SUPPLEMENTAL material: Trans-ethnic, genome-wide….Debiec et al

Table of contents

Acknowledgements

Research of the authors was supported by European Research Council Grant ERC-2012-ADG_20120314 (Grant Agreement 322947) and by French National Research Agency (Genetransnephrose, ANR-16-CE17-0004-02).

M.S. is supported by the Charles Woodson Clinical Research Fund, the Ravitz Foundation and by National Institutes of Health RO1-DK108805

The Nephrotic Syndrome Study Network Consortium (NEPTUNE), U54-DK-083912, is a part of the National Center for Advancing Translational Sciences (NCATS) Rare Disease Clinical Research Network (RDCRN), supported through a collaboration between the Office of Rare Diseases Research (ORDR), NCATS, and the National Institute of Diabetes, Digestive, and Kidney Diseases. RDCRN is an initiative of ORDR, NCATS. Additional funding and/or programmatic support for this project has also been provided by the University of Michigan, NephCure Kidney International, and the Halpin Foundation.

The authors acknowledge the NIH data repository for the use of the Population Reference Samples (POPRES) dataset (phs000145.v4.p2)[1].

Institutional approvals

The studies were approved by the relevant institutional boards in the different countries and were conducted according to the principles of the Declaration of Helsinki. Written informed consent was obtained from children's parents or adult patients. Biobanking was associated with an anonymized clinical and biological dataset.

Supplemental Note #1: Definitions of SSNS and Patient Phenotyping

Definitions. The following definitions were used for the GWAS cohorts. Nephrotic syndrome was defined as proteinuria >0.25 g/mmol and serum albumin <25 g/L (<30 in France). Steroid sensitivity was defined as full response (proteinuria trace or negative by dipstick) within 4 weeks of oral prednisone or prednisolone at a dose of 60 mg/m^2/day (± three iv pulses of methylprednisolone).

GWAS cohort patient phenotyping. All patients were carefully phenotyped (Table 1).Children with nephrotic syndrome and an age at onset between 1 and 18 years were eligible for the study. Patients with nephrotic syndrome secondary to infectious agents, malignancies, medications, and other conditions associated with nephrotic syndrome were excluded. The following clinical data were obtained: basic demographic data (including age, sex, race, and ethnicity), family history of renal disease, age at onset of symptoms, and laboratory data (including urinalysis, 24-hour urine protein excretion or spot urine protein-to-creatinine ratio, and serum creatinine). Treatment received and pattern of response to therapy were recorded.

Supplemental Note #2: Quality controls, imputation, & ancestry analysis

Quality controls. Data quality control was performed using a standard procedure[2]. In brief, per-individual quality control was first carried out to remove (i) individuals with discordant sex information, (ii) individuals with outlying missing genotype (>2%) or heterozygosity rates (4 standard deviations above or below the mean for IDF cases, 3 standard deviations above or below the mean for other samples) and (iii) duplicated or related individuals based on pairwise identity by state estimates. Per-marker quality control eliminated (i) SNPs with an excessive missing rate (>3%), (ii) SNPs showing a significant deviation from Hardy-Weinberg equilibrium in the controls (P< 1e-5), (iii) SNPs with significantly different missing

genotype rates between cases and controls (P< 1e-5), and (iv)markers with a very low minor allele frequency (< 1%). Per-marker quality control was carried out independently in each case-control cohort, and the number of SNPs kept for analysis is indicated in Table S2. In addition, a handful of isolated SNPs met genome-wide significance in the IDF-AFR GWAS without any SNP in linkage disequilibrium supporting the associations. These SNPs were considered artifactual and removed from the analysis. All quality controls were performed using PLINK[3].

Genotype imputation. We performed genotype imputation of each series using the Sanger Imputation Server, with the Haplotype Reference Consortium panel as reference for European and Maghrebian cohorts[4], and the 1000 Genomes Phase 3 data for the African cohort (Table S2). We used the recommended pipeline, including pre-phasing with EAGLE2 v2.0.5[5] and imputation with EAGLE2+PBWT[6]. Case and control data sets were imputed together taking only SNPs passing QC as input.

Ancestry analysis. Due to a high immigration rate in the Ile-de-France area, the Nephrovir cohort of patients is multi-ethnic, including patients of European, Maghrebian, African, East Asian and South Asian origin. We thus used genotype data to infer the ancestry of each patient and constitute ancestry-matched case-control cohorts. To do so, we analyzed the three patient cohorts (331 French, 98 Italian, and 35 Spanish patients) together with several control cohorts of diverse ancestries. These control panels included the French 3 Cités cohort (n=2,000)[7], multi-ethnic cohorts from the 1000 Genomes[8], (phase 3, n=3,129) and POPRES[1] projects (n=4,077), and our own series of Maghrebian controls (n=272). We used PLINK software to compute the pairwise identity by state (IBS) between each pair of samples. We then performed a principal component analysis to visualize the distribution of cases as compared to the control population panels (Fig. S1A), and retrieved the 10 first principal components for adjusting fine population structure in association studies. We then performed

a hierarchical clustering using 1-IBS as distance metric and Ward linkage to classify samples according to their identity by state scores. This clustering revealed 7 clusters, corresponding to Europeans, Africans, South Asians, East Asians, Maghrebians, and 2 American groups (Fig. S1B). We assigned to each patient the ancestry of the reference group with which it clustered. We used the reference samples to estimate the reliability of this approach. Out of 8,473 samples of European, African, Asian or Maghrebian ancestry in the 3 Cités, 1000 Genomes, POPRES and Maghrebian control cohorts, only 14 were misclassified in the clustering, corresponding to a misclassification rate of only 0.17%. Principal component analysis and hierarchical clustering were performed using the R statistical software.

Supplemental Note#3: Imputation of rs1063348 in NEPTUNE cohort

NEPTUNE patients underwent "low-pass" whole genome sequencing, a strategy that has been shown to result in accurate genotype calls for common variants[9]. On analysis of the NEPTUNE WGS data, rs1063348 did not pass our stringent quality control thresholds. To address this, we imputed it, and found that it had an estimated R^2 of 0.98. We also calculated empirical R^2 for nearby directly genotyped SNPs (100 bp flanking) with similar allele frequencies. These had an average empirical R^2=0.95. Together, these metrics of high quality imputation provided confidence in using these imputed genotypes in subsequent analyses.

Supplemental Note #4: GWAS methods

Power estimates. We used the Genetic Association Study (GAS) power calculator from MichiganUniversity(http://csg.sph.umich.edu/abecasis/cats/gas_power_calculator/index.html) to estimate the genotype relative risk detectable with an 80% power as a function of the minor allele frequency for the cohort sizes obtained in this study. The results are presented below.

	IDF_EUR cases (n=132) vs 3C controls (n=2000)	IDF_AFR cases (n=56) vs 1000G AFR controls (n=454)	IDF_MAG cases (n=85) vs Moroccan controls (n=261)	ItSpa cases (n=112) vs 1000G EUR controls (n=552)
MAF=0.05	5.07	6.25	6.41	5.82
MAF=0.1	3.45	4.16	4.31	3.91
MAF=0.25	2.45	2.89	3.03	2.74
MAF=0.5	2.2	2.63	2.83	2.49

Genome-wide association studies. We conducted genome-wide association analyses in each ancestry-matched case-control series. Significance was assessed using a logistic regression model incorporating the first *n* principal components of an ancestry analysis to correct for fine ethnic differences. The number of components to correct for, *n*, was chosen by inspecting the eigen values of each principal component analysis[10]. We included 3 components in the logistic regression of the IDF-EUR series, 6 for the IDF-AFR series, 4 for the IDF-MAF series and 3 for the ItSpa series. We used log quantile-quantile P value plots (Q-Q plots) to demonstrate the adherence of P values to the null-hypothesis over the major part of the distribution. Association tests were performed using PLINK, and the Manhattan and Q-Q plots were generated using R statistical software (Fig. S2).

Supplemental Note #5: eQTL analyses

Calculation of PEER factors and principal components for the eQTL analysis. PEER factors were created utilizing the PEER framework[11, 12] with patient age, sex, and microarray batch as covariates, across all individuals for whom we had genotype, expression, and clinical covariates (N=187). These PEER factors account for non-genetic confounders present in the expression data. We chose the number of PEER factors as covariates based on the number that maximized significant eQTLs at an arbitrary p-value threshold. Principal components were calculated using EPACTS[13] on LD-pruned whole genome sequence data across the same 187 individuals.

Framework for use of FDR method in eQTL analysis. The FDR correction was performed with respect to the hypothesis that for a SNP of interest, there was at least one gene for which that SNP was an eQTL. Therefore, the correction was made across all *cis*-genes for each SNP individually (Table S11).

Supplemental Note #6: Multi-SNP risk and epidemiologic analyses

Odds of disease. Upon identification of SSNS associated markers, we set out to determine the combined risk burden that these loci produce. We classified each of the 385 cases (244 European, 56 African, 85 Maghrebian) and 3267 controls (2552 European, 454 African, 261Maghrebian) as belonging to one of a risk group classified by the sum of risk alleles across the top two independent risk SNPs associated with SSNS: rs1063348 and rs28366266. There were 5 possible unique risk groups, ranging between zero and four risk alleles. We evaluated the odds of being a case given the burden of risk alleles for these SNPs. We calculated odds ratios, confidence intervals, and p-values using Fisher's exact test. We adjusted for multiple testing using FDR methods and found that for 0 or 1 risk alleles, there were significantly decreased odds of being a case, whereas for 3 or 4 risk alleles there were significantly increased odds of being a case (Fig. 4B).

Age of onset. In the GWAS and NEPTUNE cohorts, we also analyzed the association of number of risk alleles (summed for rs1063348 and rs28366266) and age of onset. We performed linear regression with age of onset as a continuous outcome and number of risk alleles as a continuous predictor variable. In the GWAS cohort, we discovered a significant association between number of risk alleles and age of disease onset (in years) ($\beta = -0.49$, $p = 0.006$). Using multivariable linear regression in the NEPTUNE cohort, including the first four principal components of genetic ancestry and histologic diagnosis, we also discovered a significant association between number of risk alleles and age of onset ($\beta = -2.0$,

$p = 4.8 \times 10^{-5}$), (Fig.5A and 5B).

Complete remission. In the NEPTUNE cohort, we examined the complete remission by number of risk alleles using multivariable logistic regression, including the first four principal components of genetic ancestry and histologic diagnosis in the model. We found a significant association between number of risk alleles (summed for rs1063348 and rs28366266) and achievement of complete remission (OR = 2.2, CI = [1.4, 10], p = 0.02), (Fig.6).

NEPTUNE case-only phenotypic associations. We reported allele frequencies and performed association tests of the genotype of rs1063348, rs28366266, and rs9348883 in relation to a number of clinical and demographic traits (Table S14).These included comparison in the binary traits (a) achievement of complete remission, (b) gender, and (c) RAAS blockade status, and comparison across multiple categories for morphologic diagnosis, patient-reported race, and PCA-estimated ancestry. Continuous traits such as estimated GFR (eGFR) and urine protein to creatinine ratio (UPC; g/g) were dichotomized by clinically relevant values, less than and greater than or equal to 90 and 2, respectively. Association of these traits with genotype was analyzed, using ANOVA to compare multinomial regression models with and without genotype for those traits with multiple categories and logistic regression for binary categories. These models were adjusted by four PCs and APOL1 risk status, where applicable, with genotype under an additive model.

Supplemental Note #7: Comparison with previously discovered SSNS SNPs

Via exome-chip association, Gbadegesin et al identified four markers significantly associated with SSNS[14], all in the HLA-region. Only rs1129740 and rs1071630, missense mutations in *HLA-DQA1*, were replicated in the original paper. In the present study, these two SNPs passed QC in all GWAS populations and in the NEPTUNE cohort. We therefore analyzed these

SNPs in our data to determine potential overlap between these previously reported SNPs and those we discovered.

Linkage Disequilibrium between SNPs. We calculated pairwise linkage disequilibrium (LD) between the SNPs we identified and rs1129740 and rs1071630 in each GWAS population separately using PLINK. We found that R^2 between rs1063348 and both rs1129740 and rs1071630 were high (>0.88 in all populations). However, LD between these SNPs and rs28366266 was much lower in all populations, on the order of 0.07 – 0.135 (Table S12). Therefore, while these SNPs may be tagging the same locus as rs1063348, the other two appear to be independent SSNS-associated risk loci.

Conditional analysis of trans-ethnic GWAS. We then performed the same trans-ethnic conditional meta-analysis, conditioning on rs1129740 andrs1071630 in turn instead of rs1063348. Our results are almost identical as when we conditioned on rs1063348, in that that this locus was no longer significant, while rs28366266 remained independently associated at a significant level (Fig. S3).

eQTL results for rs1129740 and rs1071630. We performed eQTL analysis in the NEPTUNE cohort for rs1129740 and rs1071630 using methods previously described (Table S13). Because rs1129740 and rs1071630 are in perfect LD in the NEPTUNE pediatric cohort, their eQTL results are identical. Overall, eQTL results for these SNPs are very similar to those for rs1063348. We find that these SNPs are significant eQTLs for *HLA-DRB1* and *HLA-DRB5,* with FDR adjusted p-values of 0.0089 and 0.017, respectively. *HLA-DQB1,* while a significant eQTL for rs1063348, just misses our significance cutoff in rs1129740 and rs1071630, with an FDR adjusted p-value of 0.062

FIGURES

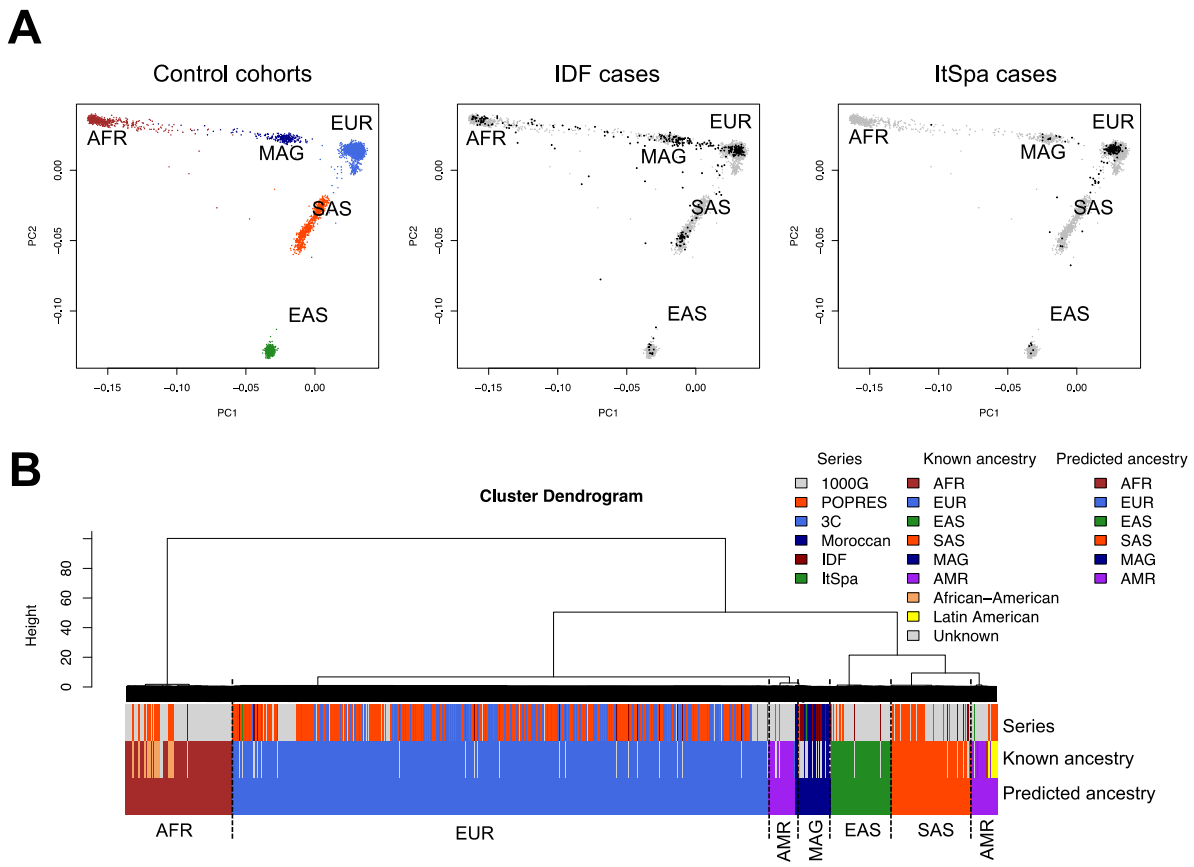

Figure S1: Population Analysis. Panel A shows the distribution of samples along the two principal components (PC1 and PC2) identified by principal component analysis of 505 cases and 9478 controls of diverse ancestries. The left panel displays the distribution of control samples from the 3 Cités, 1000 Genomes, POPRES, and Maghrebian control cohorts. A color code indicates the ancestry of each donor, which is also labeled next to each cloud of dots. AFR: African; MAG: Maghrebian; EUR: European; SAS: South Asian; EAS: East Asian. On the middle and right panels, points corresponding to the NEPHROVIR (IDF, middle) and ItSpa (right) cases are represented in black, on top of the grey dots corresponding to control samples of the left panel. Panel B shows the classification of cases and controls by hierarchical clustering. The series, known ancestry and predicted ancestry of each donor is

indicated by a color code below the dendogram. Seven clusters were defined and used to infer the ancestry of each case, according to the ancestry of controls in the same cluster.

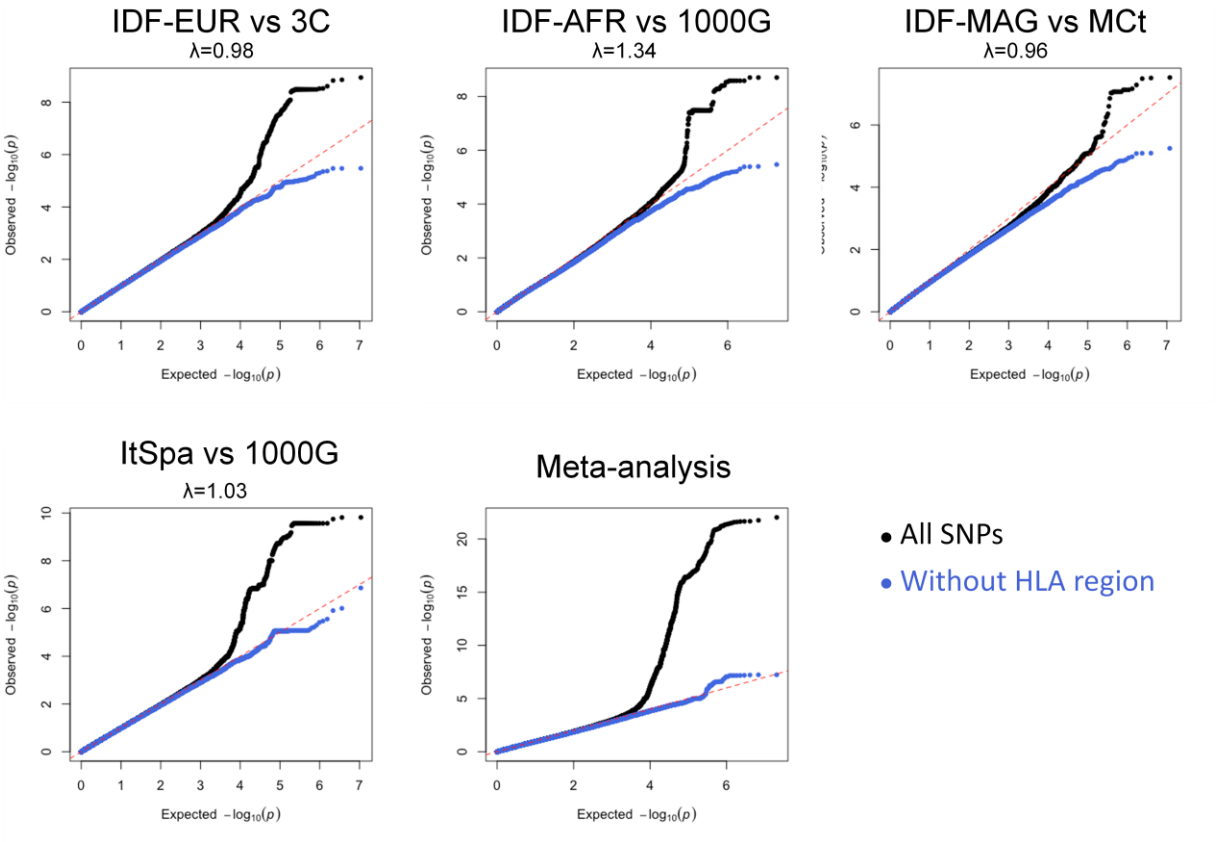

Figure S2: Quantile-Quantile Plots for Each GWAS Represented in Figure 1. The observed P value quantiles are compared with the expected distribution for each case-control series. Black points represent the quantiles when considering all SNPs; blue points represent the quantiles after removing significant SNPs in the HLA region. The plots are showing the QQ-plots post adjustment if $\lambda > 1$

Genomic inflation factors (λ) are indicated for each GWAS

.

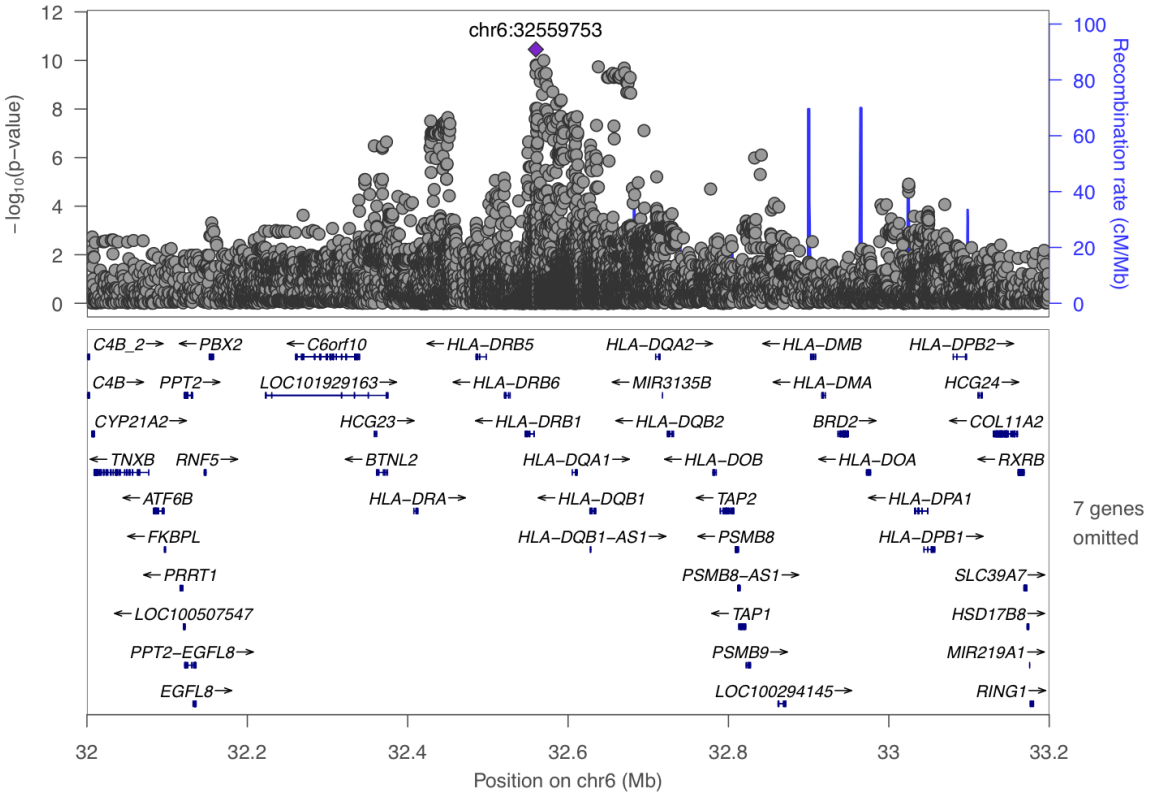

Figure S3: Conditional Analysis SNP-rs1129740/rs1071630.

Trans-ethnic meta-analysis GWAS p-values in the 6p21 region, conditioned on rs1129740, missense SNP identified by Gbadegesin et al[14] (LD=1 with rs1071630). Highlights the top SNP from this analysis, rs28366266 (6:32559753) with p-value 3.5×10^{-11} (I^2 = 5.6%, p-value of Q = 0.3651), which is the same top SNP as that found by conditional analysis on rs1063348.

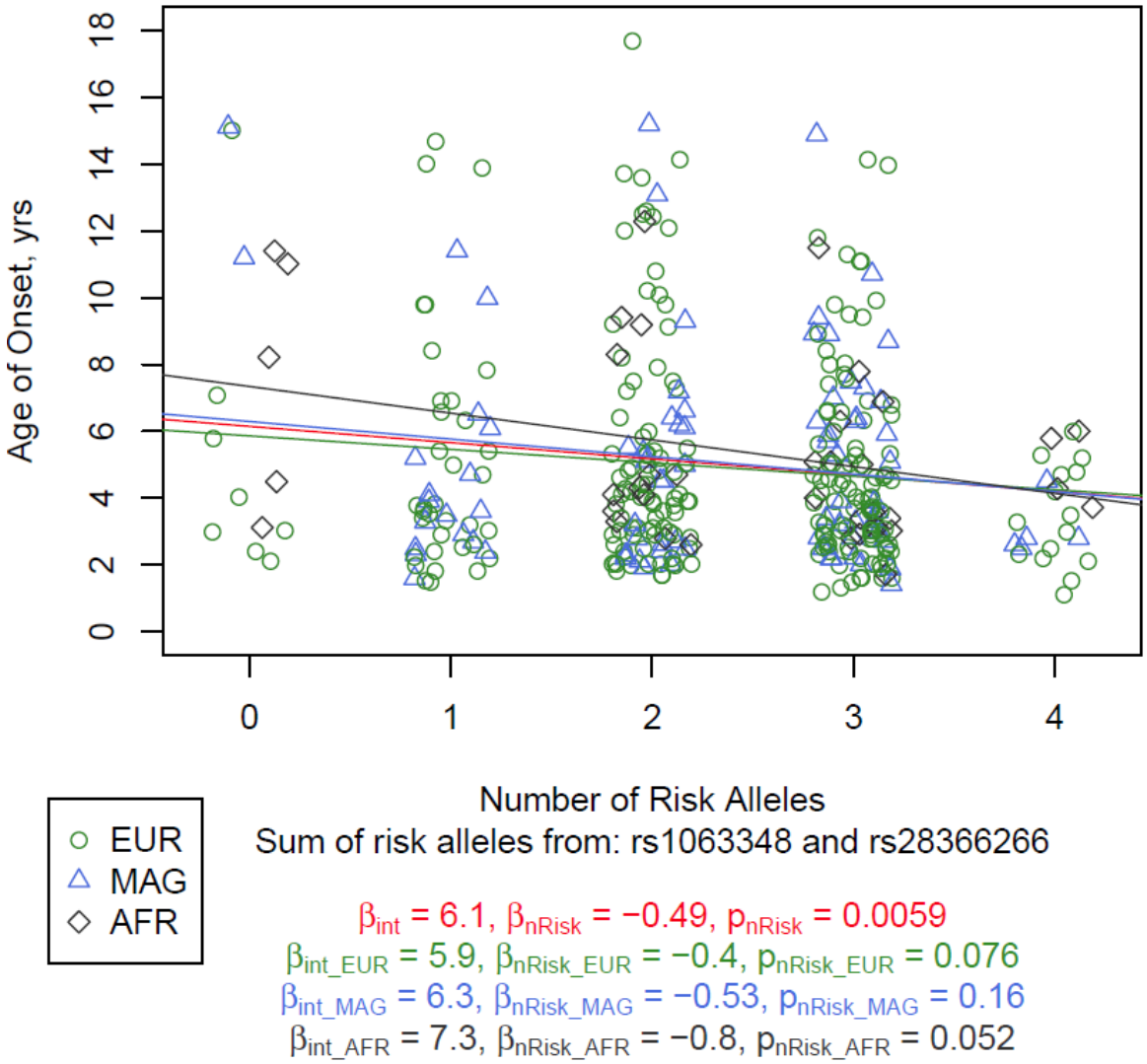

Figure S4: Age of Disease Onset by Number of Risk Alleles Stratified by Ethnicity (GWAS Cohort)

Linear regression of age of onset by summed risk alleles from rs1063348 and rs28366266 in GWAS cohort, stratified by ethnicity. Lines and values correspond to linear regression predictions, with colors corresponding to overall regression (red), regression in Caucasian European (EUR), (green), Maghrebian (MAG), (blue), and African (AFR) patients (black). Note similar, suggestive correlations of lower age of onset with increased number of risk alleles among ethnicities as in the whole GWAS cohort.

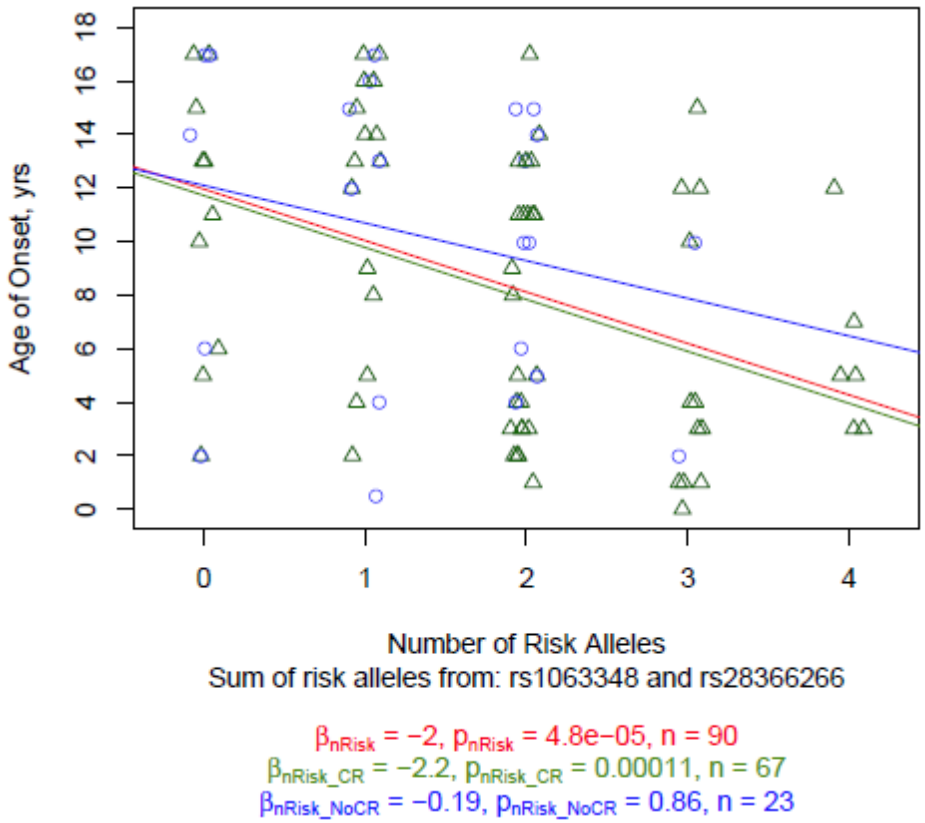

Figure S5: Age of Disease Onset by Number of Risk Alleles and Stratified by Complete Remission (NEPTUNE Cohort)

Linear regression of age of onset by summed risk alleles from rs1063348 and rs28366266 in NEPTUNE cohort, stratified by achievement of complete remission. Lines and values correspond to linear regression predictions, with colors corresponding to overall regression (red), regression in patients with (green) and without (blue)complete remission. Overall, among the various diagnoses, we see a significant relationship with more risk alleles and lower age of onset in those with complete remission.

TABLES

Table S1 Characteristics of NEPTUNE eQTL (39) and Epi (97) cohorts

Characteristics	Pediatric w/Genotypes (n=97)	Pediatric with Genotypes & Expression Data (n=39)	P Value
Age (y)	12 (6-14)	12 (6-15)	0.78
Age of Onset (y)	10 (4-13)	7.5 (3.8-14)	0.97
NS duration at biopsy (y)	0.3 (0.1-1.1)	0.4 (0.2-1.5)	0.34
Male	57 (59)	26 (67)	0.44
Histopathology			0.79
FSGS	30 (31)	10 (26)	
MCD	45 (46)	19 (49)	
MN	1 (1)	1 (3)	
Other	21 (22)	9 (23)	
PC Ancestry			0.72
AFR	47 (48)	19 (49)	
AMR	11 (11)	2 (5)	
ASN	4 (4)	2 (5)	
EUR	35 (36)	16 (41)	
Clinical characteristics at biopsy			
RAAS blockade	53 (55)	19 (49)	0.57
eGFR (ml/min/1.73m^2)	95.9 (74.7-114)	94 (72.5-114.4)	0.96
UPCR (g/g)	1.4 (0.3-4.5)	1.1 (0.1-5)	0.36
Complete Remission	69 (71)	30 (77)	0.53

Data presented as median (interquartile range) for quantitative variables and as count (%) for categorical characteristics. The P values were determined using Wilcoxon rank-sum test or Fisher's exact test.

Table S2: Summary of ancestry-matched_GWAS

Ancestry	Case cohort	Number of cases	Control cohort	Number of controls	Number of genotyped SNPs passing QC	Imputation panel	Number of SNPs passing QC after imputation	Number of significant SNPs	Lead SNP	Lead SNP p-value
European (French)	IDF-EUR	132	3Cités	2000	505979	Haplotype Reference Consortium (release 1.1)	7,529,344	74	rs9274761	1.2×10-9
African	IDF-AFR	56	1000 Genomes - AFR	454	633883	1000 Genomes Phase 3	16,514,340	127	rs2076523	4.1×10-12
Maghrebian	IDFMAG	85	Moroccancontrols	261	666757	Haplotype Reference Consortium (release 1.1)	9,859,054	3	rs28383254	3.0×10-8
European	ItSpa	112	1000 Genomes - EUR	552	1307537	Haplotype Reference Consortium (release 1.1)	7,713,329	136	rs28366276	8.2×10-11

Table S9: Multi-SNP risk analysis in the trans-ethnic cohort, combining risk alleles from SNPs rs1063348 and rs28366266.

Number of Risk Alleles	Case Information			Control Information			Statistics and p-values - within Risk Category					
	N Cases	Tot Cases	Prop Cases	N Controls	Tot Controls	Prop Controls	OR	SE	ciL	ciU	P	FDR
0	15	385	0.039	577	3267	0.18	0.19	0.27	0.1	0.32	**4.2E-15**	**7.0E-15**
1	66	385	0.17	1194	3267	0.37	0.36	0.14	0.27	0.47	**2.4E-15**	**6.0E-15**
2	128	385	0.33	992	3267	0.3	1.14	0.11	0.9	1.44	0.24	0.24
3	152	385	0.39	436	3267	0.13	4.23	0.12	3.35	5.35	**2.8E-32**	**1.4E-31**
4	24	385	0.062	68	3267	0.021	3.13	0.24	1.85	5.12	**1.6E-05**	**2.0E-05**

N Cases-number of cases with given number of risk alleles

Tot Cases-total number of SSNS cases

Prop Cases-proportion of cases with given # of risk alleles

N Controls-number of controls with given number of risk alleles

Tot Controls-total number of SSNS controls

Prop Controls-proportion of controls with given # of risk alleles

OR: Odds Ratio

SE: Standard Error

ciL: Lower bound of confidence interval

ciU: Upper bound of confidence interval

The first column "Number of Risk Alleles" indicates the summed risk alleles from the two lead SNPs rs1063348 and rs28366266.

Table S10a: Top allelotypes and haplotypes at the HLA DQ-DR locus (P < 0.01) in the IDF_EUR cohort

Top allelotypes at the HLA DQ-DR locus (P < 0.01) in the IDF_EUR cohort

CHR	Allelotype	Frequency in controls	Frequency in cases	OR	95% OR confidence interval - lower bound	95% OR confidence interval - upper bound	P-value
6	HLA_DQA1_0102	0.1366	0.07955	0.5464	0.17	0.6176	0.0006162
6	HLA_DQA1_0201	0.1396	0.2348	1.892	1.327	2.966	0.0008462
6	HLA_DRB1_0701	0.1416	0.2348	1.861	1.302	2.898	0.001134
6	HLA_DQB1_0202	0.1093	0.2045	2.095	1.328	3.149	0.001155
6	HLA_DQB1_0602	0.0918	0.03788	0.3895	0.07396	0.5257	0.001174
6	HLA_DRB1_1501	0.09355	0.04924	0.5019	0.1001	0.5923	0.001842

Top haplotypes at the HLA DQ-DR locus (P < 0.01) in the IDF_EUR cohort

CHR	HLA.DQA1	HLA.DQB1	HLA.DRB1	Frequency in controls	Frequency in cases	OR	95% OR confidence interval - lower bound	95% OR confidence interval - upper bound	P-value
6	HLA-DQA1*02:01	HLA-DQB1*02:02	HLA-DRB1*07:01	0.1037	0.2	2.136	1.376	3.316	0.0007236
6	HLA-DQA1*05:01	HLA-DQB1*03:01	HLA-DRB1*03:01	0.0007756	0.007692	29.21	4.127	206.7	0.000725
6	HLA-DQA1*01:02	HLA-DQB1*06:02	HLA-DRB1*15:01	0.0804	0.03077	0.2279	0.08442	0.6152	0.003514
6	HLA-DQA1*03:01	HLA-DQB1*03:02	HLA-DRB1*04:02	0.01732	0.03846	2.844	1.29	6.271	0.009551

Table S10b: Top allelotypes and haplotypes at the HLA DQ-DR locus (P < 0.01) in the ItSpa cohort

Top allelotypes at the HLA DQ-DR locus (P < 0.01) in the ItSpa cohort

CHR	Allelotype	Frequency in controls	Frequency in cases	OR	95% OR confidence interval - lower bound	95% OR confidence interval - upper bound	P-value
6	HLA_DQB1_0202	0.09364	0.2682	3.547	3.453	11.64	2.58E-09
6	HLA_DRB1_0701	0.1264	0.3119	3.134	2.891	8.681	9.22E-09
6	HLA_DQA1_0201	0.1273	0.3119	3.109	2.89	8.68	9.30E-09
6	HLA_DQA1_0102	0.15	0.04018	0.2372	0.1302	0.6409	0.002256

Top haplotypes at the HLA DQ-DR locus (P < 0.01) in the IDF_EUR cohort

CHR	HLA.DQA1	HLA.DQB1	HLA.DRB1	Frequency in controls	Frequency in cases	OR	95% OR confidence interval - lower bound	95% OR confidence interval - upper bound	P-value
6	HLA-DQA1*02:01	HLA-DQB1*02:02	HLA-DRB1*07:01	0.08675	0.2546	5.695	3.14	10.33	1.03E-08

Table S10c: Logistic regression models including the first 3 principal components of genetic ancestry together with the 3 top GWAS SNPs and the risk haplotype (DQA1*02:01-DQB1*02:02-DRB1*07:01).

IDF_EUR vs BC cohort

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-6.755456436	0.984510603	-6.861740658	6.80E-12
PC1	-11.17464693	19.48555363	-0.573483676	0.566317249
PC2	-197.341719	21.70121489	-9.093579322	9.58E-20
PC3	-153.3178757	24.18247214	-6.340041451	2.30E-10
risk haplotype	-11.52538047	488.9389207	-0.02357223	0.981193824
rs1063348	0.942763965	0.20045756	4.703060172	2.56E-06
rs28366266	0.601444796	0.193448527	3.109068879	0.00187678
rs9348883	1.024820274	0.469506787	2.182759232	0.029053545

ItSpa vs 1000G cohort

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-8.116191702	1.231182973	-6.592189693	4.33E-11
PC1	-99.09868141	22.45519989	-4.41317298	1.02E-05
PC2	-193.0789203	28.92118997	-6.676036513	2.45E-11
PC3	19.90727141	33.52392967	0.593822729	0.552630685
risk haplotype	0.236605592	0.468287166	0.505257477	0.613378005
rs1063348	1.079049998	0.299135938	3.607222872	0.000309492
rs28366266	1.270711153	0.430721731	2.950190485	0.003175781
rs9348883	1.523752826	0.523016044	2.913395952	0.00357521

Table S11: Glomerular eQTL results of GWAS SNPs

rs1063348 (6:32627923 Risk Allele: G, Other Allele: A)					
HLA-DRB1	major histocompatibility complex, class II, DR beta 1	-0,70118	-4,97039	0,00003	0,00099
HLA-DRB5	major histocompatibility complex, class II, DR beta 5	-0,72382	-4,46926	0,00012	0,00195
HLA-DQB1	major histocompatibility complex, class II, DQ beta 1	-0,54727	-3,64615	0,00108	0,01183
RNF5	ring finger protein 5	0,35577	2,12690	0,04238	0,34960
HLA-DPB1	major histocompatibility complex, class II, DP beta 1	-0,28737	-1,80947	0,08113	0,48276
HLA-DQB2	major histocompatibility complex, class II, DQ beta 2	-0,24928	-1,76907	0,08778	0,48276
PPT2	palmitoyl-protein thioesterase 2	-0,31396	-1,52864	0,13757	0,64856
GPSM3	G-protein signaling modulator 3	-0,25105	-1,44404	0,15982	0,65927
PSMB8-AS1	PSMB8 antisense RNA 1 (head to head)	0,23683	1,17441	0,25012	0,80446
TAP2	transporter 2, ATP binding cassette subfamily B member	-0,13727	-1,07466	0,29170	0,80446
BRD2	bromodomain containing 2	0,10684	1,07058	0,29350	0,80446
HCG23	HLA complex group 23 (non-protein coding)	0,14205	1,05405	0,30087	0,80446
C6orf10	chromosome 6 open reading frame 10	-0,17542	-0,90490	0,37324	0,80446
BTNL2	butyrophilin like 2	-0,17202	-0,89037	0,38085	0,80446
HLA-DRB6	major histocompatibility complex, class II, DR beta 6 (pseudogene)	0,13014	0,85223	0,40132	0,80446
HLA-DMA	major histocompatibility complex, class II, DM alpha	-0,11578	-0,77416	0,44532	0,80446
HLA-DRA	major histocompatibility complex, class II, DR alpha	-0,10845	-0,74008	0,46541	0,80446
HLA-DQA1	major histocompatibility complex, class II, DQ alpha 1	-0,08658	-0,68444	0,49932	0,80446
AGPAT1	1-acylglycerol-3-phosphate O-acyltransferase 1	-0,13102	-0,67807	0,50329	0,80446
HLA-DPA1	major histocompatibility complex, class II, DP alpha 1	-0,09713	-0,65212	0,51964	0,80446
NOTCH4	notch 4	0,10531	0,63434	0,53101	0,80446
HLA-DPB2	major histocompatibility complex, class II, DP beta 2 (pseudogene)	0,09386	0,61786	0,54166	0,80446
AGER	advanced glycosylation end-product specific receptor	-0,07431	-0,58885	0,56069	0,80446
MIR3135B	microRNA 3135b	-0,05897	-0,43508	0,66684	0,91690

HLA-DQA2	major histocompatibility complex, class II, DQ alpha 2	0,03944	0,31103	0,75808	0,96777
PSMB8	proteasome subunit beta 8	0,04491	0,25514	0,80048	0,96777
HLA-DMB	major histocompatibility complex, class II, DM beta	-0,03246	-0,22055	0,82704	0,96777
PSMB9	proteasome subunit beta 9	0,02733	0,14832	0,88315	0,96777
LOC100294145	uncharacterized LOC100294145	0,02473	0,14491	0,88582	0,96777
TAP1	transporter 1, ATP binding cassette subfamily B member	-0,02048	-0,12364	0,90249	0,96777
PBX2	PBX homeobox 2	-0,00967	-0,05916	0,95324	0,96777
HLA-DOA	major histocompatibility complex, class II, DO alpha	0,00892	0,05204	0,95887	0,96777
HLA-DOB	major histocompatibility complex, class II, DO beta	-0,00732	-0,04077	0,96777	0,96777

rs28366266 (6:32559753 Risk Allele: C, Other Allele T)

HLA-DRB1	major histocompatibility complex, class II, DR beta 1	-0,58450	-1,69676	0,10083	0,98332
PSMB8-AS1	PSMB8 antisense RNA 1 (head to head)	0,60886	1,65206	0,10969	0,98332
HLA-DPB1	major histocompatibility complex, class II, DP beta 1	-0,47425	-1,57746	0,12592	0,98332
MIR3135B	microRNA 3135b	-0,35279	-1,43832	0,16143	0,98332
RNF5	ring finger protein 5	0,46736	1,43725	0,16173	0,98332
BRD2	bromodomain containing 2	0,25046	1,35876	0,18507	0,98332
AGPAT1	1-acylglycerol-3-phosphate O-acyltransferase 1	-0,47818	-1,35569	0,18603	0,98332
HCG23	HLA complex group 23 (non-protein coding)	0,29971	1,19637	0,24158	0,98502
HLA-DOA	major histocompatibility complex, class II, DO alpha	0,36426	1,16348	0,25446	0,98502
HLA-DQA1	major histocompatibility complex, class II, DQ alpha 1	0,21136	0,89925	0,37619	0,98502
HLA-DRB5	major histocompatibility complex, class II, DR beta 5	-0,35076	-0,89788	0,37691	0,98502
C6orf10	chromosome 6 open reading frame 10	-0,32477	-0,89609	0,37784	0,98502
HLA-DQA2	major histocompatibility complex, class II, DQ alpha 2	0,20896	0,89250	0,37973	0,98502
GPSM3	G-protein signaling modulator 3	-0,28877	-0,86879	0,39235	0,98502
BTNL2	butyrophilin like 2	0,29700	0,82073	0,41873	0,98502
TAP2	transporter 2, ATP binding cassette subfamily B member	-0,18936	-0,78575	0,43861	0,98502
HLA-DRB6	major histocompatibility complex, class II, DR beta 6 (pseudogene)	0,17082	0,59460	0,55689	0,98502
ATF6B	activating transcription factor 6 beta	0,16844	0,57883	0,56733	0,98502

NOTCH4	notch 4	0,17797	0,57278	0,57136	0,98502
HLA-DOB	major histocompatibility complex, class II, DO beta	-0,17596	-0,52661	0,60261	0,98502
PSMB8	proteasome subunit beta 8	0,16481	0,50262	0,61916	0,98502
FKBPL	FK506 binding protein like	-0,16714	-0,50034	0,62075	0,98502
AGER	advanced glycosylation end-product specific receptor	0,11486	0,48601	0,63075	0,98502
HLA-DRA	major histocompatibility complex, class II, DR alpha	0,11435	0,41473	0,68150	0,98502
HLA-DPA1	major histocompatibility complex, class II, DP alpha 1	-0,11501	-0,41126	0,68401	0,98502
TNXB	tenascin XB	-0,12092	-0,37828	0,70808	0,98502
TAP1	transporter 1, ATP binding cassette subfamily B member	0,10009	0,32384	0,74846	0,98502
HLA-DQB1	major histocompatibility complex, class II, DQ beta 1	-0,10789	-0,31724	0,75341	0,98502
PBX2	PBX homeobox 2	0,07761	0,25442	0,80103	0,98502
HLA-DMB	major histocompatibility complex, class II, DM beta	0,04875	0,17713	0,86068	0,98502
LOC100294145	uncharacterized LOC100294145	-0,05087	-0,15949	0,87443	0,98502
HLA-DQB2	major histocompatibility complex, class II, DQ beta 2	0,03788	0,13644	0,89245	0,98502
LOC100507547	uncharacterized LOC100507547	-0,03807	-0,10734	0,91528	0,98502
PPT2	palmitoyl-protein thioesterase 2	-0,04062	-0,10168	0,91973	0,98502
HLA-DMA	major histocompatibility complex, class II, DM alpha	-0,01370	-0,04848	0,96168	0,98502
PRRT1	proline rich transmembrane protein 1	-0,01009	-0,03078	0,97566	0,98502
PSMB9	proteasome subunit beta 9	0,00653	0,01895	0,98502	0,98502

rs9348883 (6:32358549 Risk Allele: T, Other Allele: A)					
SKIV2L	Ski2 like RNA helicase	1,10381	-3,62238	0,00115	0,04924
AGER	advanced glycosylation end-product specific receptor	-0,68211	2,48390	0,01925	0,41384
HLA-DRB6	major histocompatibility complex, class II, DR beta 6 (pseudogene)	0,75351	-2,20989	0,03546	0,48200
MIR1236	microRNA 1236	0,95704	-2,10027	0,04484	0,48200
HLA-DQB2	major histocompatibility complex, class II, DQ beta 2	0,63699	-1,90756	0,06675	0,52642
CYP21A2	cytochrome P450 family 21 subfamily A member 2	0,99654	-1,78049	0,08585	0,52642
HLA-DOB	major histocompatibility complex, class II, DO beta	0,70929	-1,73957	0,09292	0,52642
C2	complement C2	0,74542	-1,71208	0,09794	0,52642

ATF6B	activating transcription factor 6 beta	-0,55921	1,55794	0,13048	0,60116
HLA-DRB1	major histocompatibility complex, class II, DR beta 1	0,65583	-1,47220	0,15212	0,60116
EHMT2	euchromatic histone lysine methyltransferase 2	0,47814	-1,45175	0,15768	0,60116
TAP1	transporter 1, ATP binding cassette subfamily B member	0,54146	-1,41614	0,16776	0,60116
NOTCH4	notch 4	-0,41713	1,06517	0,29590	0,94345
TAP2	transporter 2, ATP binding cassette subfamily B member	0,27898	-0,90901	0,37111	0,94345
HCG23	HLA complex group 23 (non-protein coding)	-0,28282	0,87316	0,39000	0,94345
TNXA	tenascin XA (pseudogene)	0,48221	-0,85806	0,39815	0,94345
FKBPL	FK506 binding protein like	-0,35302	0,83325	0,41175	0,94345
HLA-DQB1	major histocompatibility complex, class II, DQ beta 1	0,33845	-0,78573	0,43863	0,94345
TNXB	tenascin XB	0,30160	-0,74353	0,46335	0,94345
GPSM3	G-protein signaling modulator 3	-0,30789	0,72176	0,47642	0,94345
C6orf10	chromosome 6 open reading frame 10	0,33503	-0,71961	0,47773	0,94345
CFB	complement factor B	0,34549	-0,67381	0,50596	0,94345
HLA-DRB5	major histocompatibility complex, class II, DR beta 5	0,30467	-0,60548	0,54974	0,94345
BTNL2	butyrophilin like 2	-0,27075	0,58193	0,56527	0,94345
LOC100507547	uncharacterized LOC100507547	0,24460	-0,54224	0,59194	0,94345
STK19	serine/threonine kinase 19	-0,27521	0,51785	0,60863	0,94345
C2-AS1	C2 antisense RNA 1	0,24444	-0,46486	0,64563	0,94345
PSMB9	proteasome subunit beta 9	-0,18741	0,42691	0,67271	0,94345
AGPAT1	1-acylglycerol-3-phosphate O-acyltransferase 1	0,19379	-0,41769	0,67936	0,94345
PSMB8	proteasome subunit beta 8	0,15433	-0,36745	0,71604	0,94345
PBX2	PBX homeobox 2	-0,14305	0,36733	0,71613	0,94345
HLA-DRA	major histocompatibility complex, class II, DR alpha	-0,12444	0,35279	0,72689	0,94345
DXO	decapping exoribonuclease	-0,12201	0,33148	0,74275	0,94345
PRRT1	proline rich transmembrane protein 1	0,12272	-0,29335	0,77142	0,94345
MIR3135B	microRNA 3135b	0,08569	-0,26409	0,79364	0,94345
PSMB8-AS1	PSMB8 antisense RNA 1 (head to head)	-0,12852	0,26075	0,79620	0,94345
HLA-DQA1	major histocompatibility complex, class II, DQ alpha 1	-0,07317	0,24035	0,81181	0,94345
ZBTB12	zinc finger and BTB domain containing 12	0,04366	-0,12540	0,90110	0,96808
NELFE	negative elongation factor complex member E	0,04163	-0,12021	0,90518	0,96808

CYP21A1P	cytochrome P450 family 21 subfamily A member 1, pseudogene	0,05440	-0,10664	0,91584	0,96808
HLA-DQA2	major histocompatibility complex, class II, DQ alpha 2	0,02957	-0,09746	0,92305	0,96808
RNF5	ring finger protein 5	-0,01001	0,02325	0,98162	0,99452
PPT2	palmitoyl-protein thioesterase 2	-0,00354	0,00692	0,99452	0,99452

Table S12: Linkage disequilibrium of new SNPs with previously discovered SSNS SNPs

LD with rs1063348 (6:32627923_A/G)		
IDFAFR	rs1129740	0,885
IDFAFR	rs1071630	0,921
IDFMAG	rs1129740	0,961
IDFMAG	rs1071630	0,961
IDFEUR	rs1071630	0,966
IDFEUR	rs1129740	0,979
ItSpa	rs1129740	0,997
ItSpa	rs1071630	0,997
LD with rs28366266 (6:32559753_T/C)		
IDFAFR	rs1129740	0,070
IDFAFR	rs1071630	0,108
IDFEUR	rs1129740	0,123
IDFMAG	rs1129740	0,124
IDFMAG	rs1071630	0,124
IDFEUR	rs1071630	0,124
ItSpa	rs1129740	0,135
ItSpa	rs1071630	0,135

LD (measured as R2) for the replicated Gbadegesin et al[14] SNPs (rs1129740 and rs1071630) with the top two independent GWAS SNP from this study (rs1063348 andrs28366266). LD was calculated within each population individually: African (IDF), Maghrebian (IDF), European (IDF) and European (ItSpa). We observe that the rs1129740 & rs1071630 are only in strong LD with rs1066348, suggesting that the other loci discovered in this study are independent of those SNPs previously discovered.

Table S13: Glomerular eQTL results of rs1129740/rs1071630

rs1129740 (6:32609105_G/A) and rs1071630 (6:32609126_T/C)					
HLA-DRB1	major histocompatibility complex, class II, DR beta 1	-0,579	-4,187	2,54E-04	0,0089
HLA-DRB5	major histocompatibility complex, class II, DR beta 5	-0,584	-3,687	0,0010	0,0169
HLA-DQB1	major histocompatibility complex, class II, DQ beta 1	-0,436	-3,024	0,005	0,062
RNF5	Ring finger protein 5	0,352	2,343	0,026	0,232
PSMB8-AS1	PSMB8 antisense RNA 1 (head to head)	0,355	2,020	0,053	0,369
HLA-DQB2	major histocompatibility complex, class II, DQ beta 2	-0,246	-1,934	0,063	0,369
GPSM3	G-protein signaling modulator 3	-0,268	-1,717	0,097	0,472
HLA-DPB1	major histocompatibility complex, class II, DP beta 1	-0,242	-1,661	0,108	0,472
PPT2	palmitoyl-proteinthioesterase 2	-0,267	-1,420	0,167	0,648
AGPAT1	1-acylglycerol-3-phosphate O-acyltransferase 1	-0,232	-1,349	0,188	0,658
LOC100507547	uncharacterized LOC100507547	-0,210	-1,250	0,222	0,705
BRD2	Bromodomain containing 2	0,104	1,150	0,260	0,758
HLA-DRB6	major histocompatibility complex, class II, DR beta 6 (pseudogene)	0,138	0,998	0,327	0,838
HLA-DQA1	major histocompatibility complex, class II, DQ alpha 1	-0,112	-0,981	0,335	0,838
HLA-DPB2	major histocompatibility complex, class II, DP beta 2 (pseudogene)	0,114	0,826	0,416	0,892
C6orf10	chromosome 6 open reading frame 10	-0,144	-0,812	0,424	0,892
PSMB8	Proteasome subunit beta 8	0,126	0,795	0,433	0,892
PSMB9	Proteasomesubunit beta 9	0,122	0,733	0,470	0,894
HLA-DMA	major histocompatibility complex, class II, DM alpha	-0,094	-0,692	0,495	0,894
MIR3135B	microRNA 3135b	-0,082	-0,666	0,511	0,894
HLA-DPA1	major histocompatibility complex, class II, DP alpha 1	-0,077	-0,567	0,575	0,933
TAP2	transporter 2, ATP binding cassette subfamily B member	-0,063	-0,536	0,596	0,933
HLA-DRA	major histocompatibility complex, class II, DR alpha	-0,066	-0,494	0,625	0,933
TAP1	transporter 1, ATP binding cassette subfamily B member	0,071	0,471	0,641	0,933
PRRT1	Proline rich transmembrane protein 1	0,069	0,436	0,666	0,933
HLA-DMB	major histocompatibility complex, class II, DM beta	0,052	0,389	0,700	0,943
BTNL2	Butyrophilinlike 2	-0,061	-0,344	0,734	0,951
PBX2	PBX homeobox 2	-0,042	-0,281	0,781	0,956
AGER	advanced glycosylation end-product specific receptor	-0,031	-0,267	0,792	0,956
HLA-DOB	major histocompatibility complex, class II, DO beta	0,033	0,199	0,844	0,985
HLA-DQA2	major histocompatibility complex, class II, DQ alpha 2	0,015	0,133	0,895	0,990
NOTCH4	notch 4	0,018	0,121	0,905	0,990
HCG23	HLA complex group 23 (non-protein coding)	-0,005	-0,037	0,970	0,991
LOC100294145	uncharacterized LOC100294145	0,004	0,023	0,982	0,991
HLA-DOA	major histocompatibility complex, class II, DO alpha	0,002	0,011	0,991	0,991

Glomerular cis-eQTL results in NEPTUNE pediatric patients for the two replicated SSNS SNPs in Gbadegesin et al.[14]. Because of their close physical distance and perfect LD, the eQTL results for these two SNPs are identical, and therefore have been combined. We see that because of the relatively high LD

between these SNPs and rs1063348, the cis-eQTL results are very similar. This is particularly true for the significant eQTLs for rs1063348; HLA-DRB1, HLA-DRB5, and HLA-DQB1. While the overall statistics for these eQTLs are less significant for rs1129740/rs1071630 (and HLA-DQB1 loses significance at FDR < 0.05), the direction and relative magnitude remain consistent with those cis-eQTL results from rs1063348

Table S14:Clinical associations with allele frequency for all NEPTUNE pediatric patients, adjusted for PCs and APOL1 status.

		rs1063348		rs28366266		rs9348883	
Category		**Risk Allele Frequency G**	**p-value**	**Risk Allele Frequency C**	**p-value**	**Risk Allele Frequency T**	**p-value**
All Alleles (194)		0.67 (130) {0.53 in 1000G}	2,60E-04	0.11 (21) {0.17 in 1000G}	0,0429	0.94 (183) {0.89 in 1000G}	0,0338
Gender	Male (114)	0.67 (76)	1	0.09 (10)	0,35	0.94 (107)	1
	Female (80)	0.68 (54)	OR=0.96 [0.5,1.85]	0.14 (11)	OR=0.6 [0.22,1.66]	0.95 (76)	OR=0.81 [0.17,3.3]
Morphology	MN (2)	1 (2)	0,52	0.5 (1)	0,11	1 (2)	1
	MCD (90)	0.62 (56)		0.1 (9)		0.94 (85)	
	FSGS (60)	0.72 (43)		0.15 (9)		0.93 (56)	
	Other (42)	0.69 (29)		0.05 (2)		0.95 (40)	
RAAS Blockade	Yes (36)	0.56 (20)	0,12	0.06 (2)	0,38	0.86 (31)	0,033
	No (158)	0.7 (110)	OR=0.55 [0.25,1.23]	0.12 (19)	OR=0.43 [0.05,1.93]	0.96 (152)	OR=0.25 [0.06,1.09]
Reported Race	White (80)	0.75 (60)	0,1	0.11 (9)	0,98	0.96 (77)	0,34
	Black (82)	0.61 (50)		0.12 (10)		0.93 (76)	
	Asian (8)	0.5 (4)		0.12 (1)		0.88 (7)	
	Multi-Race (16)	0.81 (13)		0.06 (1)		1 (16)	
PCA Ancestry	EUR (70)	0.76 (53)	0,21	0.14 (10)	0,25	0.96 (67)	0,36
	AFR (94)	0.62 (58)		0.11 (10)		0.93 (87)	
	ASN (8)	0.75 (6)		0.12 (1)		0.88 (7)	
	AMR (22)	0.59 (13)		0 (0)		1 (22)	
Complete Remission	Yes (138)	0.72 (99)	0,042	0.14 (19)	0,042	0.93 (129)	0,52

	No (56)	0.55 (31)	OR=2.04 [1.02,4.09]	0.04 (2)	OR=4.29 [0.98,39.27]	0.96 (54)	OR=0.53 [0.05,2.69]
MCD - Complete Remission	Yes (78)	0.67 (52)	0,051	0.1 (8)	1	0.94 (73)	1
	No (12)	0.33 (4)	OR=3.93 [0.95,19.56]	0.08 (1)	OR=1.25 [0.14,60.77]	1 (12)	OR=NA
FSGS - Complete Remission	Yes (36)	0.78 (28)	0,25	0.22 (8)	0,072	0.92 (33)	0,64
	No (24)	0.62 (15)	OR=2.07 [0.58,7.67]	0.04 (1)	OR=6.41 [0.76,303.56]	0.96 (23)	OR=0.48 [0.01,6.46]
eGFR	<90 (84)	0.57 (48)	0,012	0.07 (6)	0,23	0.96 (81)	0,52
	>=90 (104)	0.75 (78)	OR=0.45 [0.23,0.86]	0.13 (14)	OR=0.5 [0.15,1.46]	0.93 (97)	OR=1.94 [0.43,12.01]
UPC	<2 (64)	0.64 (41)	0,63	0.11 (7)	1	0.94 (60)	1
	>=2 (128)	0.68 (87)	OR=0.84 [0.43,1.67]	0.1 (13)	OR=1.09 [0.35,3.12]	0.95 (121)	OR=0.87 [0.21,4.21]

REFERENCES

1. Nelson MR, Bryc K, King KS, et al. The Population Reference Sample, POPRES: A Resource for Population, Disease, and Pharmacological Genetics Research. *Am J Hum Genet* 83:347-358, 2008

2. Anderson CA, Pettersson FH, Clarke GM, et al. Data quality control in genetic case-control association studies. *Nat Protoc* 5:1564-1573, 2010

3. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559-575, 2007

4. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics* 48:1279-1283, 2016

5. Loh P-R, Danecek P, Palamara PF, et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nature Genetic* 48:1443-1448, 2016

6. Durbin R. Efficient haplotype matching and storage using the positional Burrows-Wheeler transform (PBWT). *Bioinformatics* 30:1266-1272

7. 3C Study Group. Vascular factors and risk of dementia: design of the Three-City Study and baseline characteristics of the study population. *Neuroepidemiology*22:316-325, 2003

8. 1000 Genomes Project Consortium et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* 491:56–65, 2012

9. Li Y, Sidore C, Kang HM, et al. Low-coverage sequencing: implications for design of complex trait associationstudies. *Genome Res* 21:940-951, 2011

10. Price AL, Patterson NJ, Plenge RM, et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38:904-909, 2006

11. Parts L, Stegle O, Winn J, Durbin R. Joint genetic analysis of gene expression data with inferred cellular phenotypes. *PLoS Genet* 7:e1001276, 2011

12. Stegle O, Parts L, Durbin R, Winn J. A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. *PLoS Comput Biol* 6:e1000770, 2010

13. Kang H. Efficient and parallelizable association container toolbox (EPACTS). 2014.

14. Gbadegesin RA, Adeyemo A, Webb NJ, et al. HLA-DQA1 and PLCG2 Are Candidate Risk Loci for Childhood-Onset Steroid-Sensitive Nephrotic Syndrome. *J Am Soc Nephrol* 26:1701-1710, 2015