Comprehensive Review of Steroid-Sensitive Nephrotic Syndrome Genetic Risk Loci and Transcriptional Regulation as a Possible Mechanistic Link to Disease Risk

Check for updates

Hillarey K. Stone[1,2], Sreeja Parameswaran[2,3], Amy A. Eapen[2,4], Xiaoting Chen[2,3], John B. Harley[2,3,5], Prasad Devarajan[1,2], Matthew T. Weirauch[2,3,6,7] and Leah Kottyan[2,3,4]

[1]Division of Nephrology and Hypertension, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA; [2]Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, Ohio, USA; [3]Center for Autoimmune Genomics and Etiology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA; [4]Division of Allergy and Immunology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA; [5]US Department of Veterans Affairs Medical Center, Cincinnati, Ohio, USA; [6]Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA; and [7]Division of Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA

Introduction: The etiology of steroid-sensitive nephrotic syndrome (SSNS) is not well understood. Genetic studies have established common single nucleotide polymorphisms (SNPs) that are associated with increased SSNS disease risk. We review previous genetic association studies of SSNS and nominate particular transcriptional regulators and immune cells as potential key players in the etiology of this disease.

Methods: A list of SNPs associated with SSNS was compiled from published genome wide association and candidate gene studies. The Regulatory Element Locus Intersection (RELI) tool was used to calculate the enrichment of the overlap between disease risk SNPs and the genomic coordinates of data from a collection of >10,000 chromatin immunoprecipitation sequencing experiments.

Results: After linkage disequilibrium expansion of the previously reported tag associated SNPs, we identified 192 genetic variants at 8 independent risk loci. Using the Regulatory Element Locus Intersection algorithm, we identified transcriptional regulators with enriched binding at SSNS risk loci ($10^{-05} < P_{corrected} < 10^{-124}$), including ZNF530, CIITA, CD74, RFX5, and ZNF425. Many of these regulators have well-described roles in the immune response. RNA polymerase II binding in B cells also demonstrated enriched binding at SSNS risk loci ($10^{-37} < P_{corrected} < 10^{-5}$).

Conclusion: SSNS is a complex disease, and immune dysregulation has been previously implicated as a potential underlying cause. This assessment of established SSNS risk loci and analysis of possible function implicates transcriptional dysregulation, and specifically particular transcriptional regulators with known roles in the immune response, as important in the genetic etiology of SSNS.

I diopathic nephrotic syndrome is the most common glomerular disease seen in the pediatric population, with an average incidence of 4.7 per 100,000 children.[1] Most children with nephrotic syndrome will respond to a course of corticosteroids and are therefore diagnosed with steroid-sensitive nephrotic syndrome (SSNS). Patients with SSNS are more likely to have minimal change histology and tend to have an excellent long-term prognosis with low risk for progression to end-stage renal disease (ESRD). In contrast, approximately 10% of children will fail to achieve remission with an 8-week course of corticosteroids and therefore are diagnosed with steroid-resistant nephrotic syndrome (SRNS). These children are more likely to have focal segmental glomerulosclerosis and are at increased risk for developing chronic kidney disease and ESRD.[2]

Advances in genetic research in the past 2 decades have led to the identification of more than 50 genes associated with SRNS (summarized in Stone *et al.*[3]). It is now possible to identify a subset of children with

nephrotic syndrome whose disease is caused by structural defects in the glomerular filtration barrier.[4,5] Patients with SSNS likely have more variable phenotypes and a more complex inheritance pattern, making the genetic etiology of this disease more difficult to identify. Current evidence implicates a primary dysregulation of the immune system leading to secondary podocyte damage, although it is unclear at this time what specific genetic and/or environmental factors lead to this dysregulation. A better understanding of the pathogenesis of SSNS could lead to improved diagnostic and prognostic techniques as well as more targeted therapies.

Several factors strongly suggest a genetic component in the etiology of SSNS. First, although the majority of patients with SSNS have no reported family history, a number of cases of siblings with SSNS have been described.[6,7] It is unclear exactly how common familial SSNS is, as large epidemiologic studies examining the frequency of SSNS in first-degree relatives are lacking. A sibling concordance study published in the 1970s suggested that approximately 3% of children with nephrotic syndrome have an affected sibling.[8] Although this number likely included cases of SRNS and therefore is likely an overestimation, it still argues for a possible role for genetics in SSNS. Unfortunately, updated sibling concordance studies and initial twin concordance studies in patients with SSNS have not been published.

Second, several reported cases of patients with monogenic causes of nephrotic syndrome who respond fully or partially to corticosteroids have been described.[9–16] Many of these genes are typically associated with SRNS but have been also reported in cases of steroid-responsive disease. Others seem to cause a predominantly steroid-sensitive form of nephrotic syndrome. Although a handful of cases have been described, the vast majority of patients with SSNS have no identifiable Mendelian cause.

An additional argument for the role of genetics in SSNS is the variation of incidence in patients of different ethnic backgrounds. For example, Banh et al.[17] performed a longitudinal study in a large cohort of ethnically diverse children with nephrotic syndrome. Over a period of 10 years, they found higher incidence of nephrotic syndrome in South Asians compared to Europeans and East/Southeast Asians, with an incidence of 15.83/100,000 versus 2.40/100,000 in South Asians and Europeans, respectively. Similar variations in incidence among ethnic groups have been described in a number of other studies and are summarized in Chanchlani and Parekh.[1] Although environmental factors could also contribute to this variation, it is likely that genetics, environment, and

gene by environment interactions play an etiological role.

Finally, many studies, including several genome-wide association studies (GWAS), have described an important role for the human leukocyte antigen (HLA) gene complex in SSNS. Gbadegesin et al.[18] genotyped cohorts of children of Asian and European ancestry. They identified variants that were statistically associated with SSNS across the HLA-DQ locus (HLA-DQA and HLA-DQB1). In a multi-ancestral replication study, they confirmed the association at HLA-DQA.[18] Debiec et al.[19] performed a GWAS study in patients of European, African, and Maghrebian ancestry. They identified genetic association at HLA-DQB1 as well as 2 independent genetic associations across the HLA locus on chromosome 6.[19] Focusing on children of African American ancestry and a replication cohort of children from South Asia, Adeyemo et al.[20] sequenced 2 HLA-DQA1 variants and identified significant association with SSNS in these populations. In a study of children of Japanese ancestry, Jia et al.[21] performed a GWAS followed by HLA type association analysis with replication of candidate genetic risk variants in a candidate study of prioritized genetic variants. They also found significant association at the HLA-DR/DQ region.[21] Most recently, Dufek et al.[22] assembled a large cohort of patients with SSNS and a matched set of healthy controls. They identified 3 genome-wide significant loci, including 2 loci outside of the HLA region — 4q13.3 and 6q22.1. Altogether, these groups have independently replicated and confirmed the importance of the HLA locus as well as several other disease risk loci in patients with SSNS.

The goal of the present study is to better understand the functional relevance of previously described SSNS genetic risk variants. We first curated the independent association signals identified by previously published genetic association studies. Then, we performed a linkage disequilibrium expansion step to identify all variants that were tagged by the associated variants reported with significant (or suggestive) P values. We used the analytical tool Regulatory Element Locus Intersection (RELI) to identify critical cell types and transcriptional regulatory molecules based on enrichment of chromatin immunoprecipitation sequencing (ChIP-seq) data at disease risk loci. This algorithm assesses the significance of intersections of genomic coordinates of disease risk loci and DNA bound by particular transcriptional regulators. The data generated by RELI can now be used to target particular transcription factors and other regulatory proteins which may be involved in disease pathogenesis and therefore are good candidates for further experimental investigation using human samples. This is the first

step in generating hypotheses to understand the functional relevance of the genetic risk loci identified by GWAS studies. Collectively, these data reveal potential gene regulatory mechanisms underlying the complex etiology of SSNS, providing avenues for future studies examining possible therapeutic strategies.

METHODS

Identification of Independent Genetic Risk Loci

A literature search was performed using PubMed and the National Human Genome Research Institute–European Bioinformatics Institute GWAS catalog[23] to identify genome-wide association and candidate gene studies reporting significant genetic association for SSNS. Eight studies met our search criteria. SNPs with $P < 10^{-4}$ were included in our analysis. Starting with the "tag" genetic variant of each reported SSNS risk variant ($P < 10^{-4}$), independent risk loci were identified using the variant pruning function of PLINKv1.90b.[24] We used an R^2 threshold of 0.2 to define independent loci. R^2 is a measure of linkage disequilibrium. Only variants on the same chromosome were assessed because variants on different chromosomes segregate completely independently in meiosis.

Linkage Disequilibrium Expansion

Each reported independent "tag" genetic variant was used to identify all genetic variants in linkage disequilibrium (LD) with the tag variant with $R^2 > 0.8$ in the ancestry of reported association. If a variant was identified in multiple ancestries, a union of the LD-expanded variants was included.

Pathway Analysis of Genes Near SSNS Risk Loci

Genes near SSNS risk loci were used to identify the biological pathways enriched in this gene set using Enrichr.[25,26]

Collection and Processing of ChIP-seq Datasets

A collection of 11,834 publicly available ChIP-seq datasets as of October 2019 were downloaded from the National Center for Biotechnology Information (NCBI) Sequence Read Archive and were analyzed using an in-house automated pipeline.[27] The broad classification of the origins of the cells and cell lines assessed in these ChIP-seq experiments are displayed in Supplementary Figure S1. Briefly, the pipeline first downloads the sequence read archive files, converts them to fastq files, and runs quality control (QC) on the files using FastQC (v0.11.2).[28] If FastQC detects adapter sequences, the pipeline runs the fastq files through Trim Galore (v0.4.2),[29] a wrapper script that runs cutadapt (v1.9.1)[30] to remove the detected adapter sequence from the reads. The quality-controlled reads are then aligned to the reference human genome (hg19/GRCh37) using bowtie2 (v2.3.4.1).[31] The aligned reads (in a .bam format) are then sorted using samtools (v1.8.0)[32] and duplicate reads are removed using picard (v1.89).[33] Finally, peaks are called using MACS2 (v2.1.0),[34] using 4 different parameter settings (MODE1: callpeak -g hs -q 0.01 -t bamfile, MODE2: callpeak -g hs -q 0.01 −broad -t bamfile, MODE3: callpeak -g hs -q 0.01 −broad −nomodel −extsize 500 -t bamfile, MODE4: callpeak -g hs -q 0.01 −broad −nomodel −extsize 1000 -t bamfile). ENCODE blacklist regions[35] were removed from the called peaks using the hg19-blacklist.v2.bed.gz file available at github.[36] For each ChIP-seq dataset, only the top-performing peak set of the 4 MACS2 modes is reported, where performance is gauged based on the RELI algorithm-corrected P value.

Identification of Particular Transcriptional Regulators Enriched at Risk Loci using RELI

We used our RELI algorithm to identify ChIP-seq datasets whose genomic coordinates significantly intersect SSNS genetic risk variants. As previously published,[37] all of the genetic polymorphisms in LD with the tag variant at each of the 8 independent SSNS risk loci were intersected with transcriptional regulator ChIP-seq peaks. RELI uses a permutation-based strategy to identify the significance of the overlap between each LD block and a given ChIP-seq dataset, which we designate the observed intersection count. As described previously,[37] this value is compared to the expected intersection distribution. In this procedure, the most strongly associated variant in the LD block is chosen as the reference variant. A distance vector is then generated providing the distance (in bases) of each variant in the LD block from this reference variant. A random genomic variant with approximately matched allele frequencies to the reference variant is then selected from NCBI's dbSNP database, and genomic coordinates of artificial variants are created that are located at the same relative distances from this random variant using the distance vector. Members of this artificial LD block are then intersected with each ChIP-seq dataset, as was done for the observed intersections. This strategy accounts for the number of variants in the input LD block and their relative distances, while prohibiting 'double counting' due to multiple variants in the block intersecting the same dataset. We repeat this procedure 2,000 times, generating a null distribution with stable P-values. The expected intersection distributions are used to calculate Z-scores and P-values for the observed intersection.

Table 1. Genetic risk variants associated with SSNS in GWAS at 8 independent risk loci

Locus	Chromosome	SNP	P value	Odds Ratio	Population	Study reference
4q13.3	4	rs10518133	2.50×10^{-8}	1.96	European	22
6p21.32	6	rs9348883	4.70×10^{-4}	2.51	Trans-ethnic	19
	6	rs28366266	4.10×10^{-20}	3.03	Trans-ethnic	19
6p21.32	6	rs1129740	1.187×10^{-6}	2.11	South Asian	18
	6	rs1071630	1.187×10^{-6}	2.11	South Asian	18
	6	rs1063348	9.3×10^{-23}	3.33	Trans-ethnic	19
6p21.32	6	rs1129740	5.7×10^{-11}	3.53	African American	20
	6	rs1071630	1.2×10^{-13}	4.08	African American	20
6p21.32	6	rs9273371	1.64×10^{-43}	3.29	European	22
6p21.32	6	rs9273542	1.59×10^{-43}	3.39	European	22
	6	rs9273529	2.87×10^{-43}	3.39	European	22
6p21.32	6	rs1140343	1.187×10^{-6}	2.11	South Asian	18
	6	rs3134996	1.72×10^{-25}	0.29	Japanese	21
	6	rs4642516	7.84×10^{-23}	0.33	Japanese	21
6q22.1	6	rs2858829	1.72×10^{-16}	0.53	European	22
	6	rs2637681	3.53×10^{-17}	0.52	European	22
	6	rs2637678	1.27×10^{-17}	0.51	European	22

GWAS, genome-wide association studies; SNP, single nucleotide polymorphism; SSNS, steroid-sensitive nephrotic syndrome.

RESULTS

Literature Review and Curation of Previous Studies

There have been limited studies assessing the genetic basis of SSNS. To identify all possible candidates, we used Pubmed and the GWAS catalog[23] to identify SSNS risk variants from genome wide association and candidate gene studies that reported significant genetic association for SSNS. A total of 8 studies met our search criteria.[18–22,38–40] Five of the candidate genetic variant[18,20,21] and GWAS studies[19,21,22] which sought to identify the origins of complex genetic risk for SSNS successfully identified loci with P values < 0.0001 (Table 1). An expanded table of all reported SSNS risk variants without filtering for statistical significance is provided in Supplementary Table S1.

Inclusion of All Possible Statistically Associated SSNS Variants

Genetic association studies report tag genetic variants that were the most significant association at a specific locus. Genetic variants are inherited as a haplotype of variants in linkage disequilibrium, so independent loci are variants that are not in strong linkage disequilibrium. Using a LD cutoff of $r^2 = 0.2$ in the ancestry of identification, we identified all independent genetic loci (Supplementary Table S2). We identified 8 independent genetic effects: 6 in the HLA region on chromosome 6, and 2 outside the HLA region on chromosomes 4 and 6. Next, we identified all common genetic variants in LD $r^2 > 0.8$ with the identified risk variants in the ancestry of discovery; this resulted in a total of 192 variants (Supplementary Table S3).

Although there are likely only one or 2 causal variants per SSNS risk locus, all of the variants on the LD-expanded list are candidates based on genetic association. Using the NCBI dbSNP database,[22,41] annotations of the position of the LD-expanded SSNS risk variants relative to gene coding regions were assessed (Supplementary Table S4). Only 10 variants from this LD-expanded list are located in coding regions. Of these, 7 are missense variants and 3 are synonymous variants. PolyPhen2[42] and SIFT[43] were used to predict the functional effects of the 7 missense variants. These analyses showed that only 2 are predicted to cause changes to the resulting protein function by PolyPhen2 (Supplementary Table S5), whereas all are considered to be benign by SIFT (Supplementary Table S6). The remaining variants, which represent the majority of identified single nucleotide polymorphisms (SNPs), are located in non-coding regions. Overall, these results are consistent with a possible role for alteration of gene regulatory mechanisms.

Identification of Genes Near Risk Loci

We identified a total of 16 genes within 50 kb of the comprehensive list of SSNS risk variants (Supplementary Tables S7 and S8). Included in this list of genes are several HLA genes, *BTNL2* (an immunoregulator involved in T-cell regulation, proliferation, and cytokine release), and *HCG23* (HLA complex group 23). These genes were significantly enriched for biological pathways defined by Gene Ontology enrichment terms associated with antigen presentation, T-cell receptor signaling, and interferon-gamma–associated signaling (Supplementary Table S9).

Table 2. Intersection between SSNS risk loci and top overlapping transcription factors

Cell line	Cell type	Molecule	Overlap	Enrichment	Corrected P value
HEK293	Kidney derived immortalized cell line	ZNF530	4	142.9	6.28×10^{-101}
Raji	Immortalized B cell line	CIITA	4	86.1	8.35×10^{-64}
CLL	Immortalized B cell line	CD74	4	42.1	2.40×10^{-31}
WA01	Pluripotent stem cell line	RFX5	4	38.7	4.10×10^{-28}
HEK 293T	Kidney derived immortalized cell line	ZNF425	4	28.7	8.99×10^{-20}

For this analysis, all of the genetic polymorphisms in linkage disequilibrium with the "tag" variant at each of the 8 independent SSNS risk loci were intersected with transcription factor ChIP-seq peaks from publicly deposited experimental data. For each ChIP-seq dataset, the cell line, cell type, and molecule are provided. The number of times that the ChIP dataset had a called peak that overlapped a SNP at an SSNS risk locus is indicated ("overlap"). Only one overlap was counted for each independent risk locus — even when multiple SNPs risk variants at the same risk locus overlapped a ChIP-seq dataset (i.e., the maximum overlap is 8). Only the top 5 unique transcriptional regulator datasets with 4 or more overlaps with independent SSNS risk loci are shown. A permutation strategy is used by RELI to identify the significance of the overlap. The P value is identified based on the permutations of RELI and is calculated from the Z-score. The corrected P value gives the P value after accounting for multiple testing of the many ChIP-seq datasets. ChIP-seq, chromatin immunoprecipitation sequencing; RELI, Regulatory Element Locus Intersection; SNP, single nucleotide polymorphism; SSNS, steroid-sensitive nephrotic syndrome.

Identification of Key Transcriptional Regulators at SSNS Risk Loci

To understand the biological mechanisms driving SSNS disease risk at the identified risk loci, we used our RELI tool to find ChIP-seq datasets that significantly overlap with the SSNS risk loci more than expected by chance. RELI measures the overlap of each ChIP-seq dataset with the SSNS risk loci and assesses statistical significance using a null model in which the SSNS risk loci are randomly repositioned across the genome while keeping the number of SNPs and size of the LD-expanded haplotype blocks consistent, as previously described.[37]

We identified a total of 101 ChIP-seq datasets involving 54 unique transcriptional regulators and histone marks that overlap at least 3 of 8 of the SSNS risk loci, with enrichment from 9 to 240 times that expected by random chance and corrected P values between 1.0×10^{-5} and 3.8×10^{-124}. Most of these transcriptional regulators have well-described roles in the regulation of the human immune response. Among the most significantly enriched are zinc finger protein 530 (ZNF530), class II major histocompatibility complex (CIITA), regulatory factor X5 (RFX5), CD74, and zinc finger protein 425 (ZNF425) (Table 2). Many of the ChIP-seq datasets with significant enrichment at SSNS risk loci were performed in B cell lines (56 of 101) and kidney-derived cell lines (17 of 101). Most of these datasets assessed used cell lines; however, 11.74% of datasets were from primary tissues and hematologic cells. Kidney and B cells represented 7.22% and 10% of the assessed datasets, respectively. The full results from our RELI analysis with corrected P values less than 1×10^{-6} are provided in Supplementary Table S10, and results from kidney-derived cells are provided in Supplementary Table S11.

RNA Polymerase II Enrichment at SSNS Risk Loci

RNA polymerase II (POLR2) is the regulatory molecule that is ultimately responsible for gene expression. Our RELI analysis identified strong enrichment, with 11 to 51 times more binding of POLR2 at SSNS risk loci than expected by chance ($10^{-37} < P_{corrected} < 10^{-5}$) in B cells (Supplementary Table S12).

DISCUSSION

Although the pathogenesis of SSNS remains unclear, current evidence suggests that immune dysregulation and genetic factors play important roles in the development of this complex disease. Recent genetic association studies have identified a number of SSNS genetic risk variants in the HLA region, further supporting these mechanisms.

GWAS studies are valuable tools for understanding the genetic basis of complex traits; however, there are several limitations to consider. First, although GWAS can be used to identify a region of the genome associated with disease risk, due to linkage disequilibrium, these loci almost always represent a number of potentially causal variants. This makes it difficult to identify specific variants with functional relevance. Furthermore, the majority of risk variants identified by GWAS are located in non-coding regions. Thus, it is often unclear how these variants directly contribute to disease pathogenesis. Further functional testing is therefore required to fully understand the contributions of risk loci in the etiology of a disease.[44] We hypothesized that SSNS disease risk variants identified by genetic association studies affect the binding of particular transcriptional regulators. We used our recently developed algorithm (RELI) to better understand the functional relevance of these risk alleles, specifically regarding their interactions with transcription factors and other regulatory molecules. Our overall goal was to identify particular transcriptional regulators which could be targeted for subsequent functional studies using specimens from patients with nephrotic syndrome.

This study has a number of limitations. First, interpreting GWAS results is challenging because the resulting genetic loci are largely located in non-coding regions; therefore, it is difficult to predict and validate their mechanistic roles. Furthermore, genetic association studies do not identify single SNPs, but rather nominate a set of variants that are in LD with one or more causal variant(s). Additionally, SSNS studies to

date have been underpowered to identify association with genome-wide significance. Genetic variants with $P < 5 \times 10^{-8}$ are typically considered genome-wide significant because the association survives multiple testing of more than 1 million independent pieces of the genome.[45] For this study, we used a less conservative cutoff given that the currently available GWAS and candidate gene studies have been small and have not always reached statistical significance at the thresholds typically used in genetic association studies. Studies using larger populations are therefore needed to identify more robust associations. In addition, the fact that most of the currently identified SSNS genetic risk loci are in the complex HLA region increases the challenging task of understanding the genetic etiology of the disease.

As hypothesized, the results from our analysis show strong enrichment for the binding of specific transcriptional regulatory proteins at established SSNS risk loci, suggesting that these proteins are involved in the pathogenesis of SSNS via altered binding to disease risk loci. Furthermore, many of the identified transcriptional regulators have important established roles in the immune response, emphasizing the role of immune dysregulation in the pathogenesis of this disease.

A growing body of evidence supports a role for B and T lymphocytes in the pathogenesis of nephrotic syndrome. The most compelling argument is the role of T-cell depleting agents in the treatment of nephrotic syndrome. Most of the immunosuppressive agents used to treat nephrotic syndrome target T cells, including corticosteroids, cyclophosphamide, calcineurin inhibitors, and antiproliferative agents[46,47] More recent evidence has supported the use of anti-CD20 antibodies, rituximab and ofatumumab, in the treatment of nephrotic syndrome. These agents have been shown to induce a prolonged remission in many patients, further emphasizing the role of B cells in the pathogenesis of this disease.[48] Other arguments for the involvement of immune cells in the pathogenesis of SSNS include reports of full remission of nephrotic syndrome induced by measles infection[49,50] and the association of nephrotic syndrome with T-cell lymphomas.[51] Finally, as described above, several genetic variants in the HLA region have been shown to associate with SSNS in multiple studies,[18,19,21,22] although the mechanisms underlying these associations have not yet been elucidated.

Our results agree with the current literature which supports the importance of immune dysregulation in the pathogenesis of SSNS, as several of the most highly enriched transcriptional regulators in our analysis have well-known roles in the immune response. For example, CIITA regulates transcription of MHC class I

and II genes.[52] It has been shown to interact with DNA binding proteins, other transcription factors, and histone modifying proteins, which together work to regulate MHC gene expression.[53] Defects in the *CIITA* gene result in abnormal expression of MHC II and a severe combined immunodeficiency in humans.[54,55] CIITA works in a complex with another essential regulator of MHC class II — RFX5,[56] which we also have identified as a key player at SSNS risk loci. RFX5 is a nuclear protein that binds to the X box of MHC-II promoters and plays a role in antigen presentation.[57] Similarly, CD74 is a chaperone that regulates antigen presentation for proper immune response.[58] Collectively, these data suggest that these and other particular regulatory proteins play important roles at SSNS genetic risk loci.

We hypothesize that the presence of a single risk locus or combination of risk loci may lead to altered binding of transcriptional regulators. For example, abnormal binding of CIITA could lead to changes in the expression of MHC I and/or MHC II genes. When combined with an antigenic/environmental stimulus, this altered expression could lead to SSNS risk genotype-dependent antigen presentation, and subsequent T-cell–mediated destruction of podocytes may ensue. This hypothesis must be tested in experimental studies assessing genotype-dependent binding of transcription factors in biological samples (e.g., B cells and kidney biopsy specimens) from patients with nephrotic syndrome.

Progress has already been made in understanding the resultant changes in gene expression associated with identified risk loci. Two groups have previously described the presence of expression quantitative trait loci (eQTLs) associated with SSNS risk variants, which further supports a model of allelic transcriptional regulation.[19,22] Specifically, 3 of the SNPs identified in a GWAS by Dufek *et al.*[22] alter expression of *CALHM6*. Debiec *et al.*[19] identified 2 SNPs which altered expression of 3 genes using transcriptomic data from the NEPTUNE cohort. Using Genotype-Tissue Expression (GTEx), they described 2 additional eQTLs which altered expression of several HLA genes, as well as one SNP which acted as an eQTL at a number of other genes in various tissues.[19]

CONCLUSION

Our results have identified several transcriptional regulators which may play a role in the etiology of steroid SSNS. This focused list of regulatory proteins can now be used in ChIP-seq experiments using cells from patients with nephrotic syndrome. This important first step supports the role of immune-associated

transcriptional regulators in the pathogenesis of this complex disease and will help guide further studies evaluating the functional relevance of genetic risk loci in SSNS.

AUTHOR CONTRIBUTIONS

Study conception and design: HKS, LCK, MTW, JBH; literature review: HKS, LCK; bioinformatic design, development, and analysis: SP, XC, LCK, MTW; data analysis/interpretation: SP, HKS, LCK, MTW; drafting of manuscript: HKS, LCK; supervision or mentorship: PD. Each author contributed important intellectual content during manuscript drafting or revision, accepts personal accountability for the author's own contributions, and agrees to ensure that questions pertaining to the accuracy or integrity of any portion of the work are appropriately investigated and resolved.

SUPPLEMENTARY MATERIAL

Supplementary File (PDF)

Figure S1. Classification of datatypes used in this study.

Table S1. Reported steroid-sensitive nephrotic syndrome (SSNS) risk loci from genetic association studies.

Table S2. Identification of independent steroid-sensitive nephrotic syndrome (SSNS) risk loci.

Table S3. Linkage disequilibrium expansion of steroid-sensitive nephrotic syndrome (SSNS) risk variants for each independent locus.

Table S4. Annotation of genetic variant function for linkage-disequilibrium expanded set of steroid-sensitive nephrotic syndrome (SSNS) risk variants.

Table S5. PolyPhen2 prediction of pathogenicity for steroid-sensitive nephrotic syndrome (SSNS) risk variants in coding regions.

Table S6. SIFT prediction of pathogenicity for steroid-sensitive nephrotic syndrome (SSNS) risk variants in coding regions.

Table S7. Genes within 50 kilobases of steroid-sensitive nephrotic syndrome (SSNS) risk loci.

Table S8. List of 16 genes within 50 kilobases of steroid-sensitive nephrotic syndrome (SSNS) risk loci. The boundary of loci was defined by the first and final linkage-disequilibrium expanded risk variant within the independent locus.

Table S9. GO Pathways enriched for genes near steroid-sensitive nephrotic syndrome (SSNS) risk loci.

Table S10. Intersection between steroid-sensitive nephrotic syndrome (SSNS) risk loci and transcription factor binding within the genome.

Table S11. Intersection between steroid-sensitive nephrotic syndrome (SSNS) risk loci and transcription factor binding in human embryonic kidney (HEK) cells.

Table S12. Intersection between SSNS risk loci and RNA polymerase II.

REFERENCES

1. Chanchlani R, Parekh RS. Ethnic differences in childhood nephrotic syndrome. *Front Pediatr*. 2016;4:39.

2. Koskimies O, Vilska J, Rapola J, Hallman N. Long-term outcome of primary nephrotic syndrome. *Arch Dis Child*. 1982;57(7):544–548.

3. Stone H, Magella B, Bennett MR. The search for biomarkers to aid in diagnosis, differentiation, and prognosis of childhood idiopathic nephrotic syndrome. *Front Pediatr*. 2019;7:404.

4. Bierzynska A, McCarthy HJ, Soderquest K, et al. Genomic and clinical profiling of a national nephrotic syndrome cohort advocates a precision medicine approach to disease management. *Kidney Int*. 2017;91(4):937–947.

5. Sadowski CE, Lovric S, Ashraf S, et al. A single-gene cause in 29.5% of cases of steroid-resistant nephrotic syndrome. *J Am Soc Nephrol*. 2015;26(6):1279–1289.

6. Chehade H, Cachat F, Girardin E, et al. Two new families with hereditary minimal change disease. *BMC Nephrol*. 2013;14:65.

7. Dorval G, Gribouval O, Martinez-Barquero V, et al. Clinical and genetic heterogeneity in familial steroid-sensitive nephrotic syndrome. *Pediatr Nephrol*. 2018;33(3):473–483.

8. White RH, The familial nephrotic syndrome. I. A European survey. *Clin Nephrol*. 1973;1(4):215–219.

9. Hinkes B, Wiggins RC, Gbadegesin R, et al. Positional cloning uncovers mutations in PLCE1 responsible for a nephrotic syndrome variant that may be reversible. *Nat Genetics*. 2006;38(12):1397–1405.

10. Lahdenkari AT, Kestila M, Holmberg C, Koskimies O, Jalanko H. Nephrin gene (NPHS1) in patients with minimal change nephrotic syndrome (MCNS). *Kidney Int*. 2004;65(5):1856–1863.

11. Gee HY, Zhang F, Ashraf S, et al. KANK deficiency leads to podocyte dysfunction and nephrotic syndrome. *J Clin Invest*. 2015;125:2375–2384.

12. Park E, Chang HJ, Shin JI, et al. Familial IPEX syndrome: different glomerulopathy in two siblings. *Pediatr Int*. 2015;57(2):e59–e61.

13. Rubio-Cabezas O, Minton JA, Caswell R, et al. Clinical heterogeneity in patients with FOXP3 mutations presenting with permanent neonatal diabetes. *Diabetes Care*. 2009;32:111–116.

14. Hashimura Y, Nozu K, Kanegane H, et al. Minimal change nephrotic syndrome associated with immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome. *Pediatr Nephrol*. 2009;24(6):1181–1186.

15. Roberts IS, Gleadle JM. Familial nephropathy and multiple exostoses with exostosin-1 (EXT1) gene mutation. *J Am Soc Nephrol*. 2008;19:450–453.

16. Gee HY, Ashraf S, Wan X, et al. Mutations in EMP2 cause childhood-onset nephrotic syndrome. *Am J Hum Genet*. 2014;94:884–890.

17. Banh TH, Hussain-Shamsy N, Patel V, et al. Ethnic differences in incidence and outcomes of childhood nephrotic syndrome. *Clin J Am Soc Nephrol*. 2016;11:1760–1768.

18. 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*. 2015;26:1701–1710.

19. Debiec H, Dossier C, Letouze E, et al. Transethnic, genome-wide analysis reveals immune-related risk alleles and phenotypic correlates in pediatric steroid-sensitive nephrotic syndrome. *J Am Soc Nephrol*. 2018;29:2000–2013.

20. Adeyemo A, Esezobor C, Solarin A, et al. HLA-DQA1 and APOL1 as risk loci for childhood-onset steroid-sensitive and steroid-resistant nephrotic syndrome. *Am J Kidney Dis*. 2018;71:399–406.

21. Jia X, Horinouchi T, Hitomi Y, et al. Strong aAssociation of the HLA-DR/DQ locus with childhood steroid-sensitive nephrotic syndrome in the Japanese population. *J Am Soc Nephrol*. 2018;29:2189–2199.

22. Dufek S, Cheshire C, Levine AP, et al. Genetic identification of two novel loci associated with steroid-sensitive nephrotic syndrome. *J Am Soc Nephrol*. 2019;30:1375–1384.

23. Buniello AMJ, Cerezo M, Harris LW, et al. The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics. 2019 Nucleic Acids Research 2019 [cited 2019 11/1/2019]. Available at: https://www.ebi.ac.uk/gwas/home. Accessed May 21, 2020.

24. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7.

25. Chen EY, Tan CM, Kou Y, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 2013;14:128.

26. Kuleshov MV, Jones MR, Rouillard AD, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. 2016;44(W1):W90–W97.

27. Github Mario Pujato NextGen Aligner. Available at: https://github.com/MarioPujato/NextGenAligner. Accessed May 21, 2020.

28. Babraham Institute. FastQC, 2007. Available at: www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed December 3, 2020.

29. Babraham Institutue. Trim Galore, 2007. Available at: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/. Accessed December 3, 2020.

30. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*. 2011;17:10–12.

31. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–359.

32. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754–1760.

33. Wang K, Duan CY, Wu J, et al. Predictive value of neutrophil gelatinase-associated lipocalin for contrast-induced acute kidney injury after cardiac catheterization: a meta-analysis. *Can J Cardiol*. 2016;32:1033.e19–1033.e29.

34. Broad Institute. Picard Toolkit, 2020. Available at: https://broadinstitute.github.io/picard/. Accessed December 3, 2020.

35. Model-based Analysis of ChIP-Seq, 2020. Available at: https://broadinstitute.github.io/picard/. Accessed December 3, 2020.

36. Github. Boyle-Lab/Blacklist. Available at: https://github.com/Boyle-Lab/Blacklist/tree/master/lists. Accessed May 21, 2020.

37. Harley JB, Chen X, Pujato M, et al. Transcription factors operate across disease loci, with EBNA2 implicated in autoimmunity. *Nat Genetics*. 2018;50:699–707.

38. Suvanto M, Jahnukainen T, Kestila M, Jalanko H. Single nucleotide polymorphisms in pediatric idiopathic nephrotic syndrome. *Int J Nephrol*. 2016;2016:1417456.

39. Ohl K, Eberhardt C, Spink C, et al. CTLA4 polymorphisms in minimal change nephrotic syndrome in children: a case-control study. *Am J Kidney Dis*. 2014;63:1074–1075.

40. Rossi GM, Bonatti F, Adorni A, et al. FCGR2A single nucleotide polymorphism confers susceptibility to childhood-onset idiopathic nephrotic syndrome. *Immunol Lett*. 2018;193:11–13.

41. Sherry ST, Ward M, Sirotkin K. dbSNP — Database for Single Nucleotide Polymorphisms and Other Classes of Minor Genetic Variation Genome Research 1999 [cited 2019 10/18/2019]. Available at: https://www.ncbi.nlm.nih.gov/snp/. Accessed May 21, 2020.

42. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7:248–249.

43. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res*. 2012;40(web server issue):W452–W457.

44. Tam V, Patel N, Turcotte M, Bosse Y, Pare G, Meyre D. Benefits and limitations of genome-wide association studies. *Nat Rev Genet*. 2019;20:467–484.

45. A haplotype map of the human genome. *Nature*. 2005;437:1299–1320.

46. Colucci M, Corpetti G, Emma F, Vivarelli M. Immunology of idiopathic nephrotic syndrome. *Pediatric Nephrol*. 2018;33:573–584.

47. Elie V, Fakhoury M, Deschenes G, Jacqz-Aigrain E. Physiopathology of idiopathic nephrotic syndrome: lessons from glucocorticoids and epigenetic perspectives. *Pediatr Nephrol*. 2012;27:1249–1256.

48. Ravani P, Bonanni A, Rossi R, Caridi G, Ghiggeri GM. Anti-CD20 antibodies for idiopathic nephrotic syndrome in children. *Clin J Am Soc Nephrol*. 2016;11:710–720.

49. Blumberg RW, Cassady HA. Effect of measles on the nephrotic syndrome. *Am J Dis Children*. 1947;73:151–166.

50. Meizlik EH, Carpenter AM. Beneficial effect of measles on nephrosis; report of three cases. *Am J Dis Child*. 1948;76:83–90.

51. Belghiti D, Vernant JP, Hirbec G, Gubler MC, Andre C, Sobel A. Nephrotic syndrome associated with T-cell lymphoma. *Cancer*. 1981;47:1878–1882.

52. Devaiah BN, Singer DS. CIITA and its dual roles in mhc gene transcription. *Front Immunol*. 2013;4:476.

53. Al-Kandari W, Koneni R, Navalgund V, Aleksandrova A, Jambunathan S, Fontes JD. The zinc finger proteins ZXDA and ZXDC form a complex that binds CIITA and regulates MHC II gene transcription. *J Mol Biol*. 2007;369:1175–1187.

54. Wiszniewski W, Fondaneche MC, Le Deist F, et al. Mutation in the class II trans-activator leading to a mild immunodeficiency. *J Immunol*. 2001;167:1787–1794.

55. Ahmed A, Reith W, Puck JM, Cheng LE. Novel mutation in the class II transactivator associated with immunodeficiency and autoimmunity. *J Clin Immunol*. 2015;35:521–522.

56. Scholl T, Mahanta SK, Strominger JL. Specific complex formation between the type II bare lymphocyte syndrome-associated transactivators CIITA and RFX5. *Proc Natl Acad Sci U S A*. 1997;94:6330–6334.

57. Garvie CW, Boss JM. Assembly of the RFX complex on the MHCII promoter: role of RFXAP and RFXB in relieving auto-inhibition of RFX5. *Biochim Biophys Acta*. 2008;1779:797–804.

58. Mensali N, Grenov A, Pati NB, et al. Antigen-delivery through invariant chain (CD74) boosts CD8 and CD4 T cell immunity. *Oncoimmunology*. 2019;8:1558663.