Genetic identification of two novel loci associated with steroid-sensitive nephrotic syndrome

1 Stephanie Dufek*, 1 Chris Cheshire*, 1 Adam P Levine*, 1 Richard S Trompeter*, 1 Naomi Issler, 1 Matthew Stubbs, 1 Monika Mozere, 1 Sanjana Gupta, 1 Enriko Klootwijk, 1 Vaksha Patel, 2 Daljit Hothi, 2 Aoife Waters, 2 Hazel Webb, 2 Kjell Tullus, 2 Lucy Jenkins, 2 Lighta Godinho, 3 Elena Levtchenko, 4 Jack Wetzels, 5 Nine Knoers, 6 Nynke Teeninga, 6 Jeroen Nauta, 7 Mohamed Shalaby, 7 Sherif Eldesoky, 7 Jameela A Kari, 8 Shenal Thalgahagoda, 8 Randula Ranawaka, 8 Asiri Abeyagunawardena, 9 Adebowale Adeyemo, 10 Mark Kristiansen, 11 Rasheed Gbadegesin, 12 Nicholas J Webb, 1 Daniel P Gale*, 1 Horia C Stanescu*, 1 Robert Kleta*, 1 Detlef Bockenhauer*

* contributed equally

Affiliations:
1. Department of Renal Medicine, University College London, London, UK
2. Great Ormond Street Hospital, London, UK
3. University Hospitals Leuven & University of Leuven, Leuven, Belgium
4. Nephrology, Radboud University Medical Center, Nijmegen, The Netherlands
5. Department of Genetics, UMC Groningen, Groningen, The Netherlands
6. Department of Pediatric Nephrology, Erasmus University Medical Centre-Sophia Children's Hospital, Rotterdam, The Netherlands
7. Pediatric Nephrology Center of Excellence, King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia
8. Department of Paediatrics, University of Peradeniya, Sri Lanka
9. NHGRI, NIH, Bethesda, MD, USA
10. UCL Genomics, Institute of Child Health, UCL, London, UK
11. Department of Pediatrics, Duke University School of Medicine, Durham, NC, USA
12. Department of Paediatric Nephrology & NIHR Manchester Clinical Research Facility, Manchester Academic Health Science Centre, Royal Manchester Children's Hospital, Manchester, UK

Corresponding Authors:
Detlef Bockenhauer, d.bockenhauer@ucl.ac.uk & Robert Kleta, r.kleta@ucl.ac.uk
UCL Department of Renal Medicine
Rowland Hill Street, London NW3 2PF, UK
phone: +44 (0)20 7314 7554; fax: +44 (0)20 7472 6476

Word count: 2638 (excluding methods)

Key words: GWAS, SSNS, HLA, CALHM6, FAM26F

Running title: Genetics of SSNS

Abstract

Background: Steroid-sensitive nephrotic syndrome (SSNS) is the most common form of nephrotic syndrome in childhood. SSNS is considered an autoimmune disease with an established classical human leukocyte antigen (HLA) association; however, the precise etiology of the disease is unclear. In other autoimmune diseases, the identification of loci outside the classical HLA region by genome-wide association studies (GWAS) have provided critical insights into their pathogenesis. Previously conducted GWAS of SSNS have failed to identify non-HLA loci achieving genome-wide significance.

Methods: Here we provide insight into the genetic architecture of SSNS by performing a GWAS on a large cohort of European ancestry comprising 422 cases and 5,642 controls.

Results: We confirm the previously reported association with the HLA-DR/DQ region (lead SNP rs9273542, $P=1.59\times10^{-43}$, OR=3.39) and identify additional loci on chromosome 6q22.1, containing the gene *CALHM6* (previously called *FAM26F*) and 4q13.3. *CALHM6* is implicated in immune response modulation and the lead SNP (rs2637678, $P=1.27\times10^{-17}$, OR=0.51) exhibits strong expression quantitative trait loci (eQTL) effects, the risk allele being associated with lower lymphocytic expression of *CALHM6*.

Conclusions: Our results suggest a genetically conferred risk of immune dysregulation as a key component in the pathogenesis of SSNS.

Significance Statement

SSNS is the most common glomerulopathy in childhood. While considered an autoimmune disease, its etiology is poorly understood. GWAS have provided important insights into other autoimmune diseases, but so far, only associations in the classical HLA region have been reported for SSNS. Here, we report the discovery of two loci outside the HLA region associated with SSNS at genome-wide significance. The locus with strongest association contains the gene *CALHM6*, which has been implicated in the regulation of the immune system. Our results suggest impaired down regulation of the immune system as a key mechanism in the pathogenesis of SSNS.

Introduction

Steroid-sensitive nephrotic syndrome (SNSS) is the most common form of nephrotic syndrome in children with an incidence of approximately one to ten per 100,000.[1] The majority of affected children experience a chronic relapsing course. The onset of disease manifestations is commonly associated with a preceding activation of the immune system, typically by an upper respiratory tract infection.[2] As the name implies, SSNS is characterized by a therapeutic response to glucocorticoids, as well as to other immunosuppressants. The apparent triggering of the disease by infection and the therapeutic effect of immunosuppressive treatment have suggested that SSNS is an autoimmune disorder.[3]

Investigating the genetic architecture of other autoimmune diseases through genome-wide association studies (GWAS) has proved successful in providing insight into their etiology and pathogenesis. Unsurprisingly, a common finding in these studies is the identification of association to the human leukocyte antigen (HLA) region which contains numerous genes critical for the immune system, in particular enabling the distinction between self and foreign.[4]

Arguably, however, it is the identification of risk loci outside the HLA locus that can provide the most informative mechanistic insights into the pathogenesis of such diseases. Prominent examples from nephrology include membranous nephropathy (MN) and IgA nephropathy (IGAN). In MN, GWAS identified association with *PLA2R1* suggesting that the antibody formation against the PLA2R1 receptor observed in MN is a causal disease

mechanism.[5, 6] Similarly, in IGAN, GWAS have highlighted the important role of the intestinal immune response, as well as IgA1 antibody glycosylation in the pathogenesis of the disease.[7-9]

Three GWAS of SSNS have recently been reported, but in all association at genome-wide significance was identified within the HLA region only.[10-12] We set out to perform a GWAS employing the largest number of ethnically homogenous cases and controls studied to date in an attempt to identify further loci associated with SSNS.

Methods

Full details of all methods can be found in the Supplements.

Cohorts

DNA from pediatric patients diagnosed with SSNS and of reported European ethnicity were utilized in this study. SSNS was defined according to standard clinical criteria.[13] DNA was acquired from the PREDNOS (EudraCT 2010-022489-29) and PREDNOS2 (EudraCT 2012-003476-39) trials,[14] as well as from collaborating clinicians at their affiliated institutions. Informed consent was obtained from all participants and ethical approval was granted by the host institutions. Ethnically matched controls were drawn from publicly available datasets (Supplementary Figure 1). For the replication of our findings, we examined the results of previously published GWAS of SSNS.[10-12]

Genotyping, Quality control and Imputation

Isolation of DNA, genotyping, quality control (QC) and imputation were performed using standard procedures. Cases were genotyped by UCL Genomics (Institute of Child Health, UCL, London, UK) on the Infinium Multi-Ethnic Global BeadChip v.A1 (Illumina, CA, USA). Controls had been genotyped on a variety of platforms (for further details see Supplements).[15-17] Stringent quality control steps for both SNPs and individuals (including missingness, heterozygosity and deviation from Hardy-Weinberg equilibrium) were carried out on cases and controls separately, as well as on the combined cohort (Supplementary Figure 1). Principal component analysis was employed to identify the subset of cases and controls of European ancestry (Supplementary Figure 4). The

genomic inflation factor (lambda) was calculated to estimate population stratification. Imputation was performed on the combined case-control dataset with Beagle 5.0 (https://faculty.washington.edu/browning/beagle/beagle.html) using only markers passing stringent QC and present in all datasets.[18] Only SNPs with a Dosage R-Squared (DR2) of >0.8 were included. Data from the 1000 Genomes Project Phase 3 were used as reference panel. Golden Helix SNP & Variation Suite version 8.8.1 (SVS, http://goldenhelix.com/products/SNP_Variation/index.html) and PLINK version 1.90 beta (https://www.cog-genomics.org/plink/1.9) were used for analysis.[19]

GWAS and conditional analysis

The primary association analysis was performed using logistic regression under an additive model with adjustment for the first ten principal components of ancestry.

Conditional analysis of the lead SNPs was performed using a logistic regression model. A genome-wide significance threshold of $P<5\times10^{-8}$ was utilized. SVS was used for association testing and conditional analysis. A power calculation is detailed in the supplements.

HLA type imputing

HLA imputation was performed using SNP2HLA v1.0.3 (http://software.broadinstitute.org/mpg/snp2hla/) with default parameters.[20] As input, a subset of 1,189 SNPs from those selected for GWAS (post QC) and overlapping with the SNP2HLA imputation HapMap European reference dataset were used. Logistic regression under an additive model with adjustment for the first ten principal components

of ancestry was used to test for association of each HLA allele with SSNS. Conditional analysis of the lead HLA alleles was performed, also using a logistic regression model.

eQTL analysis

Publicly available expression quantitative trait loci (eQTL) databases were queried to ascertain whether variants significantly associated with SSNS were known to influence the expression levels of corresponding gene products in multiple tissues.

Results

Study cohort

DNA from a total of 712 anonymous SSNS cases of reported European ethnicity were available to our project. The combined control dataset consisted of 6,126 individuals of reported European ethnicity. After stringent quality control (QC) and selection for European ancestry by principal component analysis, 422 cases and 5,642 controls remained (Supplementary Figure 1) with a total of 158,217 overlapping SNPs. These were imputed to 5,216,266 high-quality genome-wide SNPs. The summary statistics are available in the supplements.

GWAS results

Our GWAS revealed three loci achieving genome-wide significance (Figure 1). Together, these loci explain approximately 14% of the genetic risk for SSNS. The strongest signal corresponded to a broad peak in the classical HLA region on chromosome 6p21.32. The three leading SNPs, in strong linkage disequilibrium (LD) with each other, are: rs9273542 ($P=1.59x10^{-43}$, OR=3.39, 95%CI 2.86-4.03) and rs9273529 ($P=2.87x10^{-43}$, OR=3.39, 95%CI=2.85-4.03) in the intronic region of the gene HLA-DQB1, and rs9273371 ($P=1.64x10^{-43}$, OR=3.29, 95%CI=2.78-3.89) intergenic between HLA-DQA1 and HLA-DQB1 (Figure 2A).

Conditional analysis on rs9273542 decreased the strength of the association such that the minimum p-value achieved at this locus changed to $P=4.29\times10^{-31}$ for rs2858317, centromeric of *HLA-DQB1* (Figure 2B). Joint conditioning on rs9273542 and rs2858317

significantly reduced the strength of the association to a minimum p-value of $P=2.40\times10^{-8}$ at rs3828799 (Figure 2C). These results indicate that the association at this locus is driven by at least two independent signals.

Our two lead SNPs (rs9273542 and rs2858317) are in strong linkage disequilibrium with with rs4642516, identified by Jia *et al*.[12] and with two markers, rs1063348 and rs28366266, identified by Debiec *et al*..[11]

The strongest association outside the HLA region was with a locus on chromosome 6q22.1. The three lead SNPs, rs2637678 ($P=1.27x10^{-17}$, OR=0.51, 95%CI=0.44-0.60), rs2637681 ($P=3.53x10^{-17}$, OR=0.52, 95%CI=0.44-0.61) and rs2858829 ($P=1.72x10^{-16}$, OR=0.53, 95%CI=0.45-0.62) are all in close LD and localize around the gene *CALHM6* (previously called *FAMF26F*) (Figure 3A). Genome-wide significance was lost by conditioning on rs2637678 (Figure 3B) indicating a single signal is responsible for driving this association.

The third locus to reach genome-wide significance was on chromosome 4q13.3. The lead SNP, rs10518133, is in the intronic region of the gene *PARM1* and was associated at $P=2.50\times10^{-8}$ with OR 1.96 (95% CI=1.57-2.45) (Figure 4A). Genome-wide significance was also lost at this locus by conditioning on the lead SNP, indicating a single signal is driving the association (Figure 4B).

A Manhattan plot, as well as locus zoom plots of the identical regions, using genotyped markers only are provided as supplemental figures 5-8 and details for the lead genotyped SNPs at the respective loci are listed in supplemental table 1.

Classic HLA type analysis

HLA allele imputation was performed for HLA class I (*HLA-A*, *-C* and *-B*) and class II (*HLA-DRB1* and *-DQB1*) genes. Genome-wide significance for SSNS was achieved for nine HLA alleles, including high resolution subtypes at, *DQA1*01*, *DQA1*02*, *DQA1*07*, *DQA1*13* and *B*8* (Table 2 and Supplementary Figure 2). The strongest association was observed with *HLA-DQA1*02:01*, which resides on a haplotype with *HLA-DRB1*07:01* (indicated by equal allele frequencies in cases and controls). The *HLA-DQA1*01* allele was protective and remained independently associated at genome-wide significance after conditioning on *HLA-DQA1*02:01* ($p=1.24\times10^{-31}$, OR=0.31, 95% CI=0.25-0.38).

eQTL analysis

The three leading SNPs at the *CALHM6* locus (rs2637678, rs2637681, rs2858829) exhibit strong cis-eQTL effects in the GTEx database.[21] For all three SNPs, the highest normalized effect size is seen in EBV-transformed lymphocytes (rs2637678: normalized effect size [NES]=0.66, $p=7.2\times10^{-10}$; rs2637681: NES=0.67, $p= 1.9\times10^{-9}$; rs2858829: NES=0.66, $p=3.2\times10^{-9}$). A strong cis-eQTL effect is also noted in the eQTLgen database from blood (Z-score=43.18, $p=3.27\times10^{-310}$), which also involves the neighboring genes *DSE*, *RWDD1* and *NT5DC1*.[22] In both databases, the minor allele is associated with increased expression of *CALHM6*. Of note, no expression of *CALHM6* is noted in whole kidney in the GTEx, NephQTL and Human Kidney Atlas Expression (HKAE) databases.[23, 24] HKAE suggests specific expression in glomerulus, but this could not be confirmed in the other databases.

No significant eQTLs were found for rs10518133 (tagging the *PARM1* locus) in any tissue in the GTEx database, but in eQTLgen, the risk allele is associated with significant downregulation of *PARM1*.

Discussion

This study has confirmed the association of SSNS with the HLA locus and identified two additional loci on chromosome 6 and chromosome 4 achieving genome-wide significance.

HLA locus

By far the most associations identified by GWAS of human diseases are located in the HLA region.[25] The vast majority of these phenotypes are autoimmune or infectious diseases. The necessity for an appropriate and ever evolving response to infection drives variation in HLA peptide binding grooves and some of these variants can lead to inappropriate responses against antigens of the host.[26] Thus, unsurprisingly and consistent with previous studies, the strongest signal for association with SSNS we identified was in the HLA region.[10-12] The lead SNP, rs9273542 (Table 1), is located within the *HLA-DR/DQ* region, specifically *HLA-DQB1*.

Imputation of HLA alleles identified that *HLA-DQA1*02:01* and *HLA-DRB1*07:01* are associated with the strongest risk, whereas *HLA-DQA1*01* appears to be independently protective. *HLA-DRB1*13* ($p=2.41\times10^{-14}$) is also protective, which is similar to the results of Jia *et al.* who showed in an Asian population that *HLA-DRB1*13:02* and *HLA-DQB1*06:04* have protective effects.[12] In the study by Debiec *et al.*, serial conditional analysis of the HLA association revealed an independently associated SNP located within *BTNL2 (rs9348883)*, although this did not reach genome-wide significance.[11] We were unable to confirm an independent association with this gene, which may reflect the

different ethnicities analyzed, as the *BTNL2* signal was primarily driven by the African cohort in that study.

Associations outside the HLA region

The adaptive immune response is triggered by HLA peptide - epitope binding, but is modulated by regulatory mechanisms, elements of which are encoded in non-HLA genes.[4] Arguably, it is the identification of these genes that can provide the most informative insights into the complex architecture of the dysregulated immune response in specific diseases. We here describe the discovery of two loci outside the HLA region achieving genome-wide significant associations with SSNS.

CALHM6 *association (6q22.1)*

The strongest signal outside the HLA region is on chromosome 6q22.1 and the lead SNP, rs2637678 (Table 1), is located very near the gene *CALHM6* (Calcium Homeostasis Modulator Family Member 6), previously also annotated as *FAM26F* or *INAM* (IRF-3-dependent NK-activating molecule).[27] Of note, this locus had been identified as a potential signal in SSNS by Debiec *et al.*, but did not reach genome-wide significance.[11] The 6q22.1 SNP reported by Debiec *et al.* is identical to one of the lead SNPs identified in this study (rs2858829) and was associated with a p-value of 6.8×10^{-8}. Importantly, the existence of this published suggestive association at rs2858829 provides independent confirmatory evidence for our genome-wide significant finding.

The apparent absence of expression of *CALHM6* in the kidney is in contrast to the high expression in lymphocytes and consistent with the notion that impaired immune regulation is a key risk factor for the development of SSNS. Indeed, CALHM6 is thought to have an important role in the regulation of the immune system, by facilitating interactions and potential synapses between immune cells.[28] It is differentially regulated in response to various immune stimuli, especially interferon-γ, with predicted interferon-stimulated response element (ISRE) and signal transducer and initiator of transcription (STAT) binding sites in its promoter.[29] Members of the CALHM family are thought to belong to the ATP-release channel superfamily.[30] *CALHM6* is highly expressed in lymphocytes and ATP is a recognized trigger for apoptosis, including of immune cells.[31, 32] A key effect of glucocorticoids on the immune system is the induction of lymphocyte apoptosis.[33] Thus, CALHM6 may play a role in mediating this effect and impaired CALHM6 function may exacerbate an exaggerated immune response, leading to SSNS, which can be suppressed by pharmacologic doses of glucocorticoids. Of note, the same locus with the same risk allele, has been previously associated with another autoimmune disease highly sensitive to steroid treatment, ulcerative colitis, providing further evidence for the importance of CALHM6 in immune response regulation.[34]

CALHM6 *eQTL analysis*

We performed an eQTL analysis to further assess potential associations between our identified variants and gene expression.[35] Of note, for the lead variants in *CALHM6*, the presence of the minor allele is protective against SSNS, so that the major allele is the risk allele and enriched in patients. Interestingly, the lead variant (rs2637678) at the *CALHM6*

locus is a known eQTL and the risk allele is associated with decreased *CALHM6* expression. This is consistent with the hypothesis that the presence of the risk allele may impair lymphocyte apoptosis and thus lead to less effective downregulation of an immune response.

However, it is important to note that GWAS identify variants that represent a haploblock rather than a specific gene associated with the disease. This is best illustrated by the initial controversy over the association of *MYH9* versus *APOL1* with non-diabetic kidney disease in African Americans.[36] Thus, it is possible that another gene in the haploblock with *CALHM6* may actually be causally associated. Indeed, in the eQTLGen database, rs2637678 has highly significant eQTL effects also on neighboring genes. *CALHM6* was prioritised based on its proximity to the GWAS signal and its known function but further studies will be required to establish with certainty if this is indeed the gene responsible for driving the observed association.

PARM1 *association (4q13.3)*

In addition, we found a genome-wide significant association with a locus on chromosome 4q13.3 (lead SNP rs10518133, $P=2.50\times10^{-8}$, OR=1.96), which is located within the gene *PARM1* (Prostate androgen-regulated mucin-like protein 1). Interestingly, this locus is near a locus reported by Debiec *et al*. as achieving suggestive evidence of association with SSNS, with their lead SNP located within *BTC* (Betacellulin), the next gene upstream of *PARM1* on chromosome 4.[11] The lead SNP at 4q13 reported by Debiec *et al*. and the *PARM1* SNP identified in this study are approximately 250 kb apart but are separated by a strong recombination hotspot (> 50cM/Mb, Supplementary Figure 3). This hot spot is at

least equally strong in African populations (supplemental table 2), so that the different ethnicities cannot explain the separation of this locus between the previous and our study.

Limitations

Our study has several limitations. First, only limited clinical information on our patients is available. However, the majority of cases were obtained through two clinical trials (PREDNOS and PREDNOS2) which recruited from more than a hundred pediatric units across the UK and it is thus highly unlikely that our patient cohort is substantially different from other SSNS cohorts. Indeed, the available data, such as the 2:1 male:female ratio, matches perfectly with published data.[13]

Next, we used publicly available genotype data for controls, so that data from different genotyping platforms needed to be combined. As detailed in supplementary figure 1, this led to a limited set of overlapping genotyped SNP and the majority of SNPs used in our final analysis were imputed. However, imputation has become an accepted tool in GWAS and the fact that the analysis with genotyped markers only (supplementary figure 5) identifies the same loci, albeit with higher p-values, provides strong evidence that these loci are genuinely associated with the phenotype and have not been identified due to imputation artefacts.[37]

The most important limitation, however, is the lack of a replication cohort. The independent identification of loci at both 6q22.1 and 4q13.3 by Debiec *et al.* provides strong confirmatory evidence for our results. Yet, as detailed above, the lead SNP at 4q13.3 identified in that study is separated from our lead SNP by a recombination hotspot. It therefore remains to be determined whether this region is truly associated with SSNS

or not and, if so, if these loci are independent from each other. Moreover, Debiec *et al.* did not provide detailed information on the risk allele at 6q22.1 and we are thus unable to assess whether the allelic effect is identical in both studies. Further independent replication is thus needed to confirm the discovery of these loci.

Conclusions

Our study identifies two loci outside the HLA-region with genome-wide significant association with SSNS and thus provides important insight into the pathogenesis of SSNS. As *CALHM6* is implicated in regulating the immune response to infection, this may provide an explanation for the typical triggering of disease onset by infections. Further studies are needed to provide independent replication of our findings and to investigate the precise mechanisms and whether these could be amenable to specific treatments.

Author contributions

All authors together generated and gathered the patient samples and genetic data, and analyzed the data. Daniel P Gale, Horia C Stanescu, Robert Kleta (overall responsible), and Detlef Bockenhauer vouch for the data and the analysis. All authors helped writing the paper, and all agreed with publication.

Acknowledgements

DB, RK and RST were supported by the Mitchell Charitable Trust, Kids Kidney Research, Kidney Research UK, Garfield Weston Foundation and the Grocers' Charity. RK was supported by the David and Elaine Potter Charitable Foundation. HCS, DPG, RK and DB were supported by St Peter's Trust for Kidney, Bladder & Prostate Research. NT and JN were supported by the Dutch Kidney Foundation. JAK, RK and DB were supported by the Deanship of Scientific Research, King Abdulaziz University, Jeddah, grant number 432/003/d, for Pediatric Nephrology research. SD, DB and RK are grateful to Dr Magdi Yaqoob for support through the William Harvey Paediatric PhD Fellowship. DB is supported by the NIHR Biomedical Research Centre at GOSH/ICH. DPG acknowledges support of the UK Medical Research Council, Kidney Research UK, and the Rosetrees Trust. RG is supported by NIH/NIDDK 5R01DK098135 and 5R01DK094987 and a Doris Duke Clinical Scientist Development Award.

We are most grateful to the team of the PREDNOS and PREDNOS2 trials (trial partners are listed in the Supplements) and to the members of the Mid West Pediatric Nephrology Consortium for their study support. We are also most grateful to Gaganjit Madhan Kaur from UCL Genomics for providing expert genotyping support as well as Drs Fairfax,

Makino and Knight from the University of Oxford for sharing ethnically matched control data.

Disclosures

None.

Table of Contents

References

1. Banh TH, Hussain-Shamsy N, Patel V, Vasilevska-Ristovska J, Borges K, Sibbald C, et al.: Ethnic Differences in Incidence and Outcomes of Childhood Nephrotic Syndrome. *Clin J Am Soc Nephrol* 11**:** 1760-1768, 2016
2. Vivarelli M, Massella L, Ruggiero B, Emma F: Minimal Change Disease. *Clin J Am Soc Nephrol* 12**:** 332-345, 2017
3. Shalhoub RJ: Pathogenesis of lipoid nephrosis: a disorder of T-cell function. *Lancet* 2**:** 556-560, 1974
4. Hu X, Daly M: What have we learned from six years of GWAS in autoimmune diseases, and what is next? *Curr Opin Immunol* 24**:** 571-575, 2012
5. Sekula P, Li Y, Stanescu HC, Wuttke M, Ekici AB, Bockenhauer D, et al.: Genetic risk variants for membranous nephropathy: extension of and association with other chronic kidney disease aetiologies. *Nephrol Dial Transplant* 32: 325-332, 2017
6. Stanescu HC, Arcos-Burgos M, Medlar A, Bockenhauer D, Kottgen A, Dragomirescu L, et al.: Risk HLA-DQA1 and PLA(2)R1 alleles in idiopathic membranous nephropathy. *N Engl J Med* 364**:** 616-626, 2011
7. Wuttke M, Kottgen A: Insights into kidney diseases from genome-wide association studies. *Nat Rev Nephrol* 12**:** 549-562, 2016
8. Kiryluk K, Li YF, Scolari F, Sanna-Cherchi S, Choi M, Verbitsky M, et al.: Discovery of new risk loci for IgA nephropathy implicates genes involved in immunity against intestinal pathogens. *Nat Genet* 46**:** 1187-1196, 2014
9. Gale DP, Molyneux K, Wimbury D, Higgins P, Levine AP, Caplin B, et al.: Galactosylation of IgA1 Is Associated with Common Variation in C1GALT1. *J Am Soc Nephrol* 28**:** 2158-2166, 2017
10. Gbadegesin RA, Adeyemo A, Webb NJ, Greenbaum LA, Abeyagunawardena A, Thalgahagoda S, 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
11. Debie H, Dossier C, Letouze E, Gillies CE, Vivarelli M, Putler RK, et al.: Transethnic, Genome-Wide Analysis Reveals Immune-Related Risk Alleles and Phenotypic Correlates in Pediatric Steroid-Sensitive Nephrotic Syndrome. *J Am Soc Nephrol* 29**:** 2000-2013, 2018
12. Jia X, Horinouchi T, Hitomi Y, Shono A, Khor SS, Omae Y, et al.: Strong Association of the HLA-DR/DQ Locus with Childhood Steroid-Sensitive Nephrotic Syndrome in the Japanese Population. *J Am Soc Nephrol* 29**:** 2189-2199, 2018
13. Noone DG, Iijima K, Parekh R: Idiopathic nephrotic syndrome in children. *Lancet* 392**:** 61-74, 2018
14. Webb NJ, Frew E, Brettell EA, Milford DV, Bockenhauer D, Saleem MA, et al.: Short course daily prednisolone therapy during an upper respiratory tract infection in children with relapsing steroid-sensitive nephrotic syndrome (PREDNOS 2): protocol for a randomised controlled trial. *Trials* 15**:** 147, 2014
15. Fairfax BP, Makino S, Radhakrishnan J, Plant K, Leslie S, Dilthey A, et al.: Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat Genet* 44**:** 502-510, 2012

16. Fairfax BP, Humburg P, Makino S, Naranbhai V, Wong D, Lau E, et al.: Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* 343**:** 1246949, 2014
17. WTCCC: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447**:** 661-678, 2007
18. Browning BL, Zhou Y, Browning SR: A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am J Hum Genet* 103**:** 338-348, 2018
19. Chang CC, Chow CC, Tellie, LC, Vattikuti S, Purcell SM, Lee JJ: Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4**:** 7, 2015
20. Jia X, Han B, Onengut-Gumuscu S, Chen WM, Concannon PJ, Rich SS, et al.: Imputing amino acid polymorphisms in human leukocyte antigens. *PloS One* 8**:** e64683, 2013
21. GTEx Consortium: The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 45**:** 580-585, 2013
22. Võsa U, Claringbould A, Westra H-J, Bonder MJ, Deelen P, Zeng B, et al.: Unraveling the polygenic architecture of complex traits using blood eQTL meta-analysis. *bioRxiv***:** 447367, 2018
23. Qiu C, Huang S, Park J, Park Y, Ko YA, Seasock MJ, et al.: Renal compartment-specific genetic variation analyses identify new pathways in chronic kidney disease. *Nat Med* 24**:** 1721-1731, 2018
24. Gillies CE, Putler R, Menon R, Otto E, Yasutake K, Nair V, et al.: An eQTL Landscape of Kidney Tissue in Human Nephrotic Syndrome. *Am J Hum Genet* 103**:** 232-244, 2018
25. Lenz TL, Spirin V, Jordan DM, Sunyaev SR: Excess of Deleterious Mutations around HLA Genes Reveals Evolutionary Cost of Balancing Selection. *Mol Biol Evol* 33**:** 2555-2564, 2016
26. Trowsdale J: The MHC, disease and selection. *Immunol Lett* 137**:** 1-8, 2011
27. Malik U, Javed A, Ali A, Asghar K: Structural and functional annotation of human FAM26F: A multifaceted protein having a critical role in the immune system. *Gene* 597**:** 66-75, 2017
28. Malik U, Javed A: FAM26F: An Enigmatic Protein Having a Complex Role in the Immune System. *Int Rev Immunol* 19**:** 1-11, 2016
29. Chmielewski S, Olejnik A, Sikorski K, Pelisek J, Blaszczyk K, Aoqui C, et al.: STAT1-dependent signal integration between IFNgamma and TLR4 in vascular cells reflect pro-atherogenic responses in human atherosclerosis *PloS One,* 9**:** e113318, 2014
30. Ma Z, Tanis JE, Taruno A, Foskett JK: Calcium homeostasis modulator (CALHM) ion channels. *Pflugers Arch* 468**:** 395-403, 2016
31. Di Virgilio F, Pizzo P, Zanovello P, Bronte V, Collavo D: Extracellular ATP as a Possible Mediator of Cell-Mediated Cytotoxicity. *Immunol Today* 11**:** 274-277, 1990
32. Nagy PV, Feher T, Morga S, Matko J: Apoptosis of murine thymocytes induced by extracellular ATP is dose- and cytosolic pH-dependent. *Immunol Lett* 72**:** 23-30, 2000
33. Banuelos J, Lu NZ: A gradient of glucocorticoid sensitivity among helper T cell cytokines. *Cytokine Growth Factor Rev* 31**:** 27-35, 2016
34. Julia A, Domenech E, Chaparro M, Garcia-Sanchez V, Gomollon F, Panes J, et al.: A genome-wide association study identifies a novel locus at 6q22.1 associated with ulcerative colitis. *Hum Mol Genet* 23**:** 6927-6934, 2014
35. Battle A, Montgomery SB: Determining causality and consequence of expression quantitative trait loci. *Hum Genet* 133**:** 727-735, 2014

36. Bockenhauer D, Medlar AJ, Ashton E, Kleta R, Lench N: Genetic testing in renal disease. *Pediatr Nephrol* 27**:** 873-883, 2012
37. Marchini J, Howie B: Genotype imputation for genome-wide association studies. *Nat Rev Genet* 11**:** 499-511, 2010

Tables

Table 1: Effect estimates for lead SNPs

Locus	Gene	SNP	DR2	Minor allele	MAF cases	MAF controls	OR	95% CI	p-value
6p21.3	*HLA-DQB1*	rs9273542	0.89	T	0.51	0.24	3.39	2.86-4.03	1.59×10^{-43}
6q22.1	*CALHM6/FAM26F*	rs2637678	0.96	C	0.26	0.40	0.51	0.44-0.60	1.27×10^{-17}
4q13.3	*PARM1*	rs10518133	0.93	A	0.12	0.06	1.96	1.57-2.45	2.50×10^{-8}

Minor allele frequencies (MAF) and odds ratio (OR) with 95% confidence intervals (95% CI) for each of the minor alleles of the lead SNPs from the three loci achieving genome-wide significance. DR2 (Dosage R-Squared) indicates the Beagle imputation quality score.

Table 2: Association of imputed classical HLA alleles with SSNS

HLA allele	MAF cases	MAF controls	OR	95% CI	p-value
HLA_DQA1*02:01	0.35	0.15	3.42	2.80-4.16	1.06×10^{-32}
HLA_DQA1*01	0.13	0.38	0.36	0.30-0.43	1.90×10^{-31}
HLA_DRB1*07:01	0.35	0.15	3.26	2.68-3.97	5.62×10^{-31}
HLA_DQB1*02	0.40	0.21	2.43	2.04-2.91	9.77×10^{-22}
HLA_DQA1*01:03	0.02	0.09	0.24	0.15-0.38	1.79×10^{-14}
HLA_DRB1*13	0.04	0.11	0.31	0.22-0.44	2.41×10^{-14}
HLA_DRB1*13:01	0.02	0.08	0.23	0.15-0.37	3.18×10^{-14}
HLA_DQA1*01:01	0.08	0.15	0.46	0.35-0.59	1.53×10^{-10}
HLA_B*08:01	0.20	0.13	2.95	2.05-4.23	9.17×10^{-09}

Minor allele frequencies (MAF) for cases and controls, odds ratio (OR) with 95% confidence interval (95% CI) for each of the HLA alleles achieving genome-wide significance.

Figure legends

Figure 1. Manhattan plot

GWAS for SSNS comparing 422 European cases with 5642 ethnically matched controls. X-axis: chromosomal position. Y-axis: log-transformed p-value; horizontal red line indicates the genome-wide significance threshold ($P=5\times10^{-8}$). Three loci surpass this threshold on chromosomes 4 and 6.

Figure 2: Locus zoom plot of the HLA-DR/DQ (6p21.32) association

X-axis indicates chromosomal position, the left Y-axis the log-transformed p-value and the right Y-axis the recombination rate. The purple diamond indicates the SNP with the smallest p-value within each region. SNPs are colored based on their pairwise LD to the lead SNP as per 1000 Genomes European reference data, according to the key. Recombination hotspots are indicated by blue vertical lines. Data are shown for 200 kb either side of the lead SNP.

A) Unconditioned analysis. The lead SNP is rs9273542, in the gene *HLA-DQB1*, with a p-value of 1.59×10^{-43}.

B) Post-conditioning on rs9273542. The lead SNP is rs2858317, centromeric of *HLA-DQB1*, with a p-value of 4.29×10^{-31}.

C) Post-conditioning on rs9273542 and rs2858317. The lead SNP from this analysis is rs3828799, centromeric of *HLA-DQB1*, with a p-value of 2.40×10^{-8}.

Note, these results indicate that the association in the HLA-DRB1 region is driven mainly by two independent HLA alleles (indicated by lead SNPs in panels A and B).

Figure 3: Locus zoom plot of the *CALHM6* (6q22.1) association

The composition of this figure is as per Figure 2.

A) Unconditioned analysis. The lead SNP is rs2637678 with a p-value of 1.27×10^{-17}.

B) Post conditioning on rs2637678. No further SNPs reached genome wide significance level, indicating that the association with *CALHM6* is driven by a single haplotype.

Figure 4: Locus zoom plot of the *PARM1* (4q13.3) association

The composition of this figure is as per Figure 2.

A) Unconditioned analysis. The lead SNP is rs10518133, in the gene *PARM1*, with a p-value of $P=2.50\times10^{-8}$.

B) Post-conditioning on rs10518133. No further SNPs reached genome wide significance level, indicating that the association with *PARM1* is driven by a single haplotype.