nature communications

Article

https://doi.org/10.1038/s41467-026-78026-6

Genome-wide studies define genetic mechanisms of IgA vasculitis

Received: 15 July 2025

Accepted: 3 September 2026

Published online: 15 October 2026

Check for updates

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

IgA vasculitis is a pediatric disease with skin and systemic manifestations. Here, we conduct genome, transcriptome, and proteome-wide association studies in 2170 IgA vasculitis cases and 5928 controls, generate IgA vasculitis-specific maps of gene expression and splicing from blood of 255 pediatric cases, and reconstruct myeloid-specific regulatory networks to define disease master regulators modulated by the identified genes. We observe significant association at the *HLA-DRB1* locus (OR = 1.55, $P = 1.1 \times 10^{-25}$) and fine-map specific amino-acid risk substitutions in DRβ1. We find two novel non-HLA associated loci: *FCAR* (OR = 1.51, $P = 1.0 \times 10^{-20}$), encoding a myeloid IgA receptor FcαR, and *INPP5D* (OR = 1.34, $P = 2.2 \times 10^{-09}$), encoding an inhibitor of FcαR signalling. The *FCAR* risk locus co-localizes with a cis-eQTL that increases *FCAR* expression; the risk alleles disrupt a PRDM1 binding motif within a myeloid enhancer of *FCAR*. Another risk locus is associated with higher genetically predicted levels of plasma IL6R. The IL6R risk haplotype carries a missense variant contributing to accelerated cleavage of IL6R into a soluble form. Applying network-based approaches to blood single cell transcriptomes, we predict, independently replicate, and experimentally validate 21 IgA vasculitis master regulators co-modulated by *FCAR*, *INPP5D* and *IL6R* signalling in myeloid cells.

IgA vasculitis (IgAV), also known as Henoch-Schönlein purpura (HSP), is the most common form of childhood vasculitis, affecting 10–27 children per 100,000 per year[1]. IgAV is characterized by IgA deposition in blood vessel walls in the skin, joints, gastrointestinal tract, and the glomerular mesangium[2,3]. IgAV can lead to serious complications, including kidney failure[4]. The exact pathogenic mechanisms of IgAV remain unclear, resulting in the general lack of targeted treatments and preventive strategies[5]. The contribution of common genetic variants to IgAV risk has not been systematically assessed, although the HLA region has been implicated in small studies[6–10].

Herein, we report the genetic study of IgAV involving 8098 individuals of European and East Asian ancestry (2170 cases and 5928 controls). By combining genome-wide association study (GWAS), transcriptome-wide association study (TWAS), and proteome-wide association study (PWAS) approaches, our study implicated myeloid cells as central to the disease pathogenesis, and prioritized *HLA-DRB1*, *FCAR*, *INPP5D*, and *IL6R* as candidate causal genes for IgAV. To assess the transcriptional consequences of the newly identified risk alleles, we performed blood RNA sequencing of 255 pediatric cases with IgAV and built a disease-context-specific atlas of expression and splicing quantitative trait loci (QTLs). Using methods proposed for cancer[11,12], we reconstructed a myeloid-specific transcriptome-wide regulatory network based on blood single-cell RNA-seq and identified proteins representing candidate master regulators of IgAV-related myeloid cell state whose activity is modulated by the candidate causal genes. We replicated these regulators in an independent cohort and validated them experimentally. In aggregate, our genetic and systems biology approach defined pathogenic mechanisms and provided genetic support for therapeutic targets in IgAV.

Results

Genome-wide association study (GWAS)

We performed a GWAS meta-analysis of three large case-control cohorts (two newly genotyped and one previously published cohort[7]) comprising a total of 8098 individuals (2170 IgAV cases and 5928 controls). The study cohorts are summarized in Table S1 with detailed

e-mail: hongzh@bjmu.edu.cn; kk473@columbia.edu

https://doi.org/10.1038/s41467-026-78026-6

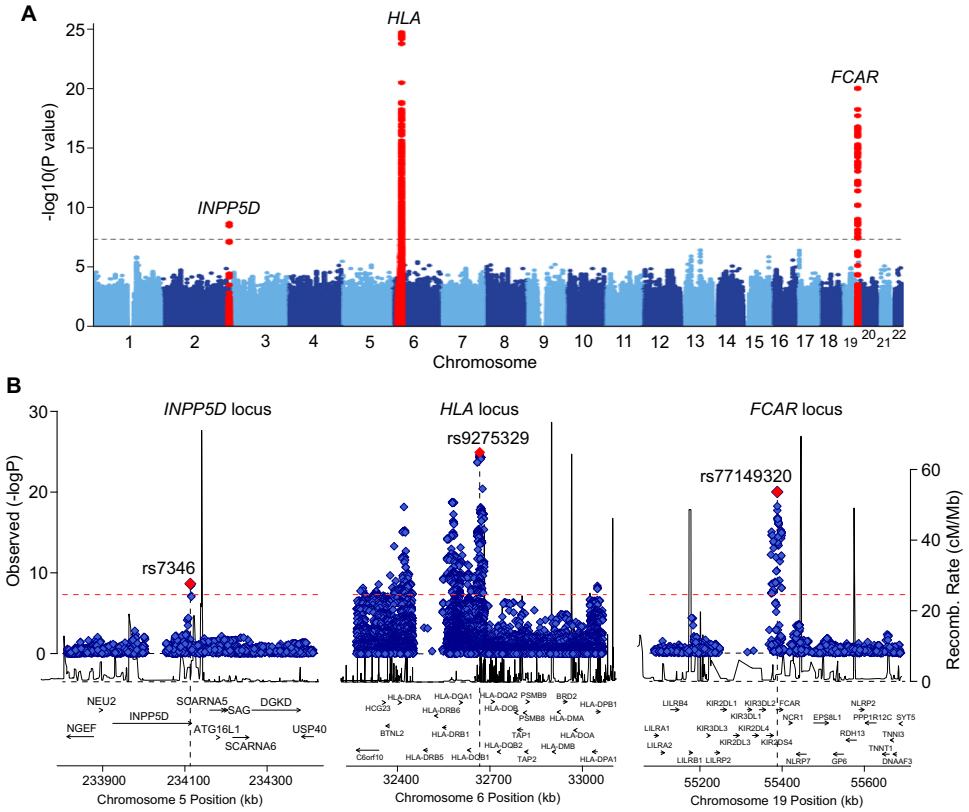


Fig. 1 | GWAS meta-analysis for IgAV identified 3 genome-wide significant regions. A Manhattan plot with genome-wide significant loci highlighted in red. The y-axis represents −log10 (*P*-value) for the fixed effects meta-analysis (two-sided, unadjusted). The dotted horizontal line indicates a genome-wide significance threshold ($P = 5 \times 10^{-8}$). **B** Regional plots for the *INPP5D* locus (left), the *HLA* locus (middle), and the *FCAR* locus (right). The x-axis shows the physical position in kilobases (kb, hg19 coordinates) and includes known gene annotations; the left y-axis presents −log10 *P*-values for association statistics (two-sided, unadjusted), and the right y-axis shows the recombination rates; the dotted horizontal line indicates a genome-wide significant threshold of 5×10^{-8}.

descriptions provided in the "Methods." In each cohort, the controls were recruited from the same ancestral populations and geographic regions as the cases and had no history of kidney or autoimmune diseases. Of the three cohorts, two were of European ancestry (1025 cases and 3173 controls), and one cohort was of East Asian ancestry (1145 cases and 2755 controls). All cohorts were genotyped with high-density SNP arrays, ancestrally matched using principal component-based methods[13] (Fig. S1A–D), and imputed using ancestry-specific genome sequence reference panels[14]. In the meta-analysis, we observed negligible genomic inflation, consistent with the absence of systematic biases ($\lambda = 1.02$, Fig. S1E). We detected three genome-wide significant loci ($P < 5 \times 10^{-8}$, Fig. 1), including the *HLA* locus (OR = 1.55, $P = 1.1 \times 10^{-25}$) and two non-HLA loci, on chr.2q37.1 (*INPP5D*, OR = 1.34, $P = 2.2 \times 10^{-09}$) and chr.19q13.42 (*FCAR*, OR = 1.51, $P = 1.0 \times 10^{-20}$). Conditional analysis identified two independent signals within the HLA region, but no secondary signals at the non-HLA loci. The independently significant SNPs are summarized in Table S2. We also detected 14 suggestive loci based on $P < 1 \times 10^{-5}$, including the *TNFSF13* and *IL6R* loci (Table S3).

Transcriptome-wide association study (TWAS) and proteome-wide association study (PWAS)

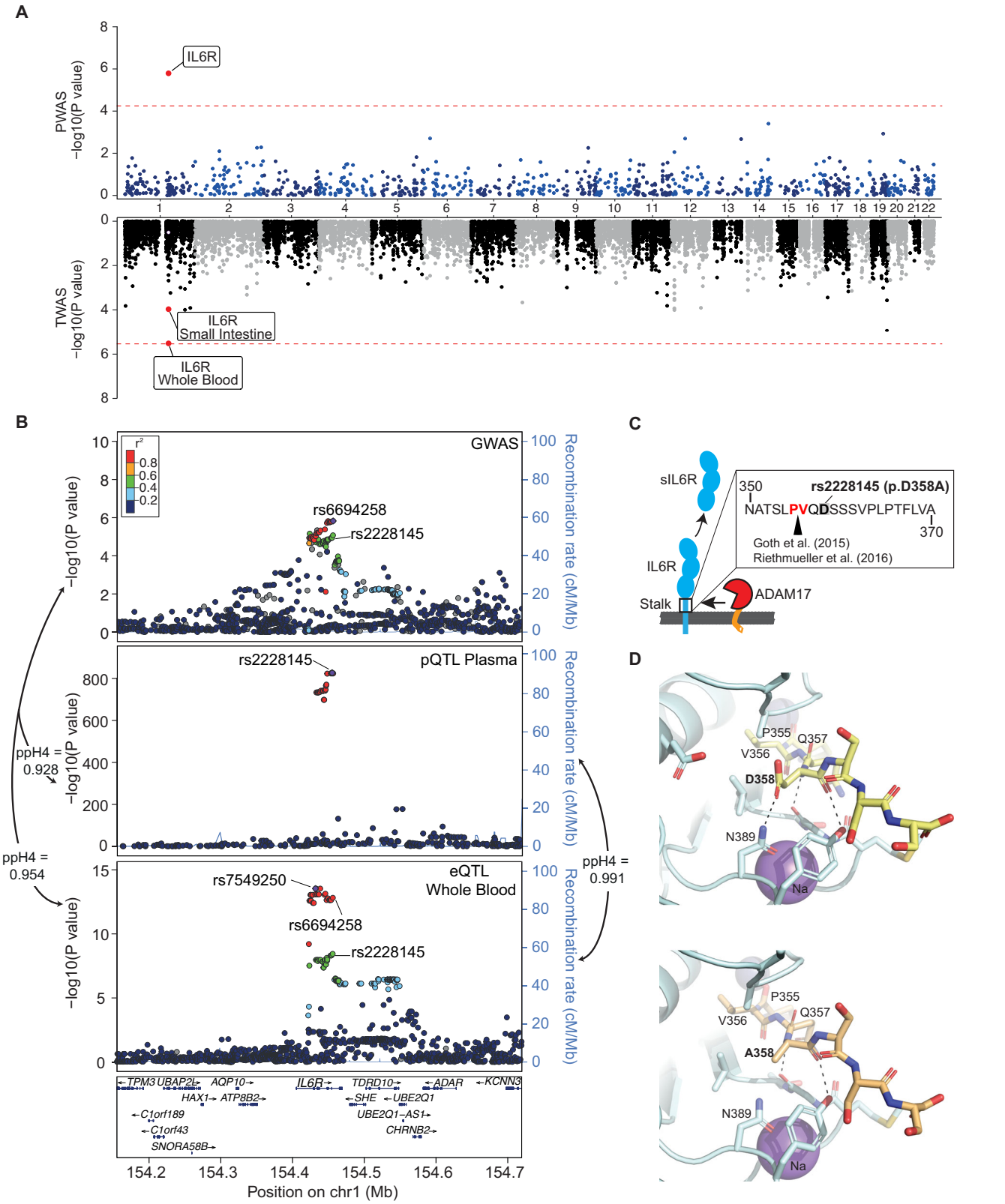
TWAS[15] and PWAS[16] represent two exploratory approaches to identify causal genes for complex traits by testing the effects of genetically determined gene or protein expression on a disease risk. Our PWAS (see "Methods") identified a single proteome-wide significant signal for soluble IL6R, with higher genetically determined plasma levels associated with greater disease risk ($P = 1.6 \times 10^{-6}$, Fig. 2A, Data S1). TWAS provided no signals of transcriptome-wide significance, although the *IL6R* signal was suggestive in blood and small intestine tissues (Data S2).

HLA-DRB1 locus

Using ancestry-specific reference panels, we imputed classical HLA alleles and individual amino-acid polymorphisms at the HLA class I and II genes ("Methods"). In the East Asian cohort, the bi-allelic analysis of all polymorphisms across the HLA region identified the strongest signal at rs660895, a non-coding SNP upstream of the *HLA-DRB1* gene; conditioning on rs660895 revealed no additional genome-wide significant signals (Data S3). The *HLA-DRB1*04* allele tagged by rs660895 ($r^2 = 0.5$, $D' = 1.00$) was most significant across all alleles in the bi-allelic analysis (OR = 1.62, $P = 3.9 \times 10^{-10}$). The gene-based analysis of all classical HLA alleles confirmed the strongest association at *HLA-DRB1* ($P = 1.5 \times 10^{-16}$, Fig. 3A and Table S4). Among the polymorphic sites in DRβ1, position 11 was most significant ($P = 2.9 \times 10^{-15}$), and there was no other significant signal after conditioning on this position (Fig. 3C and Table S5). Specifically, valine at DRβ1 position 11, tagging *HLA-DRB1*04*, conveyed the strongest risk (OR = 1.65, $P = 3.4 \times 10^{-12}$, Table S6). In contrast, proline at DRβ1 position 11, tagging *HLA-DRB1*1501*, conveyed a protective effect (OR = 0.73, $P = 4.5 \times 10^{-6}$, Table S6). The same residues with concordant effects have been previously reported in the GWAS for IgA nephropathy (IgAN)[17], suggesting shared HLA susceptibility between IgAV and IgAN.

Consistent with the East Asian cohort analysis, *HLA-DRB1* was also genome-wide significant in European cohorts (Fig. 3B and Table S7). Similarly, the analysis of polymorphic sites identified the strongest association of DRβ1 position 11, with no other independent signals after conditioning on position 11 (Fig. 3D and Table S8). We observed

https://doi.org/10.1038/s41467-026-78026-6



that leucine at position 11 was a strong risk substitution in Europeans (OR = 1.72, $P = 1.7 \times 10^{-10}$, Table S6). This substitution is rare in East Asian (control freq. 0.03) compared to European populations (control freq. 0.11), explaining the ancestry-specificity of this signal. Consistent with the findings in East Asians, we also replicated the risk effect of valine (OR = 1.42, $P = 2.6 \times 10^{-5}$) and the protective effect of proline (OR = 0.65, $P = 4.8 \times 10^{-6}$) in Europeans. These results were also consistent with the bi-allelic analysis of all polymorphisms across the HLA region (Data S4), in which leucine or valine substitution at *DRB1* position 11 conveys the strongest risk (OR = 1.71, $P = 7.8 \times 10^{-16}$).

Next, we performed a detailed analysis of the potential impact of the associated amino acid risk residues on the structure and potential

Fig. 2 | Effects of the *IL6R* locus. A Miami plot for PWAS and TWAS analyses. The red lines indicate the Bonferroni adjusted significance thresholds (5.6×10^{-5} for PWAS and 2.5×10^{-6} for TWAS). The y-axis indicates −log10 (*P*-value) from two-sided tests without multiple-testing adjustment. The source data are provided in Data S1 and S2. **B** Regional association plots for the *IL6R* locus. SNPs are plotted by position (hg19, 250 kb window) versus −log10 (*P*-values), calculated from two-sided tests with no adjustment for multiple testing, from GWAS of IgAV (top), pQTL of plasma sIL6R levels (middle), and eQTL of *IL6R* expression in whole blood (bottom). The purple diamond highlights the most significant SNP for each association. SNPs are color-coded to reflect their LD with this SNP using LocusZoom with HapMap CEU reference. The missense variant *IL6R* p.Asp358Ala (rs2228145) is indicated in each plot. **C** The p.Asp358Ala substitution in relation to the cleavage site of the membrane-bound IL6R. **D** Structure of the ADAM17 catalytic domain/IL-6R complex. The structure of ADAM17 (pale cyan ribbon render) was modeled with bound peptide 355-PVQDSSS-361 (yellow sticks in upper panel) or 355-PVQASSS-361 (pale orange sticks in lower panel) corresponding to a segment of the stalk region of IL-6R including the cleavage site (P355/V356). Residues of ADAM17 within 5 angstroms of D358 or A358 are shown as sticks. Polar interactions between residue 358 and ADAM17 are noted by dashed lines. The change D358A eliminates the interaction with N389 of ADAM17, potentially reducing the affinity of IL-6R for ADAM17. Zinc and Sodium ions are shown as slate and purple spheres, respectively. Images generated with PyMOL (Molecular Graphics System, Version 2.0, Schrödinger, LCC).

function of DRβ1. Valine at position 11 is in tight LD with histidine at position 13 ($r^2 = 0.99$) and tyrosine at position 96 ($r^2 = 0.98$), forming a distinct DRβ1 risk haplotype (V11-H13-Y96). The second (European-specific) DRβ1 risk haplotype (L11-F13-E96) is formed by leucine at position 11, in LD with phenylalanine at position 13 ($r^2 = 0.99$) and glutamic acid at position 96 ($r^2 = 0.99$). The positions 11 and 13 map to the P6 and P4 pockets of the DRβ1 antigen-binding groove, while position 96 is at the surface-exposed on the DR-β2 domain (Fig. S2 and Fig. 3). Comparisons of protein structures between the protective and risk haplotypes showed two features. First, the crevasse between pockets 6 and 4 with the protective haplotype (Fig. S2B, left) is less accessible compared to the two risk haplotypes (Fig. S2B, middle and right). Second, the protective substitution with arginine at position 13 provides a distinct electronegative character to the binding groove (Fig. S2C, left). These landscape and electrostatic changes may alter which peptides are accommodated by DRβ1[18]. Moreover, position 96 is involved in the DRβ1 interactions with the HLA-DM complex in the thymic antigen-presenting cells[19], and thus substitutions at this position may additionally alter the presentation of self-antigens for clonal deletion[20,21] (Fig. S3). As shown in Fig. S3B1–3, residue 96 is in the binding-interface with HLA-DM β-chain. All three residues Q96 (protective), E96, and Y96 (both risk) showed differing degrees of association with DM and the aqueous environment. As this is one of several residues that contribute to the interface with DM, future studies will be needed to define the mechanism of this association. Additionally, HLA-DO, a nonclassical MHC-II protein that appears to promote tolerance to self-antigens[22,23], binds to HLA-DM through this same interface to competitively inhibit interaction with DRβ1[24]. Thus, residues such as 96 could modulate the interaction of HLA-DM with other factors involved in antigen presentation.

INPP5D locus

The lead SNP (rs7346) at chr.2q37.1 locus is a synonymous variant in *INPP5D* (encoding Inositol Polyphosphate-5-Phosphatase D, a.k.a. SHIP-1) and is in tight LD with rs9247 ($r^2 = 0.99$), a missense variant p.His1157Tyr in *INPP5D* (Table S9). Although there is no clear deleterious prediction for this substitution by PolyPhen2, CADD, and REVEL, this candidate variant is located within a compositionally biased region of SHIP-1 that is enriched in charged residues, and the substitution of histidine (positively charged) with tyrosine (neutral) may alter its binding energy. SHIP-1 is expressed in myeloid cells, where its activation is mediated by tyrosine phosphorylation. Once phosphorylated, SHIP-1 functions as a negative regulator of FcαR signaling, inhibiting myeloid cell proliferation, chemotaxis, and degranulation[25,26]. Phosphorylation of SHIP-1 also occurs by simultaneous engagement of IgA and IgE Fc receptors, suppressing allergic responses[27]. In mice, deletion of *Inpp5d* leads to autoinflammatory disease with neutrophilia and gastrointestinal involvement[28,29]. In GWAS, this locus has previously been associated with allergy[30,31], and increased neutrophil[32], eosinophil[33,34], and monocyte counts[32]. In our meta-PheWAS across major biobanks involving 786,138 individuals (Fig. S4A and Data S5), the IgAV-risk allele at this locus was additionally associated with hypothyroidism (OR = 1.05, $P = 2.5 \times 10^{-10}$) and dermatitis herpetiformis (a skin rash due to IgA response against epidermal transglutaminase and involving FcαR-mediated neutrophil degranulation[35], OR = 1.95, $P = 8.7 \times 10^{-7}$).

FCAR locus

The index SNP at 19q13.42 (rs77149320) maps to the intronic region of *FCAR*, encoding IgA receptor FcαR (CD89), a transmembrane glycoprotein expressed specifically on myeloid cells. FcαR interacts with IgA immune complexes, triggering phagocytosis, cytotoxicity, and release of inflammatory mediators. In prior GWAS, this locus has been associated with IgAN[17] and inflammatory bowel disease[36]. In our meta-PheWAS, the most significant association of this locus was with a higher risk of "nephritis and nephropathy with pathological lesion" (Fig. S4B). The index SNP is in tight LD ($r^2 > 0.8$) with seven variants intersecting a putative cis-regulatory region in myeloid cells defined by ABC models[37], confirmed by FUN-LDA functional scores[38], and intersecting open chromatin by monocyte-specific ATAC-seq in 25 healthy donors[39] (Fig. 4A and Table S10). Two of the seven top risk alleles in LD (rs4806602-G and rs73065472-T) disrupt the predicted binding motif for PRDM1[40], a transcriptional repressor triggered by infections via Toll-like receptor and STAT signaling (Fig. 4A). PRDM1 activates NF-κB/TNF-R signaling by repressing NLRP2[41], maintains effector functions of B- and T-cells[41,42], and promotes myeloid cell differentiation[42]. In agreement with these findings, our network-based differential activity analysis demonstrated higher regulatory activity of PRDM1 in IgAV cases versus controls (FDR = 0.0037, see below).

IL6R locus

Soluble IL6R (Interleukin 6 receptor) reached a proteome-wide significance in PWAS and was supported by a suggestive signal in GWAS. Interestingly, the top GWAS risk allele (rs6694258-A, OR = 1.21, $P = 1.49 \times 10^{-6}$, Table S3) was associated with higher plasma protein levels of IL6R, but lower *IL6R* transcript levels in blood (Fig. 2B). The GWAS locus co-localized with both expression QTL (posterior probability 95%) and protein QTL (posterior probability 93%) supporting a shared casual haplotype (Fig. 2B and Data S6). Four SNPs in strong LD with rs6694258 ($r^2 \geq 0.8$) intersected *IL6R* enhancers predicted by the ABC model[37] and confirmed by ATAC-seq in monocytes[39] (Fig. S5 and Table S10). The observed protein QTL effect of this locus was particularly strong, with the top SNP (rs4129267) alone explaining 25.9% of variance in plasma IL6R levels. The second top variant (rs2228145, $r^2 = 0.998$ with rs4129267) is a missense variant resulting in p.Asp358Ala substitution in the stalk region, close to the ADAM17 cleavage site of IL6R (between position 355 and 356; Fig. 2C)[43], and demonstrated to increase soluble IL6R levels in prior studies[44–46]. We modeled IL6R-ADAM17 structures based on published studies[47] to demonstrate that Asp358 forms a hydrogen bond with Asn389 of ADAM17, whereas p.Asp358Ala substitution in IL6R results in the loss of this interaction (Figs. 2D and S6). This may decrease affinity between

https://doi.org/10.1038/s41467-026-78026-6

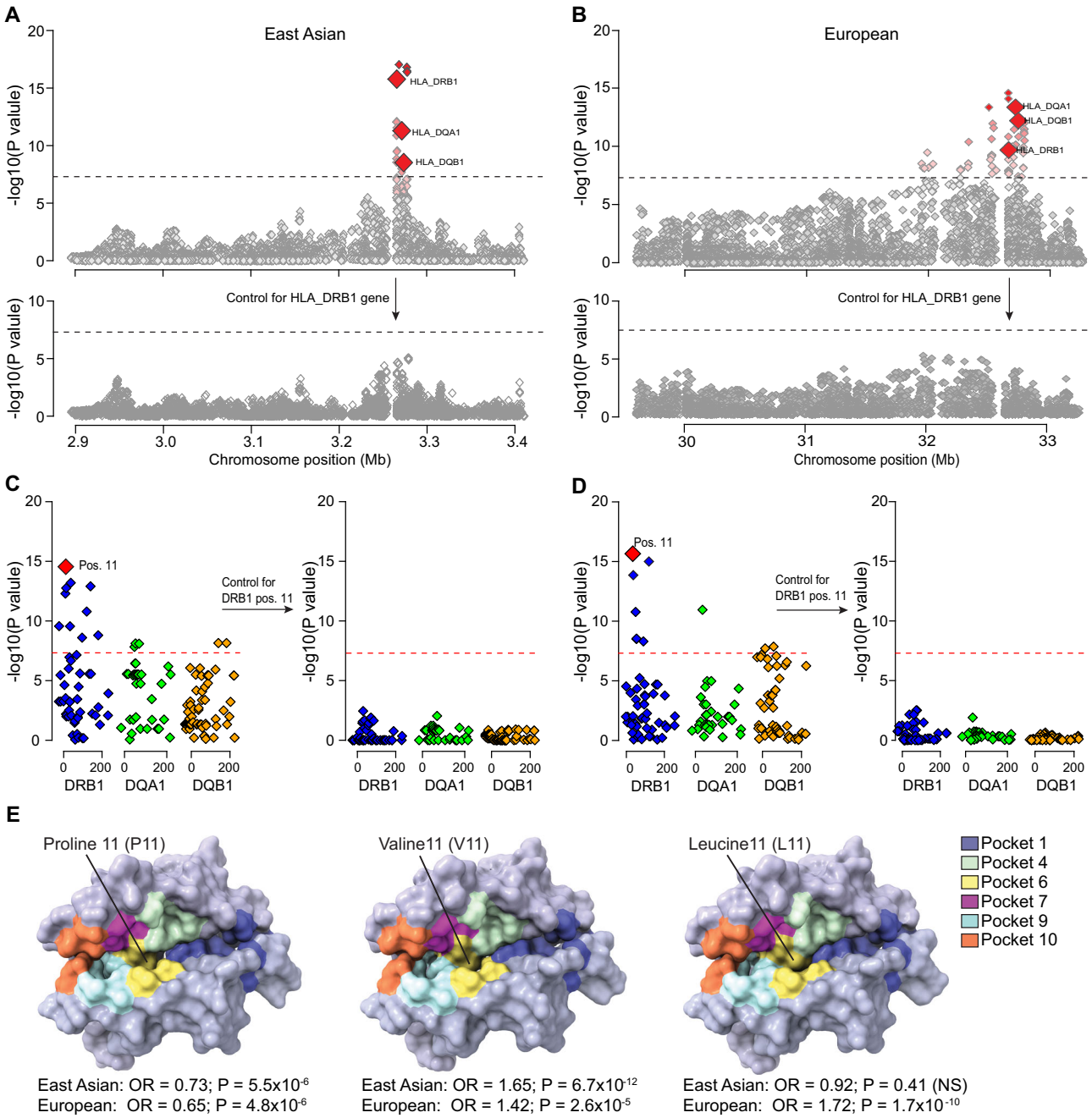


Fig. 3 | Fine mapping of the *HLA* region in East Asian and European ancestry cohorts. A Regional plot for the East Asian cohort with bi-allelic and multi-allelic association statistics for all imputed variants and classical HLA alleles; the strongest association signal was for *HLA-DRB1* (upper panel); the results for *HLA-DQA1* and *HLA-DQB1* are highlighted for reference; after controlling for *HLA-DRB1* classical alleles (black arrow), there was no residual association across the entire 3-Mb region (lower panel). The y-axis indicates −log10 (*P*-value) calculated from the χ^2 test. The source data are provided in Data S3 and Table S4. **B** Regional plot for the European cohorts with bi-allelic and multi-allelic association statistics for all imputed variants and classical HLA alleles; after controlling for *HLA-DRB1* classical alleles, there were no significant associations in the region (lower panel). The dotted horizontal line indicates the genome-wide significance threshold ($\alpha = 5 \times 10^{-8}$). The y-axis shows −log10 (*P*-value) from the χ^2 test. The source data are provided in Data S4 and Table S7. **C** East Asian analysis of polymorphic amino acid positions within DRβ1 (blue), DQα1 (green), and DQβ1 (orange) using conditional-haplotype tests; the horizontal dashed line indicates the genome-wide significance level. The most strongly associated polymorphic site was position 11 in DRβ1 (left panel); after controlling for this position, there was no residual association (right panel). The y-axis shows −log10 (*P*-value) from the χ^2 test. The source data are provided in Table S5. **D** European analysis of polymorphic amino acid positions within DRβ1 (blue), DQα1 (green), and DQβ1 (orange); consistent with East Asian analysis, the DRβ1 amino acid position 11 (left panel) provided the strongest signal; no additional independent positions were found upon conditioning on this position (right panel). The *P*-value is calculated from the χ^2 test. The source data are provided in Table S8. **E** Structure of the DRα1-DRβ1 complex. The protective variant (P11, left) and risk variants (V11, middle, and L11, right) are shown in surface model viewed into the peptide binding groove. Models are colored in a background of light blue shade, with individual peptide amino acid binding pockets shaded according to the provided color key.

https://doi.org/10.1038/s41467-026-78026-6

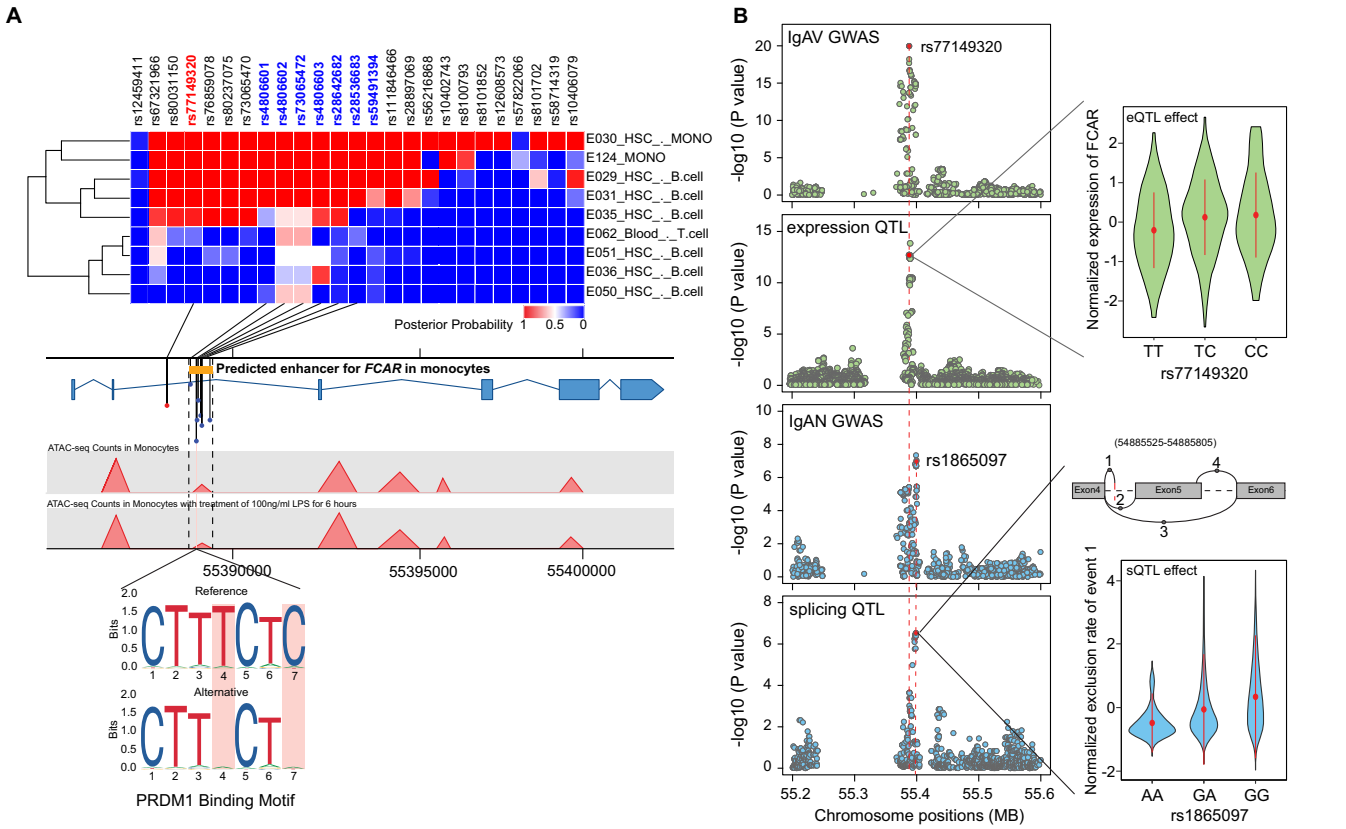


Fig. 4 | Functional annotations of the *FCAR* locus. A The heatmap of FUN-LDA functional scores for the lead SNP (rs77149320, red) and its high LD proxies ($r^2 > 0.8$, only SNPs and cell types with positive scores are depicted). The color gradient in the heatmap represents posterior probability of a genetic variant being functional in a specific cell type or tissue by FUN-LDA, ranging from high (red) to low (blue). The 7 SNPs in LD (names in blue) intersected an ABC model-predicted intronic enhancer of *FCAR*. This regulatory region was further confirmed by an ATAC-seq peak in monocytes before and after LPS treatment. The prioritized SNPs were analyzed against transcription factor binding motifs predicted by JASPAR, and two risk alleles (rs4806602-G and rs73065472-T in perfect LD and two base pairs apart) were found to disrupt the PRDM1 binding motif. **B** Colocalization of the *FCAR* locus between GWAS for IgAV and blood eQTL for *FCAR* in IgAV (top two panels), and GWAS for IgAN and blood sQTL for *FCAR* in IgAV (bottom two panels). The eQTL and sQTL analyses were conducted using the data from 255 IgAV patients. The y-axis depicts −log10 (*P*-value) (two-sided and unadjusted). The eQTL effect shown in the right upper panel (x-axis: genotypes, y-axis: normalized expression of *FCAR*). The sQTL effect shown in the right bottom panel (x-axis: genotypes, y-axis: normalized exclusion rate of chr19:54885525-54885805 region based on hg38 annotations, the splice model depicted above). The red error bars correspond to 95% confidence intervals around the mean.

IL6R and ADAM17, potentially allowing for faster release upon cleavage and a higher turnover rate. Prior studies showed that a peptide harboring the Asp358Ala substitution is cleaved more rapidly by ADAM17 than the peptide with the Asp358 residue[46]. In our Meta-PheWAS of rs2228145, the risk allele for IgAV was additionally associated with a higher risk of asthma and urticaria, but a reduced risk of atrial fibrillation/flutter and coronary syndrome (Fig. S4C and Data S7).

Shared and distinct loci with IgAN

To examine the genetic relationship between IgAV and IgAN, we tested genome-wide genetic correlation using results from the latest GWAS for IgAN[17]. We observed a significant positive genetic correlation between IgAV and IgAN ($r_g = 0.77$, $P = 2.8 \times 10^{-12}$) that also present after excluding the HLA region ($r_g = 0.69$, $P = 0.016$). To better understand the shared and distinct loci between IgAN and IgAV, we performed a cross-phenotype meta-GWAS (see "Methods"). We observed 21 genome-wide significant loci, including 3 novel loci, near *PAID4*, *WLS*, and *ANKRD55* (Fig. S7 and Table S11). All three IgAV-risk loci (*HLA*, *INPP5D*, and *FCAR* loci) became more significant in the cross-phenotype analysis (Fig. S7), but at the *FCAR* locus, significant heterogeneity was detected between IgAV and IgAN ($I^2 = 94.98$, $P < 0.0001$). While the *INPP5D* locus perfectly co-localized between the two traits (posterior probability of shared causal variant 0.99), the *FCAR* locus did not colocalize (posterior probability of different causal variants 0.98, Table S12), implying distinct genetic mechanisms between IgAV and IgAN at this locus.

Maps of blood expression and splicing QTLs in IgAV

To better interpret the effects of IgAV-risk alleles, we performed blood RNA-seq in a cohort of 255 IgAV cases and generated genome-wide maps of disease-context-specific expression and splicing QTLs. In total, we identified 9,401 cis-eGenes (13,560 independent eQTLs) at FDR < 0.05 (Data S8 and S9). Among them, 2,359 eGenes were IgAV context-specific when compared to GTEx v7 whole-blood dataset (Fig. S8). These IgAV-specific eGenes were enriched in immune-related pathways and mouse phenotypes, including T-cell activation and differentiation as well as cytokine and interleukin signaling (Fig. S8). We also identified 1,294 sGenes (Data S10), including 303 IgAV-specific sGenes enriched in the biological processes related to immunity, such as immune response-regulating signaling pathway and adaptive immune response (Table S13). The eQTL and sQTL results were then used to investigate the consequences of IgAV-risk alleles.

Based on these maps, we observed strong colocalization between the *FCAR* locus for IgAV and *FCAR* cis-eQTL (posterior probability of shared causal variant of 0.96). IgAV-risk allele was associated with increased blood *FCAR* mRNA expression (Fig. 4B). Differential gene expression analysis confirmed higher *FCAR* gene expression in IgAV cases compared to controls (fold change = 1.3, $P = 6.84 \times 10^{-06}$). In

contrast, the distinct *FCAR* association signal for IgAN did not colocalize with either eQTL or GWAS for IgAV, but it did colocalize strongly with an sQTL for *FCAR* (posterior probability of shared causal variant of 0.99). One of the top IgAN-risk alleles (rs1865097-A) is a deep intronic variant associated with a lower exclusion rate of the chr19:54885525-54885805 region (Fig. 4B). This variant, located at the end of the alternative splicing event (chr19:54885805), functions as a cryptic splice acceptor site. The protective exclusion of the chr19:54885525-54885805 region results in the transcript without exon 5 of *FCAR* encoding the transmembrane domain.

IgAV transcriptional regulators in myeloid cells

Given that all identified candidate causal genes are expressed specifically in a myeloid cell lineage, we generated a myeloid-specific transcriptome-wide regulatory gene network based on blood single-cell RNA-seq data from 163 reference individuals (Figs. 5, 9 and S10, and "Methods"). To overcome the sparseness of the single-cell gene expression matrix, we created a pseudo-bulk gene expression profile for the myeloid cell cluster. We then used a mutual-information-based network reconstruction method, ARACNe3, to identify 812,995 significant pairwise regulatory relationships between 2212 transcriptional regulators and their targets. Next, we generated the differential gene expression signature using the bulk RNA-seq data from our dataset of 255 IgAV patients and 38 healthy controls controlling for age, sex, genetic ancestry, and CIBERSORTx[48]-deconvolved cell fractions (Fig. 5B). Based on the myeloid-specific ARACNe3 regulons (Fig. 5C) and the adjusted differential gene-expression signatures, we inferred differentially activated regulators in IgAV cases compared to controls using the NaRnEA algorithm[49]. This analysis identified 773 regulators with increased activity and 916 with decreased activity in IgAV cases compared to controls based on FDR < 0.05 (Fig. 5D and Data S11). The regulators activated in the disease state were significantly enriched in the TGFβ signaling pathway (Table S14). We have previously identified genetic support for the involvement of transcriptional regulators *LITAF*, *HDAC7*, *RUNX2*, and *RUNX3* in the determination of IgA levels[50]. In IgAV cases, both *LITAF* and *HDAC7* were inferred to have increased activity in myeloid cells (FDR = 1.5×10^{-10} and 4.1×10^{-8}, respectively), while *RUNX2* and *RUNX3* were inferred to have suppressed activity compared to controls (FDR = 1.5×10^{-3} and 5.9×10^{-14}, respectively). These observations are direction-consistent with the previously reported genetic effects of these regulators on IgA levels[50], supporting global activation of the regulatory programs that promote IgA production in IgAV.

Master regulators modulated by IgAV candidate genes

Next, we hypothesized that *FCAR*, *INPP5D*, and *IL6R* transcripts may modulate the activity of proteins representing Master Regulators (MRs) of IgAV-related myeloid cell state. To test this hypothesis, we used MINDy[51], an algorithm that identifies potential modulators of transcriptional interactions by analyzing statistical dependencies between a transcription factor and its targets conditioned on the expression level of a candidate modulator. By applying this approach to all myeloid regulators using GWAS candidate genes as potential modulators and myeloid-specific ARACNe3 regulons, we identified 1194 regulators modulated by *FCAR* (Data S12), 1466 by *INPP5D* (Data S13), and 1302 by *IL6R* (Data S14) at FWER < 0.05 (Fig. 5E). A total of 1072 regulators were co-modulated by both *FCAR* and *INPP5D* with a strong negative correlation in biological activity, indicating that the regulator activities enhanced by *FCAR* were generally repressed by *INPP5D* (Spearman's r = -0.52, $P < 2.2 \times 10^{-16}$). This phenomenon is fully consistent with the established role of *INPP5D* as a negative regulator of FcαR signaling. Moreover, a total of 1009 regulators were co-modulated by *FCAR* and *IL6R*, with strong positive correlation of their biological activity (Spearman's r = 0.78, $P < 2.2 \times 10^{-16}$), suggesting concordant effects on gene regulation in myeloid cells. To better understand which specific myeloid subpopulation represents the most likely effector cell type, we superimposed the expression of *FCAR*, *INPP5D*, and *IL6*R on the pseudotime analysis of the myeloid lineage (Fig. S11). Our analysis revealed that all three candidate genes were strongly expressed in classical monocytes (Fig. S11), the most abundant myeloid cell type found in the circulating blood.

Lastly, to identify MRs affected by our genetic risk loci, we intersected the 50 most activated and 50 most inactivated IgAV MRs with the genes co-modulated by *FCAR*, *INPP5D*, and *IL6R* transcripts. This intersection, which represents convergent statistical evidence between MR and MINDy analyses, nominated 28 high-priority MRs ranked based on their MR activity from the most to the least active (Table S15). The targets of these MRs were significantly enriched in key immune-related pathways, including cytokine, chemokine, and nucleotide-binding oligomerization domain (NOD)-like receptor signaling pathways (Fig. 5E).

Validation of the prioritized master regulators

To replicate activation of the 28 prioritized MRs in IgAV patients, we performed independent validation using bulk blood RNA-seq data from the CureGN (Cure Glomerulonephropathy) cohort[52]. We first generated a differential gene expression signature adjusted for age, sex, ancestry, and cell fractions for 123 CureGN patients with IgAV against 1,300 glomerular disease controls (450 with focal segmental glomerulosclerosis, 408 with minimal-change disease, and 442 with membranous nephropathy). The NaRnEA algorithm[49] was then used to estimate the activity of the 28 MRs based on the myeloid-specific regulons. In total, 26 of the 28 (93%) MRs had significantly increased activity in IgAV patients compared to disease controls (Fig. 6A, B), providing independent replication of our initial observations.

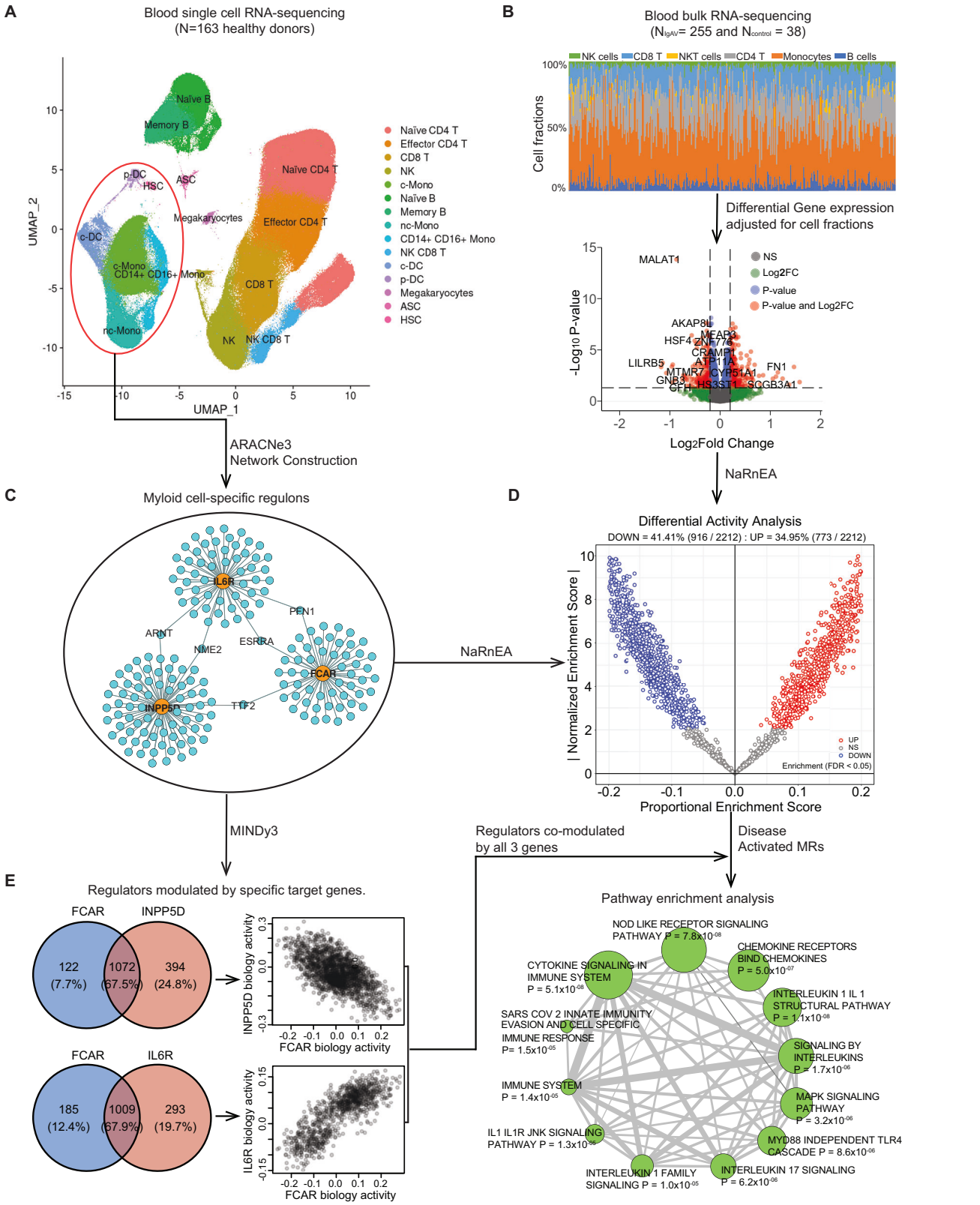
Because the prioritized MRs were predicted to be modulated by FcαR signaling, we additionally examined their activity in the monocyte cell line U937 after in vitro FcαR stimulation with polymeric IgA1 vs. control, in the presence and absence of SHIP-1 activator [AQX-1125]. The cells were collected for each experimental condition, and their transcriptomes were sequenced 60 min after stimulation (see "Methods"). Using NaRnEA[49], we inferred MR activities based on differential gene expression signatures and myeloid-specific regulons (Fig. 5C). In cells stimulated with polymeric IgA1, 22 of the 28 (79%) prioritized MRs were significantly activated compared to controls. When AQX-1125 was added alongside polymeric IgA1, 18 of these 22 MR activities (82%) were suppressed compared to polymeric IgA1 alone (Fig. 6C, D). Thus, our in vitro experiments confirm that the predicted disease master regulator activities are modulated by FcαR and SHIP-1 signaling.

Discussion

By combining hypothesis-free genome, transcriptome, and proteome-wide association approaches with gene expression and splice QTL analyses and network-based analyses, our study uncovered genetic mechanisms of IgAV. We fine-mapped *HLA-DRB1*, *FCAR*, *INPP5D*, and *IL6R* as the candidate causal genes for IgAV and characterized their effects on the disease-related regulons in myeloid cells. We also demonstrated shared genetic susceptibility between IgAV and IgAN and extensively characterized the shared and distinct susceptibility loci in combined and separate analyses of both conditions.

Our fine-mapping studies revealed that the HLA signal was largely explained by the polymorphic position 11 in DRβ1, which is in tight LD with additional substitutions at positions 13 and 96. Our haplotype-based analyses defined a protective haplotype (P11-R13-Q96) and two risk haplotypes (L11-F13-E96 and V11-H13-Y96). While positions 11 and 13 mapped to the surface of pockets 4 and 6 of peptide binding groove of DRβ1, position 96 may be involved in the interaction between DRβ1 and HLA-DM/DO, potentially affecting antigenic peptide loading to DRβ1. However, experimental studies are needed to confirm antigen presentation defects suggested by these associations.

https://doi.org/10.1038/s41467-026-78026-6



Notably, the valine risk substitution at DRβ1 position 11 tags the *DRB1*04* classical allele, while the protective proline substitution tags the *DRB1*1501*. The risk allele *DRB1*04* has previously been associated with increased risk of rheumatoid arthritis[53] and giant cell arteritis[54], while the protective DRB1*1501 allele has previously been reported as protective from type 1 diabetes[55]. The same two alleles also have a direction-consistent effect in IgA nephropathy[17]. Shared association of DRB1 alleles across clinically distinct diseases likely reflects a common upstream role of HLA class II molecules in antigen presentation and CD4+ T-cell activation rather than disease-specific biology. In this model, HLA-DRB1*04 may predispose to pathogenic adaptive immune responses across multiple conditions,

https://doi.org/10.1038/s41467-026-78026-6

Fig. 5 | Master regulators of IgA vasculitis modulated by the GWAS candidate genes. A The UMAP plot for the reference single-cell RNA-seq data from 163 healthy individuals. The blood cells were clustered and annotated into 15 different immune cell types highlighted in different colors. **B** Deconvolution and differential gene expression analysis based on blood bulk RNA-seq data from 255 IgAV cases and 38 health controls. The upper panel shows the distribution of different cell-type fractions across all 293 individuals. The volcano plot (lower panel: x-axis denotes the log2 Fold Change and the y-axis indicates the −log10 (*P*-values) (two-sided; multiple-testing adjusted) with differentially expressed genes in red ($P_{adjusted}$ < 0.05 and log2 Fold Change > 0.2) after adjusting for cell-type fractions. **C** The myeloid regulons reconstructed transcriptome-wide based on the myeloid lineage cells (circled in red above) using ARACNe3 software (a subset of only 3 regulons shown to illustrate this step). **D** Differential activity analysis based on the myeloid regulons and cell fraction-adjusted differential gene expression signature in IgAV cases compared to controls. The x-axis indicates the proportional enrichment score (effect size for gene set enrichment that was normalized by set's size, parameterization, and gene expression signature). The y-axis shows the normalized score compared to the expectation under the null hypothesis. The red and blue colors indicate the regulators with higher and lower activity, respectively, in IgAV cases compared to healthy controls. **E** Regulators co-modulated by the GWAS candidate genes: 1072 regulators co-modulated by *FCAR* and *INPP5D* with mostly opposed effects on their biological activities (upper panel) and 1009 regulators co-modulated by *FCAR* and *IL6R* with direction-consistent effects on their biological activities (lower panel). The targets of the top master regulators co-modulated by all 3 genes were enriched in multiple immune-related pathways based on FDR < 0.05 (depicted as enrichment map with a node size proportional to the statistical significance of enrichment and edge thickness proportional to gene set overlaps). The unadjusted *P*-values were calculated from one-sided tests. The source data are provided in Data S11, Data S12, Data S13, and Data S14, and Table S14.

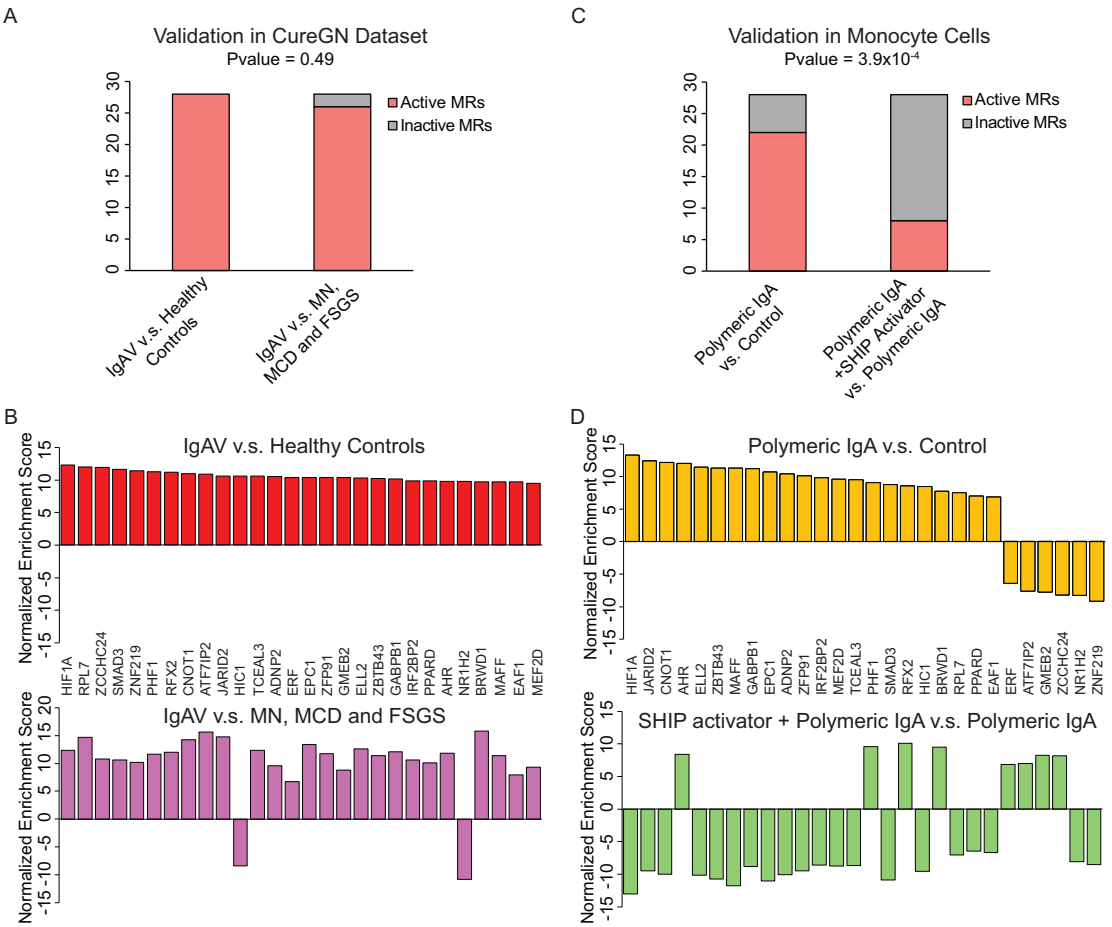


Fig. 6 | Validation of IgA vasculitis (IgAV) master regulators. A Replication in the CureGN study: the bar plots showing that 26 of the 28 (93%) prioritized disease MRs exhibited increased activity in the CureGN replication cohort (123 cases and 1300 disease controls) compared to the discovery cohort; the *P*-value is unadjusted and based on a one-sided test. **B** The top panel depicts the ranking of the 28 active MRs by their normalized enrichment score (NES) based on the original comparison of 255 cases and 38 healthy controls (discovery), while the bottom panel shows the corresponding NES based on the CureGN cohort (replication). **C** Validations by in vitro studies: the bar plots depicting 22 out of the 28 MRs were activated by polymeric IgA1 stimulation of U937 monocytes, and 22 of the 28 MRs' activities were reversed by adding the SHIP-1 activator alongside polymeric IgA1. The *P*-value is unadjusted and based on a one-sided test. **D** The top panel depicts the ranking of the 28 MRs by their NES based on the comparison of the U937 monocyte cell line stimulated with polymeric IgA1compared to control, while the bottom panel shows the corresponding NES after exposure to polymeric IgA1 with AQX-1125 compared to polymeric IgA1 alone; 22 of 28 MRs demonstrated the expected reversed pattern of activation with AQX-1125 treatment.

whereas HLA-DRB1*15:01 may be protective by modifying peptide binding or promoting more tolerogenic responses. The eventual disease phenotype is then shaped by additional non-HLA genetic factors, environmental exposures, and tissue-specific inflammatory context.

The two novel non-HLA risk loci, *FCAR* and *INPP5D*, converge on the same pathway in myeloid cells, highlighting the central role of the FcαR (CD89) signaling in the pathogenesis of IgAV. In prior studies, elevated serum levels of IgA1-CD89 complexes have been reported in children with IgAV[56] and IgAN[57]. However, our results demonstrate

distinct genetic mechanisms underlying genetic associations of the *FCAR* locus in IgAV versus IgAN. In IgAV, the risk alleles disrupt a binding motif of *PRDM1* in a monocyte-specific *FCAR* enhancer and are associated with increased *FCAR* gene expression. In IgAN, the risk alleles are associated with a lower rate of a splicing event that excludes a transmembrane domain of FcαR, leading to increased expression of the full-length isoform.

In contrast, the risk variant in *INPP5D* encoding SHIP-1, a known negative regulator of *FCAR* signaling[26,27], had the same effect in IgAV and IgAN. The inhibition of SHIP-1 is known to generate molecular profiles consistent with activation of inflammasome[58] and increased phagocytosis[59,60]. We hypothesize that p.His1169Tyr substitution may disrupt the function of SHIP-1, but further studies are needed to understand the precise effect of this variant.

We also demonstrated that the PWAS approach could effectively implicate a candidate causal gene when GWAS by itself is underpowered to establish a significant association. The PWAS signal for soluble IL6R in IgAV was supported by a colocalization between GWAS and protein QTL. About 85% of circulating sIL6R is generated by proteolysis[61], and one of the top protein QTL variants was p.Asp358Ala, a substitution mapping to the proteolytic cleavage site in the stalk region of IL6R[62], and previously associated with elevated serum levels of sIL6R[63] and reduced membrane-bound IL6R[64,65]. Once cleaved, sIL6R can interact with circulating IL6 and activate gp130, a signal transducer expressed on multiple cell types, resulting in systemic pro-inflammatory effects[63–65]. In aggregate, these results suggest a pathogenic role of sIL6R, potentially providing an opportunity to repurpose Tocilizumab, a monoclonal antibody blocking IL6R, for treatment of IgAV. The association of sIL6R with IgAV will require independent validation, and properly designed clinical studies are needed to test the effects of Tocilizumab in children with IgAV.

Our study adds to the growing evidence implicating myeloid cells in the pathogenesis of IgAV. IgA-containing immune complexes deposited in small vessels have been linked to activation of FcαR/CD89 on myeloid cells, promoting neutrophil adhesion, release of inflammatory mediators, and NETosis, thereby providing a plausible mechanistic bridge between systemic IgA dysregulation and tissue injury[66–69]. Within this framework, genetic prioritization of *FCAR*, *INPP5D*, and *IL6R* is biologically consistent with the current model of IgAV pathogenesis. *FCAR* represents the most direct mechanistic candidate, as it encodes the canonical myeloid IgA receptor through which pathogenic IgA complexes drive myeloid cell activation and endothelial injury[68], whereas SHIP1 is more plausibly interpreted as a negative regulator of Fc receptor signaling, and IL6R as a mediator of inflammatory propagation once the response has been initiated. These observations support a model in which FcαR-mediated recognition of IgA immune complexes contributes to myeloid activation, while SHIP1- and IL6R-dependent pathways modulate the magnitude and persistence of this response.

The effects of our GWAS candidate genes may also differ by tissue compartment: *IL6R* may be more relevant to systemic immune activation, consistent with the association of the IgAV-risk allele with higher genetically predicted plasma IL6R and a broader role in inflammatory amplification, whereas *FCAR* may be particularly important at the local tissue level, including in the kidney and inflamed vascular beds, where deposited IgA immune complexes could engage infiltrating or marginating myeloid cells and promote FcαRI-dependent vascular injury[66,68]. SHIP1 effects are likely to occur within activated or tissue-infiltrating myeloid cells, where altered signaling thresholds could influence the intensity of Fc receptor-driven inflammatory output. However, mechanistic studies in relevant human tissues and experimental models are needed to confirm these proposed gene- and compartment-specific effects.

Our methods for reverse engineering of myeloid-specific regulatory networks using single-cell transcriptomics allowed us to define the key transcriptional regulators of IgAV. To differentiate regulatory disease drivers from reactive changes, we prioritized differentially active regulators that were also co-modulated by the transcripts supported by genetic evidence. Among the 50 most differentially activated and 50 most inactivated regulators, 28 were significantly co-modulated by *FCAR*, *INPP5D*, and *IL6R*, and their targets were significantly enriched in key immune-related pathways. These regulators appear to drive the disease process and thus may represent attractive therapeutic targets.

Lastly, our cross-phenotype GWAS highlights shared genetic susceptibility between IgAV and IgAN, including strong genome-wide genetic correlation. Most of the known IgAN-risk loci were associated with a higher risk of IgAV, including the *TNFSF13* locus encoding APRIL. While this manuscript was under review, the FDA granted accelerated approval for the first APRIL inhibitor for the treatment of IgAN[70]. Our results provide genetic support for APRIL's involvement also in IgAV, highlighting another potential drug repurposing opportunity. Despite these commonalities, clear differences in the disease associations were detected at the *FCAR* and *IL6R* loci, and these differences could potentially contribute to the unique systemic manifestations of IgAV.

Several important limitations of our study need to be acknowledged. First, for consistency across cohorts, we used the 1000 Genomes reference panel for imputation, which may limit our genomic resolution. Second, resolving the most likely causal alleles at the HLA locus is challenging due to extended LD across the HLA region. Specifically, conditioning on multiple *DRB1* classical alleles and/or multiple amino-acid substitutions may potentially mask other causal alleles that reside on the same haplotypes. Third, our TWAS and PWAS models were derived based on European datasets, and these models may be less powerful when applied to a meta-analysis involving both East Asian and European cohorts. Our TWAS models are additionally limited by the modest sample size of transcriptomic reference datasets and the use of bulk RNA-seq that may obscure signals from relevant cell types. Fourth, our network analysis combines all blood cells of myeloid lineage to maximize power for edge detection, and it does not resolve specific myeloid sub-populations or kidney or skin-resident myeloid cells. Moreover, our network analysis does not directly incorporate genotype information from blood cell donors. Higher resolution datasets that combine genotypes with larger single-cell transcriptomic reference panels from blood, kidney, and other disease-relevant tissues would improve cell-type resolution and empower direct testing of GWAS variants on transcriptomic networks. Lastly, our GWAS discovery cohorts are modest in size and include individuals of only East Asian and European ancestries. Additional studies are still needed in larger and more ancestrally diverse cohorts to fully understand the human genetic architecture of IgA vasculitis.

Methods

Ethics

The design and conduct of the study abided by the Declaration of Helsinki and complied with all local regulations regarding the use of human participants in genetic studies. The study was approved by the Columbia IRB under protocol number AAAC7385; this umbrella protocol included coordination of data and bio-sample collections across all GIGA-kids study sites, as well as generation and analysis of genetic and transcriptomic datasets. Additionally, each individual GIGA-kids recruitment site obtained a site-specific IRB approval for participant recruitment for this study. The genetic studies of the Beijing Cohort were approved by the Ethics Committee of Peking University First Hospital (approval number 2019-03). Parental/guardian consent was provided on behalf of any participating children.

Cohort descriptions

We performed a GWAS meta-analysis of three large case-control cohorts comprising a total of 8098 individuals (2170 IgAV cases and

5928 controls). The study cohorts are summarized in Table S1 with detailed cohort descriptions below.

Beijing cohort

The cases and controls were recruited from the clinics of Peking University First Hospital, Beijing, China. We used standard diagnostic criteria for IgAV that included a palpable purpura in association with at least one of the following: diffuse abdominal pain, arthritis or arthralgia, renal involvement (hematuria and/or proteinuria), and/or IgA deposition in a biopsy specimen (skin, intestinal tract, or kidney). The controls had no history of purpura or kidney disease. All participants provided informed consent to participate in genetic studies. Genomic DNA was extracted from whole-blood samples using standard procedures, and genotyping was performed with Illumina Omni 2.5, Omni 2.5 Exome, and Omni ZhongHua chips. The analysis of intensity clusters and genotype calls was performed in Illumina Genome Studio software 2.0. All SNPs were called on forward DNA strand, and the overlapping SNPs across the three platforms were extracted for further analysis. The standard quality control (QC) included per-SNP genotyping rate >95%, per-individual genotyping rate >90%, MAF > 0.01, and HWE test $P > 1 \times 10^{-5}$ in controls. We excluded duplicates and cryptically related individuals based on the estimated pairwise kinship coefficients >0.05 using KING software[71]. Gender of each sample was imputed based on the analysis of sex chromosome markers using PLINK v1.9[72], and individuals with gender mismatch against study records were excluded. The ancestry was evaluated using principal component analyses (PCA)[13] using 46,494 high-quality (genotype rate > 99%), common (MAF > 1%), and independent (pairwise $r^2 < 0.05$) markers. The regions of high LD were excluded from this analysis (see below). Ancestry outliers were excluded based on visual inspection of PCA plots. The final PCA produced only two significant PCs by the Tracy-Widom test and a well-matched distribution of cases and controls (Fig. S1A). After QC, the final dataset comprised 3900 individuals (1145 cases and 2755 controls) genotyped with 722,465 markers with an overall genotyping rate of 0.99. The cases were 55.3% males and 44.7% females, with an age range of 6–78 years (mean age of 30 years). We carried out imputation analysis with Minimac3[73] and pre-phasing with Eagle v2.4.1[74] using East Asian reference populations from the 1000 Genomes Project (Phase 3). A total of 6,210,008 high-quality common markers (imputation $R^2 > 0.8$ and MAF > 0.01) were imputed and used in downstream analyses.

GIGA-kids (genomics of IgA-related disorders in kids) cohort

This cohort included a multicenter GIGA-kids Study, as well as participants recruited by the Columbia University CKD Biobank and additional cases and controls referred through the GIGA-Europe network including collaborating sites in Italy, Poland, and France. The IgAV diagnostic criteria were the same as for the Beijing cohort and included a palpable purpura in association with at least one of the following: diffuse abdominal pain, arthritis or arthralgia, renal involvement (hematuria and/or proteinuria), and IgA deposition in a biopsy specimen (skin, intestinal tract, or kidney mesangium). The controls had no history of purpura or kidney disease. All participants provided informed consent to participate in genetic studies. DNA of whole-blood samples was extracted using standard procedures and genotyped using Illumina MEGA chips. The analysis of intensity clusters and genotype calls was performed using Illumina Genome Studio 2.0; all SNPs were called on forward DNA strand. Standard QC filters included per-SNP genotyping rate >95%, per-individual genotyping rate >90%, MAF > 0.01, and HWE test $P > 1 \times 10^{-5}$ in controls as implemented in PLINK v1.9[72]. We excluded duplicates and cryptically related individuals with pairwise kinship coefficients >0.05[71]. The final post-QC dataset consisted of 2908 individuals genotyped for 515,946 markers with an average genotyping rate of 99.9%. This cohort was further divided into two non-overlapping sub-cohorts: IgAV without nephritis sub-cohort of 316 cases and 924 controls and IgAV with nephritis (IgAVN) sub-cohort of 425 cases and 1243 controls. The cases were 58.8% male and 41.1% female, with an age range of 0–84 years (mean age of 18 years). For ancestry principal component analysis (PCA)[13], we used 153,078 high-quality independent (pairwise $r^2 < 0.05$) SNPs. The PCA produced 7 significant PCs in each case-control sub-cohort, and the visual inspection of PC plots confirmed comparable genetic ancestry between cases and controls without clear outliers (Figs. S1B and 1C). The imputation analysis was carried out with Minimac3[73] after pre-phasing in Eagle V2.34[74], and using 1000 Genomes Europeans (Phase 3) population reference. A total of 7,192,202 and 7,200,332 common high-quality markers (imputation $R^2 > 0.8$ and MAF > 0.01) were used in downstream analyses of IgAV and IgAVN sub-cohorts, respectively.

Spanish IgAV cohort

This cohort has been described in detail before[7]. Briefly, the cohort comprised 288 cases (49.0% males and 51% females) and 1596 controls. The participants were recruited in Spain and were of European ancestry. All participants provided informed consent to participate in genetic studies. The age at disease onset ranges from 5 to 18 years, with a mean age of 7 years. Genomic DNA was extracted from whole blood using standard methods and genotyped using Infinium HumanCore BeadChip (Illumina, Inc). The standard QC was performed as described before, followed by imputation using the 1000 Genomes Europeans (Phase 3) population reference. The final dataset included 5,662,805 high-quality markers (imputation $R^2 > 0.8$ and MAF > 0.01). The PCA produced 6 significant PCs, and visual inspection of PC plots confirmed comparable ancestry between cases and controls without ancestry outliers (Fig. S1D).

Genome-wide association studies (GWAS)

Within each cohort, primary association scans were performed using logistic regression under additive coding of dosage genotypes, and with adjustment for cohort-specific significant principal components of ancestry using PLINK v1.9[72]. No substantial genomic inflation was observed in any individual cohort (lambda < 1.05 consistently for each individual cohort). Subsequently, a fixed effects meta-analysis was performed to combine the results across all cohorts using METAL[75]. Genome-wide distribution of *P*-values was examined visually using quantile-quantile plots for the combined analysis, and the genomic inflation factor was estimated at 1.02 for the overall meta-analysis (Fig. S2). To declare genome-wide significance, we used the generally accepted *P*-value threshold of 5.0×10^{-8}, while signals with $P < 1.0 \times 10^{-6}$ were considered as suggestive. For sensitivity analysis, we also performed a trans-ethnic meta-analysis under a random effects model using TransMeta software[76], but this analysis did not change our results or the number of significant loci. To detect independent signals at genome-wide significant loci, we also performed stepwise conditional analyses of each locus in each cohort using GCTA[77], and the conditioned summary statistics were combined across cohorts using fixed effects meta-analysis[75].

Transcriptome-wide association studies (TWAS)

TWAS was performed following the FUSION workflow[78] based on weights developed from the European ancestry subset of the GTEx v8 tissues that were considered relevant to IgAV (kidney cortex, liver, spleen, whole blood, small intestine-terminal ileum, and EBV-transformed lymphocyte cells). Prediction models were based on elastic net modeling[78]. Multiple testing was accounted for by a Bonferroni adjustment for 17,126 unique number of protein-coding transcripts modeled across tissues ($P < 0.05/17{,}126 = 2.9 \times 10^{-6}$).

Proteome-wide association study (PWAS)

For PWAS, we adapted the FUSION workflow[78] using plasma protein weights from the European ancestry participants of the

Atherosclerosis Risk in Communities Study (ARIC) study[79]. These weights were based on plasma proteomic quantifications using Somascan V.4.1. Prediction models were based on elastic net modeling[78]. Proteome-wide significance level was defined using a Bonferroni correction for the total number of circulating proteins tested ($P < 0.05/884 = 5.7 \times 10^{-5}$).

Functional annotations of significant loci

We first defined each GWAS locus by ±400 kb of the genome-wide significant index SNP. We annotated all transcripts within these intervals using hg19 to create sets of positional candidate genes for each locus. Using ANNOVAR software[80], we annotated all variants within the region that were in LD ($r^2 > 0.5$) with the top SNP, including all known coding, splice, and 3′UTR and 5′UTR variants. We then performed FUN-LDA analysis to estimate the posterior probability of these variants being functional across various cell types[38]. Lastly, the variants were intersected with the predicted enhancers and mapped to regulated genes using the Activity-by-Contact (ABC) model[37]. The myeloid-specific regulatory regions were confirmed by the analysis of monocyte ATAC-seq peaks before and after LPS treatment[39]. The top prioritized variants were screened against transcription factor binding motifs using JASPAR[40]. To additionally prioritize target genes, we performed locus colocalization analyses with whole-blood gene expression and splicing QTLs quantified from bulk RNA-seq of 255 IgAV cases. After harmonization of effect alleles, we identified all co-localized QTLs mapping to the region of the index SNP ± 400 kb using Coloc package in R[81]. For the *IL6R* locus supported by PWAS, we additionally performed colocalization with pQTL summary statistics from participants of the ARIC study[79]. Because the pQTL signal exhibited seven conditionally independent signals (defined by the COJO-Slct algorithm[77]), we computed conditional pQTL summary statistics by conditioning on all other independently significant SNPs using COJO-Cond[77]. Subsequently, colocalization analyses were conducted for all pairwise combinations of the GWAS and conditionally independent pQTL summary statistics using an adapted version of the Giambartolomei colocalization method as implemented in the 'coloc.fast' function with default parameters[81]. Any colocalization with PP4 > 0.8 was considered as strong evidence in support of a shared causal variant. Only one of the seven conditionally independent sIL6R cis-pQTL signals co-localized with the IL6R GWAS and eQTL loci.

HLA imputation and testing

We performed imputation of classical HLA alleles at two- and four-digit resolution, as well as individual amino-acid polymorphisms at class I (*HLA-A*, *-B*, and *-C*) and class II (*HLA-DQB1*, *-DQA1*, and *-DRB1*) loci using SNP2HLA software[82]. The European and East Asian cohorts were imputed separately, using ethnicity-specific reference panels. For European reference, we used the pre-phased HLA reference dataset generated by the Type 1 Diabetes Genetics Consortium (T1DGC, 5225 individuals)[82]. For our East Asian cohorts, we used the Pan-Asian HLA Reference Panel (268 individuals)[83]. In the association analyses, we included only common HLA alleles (MAF > 0.01) that were imputed with high certainty ($R^2 > 0.8$). Because frequencies of HLA alleles are known to vary by ancestry, the association testing was performed separately in European and East Asian cohorts. We used logistic regression models to test the additive effects of HLA allele dosages with adjustment for significant PCs of ancestry. For multi-allelic loci, we used the following logistic regression model:

$$\log(odds_i) = \beta_0 + \sum_{j=1}^{m-1} \beta_j x_{j,i} + \sum_{k=1}^{n} \beta_k P_{k,i} \quad (1)$$

where m denotes a total number of alleles at a specific multi-allelic locus, j indicates a specific allele being tested, and $x_{j,i}$ is the imputed dosage for allele j for individual i; β_0 represents the intercept and β_j represents the additive effect of an allele j; $P_{k,i}$ denotes the value for k^{th} 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. We compared log-likelihoods of the two tested models: the full model containing the test locus and relevant covariates with the reduced model without the test locus, but with the same set of covariates. The deviance was defined as $-2 \times$ log likelihood ratio, which follows a X^2-distribution with $m - 1$ degrees of freedom, from which we calculated *P*-values. In addition to multi-allelic tests, we also performed bi-allelic tests of association for all imputed SNPs, classical HLA alleles, and individual amino acid residues in HLA molecules. All analyses were performed using the dosage method under additive coding. Stepwise conditioning analyses across the HLA region were performed using both multi-allelic and bi-allelic coding of HLA variants. In each round of stepwise conditioning, we first included the most significant variant as the covariate in the logistic regression model. If additional independently associated markers were detected, they were included as covariates in subsequent models. We repeated these analyses until no residual associations across the entire locus were observed.

Analysis of polymorphic amino-acid sites in HLA genes

To test the effects of individual amino acid substitution sites within the three significant *HLA* genes (*HLA-DRB1*, *HLA-DQA1*, and *HLA-DQB1*), we applied a conditional haplotype analysis using fully phased haplotypes. 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 k additional unique haplotypes were generated and tested over the null model using the likelihood ratio test with k 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 statistically significant positions were observed.

Protein structure modeling

Models of DRα1 and DRβ1 were generated based on PDB ID: 5V4M[84,85], which harbored protective residues P11 and R13. Risk variants L11/F13 and V11/H13 were generated with COOT[86]. After removing the peptide, all three models were subjected to molecular dynamics minimization protocols with YASARA[87]. Images of DRα1 and DRβ1structures were generated with ChimeraX[88]. Modeling of the IL6R-ADAM17 peptide structures followed previously published methods[47]. Briefly, ADAM17 and N-terminal residues (CTCSPSH) of the bound TIMP3 were extracted from PDB ID: 3CKI[89]. Peptide residues were mutated to 355-PVQDSSS-361 or 355-PVQASSS-361 of the IL6R stalk region, and complexes were minimized in YASARA[87]. Models of HLA-DM-HLA-DR1 were generated with PDB ID: 4GBX[90]. The parent coordinates have glutamic acid at residue 96. Variants Q96 and Y96 were generated in COOT[86] and minimized in YASARA[87].

Meta-phenome-wide association study (Meta-PheWAS)

The meta-PheWAS for individual variants was performed in 786,138 participants across three biobanks: the UK Biobank (UKBB, *N* = 460,363), Electronic Medical Records and Genomics-III (eMERGE-III, *N* = 102,138), and All-of-Us (AoU, *N* = 223,637). For meta-analysis, we harmonized coded diagnoses across biobanks by converting all available ICD-10 codes to the ICD-9-CM system. After conversions, we identified 1816 unique phecodes. These phecodes were then tested using logistic regression adjusted for age, sex, study site, imputation batch, and 3 PCs of ancestry using the PheWAS R package[91]. Meta-PheWAS was performed using metagen software under fixed effects

using imputed genotype data for each biobank. For phenome-wide significance, we used the Bonferroni-corrected $P < 2.75 \times 10\text{-}5$ to account for 1816 independent phecodes tested.

Genetic correlations

To investigate the shared genetic architecture between IgAV and IgAN, we first collected the summary statistics from the latest IgAN GWAS (10,146 cases and 28,751 controls)[17]. The genetic correlations were estimated using bivariate LD score regression[92] with ancestry-matched 1000 Genome Project Phase 3 used for LD reference.

Blood RNA sequencing

Blood samples were collected using PAXGene RNA tubes from the participants of the GIGA-kids study. To ensure complete lysis of blood cells and stabilization of RNA, the collected blood samples were incubated at room temperature for a minimum of 2 h. The stabilized blood samples were further centrifuged to pellet cellular debris. Supernatant of the samples was lysed, digested, and homogenized using QuickGene RNA blood cell kit S (Catalog #: RB-S). Automated RNA extraction of lysate was done using the QuickGene 810 Nucleic Acid Isolation System. RNA samples were purified with the Zymo Research RNA Clean & Concentrator-5 kit (Catalog #: R1014). We use poly-A pull-down to enrich mRNAs from extracted total RNA samples with RIN > 8 (200 ng–1 μg per sample) and performed library preparation using Illumina TruSeq RNA prep kit. Libraries were then sequenced using Illumina HiSeq2500/HiSeq4000 at Columbia University Genome Center. Sequencing generated 100 bp single/pair-end reads, with a median achieved total reads per sample at 23.5 M. Raw sequencing data were processed using RTA (Illumina) for base calling and bcl2fastq2 (version 2.17) for converting BCL to fastq format, coupled with adapter trimming. RNA-seq data were aligned to the human reference genome hg38 with STAR software[93]. Following GTEx pipelines[94], we identified and excluded low-quality samples with <10 million mapped reads, read mapping rate <0.2, intergenic mapping rate >0.3, base mismatch rate >0.01, and rRNA mapping rate >0.3. The outlier samples were further identified based on expression profile using correlation-based statistics and sex incompatibility checks following published methods[95]. After individual sample QC, we quantified the gene expression across samples using RNA-SeQC[96] as read counts and transcripts per million (TPM). Only transcripts with TPM > 0.1 and 6 read counts in at least 20% of samples were used for downstream analyses. The final post-QC dataset included 294 individuals (255 IgAV cases and 38 healthy controls) and 17,790 transcripts.

Cell-type deconvolution and differential gene expression analyses

We used CIBERSOFTx[48] to deconvolve immune cell type abundance from bulk blood RNA-seq of 255 IgAV cases and 38 controls. We optimized this pipeline for 6 different major immune cell types, including CD4 and CD8 T-cells, NK cells, monocytes, B cells, and dendritic cells, using published reference panels generated by single-cell RNA sequencing of 25,000 blood cells from 45 healthy donors[97]. The deconvolution was conducted based on the default parameters. For differential gene expression, we used DEseq2 R package[98]. This analysis was adjusted for age, sex, ancestry PCs, and deconvoluted cell-type fractions.

Expression QTL (eQTL) mapping

For eQTL mapping, gene counts were normalized across samples using Trimmed Mean of M-values (TMM) method[99] across 255 IgAV cases, and the expression values for each gene were inverse-normal transformed. The cis-eQTLs were identified using linear regression as implemented in tensorQTL[100], adjusting for age, sex, top 5 ancestry PCs, and 45 PEER factors[101]. A false discovery rate (FDR) threshold ≤0.05 was applied to identify genes with at least one significant cis-eQTL (eGenes). To identify the list of all significant variant-gene pairs associated with cis-eGenes, a genome-wide empirical *P*-value threshold was defined as the empirical *P*-value of the gene closest to the 0.05 FDR threshold[102]. For each gene, variants with a nominal *P*-value below the gene-level threshold were considered significant and included in the final list of variant-gene pairs.

Splicing QTL (sQTL) mapping

The RNA-seq data were re-aligned to the hg38 reference genome using the --WASP parameter with STAR to avoid potential artifacts caused by allelic mapping bias in sQTL analyses. Splicing quantification was performed based on the intron excision levels across samples computed by LeafCutter[103]. Introns with few counts or with insufficient variability across samples were removed as described previously[94]. The filtered counts were normalized using the LeafCutter[103], and the resulting splicing quantification files were used for sQTL mapping. Mapping of cis-sQTLs was performed in 255 IgAV cases using tensorQTL[100] controlling for age, sex, top 5 ancestry PCs, and 15 PEER factors[104]. To identify cis-sGenes, FDR was calculated in the same manner as for cis-eQTL, and an sGene-level nominal *P*-value threshold was determined for all significant variant-intron excision pairs.

Single-cell sequencing and data processing

The reference blood single-cell RNA-seq dataset was constructed using samples of volunteers recruited by the Reference Ability Neural Network (RANN) study through random mailing in Northern Manhattan[105]. Briefly, Single-cell RNA-seq libraries of peripheral blood mononuclear cells (PBMC) were constructed using the 10X Genomics Chromium Next GEM Single Cell 3' Reagent Kits[106]. An estimated 20,000 cells pooled from 3–4 donors were loaded on the 10X Genomics Chromium instrument. A total of 41 sequencing libraries (163 donors) were pooled and sequenced on the Illumina NovaSeq 6000 with S4 flow cell (Illumina, San Diego, CA) using paired-end, dual-index sequencing with 28 cycles for read 1, 10 cycles for the i7 index, 10 cycles for the i5 index, and 90 cycles for read 2. Gene expression profiles were generated for a total of 984,441 cells from 163 donors. FASTQ files were preprocessed using the count command of the CellRanger software (v6.0.0, 10x Genomics) with a human reference transcriptome GRCh38–2020-A. To de-multiplex cells from pooled samples, we took advantage of SNPs in the RNA-seq reads. Cells in the filtered Unique Molecular Identifier (UMI) count matrix were grouped into four clusters (each representing one participant) based on RNA genotypes using the freemuxlet software (downloaded on May 5th, 2021, from https://github.com/statgen/popscle). Cells were excluded if freemuxlet called them as doublets (N_{cells} = 133,117). We further excluded cells with <200 or >4500 transcripts (N_{cells} = 1479), percentage mitochondrial transcripts >10% (N_{cells} = 44,737), and the percentage of ribosomal RNA > 5% (N_{cells} = 12,706). We additionally removed hemoglobin genes (N_{genes} = 12) and genes expressed in ≤3 cells (N_{genes} = 681). After QC, we included 710,654 cells and 26,970 unique transcripts from 163 donors in downstream analyses. Using the Seurat R package, we performed normalization of gene expression across all 710,654 cells using the LogNormalize method, followed by identification of variable transcripts, data scaling, and principal component analysis. We performed cell-type clustering analysis based on the top 15 principal components. We then annotated 16 different cell types based on known blood cell markers[97], including HSC (hematopoietic stem cells), ASC (antibody-secreting cells), Megakaryocytes, p-DC (plasmacytoid dendritic cell), c-DC (conventional dendritic cells), NK (natural killer) cells, NK CD8 T-cells, transitional or precursor cells, CD14 + CD16+ (classical) monocytes, nonclassical monocytes, memory B cells, naïve B cells, effector CD4 T-cells as well as Naïve CD4 T-cells (Fig. 5A).

Myeloid-specific regulatory network construction

For reverse engineering of regulatory networks in myeloid cells, we used ARACNe3 (Algorithm for the Reconstruction of Accurate Cellular Networks version 3). This is an algorithm based on mutual information (MI) that identifies direct regulatory interactions mediated by transcription factors[49,107]. Using ARACNe3, we reconstructed a transcriptome-wide regulatory network based on the myeloid cluster transcriptome generated using our reference blood single-cell RNA-seq data from 163 donors. Briefly, after QC filtering of poor-quality cells, we used 143,192 myeloid cells with UMI/cell ≥3000 across all 163 samples. We then excluded any donors with less than 500 high-quality myeloid cells, resulting in a total of 111 donors for network analysis. To address the sparsity of the single-cell expression matrix, we created a pseudo-bulk expression matrix by aggregating UMI counts across all myeloid cells (Fig. S11). The ARACNe3 network was inferred using this pseudo-bulk matrix based on FDR < 0.05. A total of 7 iterations were performed by randomly selecting 63.2% of all samples to achieve >50 targets per regulator. The final myeloid regulatory network involved 26,196 genes and 2,212 regulators based on the Gene Oncology (GO) annotations[107].

Master regulator (MR) analysis

We used NaRnEA (Nonparametric analytical Rank-based Enrichment Analysis)[49] to define differential activity of regulators based on the ARACNe-3 inferred myeloid regulons and differential gene expression signatures. NaRnEA is a gene set enrichment method that leverages a null model derived under the principle of maximum entropy. For each transcription factor, NaRnEA tests for significant enrichment of its ARACNe3-inferred regulons in differential gene expression signature compared to randomly selected non-overlapping gene sets. This is important, as master regulators may be differentially active but not necessarily differentially expressed. The direction of effect was derived for each regulator (activated or suppressed in IgAV cases compared to controls), and the statistical significance was defined based on FDR < 0.05. The top 50 most significant activated regulators in cases were defined as "IgAV Master Regulators."

Modulation of master regulators by candidate genes

To test how our GWAS candidate genes (*INPP5D*, *FCAR*, and *IL6R*) co-modulate IgAV Master Regulators, we performed MINDy analysis based on the myeloid regulons. MINDy is an algorithm designed to uncover modulator genes influencing transcriptional interactions. This is accomplished by the analysis of ΔMI (the difference in mutual information between a TF and its target genes) after conditioning on the highest and lowest expression of any candidate modulator. The statistical significance of a particular ΔMI for a given TF-target gene pair and a given modulator was determined based on FDR < 0.05[108].

Pseudotime trajectory analysis

Pseudotime trajectory analysis in myeloid cell population to order cells along a continuous differentiation axis and infer lineage structure was performed using Slingshot (v2.14.0) implemented in R (CRAN). Briefly, single-cell RNA-seq data were preprocessed, including normalization, dimensionality reduction, and cell-type annotation using Seurat package, as described above. Slingshot was then used to infer developmental paths by first representing each cell cluster as a node in a connectivity graph, then linking these clusters with a minimum spanning tree (MST) to capture the most parsimonious relationships among clusters. The lineages were obtained by following the resulting MST from a starting root cluster, resulting in an ordered sequence of clusters along each path. In this analysis, hematopoietic stem cells (HSC) were defined as the root cluster by Slingshot software. Next, for each cell, Slingshot computed pseudotime values using principal curves, as previously described[109].

Validation of master regulators in the CureGN cohort

The CureGN (Cure Glomerulonephropathy) is a prospective cohort of over 2400 patients with a kidney biopsy diagnosis of minimal change disease (MCD), focal segmental glomerulosclerosis (FSGS), membranous nephropathy (MN), IgA nephropathy (IgAN), or IgA vasculitis with nephritis (IgAVN). The study collected longitudinal clinical data and performed whole-genome sequencing (WGS) and bulk blood transcriptome profiling. In our replication analysis, we compared IgAV ($N = 123$) to other disease controls (MCD, MN, and FSGS; $N = 1300$) excluding IgAN. Briefly, RNA was collected in PAXgene Blood RNA tubes and extracted using standard methods. Poly-A pull-down was used to enrich mRNA from total RNA samples with an RIN > 8, followed by library preparation using the Illumina TruSeq RNA Prep Kit. RNA sequencing was performed using the Illumina NovaSeq 6000 platform, generating 100 bp paired-end reads. The RNA-seq processing and QC filtering were performed based on the same pipeline as in the discovery dataset. The final dataset includes 19,437 transcripts across 1423 samples. The differential gene expression signature was generated by DEseq2 R package[98] controlling for age, sex, ancestry PCs, and deconvoluted cell-type fractions. Cell-type deconvolution and master regulator analyses were performed using the same methods as for the discovery dataset described above.

Testing the effects of FcαR (CD89) and SHIP1 activation on MR activity in monocytes

Monocyte cell line U937 (American Type Culture Collection, ATCC, Rockville, MD, USA) was grown in RPMI 1640 medium supplemented with 10% fetal bovine serum and antibiotics[110]. Cell density was maintained between 0.2×10^6 and 2×10^6 cells/ml in T75 or T175 flasks. Expression of CD89 was assessed by flow cytometry with anti-CD89 antibody labeled with PE (Invitrogen cat # MA5-28547, Fig. S12). The CD89-PE antibody was added at a 1:10 dilution with the resuspended cell pellets to reach a F.C. of 4.5 μg/ml. Unstained cells and isotype control antibody (Invitrogen # 12-4714-82) served as negative controls. This pilot experiment confirmed that U937 cells expressed CD89. The experiment was performed in 6-well plates, using 0.5×10^6 cells per ml in 3 ml of RPMI 1640 medium (with 10% fetal bovine serum and antibiotics) in each well. The cells were incubated under the following conditions:

1. Medium (Control)
2. Medium (Control) + SHIP1 activator [AQX-1125] 10 μM final concentration
3. Polymeric IgA1 (Ale) 50 μg/ml (Stimulated)
4. Polymeric IgA1 (Ale) 50 μg/ml (Stimulated) + SHIP1 activator [AQX-1125] 10 μM final concentration

SHIP1 activator [AQX-1125] was obtained from MedChemExpress. Polymeric IgA1 (Ale) was isolated as described before[111]. U937 cells were collected 60 min after stimulation by centrifugation. Each sample had 3 replicates. After centrifugation, the cells were resuspended in RNAlater (Invitrogen cat #AM7020) and processed per the manufacturer's protocol. Total RNA was extracted using the QuickGene RNA Cultured Cell Kit (catalog # FK-RCS2) following the manufacturer's protocol. After extraction, RNA concentration and integrity were assessed using an Agilent 4200 TapeStation. RNA sequencing (STRDPOLYA) was performed at the Columbia University Sulzberger Genome Center.

Reporting summary

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

Data availability

GWAS and QTL summary statistics are freely available for download through the Kiryluk Lab website [https://www.columbiamedicine.org/

divisions/kiryluk/study_gwas_stat.php]. Owing to participant privacy and informed consent restrictions, genetic data are available only through controlled access data repositories. The genotype and phenotype data for the Chinese IgAV cohort are deposited in the Genome Variation Map (GVM) controlled access database maintained by China's National Genomics Data Center (NGDC) under the accession number GVM001521. Genotype, phenotype, and sequence data generated for the Columbia/GIGA-kids cohort can be accessed through the controlled access dbGAP repository maintained by the National Institute of Health (NIH), accession number phs004179.v1.p1 [https://dbgap.ncbi.nlm.nih.gov/study/phs004179.v1.p1/#study]. The CureGN data used for validation analysis are also available through dbGAP, accession number phs002480.v5.p4 [https://dbgap.ncbi.nlm.nih.gov/study/phs002480.v5.p4/#study]. For the Chinese cohort, access will be granted to qualified researchers in accordance with the regulations of NGDC and Peking University. All requests should be directed to Dr. Hong Zhang (email: hongzh@bjmu.edu.cn); all requests will be processed promptly, and applicants can typically expect a reply within two weeks of submission. A data use agreement will be required to download the controlled access data. For GIGA-kids and CureGN cohorts, qualified investigators may obtain access by submitting a data access request to dbGaP [https://dbgap.ncbi.nlm.nih.gov/] in accordance with NIH policies and the NIH Data Access Committee (DAC) requirements. Access may be granted to qualified investigators affiliated with a recognized research institution whose proposed use is consistent with the participants' informed consent and who obtain approval from the appropriate NIH DAC through submission of a Data Access Request and Data Use Certification. Data requests are typically reviewed within approximately two weeks after submission of a complete application. Additionally, publicly available biobank datasets were used for Meta-PheWAS analyses and included: the eMERGE-III cohort (dbGAP accession number phs001584.v2.p2 [https://dbgap.ncbi.nlm.nih.gov/beta/study/phs001584.v2.p2/#study]); the UKBB data (available through the UKBB web portal [https://www.ukbiobank.ac.uk/]), and the All of Us Research Program data (available through the All of Us web portal [https://allofus.nih.gov]). The specific data types used include SNP array-based genotype data and ICD code data. The 460,363 participants from UKBB and 223,637 participants from All of Us, with both SNP genotyping and ICD code data available, were included in the Meta-PheWAS. The 1000 Genomes data used for reference are available publicly through https://www.internationalgenome.org/category/data-access/. The 1000 Genomes (Phase 3) sequence data were used as reference for imputation and genetic correlation analyses. For the Chinese cohort, we carried out imputation analysis using the East Asian reference populations, while for European cohorts, we used European populations from the Phase 3 1000 Genomes Project. The same populations were used for reference in genome-wide genetic correlation analyses. The ATAC-seq data of monocytes are available through Gene Expression Omnibus accession number GSE118189. FUN-LDA scores used for functional variant annotation are available at http://www.columbia.edu/~ii2135/funlda.html. Models of DRα1 and DRβ1 were generated based on PDB structure 5V4M. ADAM17 and N-terminal residues (CTCSPSH) of the bound TIMP3 were extracted from PDB structure 3CKI. Models of HLA-DM-HLA-DR1 were generated based on PDB structure 4GBX.

References

1. Hetland, L. E., Susrud, K. S., Lindahl, K. H. & Bygum, A. Henoch-Schonlein purpura: a literature review. *Acta Derm. Venereol.* **97**, 1160–1166 (2017).
2. Jennette, J. C. et al. 2012 Revised International Chapel Hill Consensus Conference nomenclature of vasculitides. *Arthritis Rheum.* **65**, 1–11 (2013).
3. Sunderkotter, C. H. et al. Nomenclature of cutaneous vasculitis: dermatologic addendum to the 2012 Revised International Chapel Hill Consensus Conference nomenclature of vasculitides. *Arthritis Rheumatol.* **70**, 171–184 (2018).
4. Magistroni, R., D'Agati, V. D., Appel, G. B. & Kiryluk, K. New developments in the genetics, pathogenesis, and therapy of IgA nephropathy. *Kidney Int.* **88**, 974–989 (2015).
5. Watts, R. A. & Scott, D. G. Epidemiology of the vasculitides. *Semin. Respir. Crit. Care Med.* **25**, 455–464 (2004).
6. Lopez-Mejias et al. Association of HLA-B*41:02 with Henoch-Schonlein purpura (IgA vasculitis) in Spanish individuals irrespective of the HLA-DRB1 status. *Arthritis Res. Ther.* **17**, 102 (2015).
7. Lopez-Mejias, R. et al. A genome-wide association study suggests the HLA class II region as the major susceptibility locus for IgA vasculitis. *Sci. Rep.* **7**, 5088 (2017).
8. Ares, O. M. et al. HLA-B and HLA-DRB1 associations with Henoch-Schonlein purpura. *Immunology* **137**, 475–475 (2012).
9. Amoli, M. M. et al. HLA-DRB1*01 association with Henoch-Schonlein purpura in patients from northwest Spain. *J. Rheumatol.* **28**, 1266–1270 (2001).
10. Amoroso, A. et al. Immunogenetics of Henoch-Schoenlein disease. *Eur. J. Immunogenet.* **24**, 323–333 (1997).
11. Chen, J. C. et al. Identification of causal genetic drivers of human disease through systems-level analysis of regulatory networks. *Cell* **159**, 402–414 (2014).
12. Paull, E. O. et al. A modular master regulator landscape controls cancer transcriptional identity. *Cell* **184**, 334–351.e20 (2021).
13. Galinsky, K. J. et al. Fast principal-component analysis reveals convergent evolution of ADH1B in Europe and East Asia. *Am. J. Hum. Genet.* **98**, 456–472 (2016).
14. 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
15. Gamazon, E. R. et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* **47**, 1091–1098 (2015).
16. Brandes, N., Linial, N. & Linial, M. PWAS: proteome-wide association study-linking genes and phenotypes by functional variation in proteins. *Genome Biol.* **21**, 173 (2020).
17. Kiryluk, K. et al. Genome-wide association analyses define pathogenic signaling pathways and prioritize drug targets for IgA nephropathy. *Nat. Genet.* **55**, 1091–1105 (2023).
18. Roche, P. A. & Cresswell, P. Invariant chain association with HLA-DR molecules inhibits immunogenic peptide binding. *Nature* **345**, 615–618 (1990).
19. Ishina, I. A. et al. MHC class II presentation in autoimmunity. *Cells* **12**, 314 (2023).
20. Sloan, V. S. et al. Mediation by HLA-DM of dissociation of peptides from HLA-DR. *Nature* **375**, 802–806 (1995).
21. Kropshofer, H. et al. Editing of the HLA-DR-peptide repertoire by HLA-DM. *EMBO J.* **15**, 6144–6154 (1996).
22. Denzin, L. K., Fallas, J. L., Prendes, M. & Yi, W. Right place, right time, right peptide: DO keeps DM focused. *Immunol. Rev.* **207**, 279–292 (2005).
23. Jensen, P. E. Antigen processing: HLA-DO-a hitchhiking inhibitor of HLA-DM. *Curr. Biol.* **8**, R128–R131 (1998).
24. Guce, A. I. et al. HLA-DO acts as a substrate mimic to inhibit HLA-DM by a competitive mechanism. *Nat. Struct. Mol. Biol.* **20**, 90–98 (2013).
25. Liu, Q. et al. SHIP is a negative regulator of growth factor receptor-mediated PKB/Akt activation and myeloid cell survival. *Genes Dev.* **13**, 786–791 (1999).
26. Park, R. K., Izadi, K. D., Deo, Y. M. & Durden, D. L. Role of Src in the modulation of multiple adaptor proteins in FcalphaRI oxidant signaling. *Blood* **94**, 2112–2120 (1999).
27. Matsui, T., Nunomura, S., Shimokawa, T., Yoshimaru, T. & Ra, C. Functionality of the IgA Fc receptor (FcalphaR, CD89) is down-

regulated by extensive engagement of FcepsilonRI. *Clin. Immunol.* **129**, 155–162 (2008).
28. Kerr, W. G., Park, M. Y., Maubert, M. & Engelman, R. W. SHIP deficiency causes Crohn's disease-like ileitis. *Gut* **60**, 177–188 (2011).
29. Park, M. Y. et al. Impaired T-cell survival promotes mucosal inflammatory disease in SHIP1-deficient mice. *Mucosal Immunol.* **7**, 1429–1439 (2014).
30. Ferreira, M. A. et al. Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. *Nat. Genet.* **49**, 1752–1757 (2017).
31. Johansson, A., Rask-Andersen, M., Karlsson, T. & Ek, W. E. Genome-wide association analysis of 350 000 Caucasians from the UK Biobank identifies novel loci for asthma, hay fever and eczema. *Hum. Mol. Genet.* **28**, 4022–4041 (2019).
32. Chen, M. H. et al. Trans-ethnic and ancestry-specific blood-cell genetics in 746,667 individuals from 5 global populations. *Cell* **182**, 1198–1213.e14 (2020).
33. Astle, W. J. et al. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* **167**, 1415–1429.e19 (2016).
34. Vuckovic, D. et al. The polygenic and monogenic basis of blood traits and diseases. *Cell* **182**, 1214–1231.e11 (2020).
35. Antiga, E. et al. Dermatitis herpetiformis: novel perspectives. *Front. Immunol.* **10**, 1290 (2019).
36. Jostins, L. et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124 (2012).
37. Fulco, C. P. et al. Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 (2019).
38. 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).
39. Calderon, D. et al. Landscape of stimulation-responsive chromatin across diverse human immune cells. *Nat. Genet.* **51**, 1494–1505 (2019).
40. Rauluseviciute, I. et al. JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **52**, D174–D182 (2024).
41. Ulmert, I., Henriques-Oliveira, L., Pereira, C. F. & Lahl, K. Mononuclear phagocyte regulation by the transcription factor Blimp-1 in health and disease. *Immunology* **161**, 303–313 (2020).
42. Kim, S. J. Immunological function of Blimp-1 in dendritic cells and relevance to autoimmune diseases. *Immunol. Res.* **63**, 113–120 (2015).
43. Mullberg, J. et al. The soluble human IL-6 receptor. Mutational characterization of the proteolytic cleavage site. *J. Immunol.* **152**, 4958–4968 (1994).
44. Rafiq, S. et al. A common variant of the interleukin 6 receptor (IL-6r) gene increases IL-6r and IL-6 levels, without other inflammatory effects. *Genes Immun.* **8**, 552–559 (2007).
45. Galicia, J. C. et al. Polymorphisms in the IL-6 receptor (IL-6R) gene: strong evidence that serum levels of soluble IL-6R are genetically influenced. *Genes Immun.* **5**, 513–516 (2004).
46. Flynn, C. M. et al. Cathepsin S provokes interleukin-6 (IL-6) trans-signaling through cleavage of the IL-6 receptor in vitro. *Sci. Rep.* **10**, 21612 (2020).
47. Riethmueller, S. et al. Cleavage site localization differentially controls interleukin-6 receptor proteolysis by ADAM10 and ADAM17. *Sci. Rep.* **6**, 25550 (2016).
48. Newman, A. M. et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**, 773–782 (2019).
49. Griffin, A. T., Vlahos, L. J., Chiuzan, C. & Califano, A. NaRnEA: an information theoretic framework for gene set analysis. *Entropy* **25**, 542 (2023).
50. Liu, L. et al. Genetic regulation of serum IgA levels and susceptibility to common immune, infectious, kidney, and cardiometabolic traits. *Nat. Commun.* **13**, 6859 (2022).
51. Giorgi, F. M. et al. Correction: Inferring protein modulation from gene expression data using conditional mutual information. *PLoS ONE* **11**, e0163402 (2016).
52. Mariani, L. H. et al. CureGN study rationale, design, and methods: establishing a large prospective observational study of glomerular disease. *Am. J. Kidney Dis.* **73**, 218–229 (2019).
53. Okada, Y. et al. Risk for ACPA-positive rheumatoid arthritis is driven by shared HLA amino acid polymorphisms in Asian and European populations. *Hum. Mol. Genet.* **23**, 6916–6926 (2014).
54. Carmona, F. D. et al. A large-scale genetic analysis reveals a strong contribution of the HLA class II region to giant cell arteritis susceptibility. *Am. J. Hum. Genet.* **96**, 565–580 (2015).
55. Hu, X. et al. Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk. *Nat. Genet.* **47**, 898–905 (2015).
56. Pillebout, E. et al. Biomarkers of IgA vasculitis nephritis in children. *PLoS ONE* **12**, e0188718 (2017).
57. Cambier, A. et al. Soluble CD89 is a critical factor for mesangial proliferation in childhood IgA nephropathy. *Kidney Int.* **101**, 274–287 (2022).
58. Chou, V. et al. INPP5D regulates inflammasome activation in human microglia. *Nat. Commun.* **14**, 7552 (2023).
59. Pedicone, C. et al. Pan-SHIP1/2 inhibitors promote microglia effector functions essential for CNS homeostasis. *J. Cell Sci.* **133**, jcs238030 (2020).
60. Kamen, L. A., Levinsohn, J. & Swanson, J. A. Differential association of phosphatidylinositol 3-kinase, SHIP-1, and PTEN with forming phagosomes. *Mol. Biol. Cell* **18**, 2463–2472 (2007).
61. Riethmueller, S. et al. Proteolytic origin of the soluble human IL-6R in vivo and a decisive role of N-glycosylation. *PLoS Biol.* **15**, e2000080 (2017).
62. Goth, C. K. et al. A systematic study of modulation of ADAM-mediated ectodomain shedding by site-specific O-glycosylation. *Proc. Natl. Acad. Sci. USA* **112**, 14623–14628 (2015).
63. Schett, G. Physiological effects of modulating the interleukin-6 axis. *Rheumatology* **57**, ii43–ii50 (2018).
64. Garbers, C. et al. The interleukin-6 receptor Asp358Ala single nucleotide polymorphism rs2228145 confers increased proteolytic conversion rates by ADAM proteases. *Biochim. Biophys. Acta* **1842**, 1485–1494 (2014).
65. Ferreira, R. C. et al. Functional IL6R 358Ala allele impairs classical IL-6 receptor signaling and influences risk of diverse inflammatory diseases. *PLoS Genet.* **9**, e1003444 (2013).
66. Mayer-Hain, S. et al. Systemic activation of neutrophils by immune complexes is critical to IgA vasculitis. *J. Immunol.* **209**, 1048–1058 (2022).
67. Chen, X. Q. et al. The involvement of neutrophil extracellular traps in disease activity associated with IgA vasculitis. *Front. Immunol.* **12**, 668974 (2021).
68. Li, Q. et al. FcαRI-mediated neutrophil activation contributed to the pathogenesis of adult IgA vasculitis with nephritis. *Nephrol. Dial. Transplant.* **41**, 1274–1283 (2026).
69. Pillebout, E. & Sunderkotter, C. IgA vasculitis. *Semin. Immunopathol.* **43**, 729–738 (2021).
70. Perkovic, V. et al. Sibeprenlimab in IgA nephropathy—interim analysis of a phase 3 trial. *N. Engl. J. Med.* **394**, 635–646 (2026).
71. Manichaikul, A. et al. Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).

72. 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).
73. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
74. Loh, P. R. et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **48**, 1443–1448 (2016).
75. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
76. 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).
77. 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).
78. Gusev, A. et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).
79. Zhang, J. et al. Plasma proteome analyses in individuals of European and African ancestry identify cis-pQTLs and models for proteome-wide association studies. *Nat. Genet.* **54**, 593–602 (2022).
80. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
81. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
82. Jia, X. et al. Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS ONE* **8**, e64683 (2013).
83. 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).
84. Ooi, J. D. et al. Dominant protection from HLA-linked autoimmunity by antigen-specific regulatory T cells. *Nature* **545**, 243–247 (2017).
85. wwPDB Consortium. Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Res.* **47**, D520–D528 (2019).
86. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
87. Krieger, E. et al. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: four approaches that performed well in CASP8. *Proteins* **77**, 114–122 (2009).
88. Meng, E. C. et al. UCSF ChimeraX: tools for structure building and analysis. *Protein Sci.* **32**, e4792 (2023).
89. Wisniewska, M. et al. Structural determinants of the ADAM inhibition by TIMP-3: crystal structure of the TACE-N-TIMP-3 complex. *J. Mol. Biol.* **381**, 1307–1319 (2008).
90. Pos, W. et al. Crystal structure of the HLA-DM-HLA-DR1 complex defines mechanisms for rapid peptide selection. *Cell* **151**, 1557–1568 (2012).
91. Denny, J. C. et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* **26**, 1205–1210 (2010).
92. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
93. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
94. Aguet, F. et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
95. Wright, F. A. et al. Heritability and genomics of gene expression in peripheral blood. *Nat. Genet.* **46**, 430–437 (2014).
96. DeLuca, D. S. et al. RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics* **28**, 1530–1532 (2012).
97. van der Wijst, M. G. P. et al. Single-cell RNA sequencing identifies celltype-specific cis-eQTLs and co-expression QTLs. *Nat. Genet.* **50**, 493–497 (2018).
98. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
99. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
100. Taylor-Weiner, A. et al. Scaling computational genomics to millions of individuals with GPUs. *Genome Biol.* **20**, 228 (2019).
101. Stegle, O., Parts, L., Durbin, R. & Winn, J. A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. *PLoS Comput. Biol.* **6**, e1000770 (2010).
102. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
103. Li, Y. I. et al. Annotation-free quantification of RNA splicing using LeafCutter. *Nat. Genet.* **50**, 151–158 (2018).
104. Pillebout, E. IgA vasculitis and IgA nephropathy: same disease? *J. Clin. Med.* **10**, 2310 (2021).
105. Stern, Y. et al. The Reference Ability Neural Network Study: motivation, design, and initial feasibility analyses. *Neuroimage* **103**, 139–151 (2014).
106. Touil, H. et al. Human central nervous system astrocytes support survival and activation of B cells: implications for MS pathogenesis. *J. Neuroinflammation* **15**, 114 (2018).
107. Margolin, A. A. et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinform.* **7**, S7 (2006).
108. Giorgi, F. M. et al. Inferring protein modulation from gene expression data using conditional mutual information. *PLoS ONE* **9**, e109569 (2014).
109. Street, K. et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).
110. Sundstrom, C. & Nilsson, K. Establishment and characterization of a human histiocytic lymphoma cell line (U-937). *Int. J. Cancer* **17**, 565–577 (1976).
111. Takahashi, K. et al. Clustered O-glycans of IgA1: defining macro- and microheterogeneity by use of electron capture/transfer dissociation. *Mol. Cell Proteomics* **9**, 2545–2557 (2010).

Acknowledgements

We are grateful to all study participants for their participation and contributions to this study.

Author contributions

All authors have read and approved the final version of the manuscript. K.K. and H.Z. conceived the study, provided overall supervision of the project, and made the decision to publish the findings. L.L. and L.Z. performed quality control, imputation, and association analyses for GWAS discovery cohorts. J. Martin and M.A.G.-G. provided GWAS summary statistics for the Spanish cohort. L.L. performed GWAS meta-analysis, fine-mapping studies, functional annotations, and e/sQTL mapping. S. Monteiro-Martins and A. Köttgen performed transcriptome-wide and proteome-wide association studies. A. Khan performed phenome-wide association studies. J.N. and T.J.G. performed structural modeling of risk variants. P.L.D. generated single-cell RNA-seq datasets, and L.L. and M.F. performed their processing and quality control analyses. J.N., T.J.G., S.H., S.Q., Q.S. and C.R. conducted experiments using

the U937 monocyte cell line, and A.D.T. performed RNA extraction and processing. L.L., A.T.G., L.V. and A.C. performed single-cell network-based master regulator and MINDy analyses. C.B. and J.Y.Z. contributed to sample and data management. All other authors were involved in the recruitment of patients for this study.

Funding

The study was supported by the NIH grants 5K01DK137031 (L.L.), 2R01DK105124 (K.K., J.N.), and 1RC2DK116690 (K.K.). J.N. and T.J.G. are supported in part by NIH grants AI149431, DK078244, and DK082753. A.T.G. was supported by the Ruth L. Kirschstein National Research Service Awards 5T32GM007367-43 and F30CA257765. The work of A. Köttgen was supported by the German Research Foundation (DFG) Project ID 431984000 (SFB 1453) and Project ID 390939984 (EXC-2189, CIBSS). The work of S.M.M. was supported by the EQUIP Program for Medical Scientists, Faculty of Medicine, University of Freiburg. A.C. was supported by the NCI Outstanding Investigator award R35CA197745. A. Khan was supported by the NIH grant 5K25DK128563. The GIGA-kids Study was additionally supported by the Pediatric Nephrology Research Consortium (PNRC) and a gift from the IgA Nephropathy Foundation. The meta-PheWAS was conducted using data from the UK Biobank (UKBB) under the application number of 41849, the All-of-Us (AoU) research program, and the Electronic Medical Records and Genomics-III (eMERGE-III) dataset. We thank the participants and staff of these programs for their contributions. The replication of master regulator activities associated with IgA vasculitis was performed using whole-blood RNA sequencing data from the Cure Glomerulonephropathy (CureGN) Study. Funding for the CureGN consortium is provided by U24DK100845, U01DK100846, U01DK100876, U01DK100866, and U01DK100867 from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), and CureGN patient recruitment is additionally supported by NephCure.

Competing interests

J.N. is a co-founder and co-owner of and consultant for Reliant Glycosciences, LLC. B.A.J. and D.V.R. are co-founders and co-owners of Reliant Glycosciences, LLC. J.N. and B.A.J. are co-inventors on US patent application 14/318,082 (assigned to the UAB Research Foundation [UABRF] and licensed by UABRF to Reliant Glycosciences, LLC). J.N. consulted for Calliditas, Travere, Novartis, and Vera Therapeutics. J.N. reports sponsored research agreements with Argenx (via UAB). K.K. consulted for Novartis, Biogen, Vera, Travere, Otsuka, and Vertex. Dr. Califano is founder, equity holder, and consultant of DarwinHealth Inc., a company that has licensed some of the algorithms used in this manuscript from Columbia University. Columbia University is also an equity holder in DarwinHealth Inc. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-026-78026-6.

Correspondence and requests for materials should be addressed to Hong Zhang or Krzysztof Kiryluk.

Peer review information *Nature Communications* thanks Mirko Treccani and the other anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2026

Lili Liu[1,74], Li Zhu[2,74], Sara Monteiro-Martins[3], Aaron Griffin[4], Lukas J. Vlahos[4], Masashi Fujita[5], Alexandria D. Taylor[1], Cecilia Berrouet[1], Francesca Zanoni[1], Maddalena Marasa[1], Jun Y. Zhang[1], Xu-jie Zhou[2], Yasar Caliskan[6], Oleh Akchurin[7], Samhar Al-Akash[8], Augustina Jankauskiene[9], Monica Bodria[10], Aftab Chishti[11], Ciro Esposito[12], Vittoria Esposito[12], Donna Claes[13], Vladimir Tesar[14], Thomas K. Davis[15], Dmitry Samsonov[16], Dorota Kaminska[17], Tomasz Hryszko[18], Gianluigi Zaza[19], Joseph T. Flynn[20], Franca Iorember[21], Francesca Lugani[22], Dana Rizk[23], Bruce A. Julian[23], Guillermo Hidalgo[24], Mahmoud Kallash[25], Luigi Biancone[26], Antonio Amoroso[26], Luisa Bono[27], Laila-Yasmin Mani[28], Bruno Vogt[28], Fangming Lin[29], Raji Sreedharan[30], Patricia Weng[31], Daniel Ranch[32], Nianzhou Xiao[33], Alejandro Quiroga[34], Raed Bou Matar[35], Michelle N. Rheault[36], Scott Wenderfer[37], Dave Selewski[38], Sigrid Lundberg[39], Cynthia Silva[40], Sherene Mason[40], John D. Mahan[41], Tetyana L. Vasylyeva[42], Krzysztof Mucha[43], Bartosz Foroncewicz[43], Leszek Pączek[44,45], Michał Florczak[43], Małgorzata Olszewska[46], Agnieszka Gradzińska[47], Maria Szczepańska[48], Edyta Machura[48], Andrzej Badeński[48], Helena Krakowczyk[48], Przemysław Sikora[49], Norbert Kwella[50], Monika Miklaszewska[51], Dorota Drożdż[51], Marcin Zaniew[52], Krzysztof Pawlaczyk[53], Katarzyna Siniewicz-Luzeńczyk[54], Andrew S. Bomback[55], Gerald B. Appel[55], Claudia Izzi[56], Francesco Scolari[56], Anna Materna-Kiryluk[57], Malgorzata Mizerska-Wasiak[58], Laureline Berthelot[59], Evangeline Pillebout[60], Renato C. Monteiro[60], Stacy Hall[23], Shihong Qiu[23], Qiao Shang[23], Colin Reily[23], Jan Novak[23], Todd Jason Green[23], William E. Smoyer[41], M. Colleen Hastings[61], Robert J. Wyatt[62,63], Raoul Nelson[64], Atlas Khan[1], Javier Martin[65],

Miguel A. González-Gay[66,67], Philip L. De Jager[5], Anna Köttgen[3,68], Andrea Califano[4,69,70,71,72,73], Ali G. Gharavi[1], Hong Zhang[2] ✉ & Krzysztof Kiryluk[1] ✉

[1]Department of Medicine, Division of Nephrology, Columbia University, Vagelos College of Physicians & Surgeons, New York, NY, USA. [2]Renal Division, Peking University First Hospital, Peking University Institute of Nephrology, Beijing, China. [3]Institute of Genetic Epidemiology, Medical Center and Faculty of Medicine, University of Freiburg, Freiburg, Germany. [4]Department of Systems Biology, Columbia University, New York, NY, USA. [5]Division of Neuroimmunology, Department of Neurology, Columbia University, New York, NY, USA. [6]Division of Nephrology, Saint Louis University, Saint Louis, MO, USA. [7]Weill Cornell Medical Center, New York, NY, USA. [8]Driscoll Children's Hospital, Corpus Christi, TX, USA. [9]Institute of Clinical Medicine, Vilnius University, Vilnius, Lithuania. [10]Primary Care Unit, Ausl Parma, Parma, Italy. [11]Division of Pediatric Nephrology, University of Kentucky, Kentucky Children's Hospital, Lexington, KY, USA. [12]Nephrology Unit, Istituti Clinici Scientifici Maugeri IRCCS, Pavia, Italy. [13]Cinncinnati Children's Hospital, Cincinnati, OH, USA. [14]Department of Nephrology, 1st Faculty of Medicine, Charles University, Prague, Czech Republic. [15]Washington University in St. Louis, St. Louis, MO, USA. [16]Maria Fareri Children's Hospital (MCF), New York Medical College, New York, NY, USA. [17]Department of Non-Procedural Clinical Sciences, Faculty of Medicine, Wroclaw University of Science and Technology, Wroclaw, Poland. [18]2nd Department of Nephrology, Hypertension and Internal Medicine, Medical University of Bialystok, Białystok, Poland. [19]Renal, Dialysis and Transplant Unit, University of Calabria, Arcavacata, Italy. [20]Division of Nephrology, Seattle Children's Hospital, Seattle, WA, USA. [21]Phoenix Children's Hospital, Phoenix, AZ, USA. [22]IRCCS Istituto Giannina Gaslini, Genoa, Italy. [23]University of Alabama at Birmingham, Birmingham, AL, USA. [24]East Carolina University, Greenville, NC, USA. [25]SUNY Buffalo, Buffalo, NY, USA. [26]University of Turin, Italy and Center for Clinical and Computational Genomics, Non-coding RNAs and RNA-based Therapeutics, Fondazione Istituto Italiano di Tecnologia (IIT), Aosta, Italy. [27]Nephrology and Dialysis, A.R.N.A.S. Civico and Benfratellio, Palermo, Italy. [28]Department of Nephrology and Hypertension, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland. [29]Division of Pediatric Nephrology, Department of Medicine, Columbia University, New York, NY, USA. [30]Medical College of Wisconsin, Madison, WI, USA. [31]Pediatric Nephrology Department, University of California, Los Angeles, CA, USA. [32]UT Health San Antonio, San Antonio, TX, USA. [33]Valley Children's Healthcare, Madera, CA, USA. [34]Helen DeVos Children's Hospital, Grand Rapids, MI, USA. [35]Cleveland Clinic, Cleveland, OH, USA. [36]University of Minnesota, Minneapolis, MN, USA. [37]Texas Children's Hospital, Baylor College of Medicine, Houston, TX, USA. [38]Mott Children's Hospital, University of Michigan, Ann Arbor, MI, USA. [39]Danderyd University Hospital, Karolinska Institute, Stockholm, Sweden. [40]Connecticut Children's Medical Center, Hartford, CT, USA. [41]The Research Institute at Nationwide Children's Hospital, Columbus, OH, USA. [42]Texas Tech University Health Sciences Center, Amarillo, TX, USA. [43]Department of Transplantology, Immunology, Nephrology and Internal Diseases, Medical University of Warsaw, Warsaw, Poland. [44]Department of Clinical Immunology, Medical University of Warsaw, Warsaw, Poland. [45]Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw, Poland. [46]Department of Dermatology, Medical University of Warsaw, Warsaw, Poland. [47]Department of Dermatology and Pediatric Dermatology, Centre of Post-graduate Medical Education, Warsaw, Poland. [48]Department of Pediatrics, Faculty of Medical Sciences in Zabrze, Medical University of Silesia, Katowice, Poland. [49]Department of Pediatric Nephrology, Medical University of Lublin, Lublin, Poland. [50]Department of Nephrology, Transplantology and Internal Diseases, University of Warmia and Mazury, Olsztyn, Poland. [51]Department of Pediatric Nephrology and Hypertension, Jagiellonian University Medical College, Krakow, Poland. [52]Department of Pediatrics, University of Zielona Góra, Zielona Góra, Poland. [53]Department of Nephrology, Transplantology and Internal Diseases, Poznan University of Medical Sciences, Poznan, Poland. [54]Department of Paediatrics, Immunology and Nephrology, Polish Mother's Memorial Hospital Research Institute, Lodz, Poland. [55]Columbia University Medical Center, New York, NY, US. [56]Department of Medical and Surgical Specialties and Nephrology Unit, University of Brescia-ASST Spedali Civili, Brescia, Italy. [57]Department of Genetics, Medical University of Poznan, Poznan, Poland. [58]Department of Pediatrics and Nephrology, Medical University of Warsaw, Warsaw, Poland. [59]Nantes University, Inserm, CR2TI Center of Research on Translational Transplantation and Immunology, Nantes, France. [60]Center for Research on Inflammation, Paris Cité University, INSERM and CNRS, Paris, France. [61]University of Tennessee Health Science Center, Memphis, TN, USA. [62]Department of Pediatrics, University of Tennessee Health Sciences Center, Memphis, TN, USA. [63]Children's Foundation Research Institute, Le Bonheur Children's Hospital, Memphis, TN, USA. [64]University of Utah, Salt Lake City, UT, USA. [65]Institute of Parasitology and Biomedicine Lopez-Neyra, Spanish National Research Council (CSIC), Granada, Spain. [66]Division of Rheumatology, IIS-Fundación Jiménez Díaz, Madrid, Spain. [67]Medicine and Psychiatry Department, University of Cantabria, Santander, Spain. [68]Centre for Integrative Biological Signalling Studies (CIBSS), University of Freiburg, Freiburg, Germany. [69]Herbert Irving Comprehensive Cancer Center, Columbia University, New York, NY, USA. [70]Department of Biochemistry and Molecular Biophysics, Vagelos College of Physicians and Surgeons, Columbia University, New York, NY, USA. [71]Department of Biomedical Informatics, Vagelos College of Physicians and Surgeons, Columbia University, New York, NY, USA. [72]Department of Medicine, Vagelos College of Physicians and Surgeons, Columbia University, New York, NY, USA. [73]Chan Zuckerberg Biohub New York, New York, NY, USA. [74]These authors contributed equally: Lili Liu, Li Zhu. ✉e-mail: hongzh@bjmu.edu.cn; kk473@columbia.edu