Multi-population genome-wide association study implicates both immune and non-immune factors in the etiology of pediatric steroid sensitive nephrotic syndrome

Alexandra Barry[1,2+], Michelle T. McNulty[1,2+], Xiaoyuan Jia[3,4+], Yask Gupta[5+], Hanna Debiec[6+], Yang Luo[7,8,9,10], China Nagano[1,2,11], Tomoko Horinouchi[11], Seulgi Jung[12], Manuela Colucci[13], Dina F. Ahram[5], Adele Mitrotti[5,14], Aditi Sinha[15], Nynke Teeninga[16], Gina Jin[5], Shirlee Shril[17,18], Gianluca Caridi[19], Monica Bodria[20], Tze Y Lim[5], Rik Westland[21], Francesca Zanoni[6,22], Maddalena Marasa[5], Daniel Turudic[23], Mario Giordano[24], Loreto Gesualdo[14], Riccardo Magistroni[25,26], Isabella Pisani[27], Enrico Fiaccadori[27], Jana Reiterova[28], Silvio Maringhini[29], William Morello[30], Giovanni Montini[30,31], Patricia L. Weng[32], Francesco Scolari[33], Marijan Saraga[34], Velibor Tasic[35], Domenica Santoro[36], Joanna A.E. van Wijk[21], Danko Milošević[23], Yosuke Kawai[3,4], Krzysztof Kiryluk[5], Martin R. Pollak[37,38], Ali Gharavi[5], Fangmin Lin[38], Ana Cristina Simœs e Silva[39], Ruth J.F. Loos[40], Eimear E. Kenny[41,42,43], Michiel F. Schreuder[16], Aleksandra Zurowska[44], Claire Dossier[45], Gema Ariceta[46], Magdalena Drozynska-Duklas[44], Julien Hogan[45], Augustina Jankauskiene[47], Friedhelm Hildebrandt[1,18], Larisa Prikhodina[48], Kyuyoung Song[12], Arvind Bagga[15], Hae II Cheong[49], Gian Marco Ghiggeri[20] , Prayong Vachvanichsanong[50], Kandai Nozu[11], Marina Vivarelli[51], Soumya Raychaudhuri[8,9,10,52,53], Katsushi Tokunaga[3,4*], Simone Sanna-Cherchi[5*], Pierre Ronco[6,54*], Kazumoto Iijima[55,56*], Matthew G. Sampson[1,2,18*]

1. Division of Nephrology, Boston Children's Hospital, Boston, MA, USA
2. Kidney Disease Initiative, Broad Institute of MIT and Harvard, Cambridge, MA, USA
3. Genome Medical Science Project (Toyama), National Center for Global Health and Medicine (NCGM), Tokyo, Japan
4. Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan
5. Division of Nephrology, Department of Medicine, Columbia University College of Physicians and Surgeons, New York, NY, USA
6. Sorbonne Université, UPMC Paris 06, Institut National de la Santé et de la Recherde Médicale, Unité Mixte de Rechereche S 1155, Paris, France
7. Kennedy Institute of Rheumatology, University of Oxford, Roosevelt Drive, Headington, Oxford, OX3 7FY, United Kingdom.
8. Center for Data Sciences, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA.
9. Divisions of Genetics and Rheumatology, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA.
10. Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA
11. Department of Pediatrics, Kobe University Graduate School of Medicine, Kobe, Japan
12. Department of Biochemistry and Molecular Biology, University of Ulsan College of Medicine, Songpa-gu, Seoul, Korea
13. Renal Diseases Research Unit, Genetics and Rare Diseases Research Division, Istituto di Ricovero e Cura a Carattere Scientifico Ospedale Pediatrico Bambino Gesù, Rome, Italy
14. Nephrology, Dialysis and Transplantation Unit, Department of Emergency and Organ Transplantation, University of Bari Aldo Moro, Bari, Italy

15. Department of Pediatrics, AIIMS, New Delhi, India
16. Department of Pediatric Nephrology, Amalia Children's Hospital, Radboud University Medical Center, Nijmegen, The Netherlands
17. Department of Medicine, Boston Children's Hospital, Boston, MA, USA
18. Department of Pediatrics, Harvard Medical School, Boston, MA, USA
19. Laboratory on Molecular Nephrology, IRCCS Instituto Giannina Gaslini, Genoa, Italy
20. Department of Nephrology and Renal Transplantation, IRCCS Instituto Giannina Gaslini, Genoa, Italy
21. Department of Pediatric Nephrology, VU University Medical Center, Amsterdam, The Netherlands
22. Division of Transplantation, Department of Surgery, University of Pennsylvania, Philadelphia, PA
23. Department of Pediatric Nephrology, Dialysis and Transplantation, Clinical Hospital Hospital Center Zagreb, University of Zagreb Medical School, Zagreb, Croatia
24. Division of Nephrology and Pediatric Dialysis, Bari Polyclinic Giovanni XXIII Children's Hospital, Bari, Italy
25. Department of Nephrology, Dialysis and Transplant Unit, University Hospital of Modena, Modena, Italy.
26. Surgical, Medical and Dental Department of Morphological Sciences, Section of Nephrology, University of Modena and Reggio Emilia, Modena, Italy
27. Unità Operativa Nefrologia, Azienda Ospedaliero-Universitaria di Parma, Dipartimento di Medicina e Chirurgia, Università di Parma, Parma, Italy
28. Department of Nephrology, Medicine and General University Hospital, Charles University, Prague, Czech Republic
29. Department of Pediatrics, ISMETT, Palermo, Italy
30. Pediatric Nephrology, Dialysis and Transplant Unit, Fondazione IRCCS Ca' Granda-Ospedale Maggiore Policlinico, Milano, Italy
31. Department of Clinical Sciences and Community Health, University of Milan, Italy
32. Department of Pediatric Nephrology, UCLA Medical Center and UCLA Medical Center-Santa Monica, Los Angeles, CA, USA
33. Department of Medical and Surgical Specialties, Radiological Sciences, and Public Health, Division of Nephrology and Dialysis, University of Brescia and ASST Spedali Civili of Brescia, Brescia, Italy
34. Department of Pediatrics, University of Split, Split, Croatia
35. Department of Pediatric Nephrology, University Children's Hospital, Skopje, Macedonia
36. Division of Nephrology and Dialysis Unit, University of Messina, Sicily, Italy
37. Division of Nephrology, Beth Israel Deaconess Medical Center, Boston, MA, USA.
38. Department of Pediatric, Division of Pediatric Nephrology, Columbia University Irving Medical Center New York-Presbyterian Morgan Stanley Children's Hospital in New York, NY, USA
39. Department of Pediatrics, Interdisciplinary Laboratory of Medical Investigation, Faculty of Medicine, Federal University of Minas Gerais, Belo Horizonte, Brazil
40. The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA
41. Institute for Genomic Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA
42. Division of Genomic Medicine, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA
43. Division of General Internal Medicine, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA
44. Department of Pediatrics, Nephrology and Hypertension, Medical University Gdansk, Poland
45. AP-HP, Pediatric Nephrology Department, Hôpital Robert-Debré, Paris, France

46. Pediatric Nephrology, Hospital Universitari Vall d'Hebron, Universitat Autónoma de Barcelona, Barcelona, Spain
47. Institute of Clinical Medicine, Faculty of Medicine, Vilnius University, Vilnius, Lithuania
48. Research and Clinical Institute for Pediatrics, Pirogov Russian National Research Medical University, Taldomskava St., 2, Moscow, Russia
49. Department of Pediatrics, Hallym University Sacred Heart Hospital, 22, Gwanpyeong-ro 170 beon-gil, Dongan-gu, Anyang-si, Gyeonggi-do, 14068 Korea
50. Department of Pediatrics, Faculty of Medicine, Prince of Songkla University, Hat-Yai, Songkhla 90110, Thailand
51. Division of Nephrology, and Dialysis, Department of Pediatric Subspecialities, Istituto di Ricovero e Cura a Carattere Scientifico Ospedale Pediatrico Bambino Gesù, Rome, Italy
52. Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA
53. Centre for Genetics and Genomics Versus Arthritis, University of Manchester, Manchester, UK
54. Department of Nephrology, Centre Hospitalier du Mans, Le Mans, France
55. Hyogo Prefectural Kobe Children's Hospital, Kobe, Japan
56. Department of Advanced Pediatric Medicine, Kobe University Graduate School of Medicine, Kobe, Japan

+ These authors contributed equally
* Co-senior authors
Please address correspondences to:
Matt Sampson, MD MSCE
Enders 509
300 Longwood Ave
Boston, MA 02115
matthew.sampson@childrens.harvard.edu

ABSTRACT

Pediatric steroid-sensitive nephrotic syndrome (pSSNS) is the most common childhood glomerular disease. Previous genome-wide association studies (GWAS) identified a risk locus in the HLA Class II region and three additional signals. But the genetic architecture of pSSNS, and its genetically driven pathobiology, is largely unknown. We conducted a multi-population GWAS meta-analysis in 38,463 participants (2,440 cases) and population specific GWAS, discovering twelve significant associations (eight novel). Fine-mapping implicated specific amino acid haplotypes in HLA-DQA1 and HLA-DQB1 driving the HLA Class II risk signal. Non-HLA loci colocalized with eQTLs of monocytes and numerous T-cell subsets in independent datasets. Colocalization with kidney eQTLs was lacking, but overlap with kidney cell open chromatin suggests an uncharacterized disease mechanism in kidney cells. A polygenic risk score (PRS) associated with earlier disease onset in two independent cohorts. Altogether, these discoveries expand our knowledge of pSSNS genetic architecture across populations and provide cellspecific insights into its molecular drivers.

pSSNS is a rare disease of glomerular filtration barrier failure. Its incidence ranges from 0.96-13.5/100,000, being most frequent in South Asian and East Asian populations[1]. pSSNS causes massive proteinuria and increased risk of thromboembolism, sepsis, and progression to chronic kidney disease (CKD)/end-stage kidney disease (ESKD)[2–7]. And those progressing to ESKD have increased odds of recurrent NS in their allograft[8]. pSSNS is impactful across the lifespan - 31-50% of those affected have relapses in adulthood[9]. Much of pSSNS's morbidity is related to side effects of the non-specific immunosuppressants that allow some to achieve remission of their proteinuria[7,10–17].

There are no monogenic forms of pSSNS to illuminate its pathobiology. However, we know that immune dysregulation is a major contributor[18,19]. But determining causal immune factors via case-control studies of cytokines profiles, cell types, and transcriptomic signatures is challenging. The dynamic responses of the immune system at different disease stages and to various stimuli make it difficult to determine whether observed differences are causal, correlated, or due to independent biological/environmental factors. And kidney tissue in children is rarely available to determine intrarenal, molecular drivers of pSSNS.

GWAS have discovered four pSSNS risk loci[20–24]. The top signal in each was in the HLA Class II region. Two other loci are plausibly immune-related, with the closest genes being Calcium Homeostasis Modulator Family Member 6 (*CALHM6*)[25] and TNF Superfamily Member 15 (*TNFSF15)*. The lead SNP of the fourth locus was within nephrin (*NPHS1)*, a fundamental glomerular gene implicated in Mendelian NS[26]. These studies are illuminating but naturally by smaller sample sizes, primarily population-specific analyses, and limited post-GWAS analysis. Here, towards discovering a fuller spectrum of disease-associated genetic variation and unraveling its pathogenesis at the interface of the immune system and kidney, we conducted a large and diverse GWAS of pSSNS.

We conducted a multi-population, fixed-effect, inverse-variance, meta-analysis across twelve GWAS datasets comprised of 2,440 cases and 36,023 controls of Admixed American, African, East Asian, European, Maghrebian and South Asian populations (**Figure 1**, **Figure S1**, **Table S1**).To account for population-driven effect heterogeneity, we also performed a meta-regression with MR-MEGA[27]. Eight loci (four new, and all outside HLA) were significant (MR-MEGA $p < 5 \times 10^{-8}$) (**Table 1**, **Figure 2A**, **Figure S2**). The lead SNPs of the novel loci were all intronic: **(1)** rs7759971 in Abelson Helper Integration Site 1*(AHI1*; $p = 4.90 \times 10^{-12}$); **(2)** rs55730955 in CD28 molecule (*CD28*; $p = 4.27 \times 10^{-10}$); **(3)** rs8062322 in C-type Lectin Domain Containing 16A **(***CLEC16A*; $p = 1.61 \times 10^{-10}$); **(4)** rs28862935 in betacellulin **(***BTC*; $p = 1.08 \times 10^{-9}$). Four other significant loci were previously reported[24,23]. Two more significant loci emerged after conditioning: **(5)** rs1794497 upstream of *HLA-DQB1,* ($p = 6.79 \times 10^{-52}$); **(6)** rs2256318 in an intron of MHC Class I Chain-related Gene A (*MICA;* $p = 9.70 \times 10^{-18}$) (**Figure 2B**, **Figure S3**). Population specific GWAS meta-analysis discovered two additional significant loci in Europeans (**Figure 2C**, **Table S2**, **Table S3**, **Figure S4**): The lead SNPs were in introns of **(7)** rs111796602 in an intron of Engulfment and Cell Motility 1 (*ELMO1;* $p = 1.72 \times 10^{-8}$) and **(8)** rs12911841 in an intron of Mortality Factor 4 Like 1 (*MORF4L1*; $p = 3.88 \times 10^{-8}$). The *CALHM6* association appeared specific to Europeans *(*$P_{anc_het} = 4.99 \times 10^{-4}$) and the *TNFSF15* and *NPHS1* associations to East Asians *(*$P_{anc_het} = 7.76 \times 10^{-4}$ and $P_{anc_het} = 2.43\ 10^{-4}$, respectively) (**Figure S5, Figure S6**). All other loci had similar effects across populations (**Figure 2D**). Finally, there were 13 novel suggestive loci (MR-MEGA $p < 5 \times 10^{-6}$) in the multi-population GWAS (**Table S4, Table S5**). On a liability scale and excluding HLA, European heritability was 0.04 [CI: -0.08,0.16] and East Asian heritability was 0.12 [CI: 0.04,0.21], with large confidence intervals likely due to small effective sample sizes.

A number of insights emerged from evaluating disease associations, functions, and expression patterns of the lead SNPs and/or the closest genes at the novel non-HLA loci. First, PheWAS using Open

Target Genetics (http://genetics.opentargets.org)[28] found that SNPs at most loci associated with GWAS of diverse white blood cell traits, atopic disorders, and autoimmune conditions. For example, among the strongest associations with the lead SNPs at the following loci include: *CLEC16A, CD28, MICA*, and *ELMO1* with eosinophil counts; *AHI* with monocyte and neutrophil counts, asthma, and hay fever (also shared by *CD28*); and *MICA* with Type 1 Diabetes (T1DM).

Second, most of these genes, while primarily known for their role in immunity, also have known roles in kidney diseases and cells. Common *AHI1* variants are associated with atopy, lupus, and diverse immune cell traits[28]. Rare *AHI1* coding variants cause the monogenic ciliopathy Joubert Syndrome, which includes cystic kidney disease[29]. *ELMO1* participates in Rac1 pathway activation and actin cytoskeletal rearrangement[30], is expressed in podocytes[31], and is associated with diabetic nephropathy[32]. *CD28*, a T-cell glycoprotein, binds a co-stimulatory molecule B7-1 (CD80) on antigen-presenting cells. B7-1 is expressed on human podocytes in some forms of nephrotic syndrome, and blocking the B7-1/CD28 interaction with a CTLA-4 immunoglobulin can ameliorate proteinuria[33]. *MICA* is expressed in kidney endothelium, binds and activates cytotoxic CD8+ T cells and NK cells, and has increased glomerular expression in lupus[34]. *BTC* contributes to inflammation by binding to epidermal growth factor receptor[35], a gene whose renal expression is upregulated after kidney injury[36]. *CLEC16A* takes part in the B cell receptor-dependent HLA-II pathway in human B cells[37], but is also significantly expressed in the human podocytes (https://atlas.kpmp.org). It is involved in autophagy, mitophagy, and endolysosomal trafficking in multiple cell types[38,39]. It is also in close proximity to *CIITA*, a master transcription factor of HLA class II genes[40] and Dexamethasone Inducible Transcript (*DEXI),* a glucocorticoid-induced gene[41].

We next turned to discovering specific variants and genes driving these GWAS signals and discerning whether they are acting in immune cells, kidney cells, or both.

First, we conducted colocalization with eQTL data from two functionally distinct kidney compartments (glomerulus and tubulointerstitium)[42], multiple tissues from GTEx[43], and immune cells from DICE[44] and Blueprint[45]. Overall, pSSNS GWAS SNPs demonstrated significant enrichment in multiple immune cell eQTLs, led by a 69x and 62x increased odds of being monocyte and CD4+ memory Treg eQTLs, respectively (**Figure 3**). On an individual gene level, considering eQTLs with a regional colocalization probability (RCP) > 0.2, nine genes – including three which were the closest gene to a GWAS SNP (*CALHM6*, *AHI1*, *TNFSF15)* – colocalized with immune cell eQTLs (Figure 3, Table S6). All three significantly colocalized with monocyte eQTLs. *AHI* had significant eQTLs across monocytes, many T-cell subsets, and naïve B cells. Finally, in CD4+ memory Treg cells, SNPs in the "Gasdermin B (*GSDMB)"* suggestive locus colocalized with two different genes – *GSDMB* and ORMDL sphingolipid biosynthesis regulator 3 (*ORMDL3).* The *GSDMB/ORMDL3* locus is associated with multiple autoimmune disorders and eosinophilic inflammation-driven asthma[46]. In asthma, higher *GSDMB* expression is correlated with increased interferon signaling and MHC class I antigen presentation[47]. Notably, there was no colocalization with kidney eQTLs despite sufficient sample sizes to do so (**Figure S7**).

We then created a 95% credible set for all non-HLA significant loci and assessed their overlap with ATAC-seq derived open chromatin data from immune[48] and kidney cells[49,50] (**Table S7**). The SNPs with the highest posterior inclusion probability (PIP) for *AHI,* rs7759971 (PIP 0.52), overlapped with open chromatin of multiple immune cell types, including CD34+ cells, common lymphoid and myeloid progenitors, hematopoietic stem cells, and multipotent progenitors. The top PIP SNP for *CLEC16A, NPHS1, CD28, CALHM6,* and *TNFSF15* had no overlap with open chromatin. However, each locus had individual SNPs with lower PIPs that overlapped with both immune and kidney cell open chromatin.

We next fine-mapped the HLA signal (**Table S8**) We first imputed across the extended MHC region using a multi-population HLA imputation panel [51], resulting in 640 classical HLA alleles, 4,513 amino acids in HLA proteins and 49,321 SNPs in the extended MHC region for association. We used population-specific and multi-population SNP-level logistic regression, to identify specific SNPs and classical alleles associated with pSSNS (**Table S9, Supplement Note 1, Table S10, Figure S8**).

We next turned to discovering specific HLA amino acid positions most associated with risk of pSSNS through logistic regression analysis of all residues at each position. Amino acid position 47 in *HLA-DQA1* was the most strongly associated with pSSNS ($P_{omnibus} = 7.73 \times 10^{-83}$) (**Table S11, Figure S9**). Arginine was the most frequent amino acid; a substitution to lysine conferred the greatest disease risk ($P = 5.70 \times 10^{-80}$; OR [95% CI] =3.62 [3.17 – 4.14]). A second association in near-perfect linkage disequilibrium was identified at *HLA-DQA1* position 52 ($P = 1.14 \times 10^{-82}$). Arginine was again the most common amino acid at this position, and a substitution to serine conferred the greatest protection from risk ($P = 1.00 \times 10^{-28}$; OR = 0.53 [0.47 – 0.59]). After conditioning, an independent association was discovered at *HLA-DQB1* position 26 ($P = 3.22 \times 10^{-13}$). A change from the most common amino acid leucine to glycine conferred the most significant protection ($P = 4.75 \times 10^{-12}$; OR = 0.64 [0.60 – 0.73]). A haplotype analysis identified the 47_{lysine}-$52_{histidine}$ haplotype associated with greatest increased odds of pSSNS (**Figure 4a**). *HLA-DQA1* position 47 is located on the outside of the peptide-binding groove and acts as a regulator of binding stability, which, when altered, has been suggested to mediate the development of autoimmune disorders[52]. Arginine at *HLA-DQA1* position 52 has been associated with autoimmune disorders, including T1DM[53].

We then used DynaMut2[54] to model the impact on protein structural stability of two amino acid haplotypes (**Figure 4b**). This is quantified by Delta Delta G (ddG), where ddG < 0 predicts unstable

structure. The haplotype consisting of lysine (47) and histidine (52) predicted the most instability (ddG = -3.64). Notably, the predicted increase in protein instability and increased odds of disease for each haplotype were concordant. This suggests that pSSNS-associated haplotypes increased the odds of disease by increasing the instability of *HLA-DQA1* and altering its ability to properly form a stable HLA-II molecule.

Finally, we generated a pSSNS PRS using summary statistics of 1,607 cases and 11,995 controls from European and East Asian populations. We first tested its association with demographic and clinical phenotypes in 233 European children with sufficient clinical data from the EU-European sub-cohort, after adjusting for genetic principal components. There was a significant negative association observed between PRS and age of disease onset ($P = 1.49 \times 10^{-4}$, **Figure 5A**, **Table S12**). We then assessed the PRS in an independent cohort of 165 children with proteinuric kidney disease enrolled in the Nephrotic Syndrome Study Network (NEPTUNE)[55]. 30%, 49%, and 21% of all participants had focal segmental glomerulosclerosis, minimal change disease, or no biopsy, respectively. Adjusting for sex, histology and genetic principal components, we found a significant association between PRS and age of onset ($P = 0.003$; **Figure 5B, Table S12, Figure S10**).

A number of important discoveries emerged from this study. First, we tripled the number of pSSNS from four to twelve. Second, we found that while the immunological connections with the lead SNPs and closest genes in these newly discovered loci are well-established, most of them also have a *bona fide*, but overall less understood role, in kidney cells and diseases. Discovering upon which cells and organ systems each of these risk loci act will be an important future step. The availability of omics data from pediatric kidney tissue will be critical, as we hypothesize that the paucity of kidney eQTL signals we observed may be due to mapping pSSNS GWAS data to signatures primarily from adult tissues. Third, we identified monocytes and eosinophils dysfunction as a potential contributor to the

pathogenesis of pSSNS. The mechanisms by which genetically driven changes in these cell types contribute to pSSNS onset is an important area of future inquiry. Fourth, we discovered specific amino acid changes in HLA-DQA1 and -DQB1 associated with pSSNS that should help will empower subsequent studies to illuminate pathomechanisms at the most significant risk locus for this disease. Finally, the association of higher PRS with younger age of onset suggests that a stronger genetic predisposition to disease lowers the threshold of an individual to develop pSSNS in the context of environmental factors and may ultimately help share clinical screening and care. In conclusion, these findings expand our knowledge of the genetic architecture of pSSNS and accelerate our understanding of its molecular underpinnings and clinical implications.

METHODS

GWAS data summary.

GWAS data from NEPHROVIR/EU. Sample collection and genotype calling were done at Sorbonne Université in Paris. Pediatric steroid sensitive nephrotic syndrome was defined as proteinuria > 0.25g/mmol, serum albumin <25 g/L (<30 in France), full response within 4 weeks of 60 mg/m^2/day of oral prednisone or prednisolone, and age of onset < 18 years old. 244 previously reported European patients from the NEPHROVIR study[21] were combined with 159 newly recruited European patients recruited from France, Lithuania, Poland, Russia, Italy. Healthy adult controls (n=300) were recruited from Lyon, France, and combined with population-matched controls from The 1000 Genomes Project Phase 3 release (n=493)[56]. There were also 56 sub-Saharan African cases with 454 population-matched controls from The 1000 Genomes Project and 85 Maghrebian cases with 261 Moroccan population-matched controls. Both were reanalyzed from a previous report[21]. There were 160 Indian cases with 93 population-matched controls. Samples were genotyped on the Illumina Human OmniExpress or Illumina Omni 2.5 arrays.

GWAS data from Columbia University (US Cohorts). Sample collection and genotype calling was done at Columbia University in New York. Cases were defined by local recruitment centers across the US, Europe, and Brazil as either minimal change disease or non-biopsied SSNS with age of disease onset <21. Five cohorts from Columbia University consisted of patients from European (n_{cases}=371, $n_{controls}$-=4359), East Asian (n_{cases}=17, $n_{controls}$=443), sub-Saharan African (n_{cases}=65, $n_{controls}$=7344), South Asian (n_{cases} =39, $n_{controls}$=534) and Admixed American (n_{cases} =109, $n_{controls}$=13266) populations. The genotyping of the cases was performed using multiple versions of MEGA (Multi-Ethnic Global Array) chips that includes MEGA 1.0, MEGA 1.1 and MEGA^{EX}. The controls that were genotyped on MEGA1.0 were downloaded from NCBI dbGAP (IDAT files) from the PAGE consortium[57]. The differences between the chips were corrected first by mapping all the SNPs to a common cluster file in Genome Studio for individual MEGA platforms and further using Snpflip (https://github.com/biocore-ntnu/snpflip) software.

GWAS data from Kobe University. Pediatric steroid sensitive nephrotic syndrome cases were defined as urine protein to creatinine ratio ≥ 2.0, serum albumin ≤ 2.5 g/dl, and complete remission with 4-6 weeks after starting 6- mg/m1^2 oral prednisolone per day and age of onset < 18 years old. Three GWAS studies of SSNS in Japanese (n_{cases}= 987, $n_{controls}$ =3206), Korean (n_{cases}= 243, $n_{controls}$=4041) and Thai (n_{cases}=65, $n_{controls}$=94) population were completed at The University of Tokyo, Japan. The Japanese and Korean GWAS data have been previously reported[22,23]. The Thai dataset was genotyped with the Axiom array.

Dataset QC, imputation and GWAS.

Quality control, imputation, and GWAS were conducted separately for each study location and population. GC lambda (GC_{λ}) was used to assess inflation in all studies. The final case and control sizes and the number of variants tested can be found in Table S1 and Figure S1. Figure S11 shows matching

of cases and controls in PCA plots. Manhattan plots and GC can be found in Figure S12 and genome-wide significant hits resulting from dataset GWAS are in table S13.

GWAS data from NEPHROVIR/EU: EU-European, EU-African, Maghrebian, Indian.

Each file was quality controlled separately to remove related individuals (IBD > 0.1875), low call rate (genotype rate < 98%), and cases with discordant sex. SNPs were quality controlled for allele frequency (MAF < 0.01), call rate (genotype rate < 98%) in all cohorts, and Hardy Weinberg equilibrium (HWE $P<1x10^{-5}$) in controls only. The EU-European datasets were generated in multiple files and were merged stepwise on the common subset of SNPs, with the previous QC procedure reapplied after each merge. PCA plots were constructed from PLINK v1.9 to identify population outliers and check for batch effects[58]. Pre-imputation QC was conducted using McCarthy Tools with the TOPMed reference panel to check strand alignment and allele assignment. Insertions and deletions were excluded prior to imputation. Each population was imputed separately and cases and controls were imputed together on the TOPMed Imputation Server with the TOPMed r2 reference panel[59–61]. The QC was repeated after imputation and SNPs with low imputation quality (rsq < 0.3) were excluded. After imputation, UCSC Liftover[62] was used to convert SNP positions from each population dataset to build GRCh37 to match the build of summary statistics from other analyses. The association tests were completed using PLINK v1.9 under an additive model with principal component adjustment to account for population stratification.

GWAS data from Columbia University (US Cohorts) US-European, US-African, US-South Asian, US-East Asian and Admixed American.

Population was assigned by KING[63] kinship analysis software and based on continental population as defined by the 1000 Genomes Project for all cases and controls[64,65]. Within each continental population (EUR, AFR, AMR, SAS and EAS), we removed variants with genotype rate < 99%, MAF < 0.01 and

HWE $P<1x10^{-5}$. Each population was imputed separately with the TOPMed r2 panel[59–61]. After imputation, we removed first-degree relatives using KING, and variants with R2 < 0.8, MAF < 0.01 and HWE $P<1x10^{-5}$. Principal components were calculated with FlashPCA[66]. For cohorts with large case/control imbalances (Admixed American and US-African), we used the SAIGE logistic model[65] for calculating P-value and generating summary statistics. Association tests for European, South and East Asian were completed using PLINK v1.9 under an additive model with principal component adjustment to account for population stratification[58].

GWAS data from Kobe University: Japanese, Korean and Thai.

Quality control and analysis of the Japanese and Korean datasets are previously described in Jia et.al[23]. For the Thai analysis, SNPs were removed that had info score < 0.9, MAF < 0.005, call rate < 97%, or HWE $P < 1x10^{-5}$. Individuals with missing rate > 3%, IBD > 0.1875 and PCA outliers were removed. Genotypes were imputed with The 1000 Genomes reference panel using SHAPEIT[67] and IMPUTE2[68]. Logistic regression was performed with Plink1.9 and p-values were adjusted for genomic control (GC).

Population-specific and multi-population meta-analysis.

For each population-specific meta-analysis and the multi-population one, we conducted an inverse-variance, fixed-effect meta-analysis using METAL with adjustment for population stratification (GC) on each input dataset and assessment for heterogeneity selected[69]. For within-population meta-analyses, we removed variants with heterogeneity P-value < 0.05. All significant associations were visually inspected and single SNPs that did not follow the expected LD trend were not taken forward for further consideration.

For the European meta-analysis, we included available summary statistics of suggestive SNPs from Dufek et.al[24], increasing the European sample size to 1,096 cases and 12,459 controls. The resulting suggestive and genome-wide significant SNPs are annotated in tables and figures.

Multi-population meta-regression with MR-MEGA.

To account for and assess heterogeneous loci, we conducted a meta-regression using MR-MEGA[27]. We included three principal components, which captured the population structure across all 12 datasets. This allowed us to stratify heterogeneity into residual heterogeneity and heterogeneity that correlates with population. We visualized each dataset's PCs from MR-MEGA with the dataset-specific log odds ratio from METAL for each variant with heterogeneity that correlated with population. We adjusted for genomic control at the study level and after meta-regression to account for population structure within and between datasets. SNPs present in less than five studies were excluded. GC lambda (GC_λ) was used to assess inflation. Results tables include summaries from both METAL and MR-MEGA analyses (Table 1). All Manhattan plots were generated with the qqman R package [doi: 10.21105/joss.00731.] and LocusZoom web tool[70]. All significant loci are > 1Mb from each other with $r^2 < 0.1$. Loci were labeled by nearest genes.

Conditional analyses.

To identify independent secondary signals at the candidate loci, we used GCTA COJO[71,72] to conduct approximate conditional analyses based on cohort-specific meta-analysis summary statistics. Conditional analysis was conducted in each dataset, with an LD reference generated from the dataset samples, due to differences in linkage disequilibrium structure between continental populations. Each cohort was conditioned for the eight independent loci identified from the initial meta-analysis. Multi-

population meta-analysis of the conditioned cohorts was repeated in METAL[69] to assess multi-population genome-wide significant secondary loci after GCTA.

Heritability estimates.

SNP-based heritability was estimated on a liability scale with LD score regression (LDSC)[73] using a population prevalence of 16/100,000 and excluding HLA [chr6:25,000,000-34,000,000]. We used non-GC corrected population-specific meta-analysis summary statistics from METAL and pre-computed LD scores generated from 1000 Genomes EUR or EAS samples (https://alkesgroup.broadinstitute.org/LDSCORE/).

Colocalization of SSNS GWAS variants and eQTLs datasets.

We used fast enrichment estimation aided colocalization analysis (fastENLOC)[74] for colocalization analysis with glomerular (n=240) and tubulointerstitial (n=311) eQTLs from nephrotic syndrome patients[55], GTEx tissues (varied sample sizes), and immune eQTLs from both Blueprint[45] (n=200) and DICE [44] (n=91) databases. Posterior probabilities for SSNS GWAS variants were calculated from MR-MEGA Z-scores using TORUS[75]. We used an LD panel from European and East Asian 1000 Genomes samples to define haplotype blocks in the pSSNS meta-analysis[56]. Enrichment of pSSNS GWAS variants in each eQTL dataset was estimated using fastENLOC and subsequently informed prior probabilities for each analysis. For colocalization with our kidney eQTLs, which had available raw data, we could identify multiple eQTL signals per gene and multiple colocalized signals at each locus. For all other data, in which only summary statistics were available, we assumed at most one colocalized signal at each locus and did not account for LD.

Open chromatin annotation of credible sets.

95% credible sets were constructed for each independent locus identified from the multi-population meta-regression from Bayes' factors reported by MR-MEGA. Posterior inclusion probability (PIP) was estimated by dividing each Bayes' factor by the summation of Bayes' factors across all variants within a +/- 1Mb from the lead locus[76].

SNPs within 95% credible sets of our genome-wide significant loci were evaluated for positional overlap based on the boundaries of known open chromatin peaks in kidney[49] and immune[77] cell types. Open chromatin peaks were identified by MACS2 peak calling algorithm and optimized by gkmQC[50].

HLA imputation and analysis.

To fine-map the HLA region, we conducted HLA imputation with the four-digit multi-ethnic v2 reference panel on Michigan Imputation Server[51]. Cohorts were imputed individually to optimize population-specific structure within the HLA region. The imputed cohorts were then merged for multi-population associations. We used HLA-TAPAS 'assoc' module to conduct a logistic regression of the HLA region of the multi-population and population-specific datasets. For population-specific analyses, we adjusted for genotype-based principal components from Plinkv1.9[58]. The population-specific principal components and continental populations were included as covariates in the multi-population analysis. HLA-TAPAS was also used to conduct a stepwise conditional analysis, conditioning on the locus with the smallest association P-value. We additionally performed an omnibus test on the population-specific and multi-population cohorts to assess significance by amino acid position.

HLA modeling.

To predict the reference (with arginine at position 47 and serine at position 52) structure of *HLA-DQA1* we extracted the sequence of *HLA-DQA1* from UNIPROT database (Uniprot ID: P01909). We used NCBI BLASTp against PDB database to find the closest structure associated with the amino acid sequence of P01909. We identified the top hit as 6PX6_A (HLA-TCR complex, E = 2e-161) for the *HLA-DQA1* sequence[78]. We extracted the PDB coordinates for chain A from the 6PX6 and visualized in PYMOL v2.5. Since the most common amino acid haplotype in the control population was arginine (47) and serine (52), we performed mutagenesis using PYMOL to model the reference protein 3-D structure[79].

In brief, we used the mutagenesis tool from PYMOL and selected the rotamer (most likely amino acid conformation) for arginine and serine which showed the minimum number of clashes with nearby atoms. Afterwards, we adjusted the conformation of nearby atoms (within 5 Angstrom) to minimum free state using "Clean" command in PyMOL which uses MMFF94 force field[80] . Though point mutations locally affect the conformation of the protein, they can result in torsion, bending and stretching of the entire molecule. Therefore, we exported the protein structure to SPDBV software for further refinement[81].

We first fixed all the side chains of all amino acids to the best rotamer conformation using the simulated annealing method. Afterwards, we performed energy minimization using GROMOS 96 force field to extract the 3D coordinates that represent the lowest minimum energy conformation[82]. The refined protein structure of *HLA-DQA1* was then assessed for changes in stability of protein for both amino acid combinations for each haplotype using "MULTIPLE MUTATION" in DynMut2 server[54]. The instability of HLA-DQA1 was evaluated using the predicted ddG parameter which measures changes in Gibbs free energy between the folded and unfolded states and the change in folding when a mutation is present. The interaction among amino acids in reference and mutated structure were predicted using Arpeggio[83] and visualized in PyMOL.

Polygenic risk score analysis.

Construction of the polygenic risk score (PRS). To investigate genetic risk across the genome, we generated a polygenic risk score (PRS) from the GWAS of European (US-European) and East Asian (Japanese, Korean, US-East Asian) populations using PRS-CSx[84]. Population-specific weights estimated by PRS-CSx were used as input, along with a testing/training dataset, for optimization of the gamma-gamma priors a and b and the global shrinkage parameter used in the PRS-CSx model. Our main objective was to test clinical associations with PRS within case cohorts. We used case/control data and prediction accuracy to choose the best PRS model parameters. We fine-tuned and estimated PRS population weights in the EU-European and NEPTUNE separately. Since NEPTUNE is a case-only cohort, we included population--matched controls from the 1000 Genomes Project. In each dataset, 80% of cases and controls were randomly selected for training and 20% for testing. We varied the PRS-CSx hyper-parameters and chose the combination that gave the best prediction accuracy. The prediction accuracy with an area under the receiver operating characteristic curve was 0.74 (95%CI: 0.72-0.77) in EU-European dataset and 0.64 (0.61-0.68) in NEPTUNE.

Clinical associations with PRS. The PRS was applied to pediatric participants from NEPTUNE[55] (n=165), and the EU-European data for which clinical data were available (n=239). For bivariate tests, we used the Wilcoxon test for binary traits, Kruskal-Wallis for categorical traits and univariate linear regression for continuous traits. Population in NEPTUNE was predicted using Peddy[85] and the 1000 Genomes reference panel. For associations with the age of onset, we conducted linear regression adjusting for sex, four genetic principal components, and histology (NEPTUNE only).

REFERENCES

1. Noone, D. G., Iijima, K. & Parekh, R. Idiopathic nephrotic syndrome in children. *The Lancet* **392**, 61–74 (2018).
2. Gipson, D. S. *et al.* Gaining the PROMIS perspective from children with nephrotic syndrome: a Midwest pediatric nephrology consortium study. *Health Qual. Life Outcomes* **11**, 30 (2013).
3. Ruth, E. M., Landolt, M. A., Neuhaus, T. J. & Kemper, M. J. Health-related quality of life and psychosocial adjustment in steroid-sensitive nephrotic syndrome. *J. Pediatr.* **145**, 778–83 (2004).
4. Kerlin, B. A. *et al.* Epidemiology and risk factors for thromboembolic complications of childhood nephrotic syndrome: a Midwest Pediatric Nephrology Consortium (MWPNC) study. *J. Pediatr.* **155**, 105–110, 110.e1 (2009).
5. Hingorani, S. R., Weiss, N. S. & Watkins, S. L. Predictors of peritonitis in children with nephrotic syndrome. *Pediatr. Nephrol. Berl. Ger.* **17**, 678–682 (2002).

6. Rheault, M. N. *et al.* AKI in Children Hospitalized with Nephrotic Syndrome. *Clin. J. Am. Soc. Nephrol. CJASN* **10**, 2110–2118 (2015).
7. Sato, M. *et al.* Prognosis and acute complications at the first onset of idiopathic nephrotic syndrome in children: a nationwide survey in Japan (JP-SHINE study). *Nephrol. Dial. Transplant. Off. Publ. Eur. Dial. Transpl. Assoc. - Eur. Ren. Assoc.* **36**, 475–481 (2021).
8. Ding, W. Y. *et al.* Initial steroid sensitivity in children with steroid-resistant nephrotic syndrome predicts post-transplant recurrence. *J. Am. Soc. Nephrol. JASN* **25**, 1342–1348 (2014).
9. Korsgaard, T., Andersen, R. F., Joshi, S., Hagstrøm, S. & Rittig, S. Childhood onset steroid-sensitive nephrotic syndrome continues into adulthood. *Pediatr. Nephrol. Berl. Ger.* **34**, 641–648 (2019).
10. Ishikura, K. *et al.* Morbidity in children with frequently relapsing nephrosis: 10-year follow-up of a randomized controlled trial. *Pediatr. Nephrol. Berl. Ger.* **30**, 459–468 (2015).
11. Fakhouri, F. *et al.* Steroid-sensitive nephrotic syndrome: from childhood to adulthood. *Am. J. Kidney Dis. Off. J. Natl. Kidney Found.* **41**, 550–557 (2003).
12. Kyrieleis, H. A. C. *et al.* Long-term outcome of biopsy-proven, frequently relapsing minimal-change nephrotic syndrome in children. *Clin. J. Am. Soc. Nephrol. CJASN* **4**, 1593–1600 (2009).
13. Skrzypczyk, P. *et al.* Long-term outcomes in idiopathic nephrotic syndrome: from childhood to adulthood. *Clin. Nephrol.* **81**, 166–173 (2014).
14. Trompeter, R. S., Lloyd, B. W., Hicks, J., White, R. H. & Cameron, J. S. Long-term outcome for children with minimal-change nephrotic syndrome. *Lancet Lond. Engl.* **1**, 368–370 (1985).
15. Aydin, M. *et al.* The long-term outcome of childhood nephrotic syndrome in Germany: a cross-sectional study. *Clin. Exp. Nephrol.* **23**, 676–688 (2019).
16. Lee, J. M., Kronbichler, A., Shin, J. I. & Oh, J. Review on long-term non-renal complications of childhood nephrotic syndrome. *Acta Paediatr. Oslo Nor. 1992* **109**, 460–470 (2020).

17. Hjorten, R., Anwar, Z. & Reidy, K. J. Long-term Outcomes of Childhood Onset Nephrotic Syndrome. *Front. Pediatr.* **4**, 53 (2016).

18. Shalhoub, R. J. Pathogenesis of lipoid nephrosis: a disorder of T-cell function. *Lancet Lond. Engl.* **2**, 556–560 (1974).

19. Iijima, K., Sako, M., Kamei, K. & Nozu, K. Rituximab in steroid-sensitive nephrotic syndrome: lessons from clinical trials. *Pediatr. Nephrol. Berl. Ger.* **33**, 1449–1455 (2018).

20. Gbadegesin, R. A. *et al.* HLA-DQA1 and PLCG2 Are Candidate Risk Loci for Childhood-Onset Steroid-Sensitive Nephrotic Syndrome. *J. Am. Soc. Nephrol. JASN* **26**, 1701–1710 (2015).

21. Debiec, H. *et al.* Transethnic, Genome-Wide Analysis Reveals Immune-Related Risk Alleles and Phenotypic Correlates in Pediatric Steroid-Sensitive Nephrotic Syndrome. *J. Am. Soc. Nephrol. JASN* **29**, 2000–2013 (2018).

22. Jia, X. *et al.* Strong Association of the HLA-DR/DQ Locus with Childhood Steroid-Sensitive Nephrotic Syndrome in the Japanese Population. *J. Am. Soc. Nephrol. JASN* **29**, 2189–2199 (2018).

23. Jia, X. *et al.* Common risk variants in NPHS1 and TNFSF15 are associated with childhood steroid-sensitive nephrotic syndrome. *Kidney Int.* **98**, 1308–1322 (2020).

24. Dufek, S. *et al.* Genetic Identification of Two Novel Loci Associated with Steroid-Sensitive Nephrotic Syndrome. *J. Am. Soc. Nephrol. JASN* **30**, 1375–1384 (2019).

25. Julia, A. *et al.* A genome-wide association study identifies a novel locus at 6q22.1 associated with ulcerative colitis. *Hum. Mol. Genet.* **23**, 6927–34 (2014).

26. Ovunc, B. *et al.* Mutation analysis of NPHS1 in a worldwide cohort of congenital nephrotic syndrome patients. *Nephron Clin. Pract.* **120**, c139-146 (2012).

27. Mägi, R. *et al.* Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum. Mol. Genet.* **26**, 3639–3650 (2017).

28. Ghoussaini, M. *et al.* Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Res.* **49**, D1311–D1320 (2021).
29. Ferland, R. J. *et al.* Abnormal cerebellar development and axonal decussation due to mutations in AHI1 in Joubert syndrome. *Nat. Genet.* **36**, 1008–1013 (2004).
30. Kukimoto-Niino, M. *et al.* Cryo-EM structure of the human ELMO1-DOCK5-Rac1 complex. *Sci. Adv.* **7**, eabg3147 (2021).
31. Sharma, K. R. *et al.* ELMO1 protects renal structure and ultrafiltration in kidney development and under diabetic conditions. *Sci. Rep.* **6**, 37172 (2016).
32. Shimazaki, A. *et al.* Genetic variations in the gene encoding ELMO1 are associated with susceptibility to diabetic nephropathy. *Diabetes* **54**, 1171–1178 (2005).
33. Yu, C.-C. *et al.* Abatacept in B7-1–Positive Proteinuric Kidney Disease. *N. Engl. J. Med.* **369**, 2416–2423 (2013).
34. Spada, R. *et al.* NKG2D ligand overexpression in lupus nephritis correlates with increased NK cell activity and differentiation in kidneys but not in the periphery. *J. Leukoc. Biol.* **97**, 583–598 (2015).
35. Rayego-Mateos, S. *et al.* Role of Epidermal Growth Factor Receptor (EGFR) and Its Ligands in Kidney Inflammation and Damage. *Mediators Inflamm.* **2018**, 8739473 (2018).
36. Ju, W. *et al.* Tissue transcriptome-driven identification of epidermal growth factor as a chronic kidney disease biomarker. *Sci Transl Med* **7**, 316ra193 (2015).
37. Rijvers, L. *et al.* The Role of Autoimmunity-Related Gene CLEC16A in the B Cell Receptor-Mediated HLA Class II Pathway. *J. Immunol. Baltim. Md 1950* **205**, 945–956 (2020).
38. Tam, R. C. Y. *et al.* Human CLEC16A regulates autophagy through modulating mTOR activity. *Exp. Cell Res.* **352**, 304–312 (2017).
39. Pearson, G. *et al.* Clec16a, Nrdp1, and USP8 Form a Ubiquitin-Dependent Tripartite Complex That Regulates β-Cell Mitophagy. *Diabetes* **67**, 265–277 (2018).

40. Steimle, V., Siegrist, C. A., Mottet, A., Lisowska-Grospierre, B. & Mach, B. Regulation of MHC class II expression by interferon-gamma mediated by the transactivator gene CIITA. *Science* **265**, 106–109 (1994).

41. Edgar, A. J., Birks, E. J., Yacoub, M. H. & Polak, J. M. Cloning of dexamethasone-induced transcript: a novel glucocorticoid-induced gene that is upregulated in emphysema. *Am. J. Respir. Cell Mol. Biol.* **25**, 119–124 (2001).

42. Han, S. K. *et al.* Mapping genomic regulation of kidney disease and traits through high-resolution and interpretable eQTLs. 2022.06.01.494352 Preprint at https://doi.org/10.1101/2022.06.01.494352 (2022).

43. GTEx Consortium *et al.* Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nat. Genet.* **50**, 956–967 (2018).

44. Schmiedel, B. J. *et al.* Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell* **175**, 1701-1715.e16 (2018).

45. Fernández, J. M. *et al.* The BLUEPRINT Data Analysis Portal. *Cell Syst.* **3**, 491-495.e5 (2016).

46. Li, X. *et al.* Genetic analyses identify GSDMB associated with asthma severity, exacerbations, and antiviral pathways. *J. Allergy Clin. Immunol.* **147**, 894–909 (2021).

47. Das, S., Miller, M. & Broide, D. H. Chapter One - Chromosome 17q21 Genes ORMDL3 and GSDMB in Asthma and Immune Diseases. in *Advances in Immunology* (ed. Alt, F. W.) vol. 135 1–52 (Academic Press, 2017).

48. Ecker, S. *et al.* Genome-wide analysis of differential transcriptional and epigenetic variability across human immune cell types. *Genome Biol.* **18**, 18 (2017).

49. Muto, Y. *et al.* Single cell transcriptional and chromatin accessibility profiling redefine cellular heterogeneity in the adult human kidney. *Nat. Commun.* **12**, 2190 (2021).

50. Han, S. K. *et al.* Quality assessment and refinement of chromatin accessibility data using a sequence-based predictive model. 2022.02.24.481844 Preprint at https://doi.org/10.1101/2022.02.24.481844 (2022).

51. Luo, Y. *et al. A high-resolution HLA reference panel capturing global population diversity enables multi-ethnic fine-mapping in HIV host response*. http://medrxiv.org/lookup/doi/10.1101/2020.07.16.20155606 (2020) doi:10.1101/2020.07.16.20155606.

52. Frommer, L., Flesch, B. K., König, J. & Kahaly, G. J. Amino Acid Polymorphisms in Hla Class II Differentiate Between Thyroid and Polyglandular Autoimmunity. *J. Clin. Endocrinol. Metab.* **105**, dgz164 (2020).

53. Badenhoop, K. *et al.* Susceptibility and resistance alleles of human leukocyte antigen (HLA) DQA1 and HLA DQB1 are shared in endocrine autoimmune disease. *J. Clin. Endocrinol. Metab.* **80**, 2112–2117 (1995).

54. Rodrigues, C. H. M., Pires, D. E. V. & Ascher, D. B. DynaMut2: Assessing changes in stability and flexibility upon single and multiple point missense mutations. *Protein Sci. Publ. Protein Soc.* **30**, 60–69 (2021).

55. Gadegbeku, C. A. *et al.* Design of the Nephrotic Syndrome Study Network (NEPTUNE) to evaluate primary glomerular nephropathy by a multidisciplinary approach. *Kidney Int.* **83**, 749–756 (2013).

56. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

57. Wojcik, G. L. *et al.* Genetic analyses of diverse populations improves discovery for complex traits. *Nature* **570**, 514–518 (2019).

58. 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).

59. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *bioRxiv* 563866 (2019) doi:10.1101/563866.

60. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).

61. Fuchsberger, C., Abecasis, G. R. & Hinds, D. A. minimac2: faster genotype imputation. *Bioinformatics* **31**, 782–784 (2015).

62. Hinrichs, A. S. *et al.* The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* **34**, D590–D598 (2006).

63. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).

64. Clarke, L. *et al.* The international Genome sample resource (IGSR): A worldwide collection of genome variation incorporating the 1000 Genomes Project data. *Nucleic Acids Res.* **45**, D854–D859 (2017).

65. Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341 (2018).

66. Abraham, G., Qiu, Y. & Inouye, M. FlashPCA2: principal component analysis of Biobank-scale genotype datasets. *Bioinforma. Oxf. Engl.* **33**, 2776–2778 (2017).

67. Delaneau, O., Marchini, J. & Zagury, J.-F. A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–181 (2011).

68. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).

69. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).

70. Pruim, R. J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinforma. Oxf. Engl.* **26**, 2336–2337 (2010).

71. 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).

72. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).

73. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

74. Wen, X., Pique-Regi, R. & Luca, F. Integrating molecular QTL data into genome-wide genetic association analysis: Probabilistic assessment of enrichment and colocalization. *PLOS Genet.* **13**, e1006646 (2017).

75. Wen, X., Lee, Y., Luca, F. & Pique-Regi, R. Efficient Integrative Multi-SNP Association Analysis via Deterministic Approximation of Posteriors. *Am. J. Hum. Genet.* **98**, 1114–1129 (2016).

76. Maller, J. B. *et al.* Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat. Genet.* **44**, 1294–1301 (2012).

77. Corces, M. R. *et al.* Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet.* **48**, 1193–1203 (2016).

78. Ting, Y. T. *et al.* A molecular basis for the T cell response in HLA-DQ2.2 mediated celiac disease. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 3063–3073 (2020).

79. Mooers, B. H. M. Shortcuts for faster image creation in PyMOL. *Protein Sci. Publ. Protein Soc.* **29**, 268–276 (2020).

80. Halgren, T. A. MMFF VII. Characterization of MMFF94, MMFF94s, and other widely available force fields for conformational energies and for intermolecular-interaction energies and geometries. *J. Comput. Chem.* **20**, 730–748 (1999).

81. Johansson, M. U., Zoete, V., Michielin, O. & Guex, N. Defining and searching for structural motifs using DeepView/Swiss-PdbViewer. *BMC Bioinformatics* **13**, 173 (2012).

82. Pol-Fachin, L., Fernandes, C. L. & Verli, H. GROMOS96 43a1 performance on the characterization of glycoprotein conformational ensembles through molecular dynamics simulations. *Carbohydr. Res.* **344**, 491–500 (2009).

83. Jubb, H. C. *et al.* Arpeggio: A Web Server for Calculating and Visualising Interatomic Interactions in Protein Structures. *J. Mol. Biol.* **429**, 365–371 (2017).

84. Ruan, Y. *et al.* Improving polygenic prediction in ancestrally diverse populations. *Nat. Genet.* **54**, 573–580 (2022).

85. Pedersen, B. S. & Quinlan, A. R. Who's Who? Detecting and Resolving Sample Anomalies in Human DNA Sequencing Studies with Peddy. *Am. J. Hum. Genet.* **100**, 406–413 (2017).

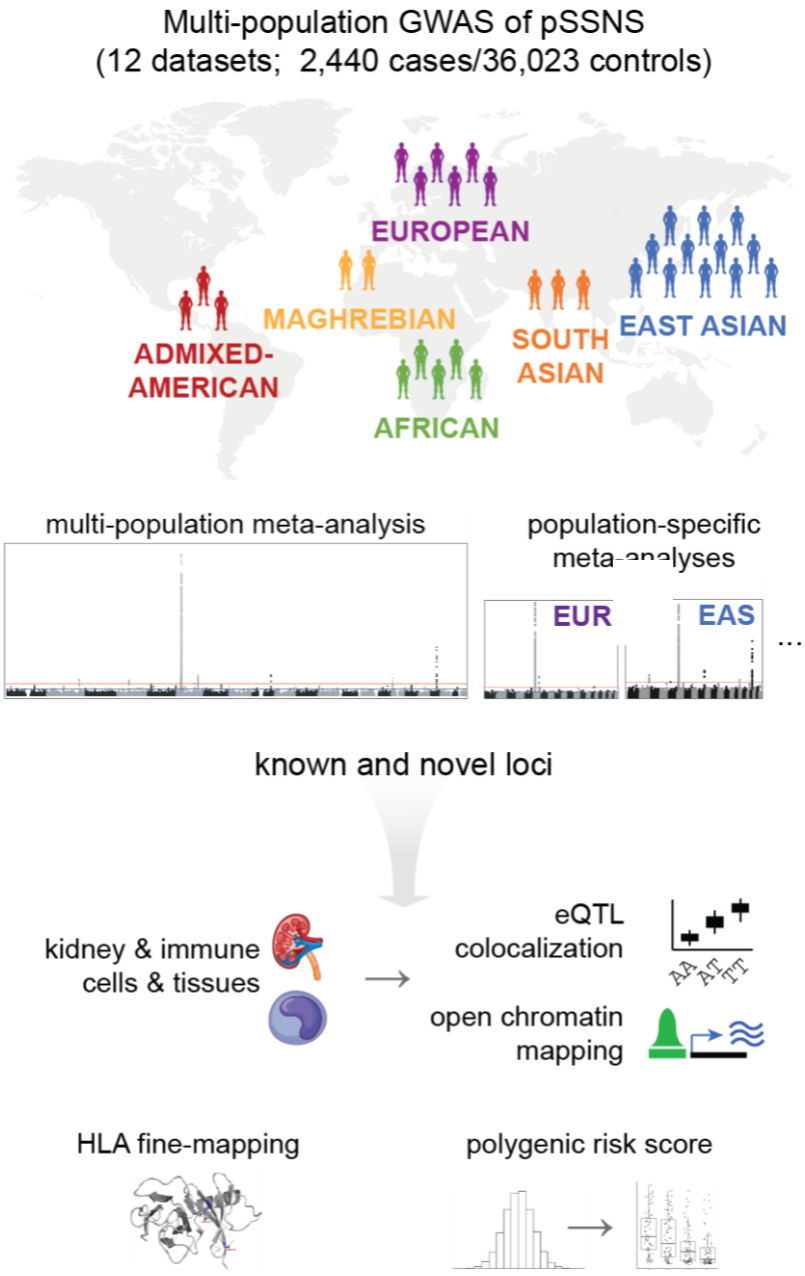

Figure 1: Flowchart of study design. pSSNS= pediatric steroid sensitive nephrotic syndrome, EUR = European, EAS = East Asian, eQTL = expression quantitative trait loci.

Nearest Gene	Top SNP	Position (hg19)	EA	NEA	Fixed-effects meta-analysis OR [95% CI]	Fixed-effects meta-analysis P-value	Fixed-effects meta-analysis Het I^2 P-value	Meta-regression (MR-MEGA) P-value	Meta-regression (MR-MEGA) Population het P-value
Discovery Meta-analysis									
HLA-DQB1	rs1063355	6:32627714	T	G	0.46 [0.42, 0.50]	4.51×10^{-81}	0.09	6.85×10^{-82}	0.26
NPHS1	rs412175	19:36342103	C	T	1.65 [1.49, 1.83]	4.15×10^{-22}	8.63×10^{-2}	2.30×10^{-24}	2.43×10^{-4}
CALHM6	rs2637678	6:116787378	C	T	0.79 [0.73, 0.85]	1.48×10^{-10}	7.30×10^{-5}	2.06×10^{-12}	4.99×10^{-4}
AHI1	rs7759971	6:135746884	T	C	1.34 [1.24, 1.44]	1.28×10^{-13}	0.52	4.90×10^{-12}	0.74
TNSFS15	rs10817678	9:117579457	G	A	0.79 [0.73, 0.85]	2.92×10^{-10}	0.04	5.57×10^{-12}	7.76×10^{-4}
CLEC16A	rs8062322	16:11092319	A	C	0.75 [0.68, 0.81]	2.78×10^{-11}	0.33	1.61×10^{-10}	0.16
CD28	rs55730955	2:204585956	A	T	0.74 [0.68, 0.81]	3.59×10^{-11}	0.62	4.27×10^{-10}	0.34
BTC	rs28862935	4:75693465	A	G	1.41 [1.27, 1.56]	1.00×10^{-10}	0.18	1.08×10^{-09}	0.24
Conditional Meta-analysis									
HLA-DQB1	rs1794497	6:32649180	T	C	0.49 [0.44, 0.53]	6.79×10^{-52}	0.03	-	-
MICA	rs2256318	6:31381519	A	G	1.49 [1.36, 1.64]	9.71×10^{-18}	4.18×10^{-4}	-	-

Table 1. Genome-wide significant SNPs from multi-population meta-analysis
EA=effect allele, NEA=non-effect allele, OR [95% CI]= Odds ratio with 95% confidence interval. MR-MEGA results are not available for the conditional analysis.

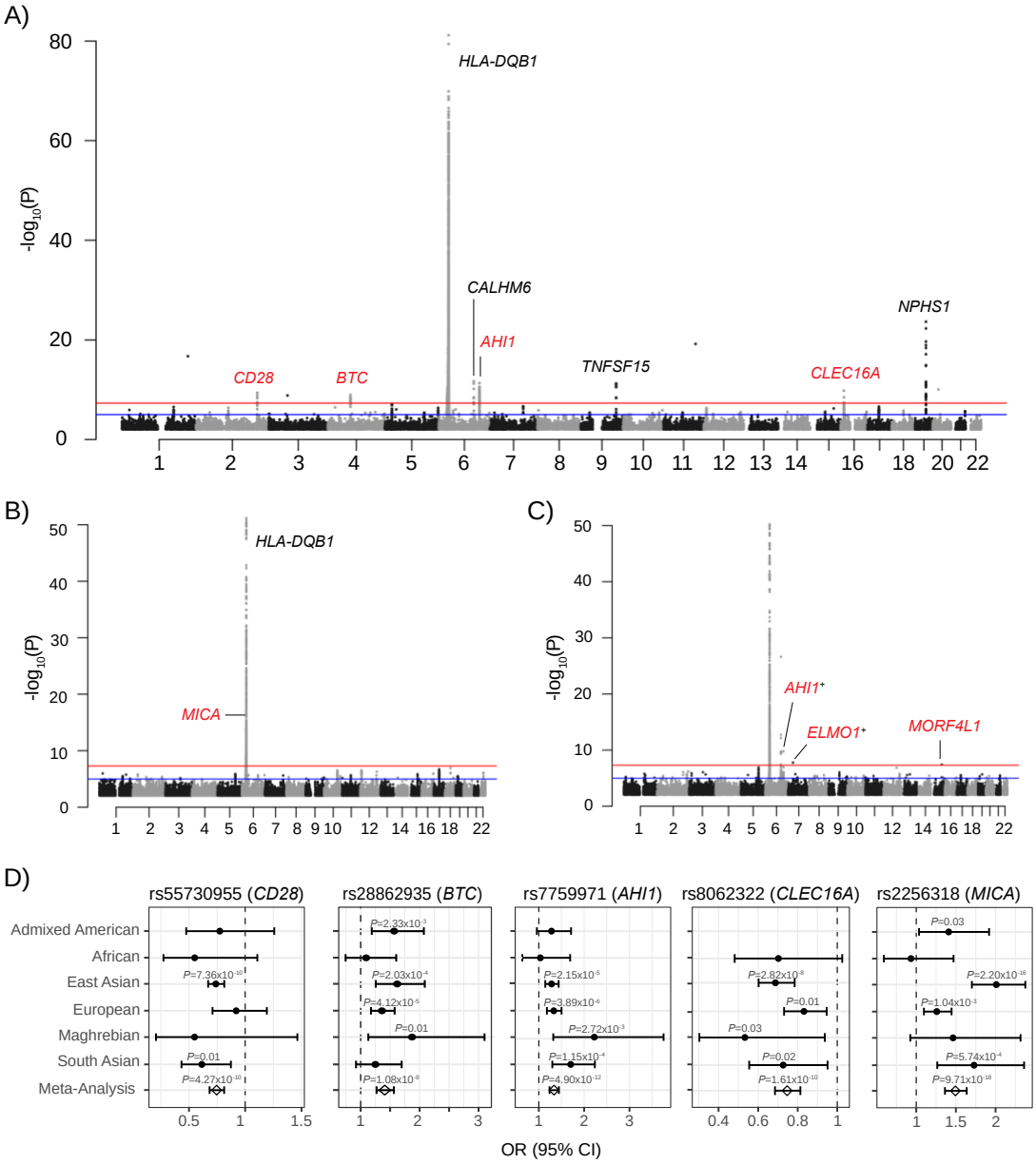

Figure 2. Manhattan plots of **A)** multi-population meta-analysis of 2,440 cases vs. 36,023 controls, **B)** multi-population conditional meta-analysis, **C)** European meta-analysis of 674 cases vs. 6,817 controls. Novel genome-wide association are indicated with in red and only novel associations are labeled in B and C. Discoveries that included the summary statistics from suggestive SNPs available from Dufek et al indicated with '+', **D)** Multi-population & single population odds ratios for novel multi-population significant SNPs discovered

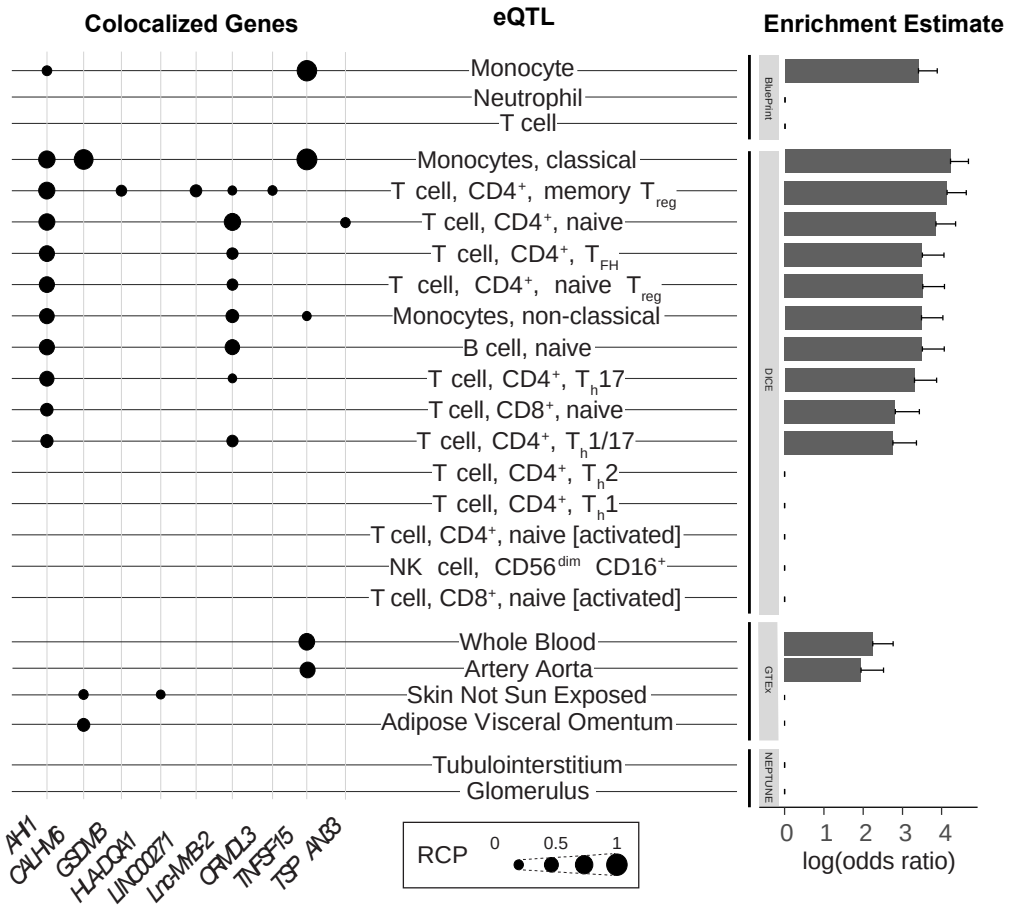

Figure 3. Colocalization of SSNS GWAS and eQTL datasets. Each eQTL data set is labeled with colocalized loci to the left and enrichment estimates to the right. Genes with regional colocalization probability (RCP) > 0.2 in at least one tissue/cell are included. pSSNS GWAS loci that colocalized with tissue/cell-type eQTLs are indicated by black dots, with larger dots indicating higher RCP. GTEx tissues without associations are excluded from this figure (see Figure S7). Enrichment estimates, from fastENLOC, are based on genome-wide summary statistics from GWAS and include a shrinkage parameter, resulting in 0 enrichment for multiple tissues/cell-types. **logOR=2 ~ OR=7.5, logOR=3 ~ OR=20.1, logOR=4 ~ OR=50.6**.

A)

Amino Acid Haplotype		Haplotype Frequency				
DQA1_47	DQA1_52	Case	Control	OR (95% CI)	P-value	Predicted Stability Change
lysine	histidine	0.28	0.12	3.56 (3.12, 4.06)	1.09×10^{-78}	-3.64
cystine	arginine	0.34	0.31	1.93 (1.71, 2.18)	8.77×10^{-27}	-2.67
glutamine	arginine	0.13	0.18	1.69 (1.45, 1.97)	2.90×10^{-11}	-0.05
arginine	serine	0.25	0.39	Reference	-	-

B)

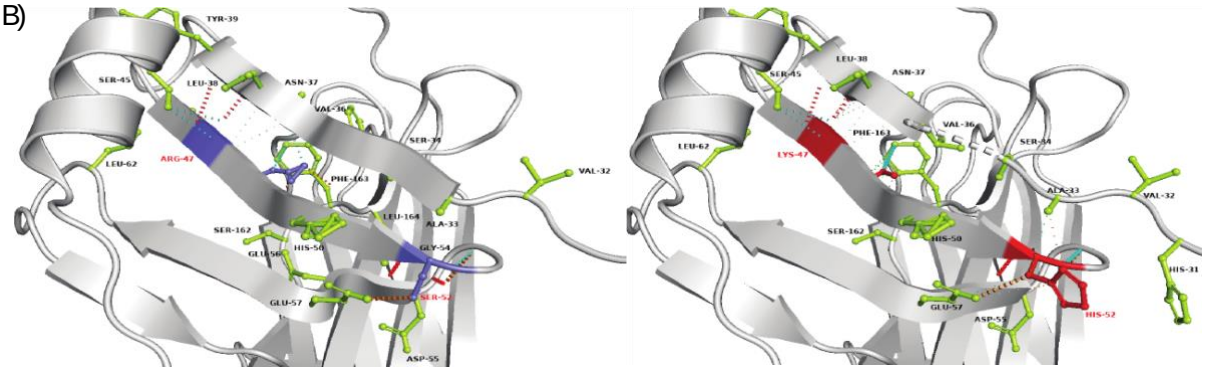

Figure 4. *HLA-DQA1* amino acid risk associations. A) Increased risk and predicted stability change of the two-amino acid residue hapltoypes at *HLA-DQA1* positions 47 and 52. Odds ratios are from a joint logistic regression with arginine$_{47}$-serine$_{52}$. The reference haplytype is the one conferring the strongest protection (i.e., odds ratio indicate increase in risk compared to arginine$_{47}$-serine$_{52}$)**.** Increasingly negative "Predicted Stability Change" indicates increasingly decreased stability. **B)** Protein structure for the reference haplotype arginine$_{47}$-serine$_{52}$ (left, blue) and lysine$_{47}$-histidine$_{52}$ (right, red). The residues in lime color displays potential interacting amino acid with mutated amino acids. The color scheme for interaction is as follows: Cyan for Van der Waals, red for hydrogen bonds, green for hydrophobic bonds, sky blue for Carbonyl bonds and orange for polar bonds. When no bonds are displayed but the amino acids are shown they were predicted to form weak VDW bonds.

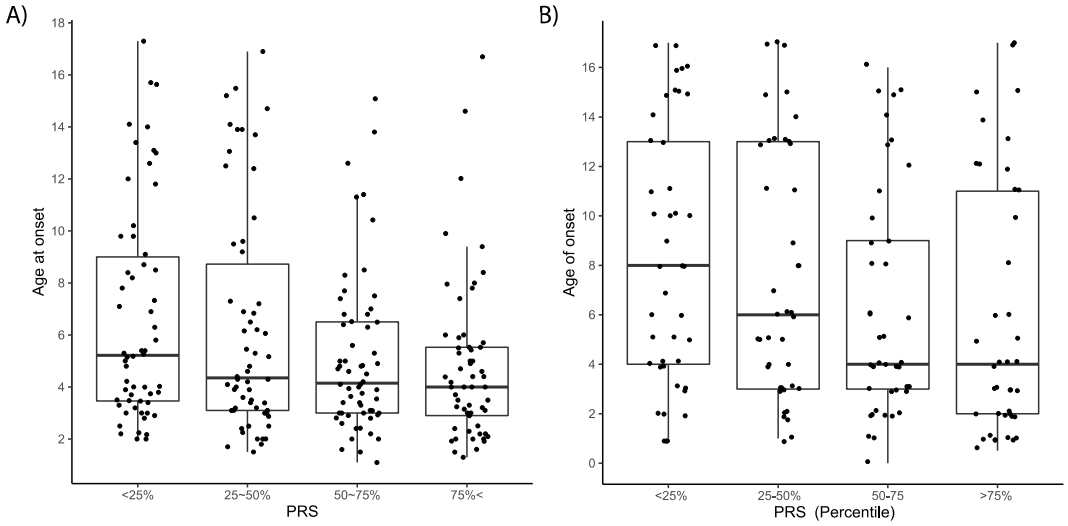

Figure 5. Association between pSSNS PRS quartiles and age of onset in pediatric SSNS patients from **A)** EU-European and **B)** NEPTUNE pediatric cohorts.