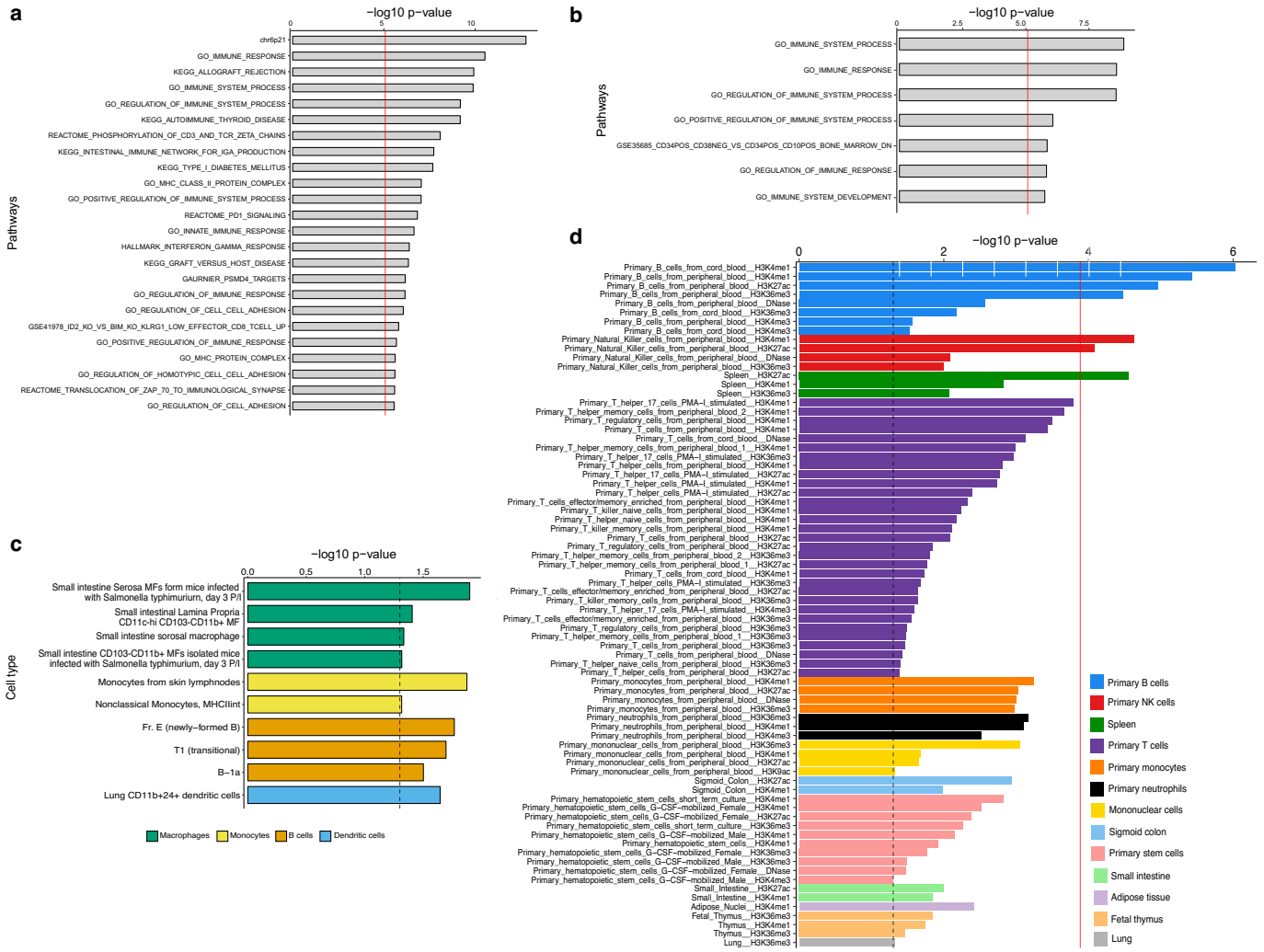
Extended Data Fig. 5 | Stepwise conditional analysis of imputed polymorphic amino-acid positions in DRβ, DQβ, and DQα peptides in East Asian and European cohorts. a, East Asian cohorts. **b**, European cohorts. Each symbol represents a polymorphic site tested for association with IgAN along the peptide sequence. The *x*-axis shows genomic position of the sequence encoding each amino acid. The *y*-axis shows global statistical significance for each polymorphic site (two-sided multiallelic Wald test without adjustment for multiple testing). The dashed horizontal line corresponds to a genome-wide significance threshold of 5.0×10^{-8}. **c**, Physical location of independently associated amino acid positions 11 (red) and 71 (orange) in the structural model of DRβ. The DR structure was visualized using UCSF Chimera v1.16 based on Protein Data Bank entry 3PDO.

https://doi.org/10.1038/s41588-023-01422-x



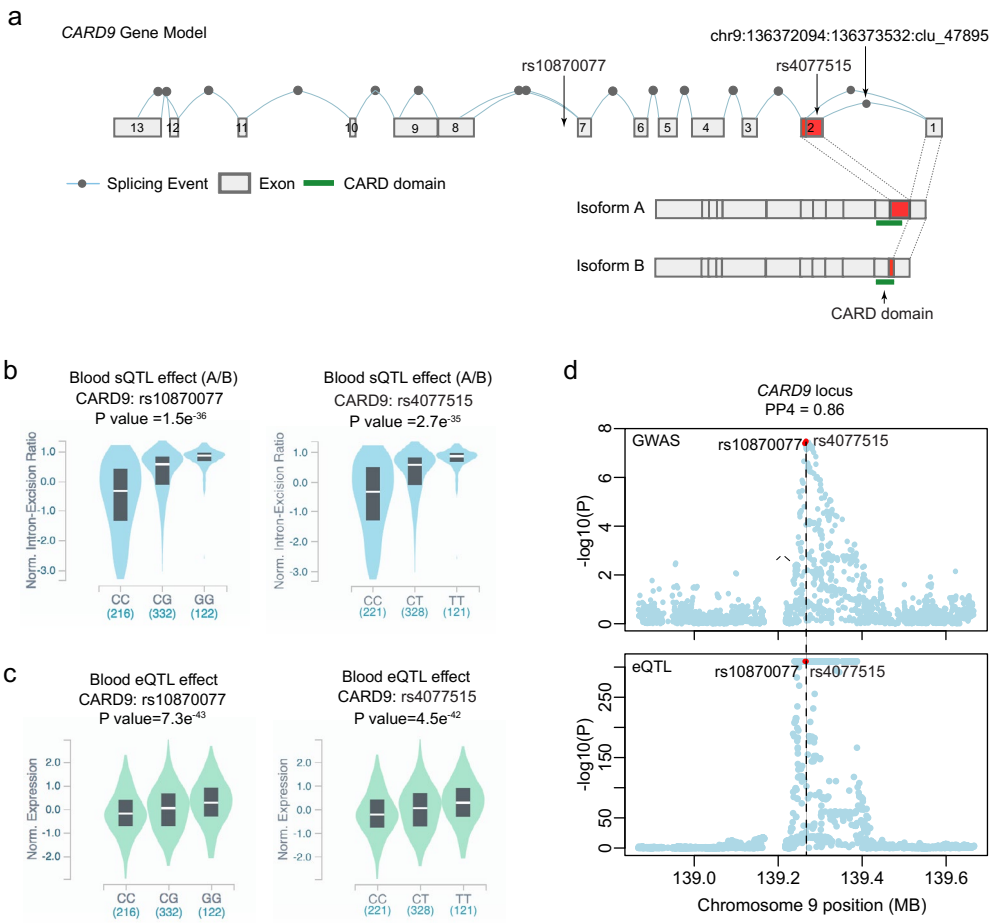
Extended Data Fig. 6 | Enrichment tests for the GWAS candidate gene set against human ortholog gene sets that when genetically manipulated in mice result in "Abnormal Antibody Levels" or "Abnormal Immunoglobulin A Levels". a, Plot for "Abnormal Antibody Levels". **b**, Plot for "Abnormal Immunoglobulin A Levels". The *x*-axis depicts gene rank by statistical significance. The *y*-axis depicts $-\log_{10}$ of two-sided *P*-value based on the enrichment test without multiple testing correction. The red horizontal line is the Bonferroni-corrected statistical significance level corrected for the number of genes tested.

https://doi.org/10.1038/s41588-023-01422-x



Extended Data Fig. 7 | Pathway, cell type and tissue enrichment analyses. **a**,**b**, MAGMA pathway enrichment analysis based on GWAS summary statistics with (**a**) and without (**b**) the HLA region. $-\log_{10}$ of two-sided *P*-values for the enrichment test (not corrected for multiple testing) are depicted along the *x*-axis. Significant pathways are listed along the *y*-axis. Red vertical lines indicate significance threshold after accounting for multiple testing. **c**, Cell type-specific heritability enrichment for functional annotations in ImmGen dataset of mouse regulatory elements and expression data demonstrating the strongest enrichment in small intestine-derived macrophages profiled three days after Salmonella infection. **d**, Cell type-specific heritability enrichment for individual functional annotations generated by Roadmap Epigenomics demonstrates significant enrichment in immune cells, especially of B-cell lineage. All enrichments at nominal two-sided uncorrected $P < 0.05$ are displayed and grouped according to cell and tissue class. A solid red line represents a stringent Bonferroni-corrected $-\log_{10}$ of the two-sided *P*-value threshold of significance for the Roadmap dataset ($P = 1.3 \times 10^{-4}$). A dotted black line represents the $-\log_{10}$ of the nominal *P*-value of 0.05. GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; MF, macrophages; P/I, post-injection.

https://doi.org/10.1038/s41588-023-01422-x



Extended Data Fig. 8 | Cis-regulatory effects at the *CARD9* locus. a, The *CARD9* gene model depicting 13 exons of *CARD9* and the splicing events from GTEx blood RNA-seq. The top IgAN risk allele (rs4077515-T) encodes S12N substitution in the second exon of *CARD9*. This SNP also exhibits a strong and statistically significant blood sQTL effect, wherein the risk allele is associated with higher rates of the chr9:136372094:136373532:clu_47895 splicing event, leading to the retention of exon 2 (red) in the coding sequence of CARD9 (isoform A), while the protective (rs4077515-C) allele is associated with the alternative splicing event that truncates exon 2 (isoform B). The functional CARD domain (green) maps to a portion of exon 2 that is intact in isoform A, but truncated in isoform B. **b**, Blood splice QTL violin plots of normalized intron excision ratios (corresponding to the ratio of isoform A to isoform B) by the genotype of rs10870077 (top blood sQTL in GTEx) and rs4077515 (top SNP in GWAS for IgAN). These two SNPs are in near perfect linkage disequilibrium ($r^2 > 0.98$). The white bar represents a median, the thick gray bar represents an interquartile range, and the blue shape reflects the distribution kernel density estimation. Numbers in parentheses under the *x*-axis indicate *n* independent samples per each genotype group. **c**, Both rs10870077 and rs4077515 are also associated with a significant cis-eQTL effect on *CARD9* mRNA levels in GTEx blood (violin plots with similar definitions as in **b**). **d**, The blood eQTL signal for *CARD9* significantly co-localizes with the GWAS signal (PP4 of 0.86). The top panel represents a regional plot for the GWAS signal (Immunochip data excluded). The bottom panel represents a blood cis-eQTL signal for *CARD9* from the QTLGen Consortium meta-analysis (*y*-axis truncated at $P < 1 \times 10^{-310}$); rs10870077 and rs4077515 are indicated in red. All *P*-values are two-sided, correspond to a logistic regression Wald test under an additive genetic model, and are not corrected for multiple testing.

nature portfolio

Corresponding author(s): Krzysztof Kiryluk and Ali Gharavi

Last updated by author(s): Mar 1, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No computer code was used for data collection. The data was manually entered into the Progeny genetic data management system.
Data analysis	No custom computer code was used to analyze data. Only open source software packages were used for data analysis, including the following: R v.3.5.2 (CRAN), PLINK v1.07 and v1.9, METAL version 2011-03-25, TransMeta, EAGLE, MINIMAC3, KING v.2.3.0, COJO in GCTA version 1.92.0beta, Cytoscape v3.7.0, ANNOVAR, FlashPCA v.1.2.6, COLOC v.5.1.0.1, CAVIAR, RELI, LDPred version 1.0.11, SNP2HLA v.1.0, LDSC v1.0.1, DEPICTv1 release 194, GARFIELD v2, MAGMA v1.09, PheWAS R package v.0.99.5-5, BiomaRt in Bioconductor release 3.16, GLay, GeNets, ClusterONE v.1.0, UCSF Chimera v1.16.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Primary genotype data for previously published cohorts is available through dbGAP under accession number phs000431.v2.p1 and the new genotype data as well as summary statistics are available under dbGAP accession number phs000431.v3.p1. Our IRB determined that the use of the primary genotype data is restricted to genetic studies of kidney disease. The PAGE consortium control genotype data is available on dbGAP under accession number phs000356.v2.p1. The Electronic

Medical Records and Genomics-III (eMERGE-III) imputed genotype and phenotype data are available through dbGAP, accession number: phs001584.v2.p2. The UK Biobank genotype and phenotype data are available through the UK Biobank web portal https://www.ukbiobank.ac.uk/. The DICE dataset is available at https://dice-database.org/; the Blood eQTL metanalysis by eQTLGen is available at https://www.eqtlgen.org/, the Kidney eQTL Atlas is available at http://susztaklab.com/eqtl; NEPTUNE eQTL Browser is available at http://nephqtl.org/; GTEx is available at https://gtexportal.org/home/; GWAS catalog is available at https://www.ebi.ac.uk/gwas; LDHub is available at http://www.nealelab.is/tools-and-software; Open Targets platform is available at https://www.targetvalidation.org; DLRP can be accessed at https://dip.doe-mbi.ucla.edu/dip/DLRP.cgi, DrugBank can be accessed at https://www.drugbank.ca; GlobalData database is available at https://www.globaldata.com/industries-we-cover/pharmaceutical; FUN-LDA scores are available at http://www.columbia.edu/~ii2135/funlda.html; MSigDB and GSEA are available at http://software.broadinstitute.org/gsea/msigdb/; STRING is available at https://string-db.org; InWeb is available at http://www.lagelab.org/resources/; the Metabolomics GWAS Server is available at http://metabolomics.helmholtz-muenchen.de/gwas/; the MGI database is available at http://www.informatics.jax.org.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	All available cases and controls consented for genetic studies with DNA or SNP array data available were included in the study to maximize sample size for discovery. All included cases had a kidney biopsy confirmed diagnosis of IgA nephropathy.
Data exclusions	Within each cohort and across all cohorts we excluded any duplicate samples, cryptically related samples, or ancestry outliers by PCA analysis as described in the methods. There were no exclusions based on age, sex, or self-reported race/ethnicity.
Replication	This study involves a global meta-analysis of all IgAN case-control cohorts. We were unable to provide independent replication of the discovery GWAS findings in additional cohorts, since all currently available cohorts were included in the global discovery meta-analysis. For testing polygenic risk scores, we use a smaller independent replication cohort (GCKD study) as described in the methods. We additionally provide independent replication in the eMERGE-III and UKBB biobanks by recovering associations with relevant phenotypes.
Randomization	Not applicable, this is an observational genetic study and not a randomized trial.
Blinding	Not applicable, this is an observational genetic study and not a randomized trial that requires blinding.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	All cases included in this study had a kidney biopsy-confirmed diagnosis of primary IgA nephropathy. Detailed breakdown of cases and controls by cohort, ancestry, and genotyping platform is provided in Supplemental Table S1. Additional cohort descriptions are provided in Supplementary Notes.
Recruitment	The subjects included in this study were recruited across multiple nephrology centers worldwide as described in Supplementary Notes. All cases had a biopsy-confirmed diagnosis of primary IgA nephropathy. We excluded any patients with suspected secondary causes such as those with an autoimmune disease, malignancy, or liver disease. We used ancestry-matched population controls without any known history of kidney disease. All participants were consented for genetic

studies. We do not suspect self-referral bias, since the cases were ascertained by nephrologists at respective clinical sites among all biopsied patients diagnosed with IgA nephropathy. However, because of the biopsy-based case ascertainment, milder cases of IgA nephropathy that may not be routinely biopsied were likely missed by our analysis. The use of population controls might have led to some control misclassification, although this problem is likely minimized by the low overall disease prevalence.

Ethics oversight

All subjects provided informed consent to participate in genetic studies, and local ethics review committees at each of the individual participating sites approved the study. This includes the following Institutional Review Boards (IRBs), Research Ethics Boards (REBs), or Ethics Committees: Columbia University IRB; University of Brescia Ethics Committee; Gaslini Institute Ethics Committee; University of Bari Ethics Committee; University of Parma Ethics Committee; University of Torino Ethics Committee; University of Warsaw Ethics Committee; Poznan Medical University Ethics Committee; University of Napoli Ethics Committee; Brotzu Hospital Ethics Committee; University Hospital of Trieste Ethics Committee; University of Modena Ethics Committee; University of Split Ethics Committee; Ospedale San Raffaele (OSR) Ethics Committee; Istanbul University Ethics Committee; Peking University First Hospital Ethics Committee; University of Buenos Aires Ethics Committee; University of Skopje Ethics Committee; Bern University Ethics Committee; Amsterdam VU University Medical Center Ethics Committee; Hospital Universitario Puerta del Hierro Majadahonda Ethics Committee; Jean Monnet University Ethics Committee; Medical University of Innsbruck Ethics Committee; University of Pécs Ethics Committee; Pirogov Russian National Research Medical University Ethics Committee; University of Aachen Ethics Committee; University of Toronto REB; University Hospitals in Leuven Ethics Committee, Yale University IRB; University of Alabama IRB; Karolinska Institute Ethics Committee; INSERM Ethics Committee; Vilnius University Ethics Committee, Seoul National University Hospital Ethics Committee; Ruijin Hospital Ethics Committee; Juntendo University Ethics Committee; Charles University Ethics Committee; Niigata University Ethics Committee; Zagreb Medical School Ethics Committee; The Aristotle University of Thessaloniki Ethics Committee; Indiana University IRB; Vanderbilt University Medical Center IRB; University of California Los Angeles IRB; the University of Michigan IRB; Nationwide Children's Hospital IRB; Boston Children's Hospital IRB; Le Bonheur Children's Hospital Foundation (University of Tennessee) IRB; Texas Tech University IRB; St. Louis Children's Hospital (Wash U) IRB; University of Minnesota IRB; Stony Brook IRB; Helen DeVos Children's Hospital (Spectrum Health) IRB; Cincinnati Children's Hospital IRB; University of Kentucky IRB; Cornell University IRB; Connecticut Children's Medical Ctr IRB; University of Texas-San Antonio IRB; Seattle Children's IRB; University of Miami IRB; NY Medical College IRB; Medical College of Wisconsin IRB; Driscoll Children's Hospital IRB; Phoenix Children's Univ. of Arizona IRB. The Institutional Review Board of Columbia University oversaw all of the human subjects research compliance related to this study and approved the umbrella protocol (number IRB-AAAC7385).

Note that full information on the approval of the study protocol must also be provided in the manuscript.