

Candidate genes for IgA nephropathy in pediatric patients: exome-wide association study

Authors:

Anastasiia A. Buianova¹, Mariia V. Proskura², Valery V. Cheranev¹, Vera A. Belova¹, Anna O. Shmitko¹, Anna S. Pavlova¹, Iuliia A. Vasiliadis¹, Oleg N. Suchalko¹, Denis V. Rebrikov¹, Edita K. Petrosyan², Dmitriy O. Korostin¹

Affiliations:

1 – Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Pirogov Russian National Research Medical University, Moscow, Russia

2 – Russian Children’s Clinical Hospital, Nephrology Department, Moscow, Russia

Correspondence to: Anastasiia A. Buianova, email: anastasiabuianova97@gmail.com

Abstract

IgA nephropathy (IgAN) is an autoimmune disorder that commonly manifests during adolescence. This disease is believed to be a non-monogenic disorder related to variations in multiple genes involved in various biological pathways.

We performed the exome-wide association study of 70 children with IgAN confirmed by renal biopsy and 637 healthy donors to identify gene associations responsible for the disease. The HLA allele frequencies between the patients and healthy donors from the bone marrow registry of the Pirogov University were compared. We tested 78,020 gene markers for association, performed the functional enrichment analysis and the transcription factor binding preference detection.

We detected 333 genetic variants, employing three inheritance models. The most significant association with the disorder was observed for rs143409664 (*PRAG1*) in case of the additive and dominant models ($P_{\text{BONF}} = 1.808 \times 10^{-15}$ and $P_{\text{BONF}} = 1.654 \times 10^{-15}$, respectively) and for rs13028230 (*UBR3*) in case of the recessive model ($P_{\text{BONF}} = 1.545 \times 10^{-9}$). Enrichment analysis indicated the strongly overrepresented “immune system” and “kidney development” terms. The HLA-DQA1*01:01:01G allele ($P = 0.0076$; OR, 2.021 [95% CI, 1.322-3.048]) was significantly the most frequent among IgAN patients.

Here we characterized, for the first time, the genetic background of the Russian IgAN patients identifying the risk alleles typical of the population. The most prominent signals were detected in previously undescribed loci.

Keywords: Autoimmune Disease; Exome Sequencing; Genetic Predisposition to Disease; IgA nephropathy; Pediatrics.

Introduction

IgA nephropathy (IgAN) is one of the most common primary glomerulonephritides in the world both in adults and children [1]. Although it has been extensively studied for more than half a century, it still remains the leading cause of the end-stage kidney disease (ESKD) [2]. The clinical presentation of the disease is highly variable: from painless microhematuria to the rapidly progressing glomerulonephritis and ESKD [3-7]. From 20% to 40% of cases of IgAN progression to ESKD within 20 years from the onset require renal replacement therapy [8-11].

IgA nephropathy is diagnosed through kidney biopsy examination. It shows the deposition of IgA-containing immune complexes in the mesangium inducing mesangial cell proliferation and extracellular matrix accumulation [1]. At present, there are two basic views on the causes of the disease. Several groups suggest that IgAN is an autoimmune disease leading to the antibody-mediated destruction of the glomerular basement membrane. IgAN pathogenesis is complex and likely to involve several different pathways forming a complex network where infections may play a triggering role. Though in some cases, IgA nephropathy can precede the infection that induces a dysregulated immune response, IgA nephropathy itself is not an infectious disease [12].

Furthermore, in IgAN, the glycosylation of O-linked glycans in the hinge region of IgA1 is disrupted resulting in the high blood level of circulating galactose-deficient IgA1 (Gd-IgA1) and its abnormal clearance [13]. Gd-IgA1 has been shown to be highly heritable, therefore, the link between IgAN and Gd-IgA1 and its role in IgAN pathogenesis might provide novel insights into the pathogenic processes involved in IgAN [14]. Levy M et al. suggested a possible contribution of genetic predisposition to IgAN [15]. Their hypothesis is based on a certain geographical prevalence of the IgAN. Based on kidney biopsy, IgA nephropathy was diagnosed in 20% and 40% of children with glomerular diseases in Europe and Asia, respectively, whereas it was not frequently registered in the African population [7, 16].

The presence of genes involved in the immunity against intestinal pathogens is thought to be responsible for the high incidence of IgA nephropathy in Asia. Certain loci are associated with the risk of inflammatory bowel disease (IBD), the state of the intestinal epithelial barrier, and response to mucosal pathogens. The genetic predisposition strongly correlates with helminth diversity, suggesting a possible role for the host–intestinal pathogen interactions in IgAN geographic variations [17]. Familial cases of IgAN are also known and described supporting the genetic role in the disease etiology [18-22].

The role of gene-candidates in IgAN development was estimated by genome-wide association study (GWAS). In familial forms, a high risk of the disease was found to correlate with the following specific loci: *6q22–23 (IGAN1)*, *4q26–31 (IGAN2)* and *17q12–22 (IGAN3)*

[23, 24]. In East Asia, the associated proteins contribute to the adaptive and innate immunity, IgA1 glycosylation, and the renin-angiotensin system [25], including haplotypes *HLA-DQ* and *HLA-DR*: HLA-DRB1*14:05:01 (belonging to DR*14), HLA-DRB1*03:01:01, HLA-DRB1*04, HLA-DQB1*03:01 [26-28]. Xia YF et al. showed the Megsin gene (*SERPINB7*) to be a major factor determining the predisposition to the disease and its progression in the Chinese population [29, 30]. Twenty-four candidate genes associated with IgAN were analyzed to identify their interactions. The cooperation between *C1GALT1-330G/T* (rs1008898) and *IL5RA31+197A/G* (rs340833) was found to be statistically significant ($P < 0.001$) for IgAN pathogenesis [31], whereas the combination of *P-selectin-2441a/G* with *CD14-159C/T* was associated with gross hematuria in IgAN patients. Moreover, the interaction of *TGF- β 1 509T/C*, *P-selectin-2441A/G*, and *MCP-1 2518A/G* was found to influence the crescent formation [32]. The interaction between two key genes, *C1GALT1* and *ST6GALNAC2*, was also shown to affect the susceptibility to IgAN and disease progression [33]. In the UK population, a significant association with IgAN was detected on the chromosome 6 in the region of the MHC ($P = 1 \times 10^{-9}$) [34]. Five loci associated with susceptibility to IgAN were identified in Chinese patients: three distinct loci in the MHC region, a common deletion of *CFHR1* and *CFHR3 1q32* loci, and a locus *22q12 (HORMAD2)* [35]. In 2014, Kiryluk et al. revealed six haplotypes associated with a high risk of IgAN: already known variants in *ITGAM*, *ITGAX*, *VAV3* and *CARD9* genes, as well as two new haplotypes, *HLA-DQB1* and *DEFA* [17]. Ming Li et al. identified genes responsible for susceptibility to the disease and associated with IgAN, located in *17p13* and *8p23* [36], including the tumor necrosis factor (*TNFSF13*) and α -defensin (*DEFA*). rs660895 (*HLA-DRB1*) was found to correlate with the IgA serum level and proteinuria level [37]. The novel genes associated with IgAN included *ST6GAL1* at *3q27.3*, *ACCS* at *11p11.2*, and *ODF1-KLF10* at *8q22.3*. The *ITGAX-ITGAM (16p11.2)* association was confirmed being moderately replicated, and previously observed genes *VAV3(1p13)* and *CARD9 (9q34)* were detected as well.

Given that the genetic aspects of IgAN have previously demonstrated high heterogeneity in terms of the identified associations, we performed exome-wide association study (EWAS) employing our own clinical samples.

Materials and Methods

Ethics statement

All legal representatives as well as patients over 15 y.o. signed the appropriate informed voluntary consent to participate in the study. Local ethics committee of Pirogov Russian National Research Medical University approved this study on the 17th of December, 2018 (Protocol No. 181).

Patient cohort

Between 2019 and 2021, 70 children at the Russian Children's Clinical Hospital had been observed at the nephrology department for at least 6 months. During laboratory tests, 4 ml of blood were sampled from the cubital vein into a test tube with EDTA for subsequent EWAS. In all patients, the diagnosis of primary IgA nephropathy was confirmed by kidney biopsy and further histological examination with immunofluorescence, electronic, and light microscopy.

gDNA extraction

DNA isolation was performed using the DNeasy Blood and Tissue Kit (Qiagen) according to the manufacturer's protocol. The extracted DNA was quantified with the Qubit dsDNA BR Assay system (Life Technologies), and its quality was assessed by 1% agarose gel electrophoresis.

Library preparation and enrichment

DNA Libraries were prepared from 500 ng of genomic DNA using the MGIEasy Universal DNA Library Prep Set (MGI Tech) according to the manufacturer's protocol. DNA fragmentation was performed by ultrasonication using Covaris S-220 with the average fragment length of 250 bp. Whole-exome enrichment of DNA library pools was performed according to a previously described protocol [38] using the SureSelect Human All Exon v7 probes (Agilent Technologies). The concentrations of DNA libraries were measured using Qubit Flex (Life Technologies) with the dsDNA HS Assay Kit following to the manufacturer's protocol. The quality of the prepared libraries was assessed using Bioanalyzer 2100 with the High Sensitivity DNA kit (Agilent Technologies) according to the manufacturer's instructions.

Sequencing

The enriched library pools were further circularized and sequenced by a paired end sequencing using DNBSEQ-G400 with the DNBSEQ-G400RS High-throughput Sequencing Set PE100 following the manufacturer's instructions (MGI Tech) with the average coverage of 100x. Fastq files were generated using the basecallLite software from the manufacturer (MGI Tech).

Control samples

We used exome sequencing data from 637 healthy Russian donors previously processed in our laboratory as a control dataset. To compare HLA allele frequencies, we used data on healthy donors from the Pirogov University register including 1849 individuals.

Raw sequencing data analysis

The quality control of the obtained paired fastq files was performed by FastQC v0.11.9 [39]. Based on the quality metrics, fastq files were trimmed using BBDuk by BBMap v38.96

[40]. Reads were aligned to the indexed reference genome GRCh37 using bwa-mem2 v2.2.1 [41, 42]. SAM files were converted into bam files and sorted using SAMtools v1.9 to check the percentage of the aligned reads [43]. The duplicates in the obtained bam files were marked using Picard MarkDuplicates v2.22.4 [44] and excluded from further analysis.

We performed the quality control analysis on marked bam files with the following Agilent all-exon v7 target file “regions.bed”. For the samples that passed quality control (width of target coverage $10x \geq 95\%$), SNVs and indels were called using the bcftools mpileup software v1.9 [45], and vcf files were obtained for each sample. After variant calling, vcf files were normalized using vt normalize v0.5772-60f436c3 [46] and filtered based on the target regions expanded by ± 100 base pairs flanking each end. Calling data were annotated using the InterVar software [47].

Exome-wide association study

We analyzed DNA samples from 70 IgA nephropathy (IgAN) patients and 637 samples from healthy donors. Variants in vcf files (ver 4.2) were filtered based on coverage (threshold = 13) and quality (threshold = 20). InDel variants were normalized using vt (ver v0.5772-60f436c3) tool. All vcf files were merged by bcftools merge (ver 1.10.2). To obtain the list of genes containing sets of false positive variations, the genomic intervals were extracted from the gtf file of genome annotation (GRCh37.p13). Then, using bedtools (ver 2.27.1) [48], variations in these genomic intervals were excluded from the merged vcf files. Data were analyzed in the PLINK v1.90b6.24 software [49]. The deviation of sample heterozygosity levels was within 3 SD from the mean. The markers (both SNVs and indels) genotyped in less than 80% of the samples as well as individuals with less than 80% of genotyped markers were filtered out. In all samples, predicted sex coincided with the actual sex. Due to a small sample size, we chose a threshold of 0.0001 and a frequency of a minor allele of 0.1 for checking the Hardy-Weinberg equilibrium. After setting the threshold value $PI_HAT = 0.02$, we did not exclude samples from the analysis. Finally, 78,020 of 793,015 markers were subjected to further analysis. We created the additive, dominant, and recessive genetic models; p-value was corrected using the Bonferroni criterion ($P_{adj} < 0.05$).

Functional enrichment analysis and transcription factor (TF) binding preference detection

The charts of gene enrichment analysis were created using the Metascape v3.5.20230101[50].

The eQTL analysis was performed using the following databases: Kidney eQTLs Atlas [51], NephQTL [52], and Database of Immune Cell eQTLs (DICE) project [53]. The cut-off for p-value was $P < 0.05$.

The STRING (<https://string-db.org>, date of access 12 April 2023) database was used to predict functional interactions of proteins. The search was restricted to “*Homo sapiens*” and the interaction score ≥ 0.4 were applied to construct the PPI networks.

We applied atSNP [54] to predict differential binding of a TF to a marker by labeling SNP-motif combinations with atSNP $pval_diff < 0.05$ as significant.

HLA-typing

HLA typing was performed on the exome data with help of three programs: HLA-HD (ver 1.7.0) [55], HLAScan (ver 2.1.4) [56], Kourami (ver 0.9.6) [57] using the allele database IPD-IMGT/HLA (version 3.51.0) [58]. The obtained data on the alleles were transformed into the corresponding G-groups, and results were merged into the final version. In case different tools provided conflicting results, we chose the alleles identified using HLA-HD or Kourami and checked bam files in IGV [59].

Statistical analysis

Principal component analysis (PCA) and allele frequency estimation were performed using PLINK v1.90b6.24 (Supplemental Figure 1). Manhattan and quantile-quantile (Q-Q) plot were generated with Python scripts. The one-way analysis of variance (ANOVA) test, Chi-squared test and Fisher’s exact test [60] were used to compare the differences between groups in RStudio 2022.02.0. The multiple comparison correction was conducted using the Bonferroni method: we multiplied p-values by the number of the alleles under consideration. Testing for the Hardy-Weinberg equilibrium was performed using the Arlequin 3.5 software [61]. All data were presented as mean \pm standard deviation (SD) or median [Q1, Q3]. Data normality was checked using the Kolmogorov-Smirnov test.

Results

Clinical Characteristics

The clinical characteristics of 70 IgAN patients are summarized in table S1.

We found no significant effects of sex on the clinical features based on a one-way analysis of variance (ANOVA). The control sample had an even distribution of sexes, with 52.04% males and 47.96% female.

Morphological analysis

Morphological data of kidney biopsy samples typical of their diagnosis are presented in Figure 1.

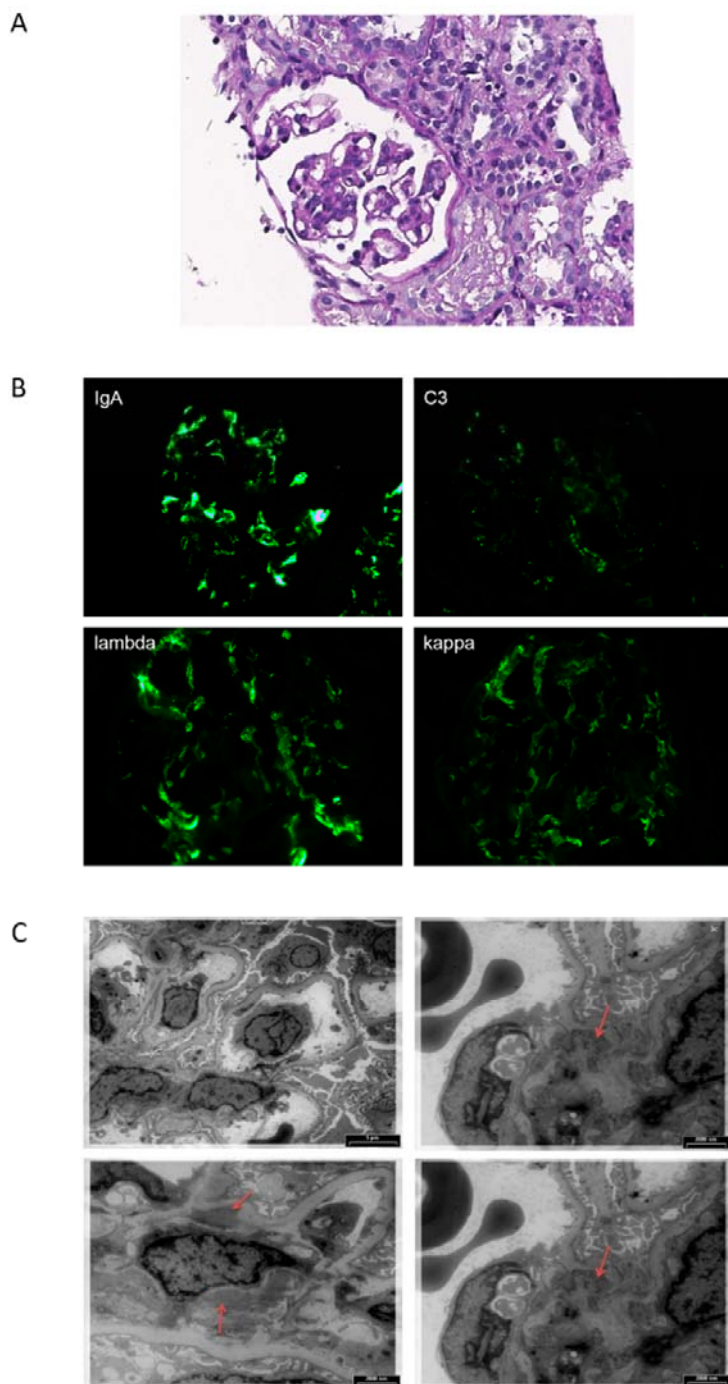


Figure 1. Renal biopsy of the typical patient with IgAN. Hematoxylin-eosin stain (magnification $\times 400$) (A). Immunofluorescence microscopy, staining for IgA, C3, lambda, kappa (magnification $\times 400$) (B). Electron microscopy. Red arrows indicate paramesangial deposits (C).

In the histological sample, matrix expansion and mesangial hypercellularity (up to 10 cells per a mesangial region) are presented (figure 1A). According to immunofluorescence: IgA, kappa, lambda in mesangium +++. IgG, IgM, C1q, fibrinogen – negative. C3 – ++ (figure 1B). The electron microscopy study reveals noticeable mesangial proliferation (4-6 nuclei) and massive multiple paramesangial deposits (figure 1C).

Description of the sequencing data

The main quality control metrics for the whole-exome sequencing of IgAN patient samples are listed in table S2. Sex chromosome karyotype and results of analysis of the SRY gene coverage coincided with the data from the patient medical cards.

Exome-wide association study of IgAN

IgAN patients were successfully matched with control subjects in case of the recessive model, as clearly shown by the small systematic deviation ($\lambda = 1.196$) of the observed distribution from the expected distribution under the null hypothesis which claimed the absence association in the Q-Q plot (figure 2A). The association p-value of multiple markers located on autosomes are shown on the Manhattan plots (figure 2B).

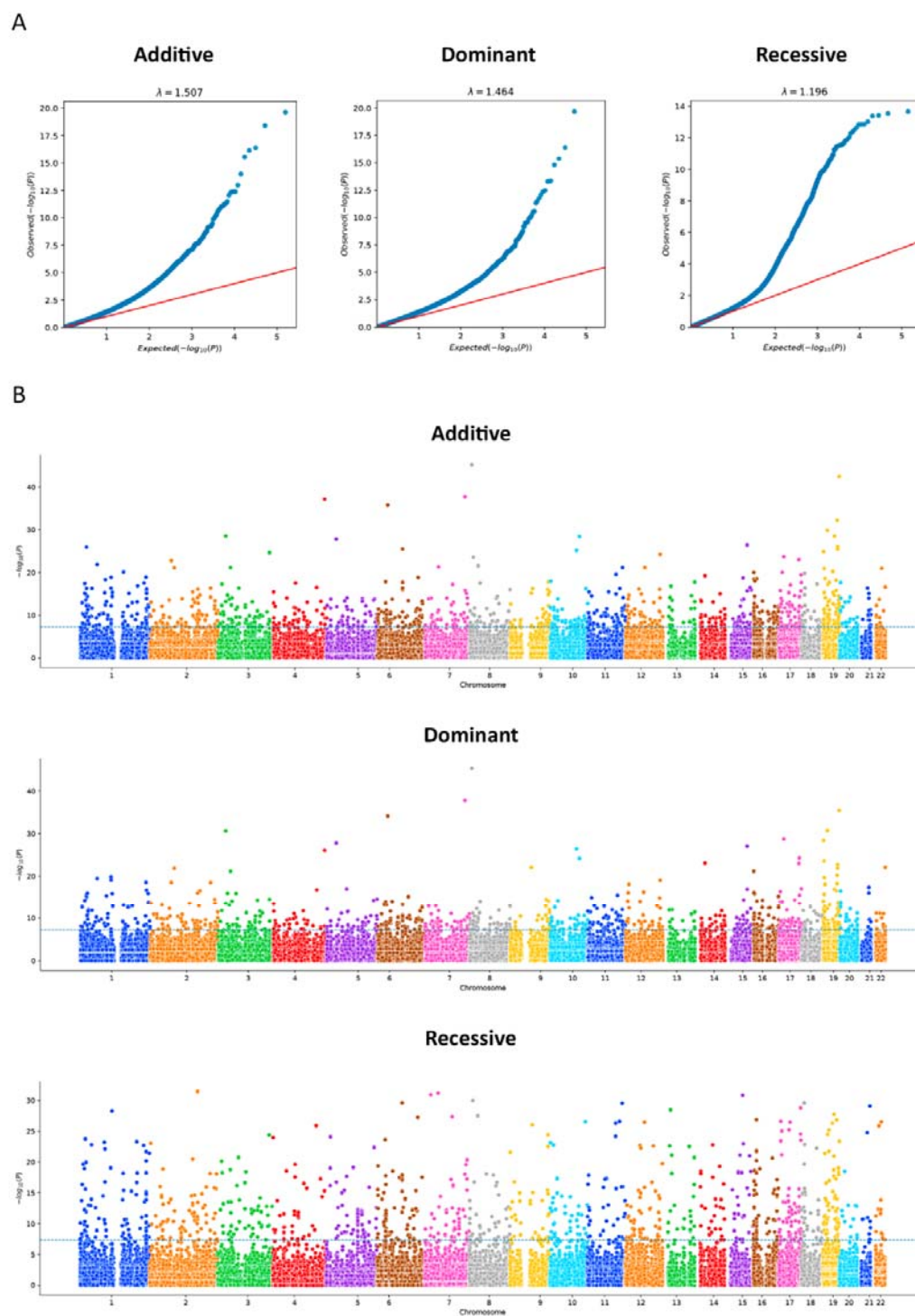


Figure 2. The Q-Q plots for association analysis. Each figure shows the expected (x-axis) and observed (y-axis) log (p-values) (A). The Manhattan plots for the exome-wide association study (EWAS) of IgA nephropathy (IgAN). The figure shows the p-value for the association with the disease (expressed as a negative logarithm of the p-value, y-axis) for each tested marker plotted

against the chromosomal position of the markers (x-axis). The blue line at 6 represents the threshold for EWAS ($-\log_{10}(1 \times 10^{-6})$) (B).

A total of 333 markers were identified after the p-value correction, including 52 insertions/ deletions. For each model, we showed the 10 most significant markers; with 24 of them not overlapping (table 1).

Table 1. The exome-wide association study analysis of IgAN patients and control groups.

GENE	CHR	POSITION	SNP	BONF	A1	F_A	F_U	A2	CHISQ	P	OR (95% CI)	SE
Additive												
<i>PRAG1</i>	8	8,176,387	rs143409664	1.808×10^{-15}	CGG GGC G	0.4571	0.07692	C	179.8	5.464×10^{-41}	10.11 (6.834 – 14.94)	0.1996
<i>ZNF787</i>	19	56,599,437	rs5828672	2.926×10^{-14}	C	0.4929	0.07849	CTC G	205.8	1.13×10^{-46}	11.41 (7.731 – 16.84)	0.1986
<i>TRBV5-4</i>	7	142,168,890	rs767448033	3.146×10^{-12}	T	0.3571	0.07849	C	103.3	2.906×10^{-24}	6.522 (4.365 – 9.745)	0.2048
<i>TENM3</i>	4	183,370,244	rs35591339	5.646×10^{-12}	A	0.5071	0.08399	AGC G	205.3	1.485×10^{-46}	11.22 (7.629 – 16.51)	0.1969
<i>PIMI</i>	6	37,138,023	rs1300416314	2.207×10^{-11}	G	0.4286	0.1148	GGC A	99.66	1.81×10^{-23}	5.784 (3.969 – 8.429)	0.1921
<i>CYTH2</i>	19	48,985,178	rs34527214	7.812×10^{-10}	TA	0.5357	0.1546	T	117.9	1.814×10^{-27}	6.308 (4.378 – 9.089)	0.1863
<i>MISP3</i>	19	14,184,713	rs35666756	8.00×10^{-9}	CG	0.3429	0.09105	C	78.01	1.028×10^{-18}	5.208 (3.499 – 7.753)	0.2029
<i>EOMES</i>	3	27,763,427	rs368178421	3.243×10^{-8}	GCG GCG C	0.3623	0.1009	G	76.89	1.805×10^{-18}	5.06 (3.418 – 7.491)	0.2001

<i>FBXO27</i>	19	39,505,111	rs4803184	3.255×10^{-8}	G	0.4786	0.1389	A	101.9	5.818×10^{-24}	5.688 (3.938 – 8.216)	0.1876
<i>SH3PXD2A</i>	10	105,428,453	rs10533306	3.656×10^{-8}	C	0.4714	0.1852	CAG	61.51	4.41×10^{-15}	3.923 (2.735 – 5.627)	0.184
Dominant												
<i>PRAG1</i>	8	8,176,387	rs143409664	1.654×10^{-15}	CGG GGC G	0.4571	0.07692	C	179.8	5.464×10^{-41}	10.11 (6.834 – 14.94)	0.1996
<i>TRBV5-4</i>	7	142,168,890	rs767448033	3.146×10^{-12}	T	0.3571	0.07849	C	103.3	2.906×10^{-24}	6.522 (4.365 – 9.745)	0.2048
<i>ZNF787</i>	19	56,599,437	rs5828672	3.25×10^{-11}	C	0.4929	0.07849	CTC G	205.8	1.13×10^{-46}	11.41 (7.731 – 16.84)	0.1986
<i>PIMI</i>	6	37,138,023	rs1300416314	1.187×10^{-10}	G	0.4286	0.1148	GGC A	99.66	1.81×10^{-23}	5.784 (3.969 – 8.429)	0.1921
<i>MISP3</i>	19	14,184,713	rs35666756	3.523×10^{-9}	CG	0.3429	0.09105	C	78.01	1.028×10^{-18}	5.208 (3.499 – 7.753)	0.2029
<i>EOMES</i>	3	27,763,427	rs368178421	3.964×10^{-9}	GCG GCG C	0.3623	0.1009	G	76.89	1.805×10^{-18}	5.06 (3.418 – 7.491)	0.2001
<i>SREBF1</i>	17	17,740,164	rs60282872	2.629×10^{-8}	G	0.2643	0.08477	GC	44.17	3.013×10^{-11}	3.878 (2.537 – 5.928)	0.2165
<i>PRTN3</i>	19	843,692	rs2301879	3.49×10^{-8}	G	0.2929	0.08948	A	53.45	2.646×10^{-13}	4.214 (2.792 – 6.361)	0.2101
<i>AMACR</i>	5	34,008,206	rs3217251	6.688×10^{-8}	C	0.2857	0.08085	CCG GCG CCA CGC CCC CAG	58.24	2.322×10^{-14}	4.548 (2.993 – 6.91)	0.2135

								CCG				
<i>CHRNA3</i>	15	78,913,067	rs751352647	1.363×10^{-7}	A	0.4214	0.1546	ACA G	60.56	7.149×10^{-15}	3.982 (2.755 – 5.755)	0.1879
Recessive												
<i>UBR3</i>	2	170,871,976	rs13028230	1.545×10^{-9}	A	0.5429	0.2951	G	35.59	2.431×10^{-9}	2.836 (1.991 – 4.039)	0.1804
<i>PKD1L1</i>	7	47,835,116	rs2348459	2.031×10^{-9}	T	0.5571	0.2363	C	66.2	4.077×10^{-16}	4.067 (2.844 – 5.815)	0.1825
<i>DNAH11</i>	7	21,639,818	rs6461586	2.649×10^{-9}	T	0.3643	0.1586	C	36.34	1.657×10^{-9}	3.041 (2.089 – 4.427)	0.1916
<i>CIAO2A</i>	15	64,381,148	rs16947748	2.861×10^{-9}	G	0.5786	0.2841	A	50.83	1.01×10^{-12}	3.459 (2.421 – 4.942)	0.1821
<i>TRMT9B</i>	8	12,870,438	rs2466264	6.754×10^{-9}	C	0.6714	0.3265	G	65.07	7.224×10^{-16}	4.215 (2.907 – 6.112)	0.1896
<i>MDN1</i>	6	90,400,292	rs954638	9.888×10^{-9}	C	0.4929	0.179	T	74.9	4.943×10^{-18}	4.458 (3.108 – 6.397)	0.1842
<i>PIEZO2</i>	18	10,757,868	rs7242408	1.003×10^{-8}	A	0.5071	0.2276	G	51.83	6.039×10^{-13}	3.491 (2.445 – 4.986)	0.1818
<i>OR10G6</i>	11	123,865,086	rs7944434	1.053×10^{-8}	A	0.4571	0.1743	C	62.56	2.583×10^{-15}	3.991 (2.777 – 5.735)	0.185
<i>PWP2</i>	21	45,547,563	rs2242954	1.658×10^{-8}	G	0.4	0.1931	C	32.15	1.427×10^{-8}	2.786 (1.933 – 4.016)	0.1865
<i>BAIAP2</i>	17	79,084,072	rs4072588	2.268×10^{-8}	T	0.5214	0.2339	G	53.88	2.132×10^{-13}	3.568 (2.499 – 5.095)	0.1817

Note: CHR – chromosome, BP – base pair based on the human reference genome GRCh37, A1 – minor allele, F_A – frequency of A1 allele in patients, F_U – frequency of A1 allele in controls, A2 – major allele, CHISQ – basic allelic test chi-square (1df), P – p-value for CHISQ, OR (95% CI) – odds ratio and confidence interval for odds ratio, SE – standard error, BONF – Bonferroni single-step adjusted p-values.

Gene set enrichment and eQTL analysis

Using the Metascape tool, we built networks showing the specificities underlying the interactions between genes.

In additive model, the terms “CAMKK2 pathway” (WP4874) ($\log[q\text{-value}] = -0.27$) and “positive regulation of catabolic process” (GO:0042176) ($\log[q\text{-value}] = -0.15$) demonstrated the most significant overexpression (figure 3A).

In dominant model, the terms “cell-cell adhesion via plasma-membrane adhesion molecules” (GO:0098742) ($\log[q\text{-value}] = -0.19$) and “ER-nucleus signaling pathway” (GO:0006984) ($\log[q\text{-value}] = -0.19$) were significantly higher enriched (figure 3B).

In recessive model, the most overrepresented terms included “response to acid chemical” (GO:0001101) ($\log[q\text{-value}] = -0.97$) and “MTOR signalling” (R-HSA-165159) ($\log[q\text{-value}] = -0.97$) (figure 3C).

Moreover, cell type signature analysis revealed kidney-related terms to be the topmost. For instance, the “lake adult kidney c5 proximal tubule epithelial cells stress inflame” (M39224) was the first and the second hit for the recessive and additive models, respectively; for the recessive model, the second hit was “lake adult kidney c26 mesangial cells” (M39245). In case of the dominant model, the least significantly enriched terms were “lake adult kidney c11 thin ascending limb” (M39230) and “lake adult kidney c15 connecting tubule” (M39234).

For each inheritance model, we detected 4 variants: rs10710110 (*NBPF3*), rs34527214 (*CYTH2*), rs4803184 (*FBXO27*), and rs71185698 (*PSMD2*) (figure 3D). Based on the eQTL analysis relying on the NephQTL database, there was no substantial difference between the number of variants affecting the expression in the glomeruli and tubulointerstitium (figure 3E); however, their number was nearly two times higher than detected by DICE. We found only two variants in the Kidney eQTLs Atlas: rs2242954 affected *PWP2* expression in renal tubules and *C21orf33* expression in the renal tubules and glomeruli, while rs4803184 influenced *FBXO27* expression in the glomeruli. rs2242954 is also known to elevate the *C21orf33* expression in macrophages ($\beta = 18.259 \pm 4.087$, $P = 1.156e^{-5}$), NK cells ($\beta = 9.086 \pm 1.943$, $P = 4.593 \times 10^{-6}$), CD4+ T cells ($\beta = 5.722 \pm 1.312$, $P = 1.822 \times 10^{-5}$), dendritic cells ($\beta = 4.964 \pm 0.769$, $P = 4.818 \times 10^{-10}$), and CD8+ T cells ($\beta = 3.611 \pm 0.689$, $P = 3.245 \times 10^{-7}$).

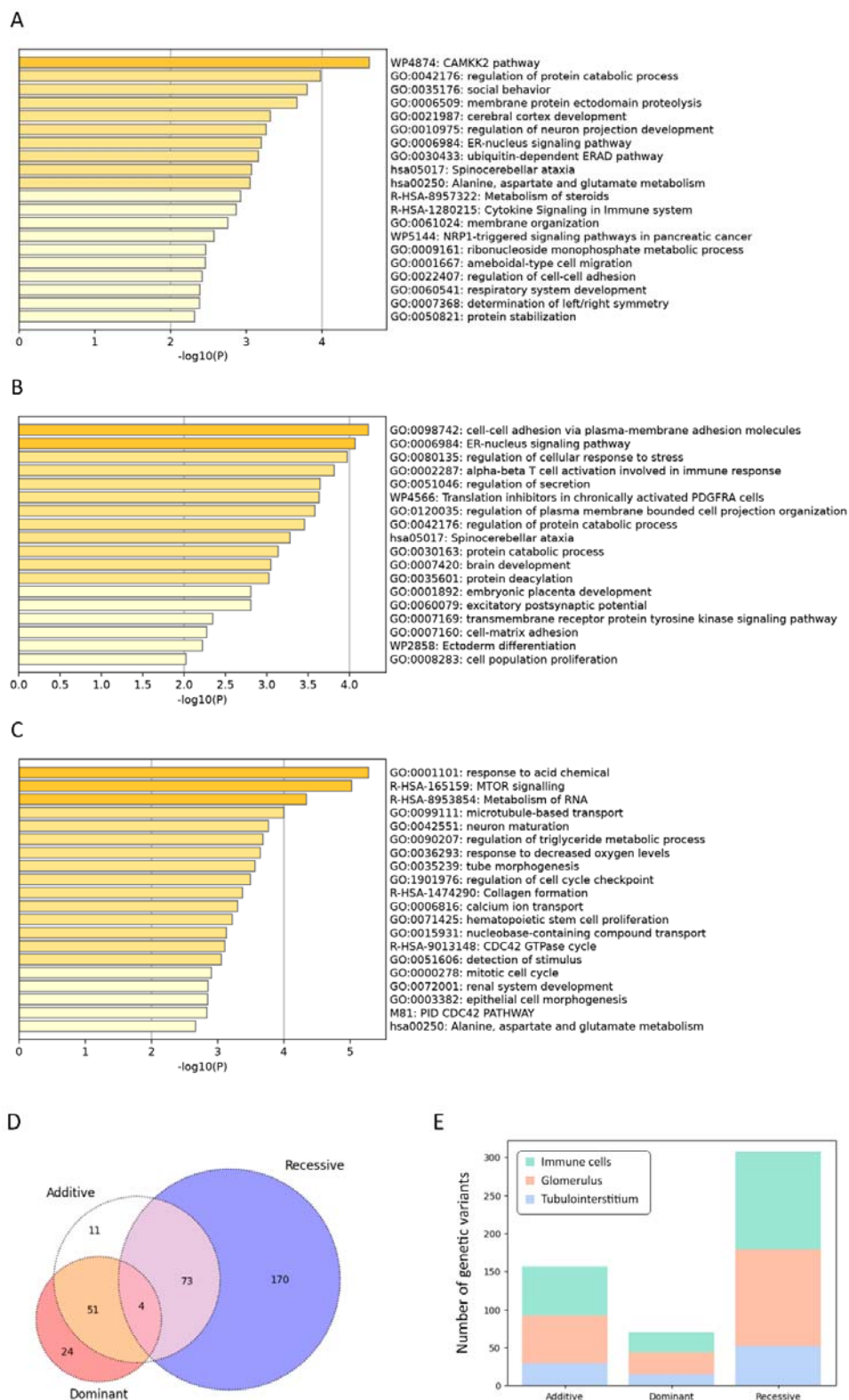


Figure 3. Statistically enriched terms using Metascape, additive (A), dominant (B) and recessive (C) logistic models. The intersection of genetic variant set based on different inheritance models,

a Venn diagram (D). Markers are interrogated against the following datasets: Database of Immune Cell eQTLs (DICE) project and NephQTL (E).

Protein-Protein Interaction (PPI) network analysis of potential targets

The PPI networks were obtained using the String online platform. None of the networks had PPI enrichment $P < 0.05$, the lowest value belonging to the recessive model with $P = 0.0613$. Its PPI contained 205 nodes, 160 edges, and the average node degree equal to 1.56. After clustering with k-means into 3 clusters, the most enriched cluster was the cluster of the recessive model with $P = 3.09 \times 10^{-14}$ (average node degree was 1.32) (figure 4).

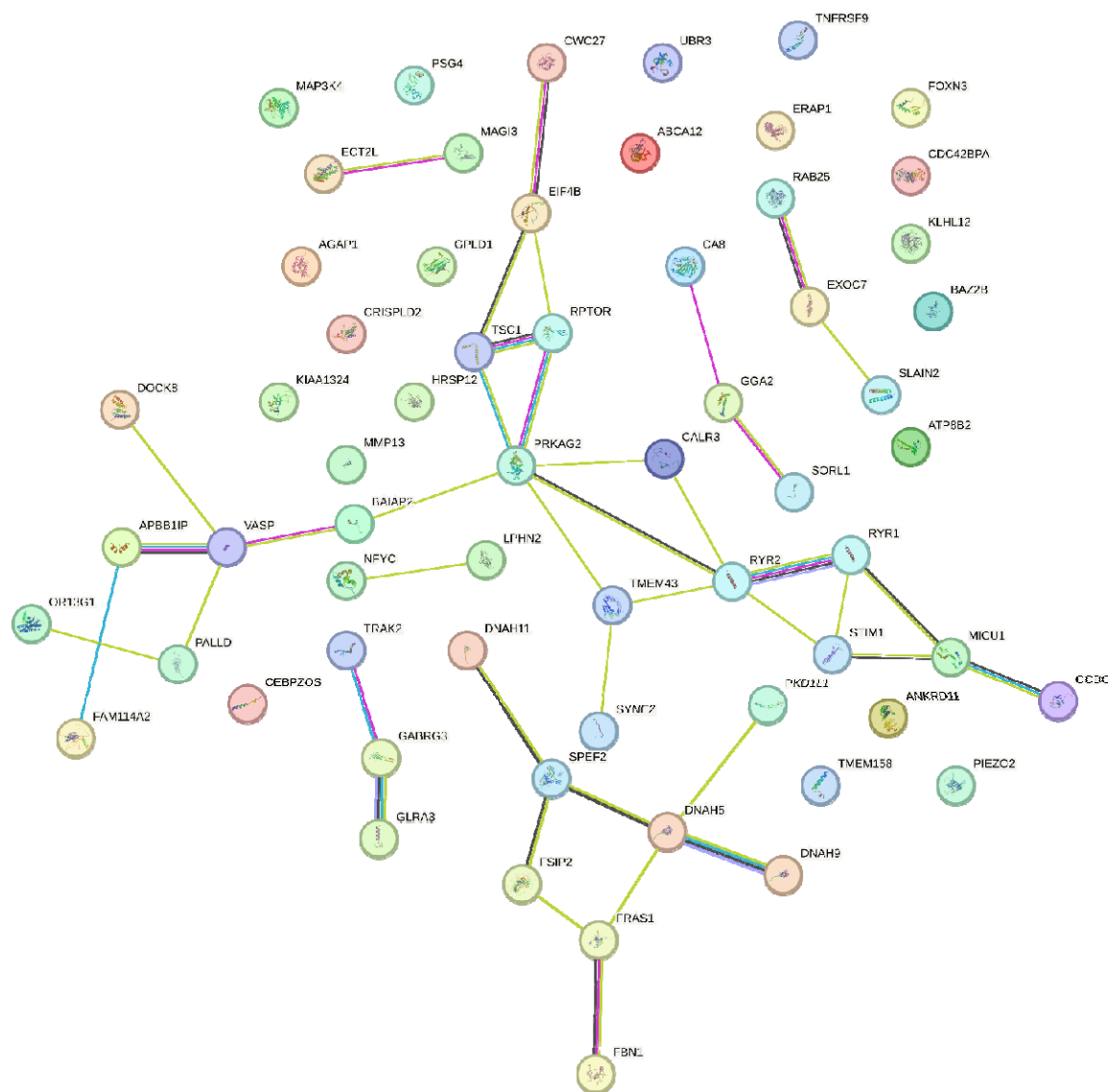


Figure 4. The STRING protein-protein interaction network based on gene interactions, recessive model, one of the three clusters.

Transcription factor (TF) binding sites' enrichment analysis

We identified numerous regions with predicted TF binding sites near SNVs that could affect gene expression (table S3). 11 SNVs enhance and 2 SNVs disrupt transcription factor binding sites.

Frequency and comparison of allele frequencies between IgAN patients and healthy donors

Table 3 shows the most frequent HLA alleles in the IgAN patient samples. Among them, we detected 142 unique alleles: 22 HLA-A, 34 HLA-B, 22 HLA-C, 28 HLA-DRB1, 7 HLA-DOA1, 13 HLA-DOB1, and 16 HLA-DPR1 alleles. The loci HLA-C (expected heterozygosity =

92.41%, observed heterozygosity = 90.00%, $P = 0.0036$) and HLA-DRB1 (expected heterozygosity = 92.65%, observed heterozygosity = 84.29%, $P = 0.0013$) exhibited statistically significant deviations from the Hardy-Weinberg equilibrium (table S4). In the healthy donors' group, the most frequent alleles were HLA-A*02:01:01G (27.13%), HLA-B*07:02:01G (11.79%), HLA-C*07:02:01G (13.22%), HLA-DRB1*07:01:01G (13.36%), HLA-DQA1*05:01:01G (24.64%), HLA-DQB1*03:01:01G (22.12%) and HLA-DPB1*04:01:01G (44.19%) (table S5). Allele HLA-B*56:01:01G (OR, 4.331 [95% CI: 1.739-9.437]) demonstrated the strongest association with IgAN, although it did not reach statistical significance. In the Russian population, this allele was observed only among 1.5% of Nizhnyi Novgorod citizens [62]. Alleles HLA-C*01:02:01G ($P = 0.029$; OR, 2.727 [95% CI, 1.477-4.738]), HLA-DRB1*01:01:01G ($P = 0.034$; OR, 2.189 [95% CI, 1.373-3.389]), HLA-DQA1*01:01:01G ($P = 0.0076$; OR, 2.021 [95% CI, 1.322-3.048]) and HLA-DQB1*05:01:01G ($P = 0.01$; OR, 2.124 [95% CI: 1.383-3.195]) were significantly more frequent among IgAN patients.

Discussion

Our study is the first Russian EWAS conducted on children with IgAN that employs the same population-based control cohort, takes insertions/deletions into consideration as well as utilizes HLA typing with up to 3-field resolution. The limitation of the study is a relatively small patient cohort size, despite applying the strict Bonferroni criterion.

Epidemiological data on the IgAN frequency in Russia are scarce. According to the statistics from the period of 1999-2019, IgAN was the most common type of immune glomerulopathies (41.5%). It was detected in 1 out of every 4 kidney biopsies, indicating a higher prevalence compared to Asia, Europe, and America [63]. The annual incidence of IgAN is approximately 8-10 cases per 1 million, and the average age of the onset is 34 ± 12 years. Therefore, the size of the children cohort prospectively recruited in Moscow for 3 years aligns with the expected size.

The most prominent signals were detected in the loci that had not been described before in IgAN patients. However, through eQTL analysis, we discovered that the variants affect gene expression in the glomeruli and renal tubules and confirmed their connection to significant Metascape categories related to immunity and kidney development. Moreover, the functional enrichment analysis revealed signaling pathways associated with nervous system development, which surprisingly had been already reported. *ANKRD16* was suggested as a candidate gene for IgAN in the Korean population [64], although its mutant protein was reported to be associated with the Purkinje cell degeneration [65] and had not been linked to the renal disorders before.

One of the examples that might help explain these findings is the *CYLD* gene. Its gain-of-function mutations was observed in the Alzheimer and Parkinson patients, while its loss-of-function mutations were discovered in the patients with the benign skin neoplasms, with no observed overlap between mutant phenotypes. Another possible explanation is phenotypic heterogeneity. We obtained the results comparable to the global data only by studying the HLA loci [17, 34]. In the patient group, we identified a slight predominance of the HLA-DRB1*03:01:01 [26] and HLA-DQB1*03:01 [28] alleles, although it was not statistically significant.

The rs143409664 variant is VUS since it produces an elongated protein while the alteration in the nucleotide sequence does not reside in the regions with repeats. In the Russian population, the rs143409664-C allele has frequencies of 0.5211 among healthy individuals and 0.5242 among the patients according to RuSeq [67]. *PRAG1* knockout in mice is lethal [68], and this gene has been described in several papers. Although its expression was detected in kidneys as well [69], there is no reliable evidence supporting the association of *PRAG1* with the disease. The frequency of the rs5828672 variant in the *ZNF787* gene is 0.9085, however, it was absent in RuSeq, which may indirectly imply the inexplicable predominance of the allele in IgAN patients. This variant did not show any linkage disequilibrium as well. *ZNF787* has been characterized to a certain extent; it possesses a proven ability (it is known to be able) to suppress the neuron growth and differentiation from iPSC [70]. The frequency of the *UBR3* gene rs13028230-G allele is 0.639, according to gnomAD, indicating the lower representation of the A allele; however, no information on it could be found in RuSeq. The highest value of r^2 , the indicator of the linkage disequilibrium, in our sample belonged to rs72891954 (0.327), which slightly deviates from the value in the global population (0.282, $P < 0.0001$) [71]. Therefore, the hypothesis suggesting that the *UBR3* allele belongs to a haplotype due to population shifts does not appear convincing. The *UBR3* mRNA levels in patients with lupus nephritis correlated with SLEDAI-2K and the index of histological activity [72]. The repertoires of the B and T cell receptors were studied in IgAN patients, and some types were found to be associated with the disease [73]. We identified the previously undescribed variant rs767448033 in the *TRBV5-4* gene. The expression of the transcription factor *EOMES* was higher in the circulating CD8 T cells (59%) compared to CD4 T cells (15%); however, in the kidney tissue, its expression in CD4 T cells was higher than in blood (32%) [74]. An elevated *SREBF1* expression was detected in patients with the chronic renal disease [75]. The *CHRNA3* gene variants correlated with eGFR, albumin-creatinine ratio, and albuminuria. Although *OR10G6* gene is related to the olfactory system, its variant rs1453654 was associated with the elevated expression of $IFN\gamma$ after smallpox vaccination [77]. In the rat model of hypertension disease, *Piezo2* activation was

observed in mesangial cells, renin cells, and perivascular mesenchymal cells implying its contribution to renal fibrosis [78]. Based on the dominant model, we identified the *ITGAM* gene among the IgAN candidate genes with repeatedly confirmed association [79]. Its variant rs60662530 had $P_{\text{BONF}} = 0.046$, however, the frequency ratio did not support the suggestion that the -CTTG allele was more frequent among patients (OR, 2.887 [95% CI: 0.1893-1.992]), as the lower CI was less than 1.

We did not study the external risk factors, therefore, we can rely only on the ratio of the allele frequencies and genotypes from the samples.

Although there are quite a few studies on genetics of IgAN patients, the role of particular genes as well as signaling pathways in the disease pathogenesis remains elusive. This complexity may be attributed to the highly heterogeneous interactions of hereditary factors or ethnospecificity of their manifestation. Answering these questions require further research on larger samples utilizing more powerful molecular and genetic techniques such as GWAS and transcriptome comparison.

Data availability statement:

The data underlying this article will be shared on reasonable request to the corresponding author.

Funding:

This research was funded by grant №075-15-2019-1789 from the Ministry of Science and Higher Education of the Russian Federation allocated to the Center for Precision Genome Editing and Genetic Technologies for Biomedicine.

Authors' contributions:

Conceptualization: MVP, EKP; formal analysis: AAB, VVC; visualization: AAB; resources: MVP, DOK; software: VVC, ASP, IAV, ONS; acquisition of data: VAB, AOS; project administration: EKP, DOK; supervision: DVR; writing – original draft preparation: AAB, MVP; writing – review & editing: EKP, DOK. Each author contributed important intellectual content during manuscript drafting or revision and agrees to be personally accountable for the individual's own contributions and to ensure that questions pertaining to the accuracy or integrity of any portion of the work, even one in which the author was not directly involved, are appropriately investigated and resolved, including with documentation in the literature if appropriate.

Conflict of interest statement:

The authors declare that they have no relevant financial interests.

References:

1. Wyatt RJ, Julian BA. IgA nephropathy. *N Engl J Med*. 2013;368(25):2402-2414. doi:10.1056/NEJMra1206793
2. Lai KN, Tang SC, Schena FP, et al. IgA nephropathy. *Nat Rev Dis Primers*. 2016;2:16001. Published 2016 Feb 11. doi:10.1038/nrdp.2016.1
3. Schena FP, Nistor I. Epidemiology of IgA Nephropathy: A Global Perspective. *Semin Nephrol*. 2018;38(5):435-442. doi:10.1016/j.semnephrol.2018.05.013
4. Rodrigues JC, Haas M, Reich HN. IgA Nephropathy. *Clin J Am Soc Nephrol*. 2017;12(4):677-686. doi:10.2215/CJN.07420716
5. Fabiano RC, Pinheiro SV, Simões E Silva AC. Immunoglobulin A nephropathy: a pathophysiology view. *Inflamm Res*. 2016;65(10):757-770. doi:10.1007/s00011-016-0962-x
6. Working Group of the International IgA Nephropathy Network and the Renal Pathology Society, Coppo R, Troyanov S, et al. The Oxford IgA nephropathy clinicopathological classification is valid for children as well as adults. *Kidney Int*. 2010;77(10):921-927. doi:10.1038/ki.2010.43
7. Coppo R. Pediatric IgA Nephropathy in Europe. *Kidney Dis (Basel)*. 2019;5(3):182-188. doi:10.1159/000495751
8. Zhou FD, Zhao MH, Zou WZ, Liu G, Wang H. The changing spectrum of primary glomerular diseases within 15 years: a survey of 3331 patients in a single Chinese centre. *Nephrol Dial Transplant*. 2009;24(3):870-876. doi:10.1093/ndt/gfn554
9. Li LS, Liu ZH. Epidemiologic data of renal diseases from a single unit in China: analysis based on 13,519 renal biopsies. *Kidney Int*. 2004;66(3):920-923. doi:10.1111/j.1523-1755.2004.00837.x
10. Pan X, Xu J, Ren H, et al. Changing spectrum of biopsy-proven primary glomerular diseases over the past 15 years: a single-center study in China. *Contrib Nephrol*. 2013;181:22-30. doi:10.1159/000348638
11. Barratt J, Feehally J. IgA nephropathy. *J Am Soc Nephrol*. 2005;16(7):2088-2097. doi:10.1681/ASN.2005020134
12. Rollino C, Vischini G, Coppo R. IgA nephropathy and infections. *J Nephrol*. 2016;29(4):463-468. doi:10.1007/s40620-016-0265-x

13. Kuroyanagi Y, Takeda A, Kinashi H, et al. Galactose-deficient IgA1 Is Involved in IgA Deposition in Renal Grafts Biopsied One Hour after Kidney Transplantation. *Intern Med.* 2023;62(11):1581-1589. doi:10.2169/internalmedicine.0141-22
14. Neugut YD, Kiryluk K. Genetic Determinants of IgA Nephropathy: Western Perspective. *Semin Nephrol.* 2018;38(5):443-454. doi:10.1016/j.semnephrol.2018.05.014
15. Levy M, Berger J. Worldwide perspective of IgA nephropathy. *Am J Kidney Dis.* 1988;12(5):340-347. doi:10.1016/s0272-6386(88)80021-0
16. O'Shaughnessy MM, Hogan SL, Thompson BD, Coppo R, Fogo AB, Jennette JC. Glomerular disease frequencies by race, sex and region: results from the International Kidney Biopsy Survey. *Nephrol Dial Transplant.* 2018;33(4):661-669. doi:10.1093/ndt/gfx189
17. Kiryluk K, Li Y, Scolari F, et al. Discovery of new risk loci for IgA nephropathy implicates genes involved in immunity against intestinal pathogens. *Nat Genet.* 2014;46(11):1187-1196. doi:10.1038/ng.3118
18. Julian BA, Quiggins PA, Thompson JS, Woodford SY, Gleason K, Wyatt RJ. Familial IgA nephropathy. Evidence of an inherited mechanism of disease. *N Engl J Med.* 1985;312(4):202-208. doi:10.1056/NEJM198501243120403
19. Wyatt RJ, Rivas ML, Julian BA, et al. Regionalization in hereditary IgA nephropathy. *Am J Hum Genet.* 1987;41(1):36-50.
20. Izzi C, Sanna-Cherchi S, Prati E, et al. Familial aggregation of primary glomerulonephritis in an Italian population isolate: Valtrompia study. *Kidney Int.* 2006;69(6):1033-1040. doi:10.1038/sj.ki.5000185
21. Scolari F, Amoroso A, Savoldi S, et al. Familial clustering of IgA nephropathy: further evidence in an Italian population. *Am J Kidney Dis.* 1999;33(5):857-865. doi:10.1016/s0272-6386(99)70417-8
22. Schena FP, Cerullo G, Rossini M, Lanzilotta SG, D'Altri C, Manno C. Increased risk of end-stage renal disease in familial IgA nephropathy. *J Am Soc Nephrol.* 2002;13(2):453-460. doi:10.1681/ASN.V132453
23. Gharavi AG, Yan Y, Scolari F, et al. IgA nephropathy, the most common cause of glomerulonephritis, is linked to 6q22-23. *Nat Genet.* 2000;26(3):354-357. doi:10.1038/81677
24. Bisceglia L, Cerullo G, Forabosco P, et al. Genetic heterogeneity in Italian families with IgA nephropathy: suggestive linkage for two novel IgA nephropathy loci. *Am J Hum Genet.* 2006;79(6):1130-1134. doi:10.1086/510135

25. Yu HH, Chu KH, Yang YH, et al. Genetics and immunopathogenesis of IgA nephropathy. *Clin Rev Allergy Immunol*. 2011;41(2):198-213. doi:10.1007/s12016-010-8232-0
26. Cao HX, Li M, Nie J, Wang W, Zhou SF, Yu XQ. Human leukocyte antigen DRB1 alleles predict risk and disease progression of immunoglobulin A nephropathy in Han Chinese. *Am J Nephrol*. 2008;28(4):684-691. doi:10.1159/000122111
27. Fennessy M, Hitman GA, Moore RH, et al. HLA-DQ gene polymorphism in primary IgA nephropathy in three European populations. *Kidney Int*. 1996;49(2):477-480. doi:10.1038/ki.1996.67
28. Raguénès O, Mercier B, Clèdes J, Whebe B, Férec C. HLA class II typing and idiopathic IgA nephropathy (IgAN): DQB1*0301, a possible marker of unfavorable outcome. *Tissue Antigens*. 1995;45(4):246-249. doi:10.1111/j.1399-0039.1995.tb02447.x
29. Xia YF, Huang S, Li X, et al. A family-based association study of megsin A23167G polymorphism with susceptibility and progression of IgA nephropathy in a Chinese population. *Clin Nephrol*. 2006;65(3):153-159. doi:10.5414/cnp65153
30. Li YJ, Du Y, Li CX, et al. Family-based association study showing that immunoglobulin A nephropathy is associated with the polymorphisms 2093C and 2180T in the 3' untranslated region of the Megsin gene. *J Am Soc Nephrol*. 2004;15(7):1739-1743. doi:10.1097/01.asn.0000130858.00688.46
31. Wang W, Sun Y, Fu Y, Yu X, Li M. Interaction of C1GALT1-IL5RA on the susceptibility to IgA nephropathy in Southern Han Chinese. *J Hum Genet*. 2013;58(1):40-46. doi:10.1038/jhg.2012.133
32. Wang W, Sun Y, Fu Y, Yu X, Li M. The effects of both single-locus and multi-locus interaction on the clinical manifestations of IgA nephropathy in Southern Han Chinese. *Nephrol Dial Transplant*. 2014;29(3):550-555. doi:10.1093/ndt/gft468
33. Zhu L, Tang W, Li G, et al. Interaction between variants of two glycosyltransferase genes in IgA nephropathy. *Kidney Int*. 2009;76(2):190-198. doi:10.1038/ki.2009.99
34. Feehally J, Farrall M, Boland A, et al. HLA has strongest association with IgA nephropathy in genome-wide analysis. *J Am Soc Nephrol*. 2010;21(10):1791-1797. doi:10.1681/ASN.2010010076
35. Gharavi AG, Kiryluk K, Choi M, et al. Genome-wide association study identifies susceptibility loci for IgA nephropathy. *Nat Genet*. 2011;43(4):321-327. Published 2011 Mar 13. doi:10.1038/ng.787
36. Li M, Yu XQ. Genetic Determinants of IgA Nephropathy: Eastern Perspective. *Semin Nephrol*. 2018;38(5):455-460. doi:10.1016/j.semnephrol.2018.05.015

37. Li M, Foo JN, Wang JQ, et al. Identification of new susceptibility loci for IgA nephropathy in Han Chinese. *Nat Commun*. 2015;6:7270. Published 2015 Jun 1. doi:10.1038/ncomms8270
38. Belova V, Pavlova A, Afasizhev R, et al. System analysis of the sequencing quality of human whole exome samples on BGI NGS platform. *Sci Rep*. 2022;12(1):609. Published 2022 Jan 12. doi:10.1038/s41598-021-04526-8
39. Andrews, S. FastQC: a quality control tool for high throughput sequence data. 2017.
40. <https://github.com/BioInfoTools/BBMap>
41. Fast and accurate short read alignment with Burrows–Wheeler transform. 2009.
42. <https://github.com/bwa-mem2/bwa-mem2>
43. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-2079. doi:10.1093/bioinformatics/btp352
44. <https://broadinstitute.github.io/picard/>
45. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011;27(21):2987-2993. doi:10.1093/bioinformatics/btr509
46. Tan A, Abecasis GR, Kang HM. Unified representation of genetic variants. *Bioinformatics*. 2015;31(13):2202-2204. doi:10.1093/bioinformatics/btv112
47. Li Q, Wang K. InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. *Am J Hum Genet*. 2017;100(2):267-280. doi:10.1016/j.ajhg.2017.01.004
48. Quinlan AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics*. 2014;47:11.12.1-11.12.34. Published 2014 Sep 8. doi:10.1002/0471250953.bi1112s47
49. <https://github.com/chrchang/plink-ng/tree/master/1.9>
50. Zhou Y, Zhou B, Pache L, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*. 2019;10(1):1523. Published 2019 Apr 3. doi:10.1038/s41467-019-09234-6
51. Qiu C, Huang S, Park J, et al. Renal compartment-specific genetic variation analyses identify new pathways in chronic kidney disease. *Nat Med*. 2018;24(11):1721-1731. doi:10.1038/s41591-018-0194-4
52. Gillies CE, Putler R, Menon R, et al. An eQTL Landscape of Kidney Tissue in Human Nephrotic Syndrome. *Am J Hum Genet*. 2018;103(2):232-244. doi:10.1016/j.ajhg.2018.07.004

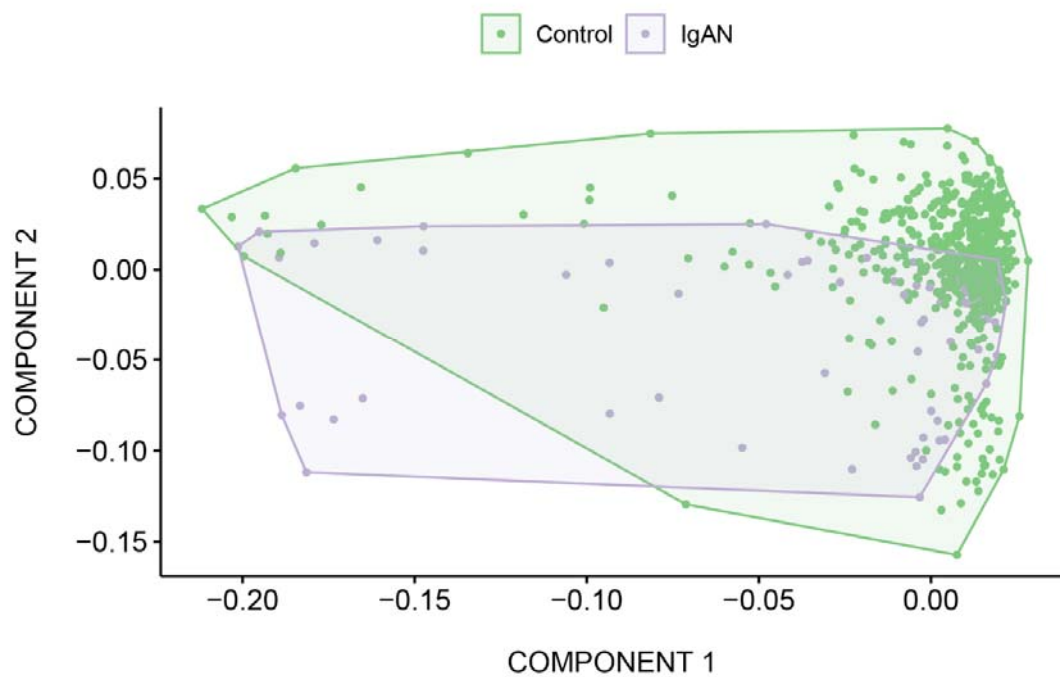
53. Schmiedel BJ, Singh D, Madrigal A, et al. Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell*. 2018;175(6):1701-1715.e16.
doi:10.1016/j.cell.2018.10.022
54. Zuo C, Shin S, Keleş S. atSNP: transcription factor binding affinity testing for regulatory SNP detection. *Bioinformatics*. 2015;31(20):3353-3355.
doi:10.1093/bioinformatics/btv328
55. Kawaguchi S, Higasa K, Shimizu M, Yamada R, Matsuda F. HLA-HD: An accurate HLA typing algorithm for next-generation sequencing data. *Hum Mutat*. 2017;38(7):788-797. doi:10.1002/humu.23230
56. Ka S, Lee S, Hong J, et al. HLAScan: genotyping of the HLA region using next-generation sequencing data. *BMC Bioinformatics*. 2017;18(1):258. Published 2017 May 12. doi:10.1186/s12859-017-1671-3
57. Lee H, Kingsford C. Kourami: graph-guided assembly for novel human leukocyte antigen allele discovery. *Genome Biol*. 2018;19(1):16. Published 2018 Feb 7.
doi:10.1186/s13059-018-1388-2
58. Robinson J, Barker DJ, Georgiou X, Cooper MA, Flicek P, Marsh SGE. IPD-IMGT/HLA Database. *Nucleic Acids Res*. 2020;48(D1):D948-D955. doi:10.1093/nar/gkz950
59. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 2013;14(2):178-192. doi:10.1093/bib/bbs017
60. Kim HY. Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. *Restor Dent Endod*. 2017;42(2):152-155. doi:10.5395/rde.2017.42.2.152
61. Excoffier L, Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour*. 2010;10(3):564-567. doi:10.1111/j.1755-0998.2010.02847.x
62. Khamaganova EG, Leonov EA, Abdrakhimova AR, et al. HLA diversity in the Russian population assessed by next generation sequencing. *Medical Immunology (Russia)*. 2021;23(3):509-522. (In Russ.) doi:10.15789/1563-0625-HDI-2182
63. Dobronravov VA, Muzhetskaya TO, Lin DI, Kochoyan ZS. Immunoglobulin A-nephropathy in Russian population: clinical and morphological presentation and long-term prognosis. *Nephrology (Saint-Petersburg)*. 2019;23(6):45-60. doi:10.36485/1561-6274-2019-236-45-60
64. Jeong KH, Kim JS, Lee YH, et al. Genome-wide association study identifies new susceptible loci of IgA nephropathy in Koreans. *BMC Med Genomics*. 2019;12(1):122. Published 2019 Aug 19. doi:10.1186/s12920-019-0568-6

65. Vo MN, Terrey M, Lee JW, et al. ANKRD16 prevents neuron loss caused by an editing-defective tRNA synthetase [published correction appears in *Nature*. 2018 Aug;560(7720):E35]. *Nature*. 2018;557(7706):510-515. doi:10.1038/s41586-018-0137-8
66. Zajicek AS, Ruan H, Dai H, et al. Cylindromatosis drives synapse pruning and weakening by promoting macroautophagy through Akt-mTOR signaling. *Mol Psychiatry*. 2022;27(5):2414-2424. doi:10.1038/s41380-022-01571-1
67. Barbitoff YA, Khmelkova DN, Pomerantseva EA, Slepchenkov AV, Zubashenko NA, Mironova IV, et al. Expanding the Russian allele frequency reference via cross-laboratory data integration: insights from 6,096 exome samples. medRxiv. Preprint posted online November 4, 2021.
68. Weaver KL, Alves-Guerra MC, Jin K, et al. NACK is an integral component of the Notch transcriptional activation complex and is critical for development and tumorigenesis. *Cancer Res*. 2014;74(17):4741-4751. doi:10.1158/0008-5472.CAN-14-1547
69. Tanaka H, Katoh H, Negishi M. Pragmin, a novel effector of Rnd2 GTPase, stimulates RhoA activity. *J Biol Chem*. 2006;281(15):10355-10364. doi:10.1074/jbc.M511314200
70. Kutsche LK, Gysi DM, Fallmann J, et al. Combined Experimental and System-Level Analyses Reveal the Complex Regulatory Network of miR-124 during Human Neurogenesis. *Cell Syst*. 2018;7(4):438-452.e8. doi:10.1016/j.cels.2018.08.011
71. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*. 2015;31(21):3555-3557. doi:10.1093/bioinformatics/btv402
72. Aguirre-Valencia D, Ríos-Serna LJ, Posso-Osorio I, et al. Expression of BAFF, APRIL, and cognate receptor genes in lupus nephritis and potential use as urinary biomarkers. *J Transl Autoimmun*. 2019;3:100027. Published 2019 Dec 17. doi:10.1016/j.jtauto.2019.100027
73. Ou M, Zheng F, Zhang X, et al. Integrated analysis of B-cell and T-cell receptors by high-throughput sequencing reveals conserved repertoires in IgA nephropathy. *Mol Med Rep*. 2018;17(5):7027-7036. doi:10.3892/mmr.2018.8793
74. van der Putten C, Remmerswaal EBM, Terpstra ML, et al. CD8 and CD4 T Cell Populations in Human Kidneys. *Cells*. 2021;10(2):288. Published 2021 Feb 1. doi:10.3390/cells10020288
75. Dorotea D, Koya D, Ha H. Recent Insights Into SREBP as a Direct Mediator of Kidney Fibrosis via Lipid-Independent Pathways. *Front Pharmacol*. 2020;11:265. Published 2020 Mar 17. doi:10.3389/fphar.2020.00265

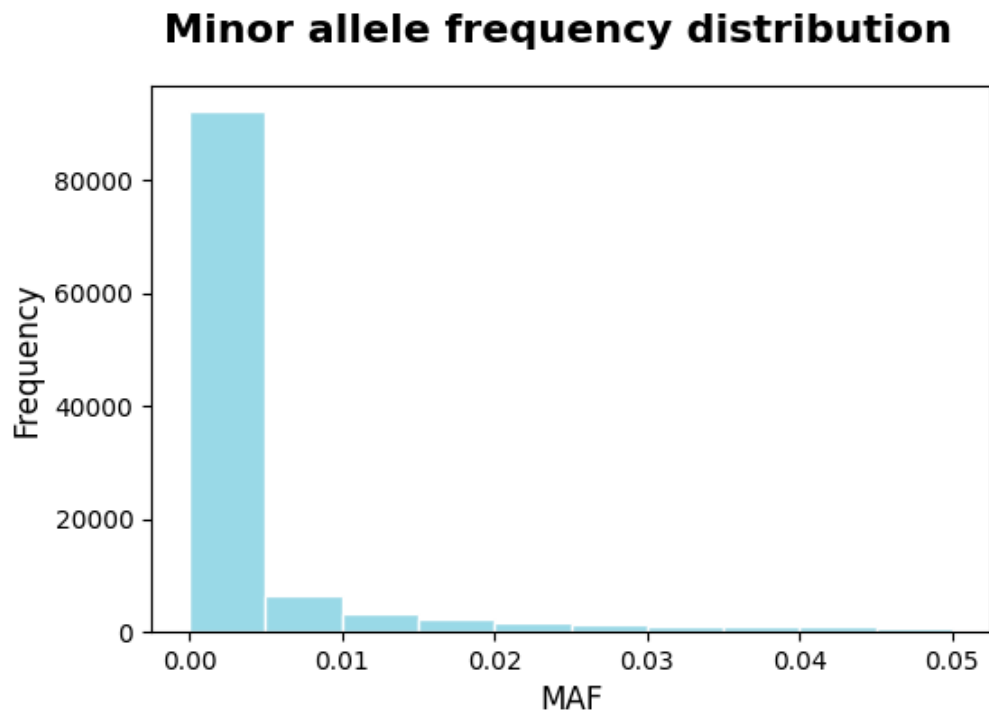
76. Zhu Y, Yang J, Li S, et al. Genetic variants in nicotinic acetylcholine receptor genes jointly contribute to kidney function in American Indians: the Strong Heart Family Study. *J Hypertens*. 2014;32(5):1042-1049. doi:10.1097/HJH.000000000000151
77. Kennedy RB, Ovsyannikova IG, Pankratz VS, et al. Genome-wide genetic associations with IFN γ response to smallpox vaccine. *Hum Genet*. 2012;131(9):1433-1451. doi:10.1007/s00439-012-1179-x
78. Ochiai K, Mochida Y, Nagase T, Fukuhara H, Yamaguchi Y, Nagase M. Upregulation of Piezo2 in the mesangial, renin, and perivascular mesenchymal cells of the kidney of Dahl salt-sensitive hypertensive rats and its reversal by esaxerenone [published online ahead of print, 2023 Feb 21]. *Hypertens Res*. 2023;10.1038/s41440-023-01219-9. doi:10.1038/s41440-023-01219-9
79. Shi D, Zhong Z, Xu R, et al. Association of ITGAX and ITGAM gene polymorphisms with susceptibility to IgA nephropathy. *J Hum Genet*. 2019;64(9):927-935. doi:10.1038/s10038-019-0632-2

Supplementary Material

Supplementary Figure 1A. Principal component analysis of the IgAN samples comprising the samples from healthy donor cohort.



Supplementary Figure 1B. Minor allele frequency distribution in IgAN patients and healthy donors.



Supplementary Table 1. Clinical characteristics of the IgAN patients.

Features		(n = 70)
Sex	Male	64.4%
	Female	35.6%
Age, years		10.3±3.8
Serum creatinine level (mg/dL)		46.5±7.1
eGFR (mL/min/1.73m ²)		97.7±26.9
Proteinuria (g/day/1.73m ²)		0.9 [0.5-2.0]
Mean blood pressure (mm Hg)		84.5±12.42
Symptoms at the debut	persistent microscopic hematuria	20.0%
	macroscopic hematuria	38.6%
	nephritic syndrome	11.5%
	nephrotic syndrome	25.7%
	acute kidney injury	2.8%
	rapidly progressive glomerulonephritis	1.4%

Note: Conversion factors for units: serum creatinine in mg/dL to $\mu\text{mol/L}$, $\times 88.4$.

Supplementary Table 2. Key quality control metrics for exome sequencing of the experimental samples from the IgAN cohort.

Metrics	Mean	Min	Max
Single reads per sample	79,494,574	41,389,840	162,054,108
Estimated library size	167,316,756	43,904,883	362,437,565
Duplicates	12.33	5.86	26.82
On-target bases	85.3%	62%	92%
Mean target coverage	89.65	46.2	140
Median target coverage	80.62	42	128
Width 10X	96.9%	94.3%	97%
Width 20X	95.2%	85%	97%
Width 30X	91.2%	70%	96%

Supplementary Table 3. The TF binding motifs affected by potentially deleterious variants.

SNP	TF	Motif	TF binding site	P
rs3087873	SMAD	AP1_disc10	Enhance	0.00212
rs3087873	ESRRA	AP1_disc10	Enhance	0.00212
rs3087873	MYC	AP1_disc10	Enhance	0.00212
rs2286239	RAD21	RAD21_disc5	Enhance	0.0034
rs33749	ETS	RAD21_disc5	Enhance	0.00348
rs622861	REST	RUNX2_1	Enhance	0.00414
rs2209956	AP1	HLTF_1	Enhance	0.00858
rs11807205	SOX10	LEF1_1	Enhance	0.01723
rs11807205	TFAP2A	LEF1_1	Enhance	0.01723
rs11807205	GATA3	LEF1_1	Enhance	0.01723
rs11807205	E2F1	LEF1_1	Enhance	0.01723
rs11807205	SETDB1	LEF1_1	Enhance	0.01723
rs12495243	EN1	EN1_2	Enhance	0.01832
rs2042367	LEF1	GATA_1	Enhance	0.02353
rs33749	ZNF143	CHD2_disc3	Enhance	0.02499
rs4535042	TFAP2A	TFAP2A_1	Enhance	0.03305
rs4535042	TFAP2A	TFAP2A_4	Enhance	0.03305
rs12495243	VDR	VDR_1	Disrupt	0.03416
rs3087873	PITX2	REST_disc8	Enhance	0.03947
rs3087873	ELF1	REST_disc8	Enhance	0.03947
rs3087873	GATA2	REST_disc8	Enhance	0.03947
rs913178	VDR	EP300_disc9	Disrupt	0.0429
rs913178	AP1	EP300_disc9	Disrupt	0.0429
rs913178	BCL	EP300_disc9	Disrupt	0.0429

Note: TF – transcription factor, Motif – binding sites with high affinity to TF. P – p-value that statistically confirms potential stimulation or loss of function of the genomic region harboring SNP in terms of TF binding.

Supplementary Table 4. Conformity to Hardy-Weinberg equilibrium.

Locus	No. of Genotypes	Observed Heterozygosity	Expected Heterozygosity	p-value
IgAN				
<i>HLA-A</i>	70	0.88571	0.86228	0.41045
<i>HLA-B</i>	70	0.92857	0.96310	0.18927
<i>HLA-C</i>	70	0.90000	0.92405	0.00358
<i>HLA-DRB1</i>	70	0.84286	0.92652	0.00125
<i>HLA-DQA1</i>	70	0.67143	0.78335	0.05311
<i>HLA-DQA2</i>	70	0.72857	0.83659	0.09385
<i>HLA-DPB1</i>	70	0.78571	0.81408	0.18547

Supplementary Table 5. The distribution of the alleles HLA-A, -B, -DRB1, -DQA1, -DQB1 and -DPB1 in IgAN patients and healthy donors.

Allele	IgAN (%)	Control (%)	P' (chi-sq)	P' (Fisher's exact test)	OR	95% CI
A*02:01:01G	30.00	27.13	> 0.05		1.151	0.777-1.682
A*24:02:01G	16.43	10.98	> 0.05		1.594	0.961-2.543
A*01:01:01G	9.29	12.14	> 0.05		0.741	0.381-1.327
A*03:01:01G	8.57	13.95	> 0.05		0.578	0.289-1.055
A*31:01:02G	5.71	2.00		> 0.05	2.967	1.210-6.327
B*35:01:01G	9.29	6.35	> 0.05		1.508	0.770-2.725
B*27:05:02G	7.14	3.87	> 0.05		1.912	0.877-3.734
B*56:01:01G	5.71	1.38		> 0.05	4.331	1.739-

						9.437
B*51:01:01G	5.71	5.19	> 0.05		1.107	0.461- 2.291
B*07:02:01G	5.71	11.79	> 0.05		0.454	0.190- 0.929
C*04:01:01G	16.43	12.41	> 0.05		1.387	0.837- 2.210
C*01:02:01G	11.43	4.52	0.029		2.727	1.477- 4.738
C*02:02:02G	10.00	6.06	> 0.05		1.723	0.901- 3.0621
C*06:02:01G	9.29	11.95	> 0.05		0.754	0.387- 1.351
C*03:04:01G	6.43	5.16	> 0.05		1.261	0.556- 2.521
DRB1*01:01:01G	40.00	10.25	0.034		2.189	1.373- 3.389
DRB1*03:01:01G	9.29	7.73	> 0.05		1.221	0.625- 2.120
DRB1*13:01:01G	8.57	6.92	> 0.05		1.260	0.626- 2.319
DRB1*07:01:01G	7.86	13.36	> 0.05		0.553	0.267- 1.033
DRB1*11:04:01G	6.43	4.87	> 0.05		1.343	0.591- 2.686
DQA1*05:01:01G	30.00	24.64	> 0.05		1.310	0.869- 1.949
DQA1*01:01:01G	28.57	16.51	0.0076		2.021	1.322- 3.048
DQA1*03:01:01G	17.86	11.40	> 0.05		1.689	1.012- 2.733
DQA1*01:03:01G	10.00	9.13	> 0.05		1.105	0.567- 2.009
DQA1*02:01:01G	7.86	13.41	> 0.05		0.551	0.262-

						1.048
DQB1*03:01:01G	26.43	22.12	> 0.05		1.265	0.837- 1.875
DQB1*05:01:01G	24.29	13.12	0.01		2.124	1.383- 3.195
DQB1*02:01:01G	14.29	18.39	> 0.05		0.740	0.433- 1.204
DQB1*06:03:01G	8.57	7.17	> 0.05		1.214	0.603- 2.233
DQB1*03:02:01G	7.14	7.22	> 0.05		0.988	0.458- 1.906
DPB1*04:01:01G	32.14	44.19	> 0.05		0.598	0.408- 0.868
DPB1*04:02:01G	22.14	13.44	> 0.05		1.831	1.174- 2.786
DPB1*02:01:02G	15.71	14.66	> 0.05		1.086	0.649- 1.740
DPB1*03:01:01G	10.00	9.84	> 0.05		1.018	0.535- 1.795
DPB1*06:01:01G	3.57	1.43		> 0.05	2.546	0.782- 6.470

Note: chi-sq – basic allelic test chi-square (1df), OR – odds ratio for Fisher's exact test, 95% CI – odds ratio and confidence interval for odds ratio for Fisher's exact test, P' – Bonferroni single-step adjusted p-values.