

Using Population Genetics to Interrogate the Monogenic Nephrotic Syndrome Diagnosis in a Case Cohort

Matthew G. Sampson,* Christopher E. Gillies,* Catherine C. Robertson,*
Brendan Crawford,* Virginia Vega-Warner,* Edgar A. Otto,* Matthias Kretzler,^{†‡} and
Hyun Min Kang[§]

Departments of *Pediatrics and Communicable Diseases, and [†]Internal Medicine, Division of Nephrology and
[‡]Department of Computational Medicine and Bioinformatics, University of Michigan School of Medicine, Ann Arbor,
Michigan; and [§]Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, Michigan

ABSTRACT

To maximize clinical benefits of genetic screening of patients with nephrotic syndrome (NS) to diagnose monogenic causes, reliably distinguishing NS-causing variants from the background of rare, noncausal variants prevalent in all genomes is vital. To determine the prevalence of monogenic NS in a North American case cohort while accounting for background prevalence of genetic variation, we sequenced 21 implicated monogenic NS genes in 312 participants from the Nephrotic Syndrome Study Network and 61 putative controls from the 1000 Genomes Project (1000G). These analyses were extended to available sequence data from approximately 2500 subjects from the 1000G. A typical pathogenicity filter identified causal variants for NS in 4.2% of patients and 5.8% of subjects from the 1000G. We devised a more stringent pathogenicity filtering strategy, reducing background prevalence of causal variants to 1.5%. When applying this stringent filter to patients, prevalence of monogenic NS was 2.9%; of these patients, 67% were pediatric, and 44% had FSGS on biopsy. The rate of complete remission did not associate with monogenic classification. Thus, we identified factors contributing to inaccurate monogenic classification of NS and developed a more accurate variant filtering strategy. The prevalence and clinical correlates of monogenic NS in this sporadically affected cohort differ substantially from those reported for patients referred for genetic analysis. Particularly in unselected, population-based cases, considering putative causal variants in known NS genes from a probabilistic rather than a deterministic perspective may be more precise. We also introduce GeneVetter, a web tool for monogenic assessment of rare disease.

J Am Soc Nephrol 27: 1970–1983, 2016. doi: 10.1681/ASN.2015050504

More than 25 genes have been discovered in which fully penetrant mutations cause monogenic forms of nephrotic syndrome (NS).¹ Published reports suggest that monogenic causes are responsible for a substantial portion of NS, and these patients are resistant to immunosuppressive treatment.^{2–6} These prevalence estimates and clinical correlates can influence decisions to screen patients for monogenic NS as well as the subsequent management. Given this and the increasing availability of sequencing technology, there is growing interest in expanding the use of genetic screening to diagnose monogenic NS.

However, concurrently, results of population-based genetic studies outside of NS are challenging beliefs about the accuracy of pathogenicity prediction of rare variants found in known disease genes

Received May 8, 2015. Accepted September 6, 2015.

Published online ahead of print. Publication date available at www.jasn.org.

Correspondence: Dr. Matthew G. Sampson, Division of Pediatric Nephrology, University of Michigan School of Medicine, Ann Arbor, MI 48109. Email: mgsamps@med.umich.edu

Copyright © 2016 by the American Society of Nephrology

as well as their clinical consequences. Exome-wide sequencing studies in thousands of population subjects have identified a substantial prevalence of predicted harmful, protein-altering genetic variants in each human genome.^{7–10} Sequencing studies of genes implicated in monogenic disease in the population and/or healthy controls have shown that the prevalence of putatively causal mutations is orders of magnitude greater than the reported prevalence of the disease.^{11–15}

Thus, three critical concerns for the broad clinical implementation of genetic sequencing in NS are (1) distinguishing bystander or incompletely penetrant variants from causal mutations, (2) setting appropriate expectations for the chance that a patient will be diagnosed with a monogenic form of NS, and (3) determining the clinical consequences for those harboring a causal variant in a known NS gene.¹⁶

To address these concerns, 21 genes previously identified as harboring causal mutations for steroid resistant nephrotic syndrome (SRNS) (monogenic NS genes) were sequenced in putative controls from the 1000 Genomes Project (1000G) and 312 North American patients with NS recruited from the prospective, longitudinal Nephrotic Syndrome Study Network (NEPTUNE).¹⁷ The prevalence of predicted causal NS mutations in the population was analyzed under a number of filtering parameters. These results were used to devise a more accurate filtering strategy that reduced false attribution of pathogenicity. This filter was subsequently applied to patients to determine the prevalence of monogenic NS in the NEPTUNE cohort and its clinical correlates.

RESULTS

In total, 312 participants from NEPTUNE (Supplemental Table 1, Table 1) and 61 ancestry-matched subjects from the 1000G¹⁸ underwent targeted sequencing of 21 genes implicated in SRNS (Table 2). A default filtering pipeline was created, similar to previously reported variant-filtering pipelines,^{2–4} to classify variants found in these genes as causal

Table 1. Clinical and demographic characteristics of 312 subjects with NS

Name	Value or Distribution
Median age, yr	34
Age range, yr	2–84
Histologic diagnosis	FSGS (30%), minimal change (22%), membranous (15%), IgA nephropathy (10%), others (23%; Supplemental Table 1)
Fraction of pediatric patients	30%
Continental ancestry	European (48%), African (29%), admixed American (15%), East Asian (9%)

Table 2. Genes implicated in SRNS undergoing sequencing

Inheritance Mode	Genes
Dominant	<i>ACTN4</i> , <i>CD2AP</i> , <i>CFH</i> , <i>INF2</i> , <i>LMX1B</i> , <i>TRPC6</i> , <i>WT1</i>
Recessive	<i>COQ2</i> , <i>COQ6</i> , <i>CUBN</i> , <i>ITGA3</i> , <i>LAMB2</i> , <i>MYO1E</i> , <i>NEIL1</i> , <i>NPHS1</i> , <i>NPHS2</i> , <i>PDSS2</i> , <i>PLCE1</i> , <i>PTPRO</i> , <i>SCARB2</i> , <i>SMARCAL1</i>

(Table 3). Using the “default filter”, 13 of 312 patients with NS (4.2%) were classified with monogenic NS; 75% (9 of 13) classified as such had putative mutations in NS genes with autosomal dominant transmission. Analysis of the same genes under the same filter in 61 ancestry-matched, sequenced subjects from the 1000G did not identify any putative mutations.

Given the rarity of these types of variants, analyzing >61 subjects would be necessary to accurately determine their population prevalence. Thus, the same analysis was then extended to the approximately 2500 subjects from Phase 3 of the 1000G,¹⁹ with a 5.8% prevalence of subjects with putatively causal variants in monogenic NS genes; 70% of these subjects had putative causal variants in NS genes with dominant transmission. Hereafter, the prevalence of variants implicated as causal in subjects from the Phase 3 1000G is referred to as the background prevalence for a given filtering pipeline.

A 5.8% background prevalence is >550 times higher than the population prevalence of NS (<0.01%), even without considering only monogenic forms of this condition. Assuming that subjects in the 1000G are not enriched for monogenic NS, this implies that a substantial fraction of the population is carrying rare variants in implicated NS genes that are classified as mutations by existing filtering strategies. A proportion of this 5.8% background prevalence is likely caused by false attribution of pathogenicity to rare but harmless variants. It is also likely that there is a greater amount of incomplete penetrance for the NS phenotype than is currently appreciated.

The discovery of substantial background prevalence under the default filter could have important research and clinical implications in terms of misclassification of pathogenicity and recognition of incomplete penetrance. Thus, with an aim to identify a more stringent filtering strategy that reduced background prevalence without inflating false negative rates, conventional and novel filtering criteria were analyzed against population-level data of known monogenic NS genes.

The Discordance between Functional Prediction Databases

The use of functional prediction programs in filtering pipelines was examined first. Requiring two of three programs to predict a variant as damaging is currently an accepted approach to pathogenicity filtering of missense variants.² The merits of this choice were investigated.

PolyPhen2,²⁰ SIFT,²¹ and MutationTaster²² pathogenicity predictions of 1220 missense variants found in the SRNS genes

Table 3. Characteristics of the default filter devised to classify variants as causal for NS

Parameter	Default
No. of genes	21
AF threshold in sequenced samples	<5%
AF threshold in the EVS	<0.5% (dominant gene); <1% (recessive gene)
AF threshold in the 1000G	Not used
Functional prediction threshold	Predicted damaging by two of three of PolyPhen2, SIFT, and MutationTaster (A or D)
Transcript specificity	Longest transcript
Conservation score (GERP)	Not used

GERP, genome evolutionary rate profiling.

in the subjects in the 1000G were computed. When requiring one, two, or three of these three programs to classify a variant as damaging, 12.6%, 5.8%, or 1.5% background prevalence, respectively, in 1000G was observed (Figure 1). When applying this approach to 312 patients with NS, the prevalence of monogenic diagnosis was 8.7%, 4.2%, or 2.2%, respectively. As expected, the most stringent threshold was most effective in reducing background prevalence but also likely increased false negatives. This was caused by both missing scores for particular variants and discordance between classification methods. Altogether, requiring two of three was found to be a compromise that substantially reduces background prevalence while allowing enough variability in the prediction algorithms to prevent excessive false negative predictions.

Population Stratification

Applying an allele frequency (AF) threshold using Exome Variant Server (EVS) data (<http://evs.gs.washington.edu/EVS/>) is frequently used to exclude noncausal variants. However, because the EVS population is only composed of European Americans and blacks, EVS-based AF thresholds will not robustly remove variants that, although rare in those of African or European ancestries, are common in other ancestries (e.g., East Asian). In addition, some variants predicted to be rare in the 1000G continental populations may, in fact, be common when considering local populations (<http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/>). To address this, an AF of <1% in any of five 1000G continental ancestries or more stringently, any of 26 local population groups was applied to the NS gene set in subjects in the 1000G (Supplemental Figure 1). By applying continental and local population AF thresholds rather than solely using EVS AF, the background prevalence in the population decreased from 5.8% (n=147) to 4.5% (n=113) and 3.7% (n=95), respectively.

Incorporating Prior Gene- and Transcript-Specific Knowledge

Negative selection is the genome-wide process of depletion of variants that are harmful and reduce an organism's ability to

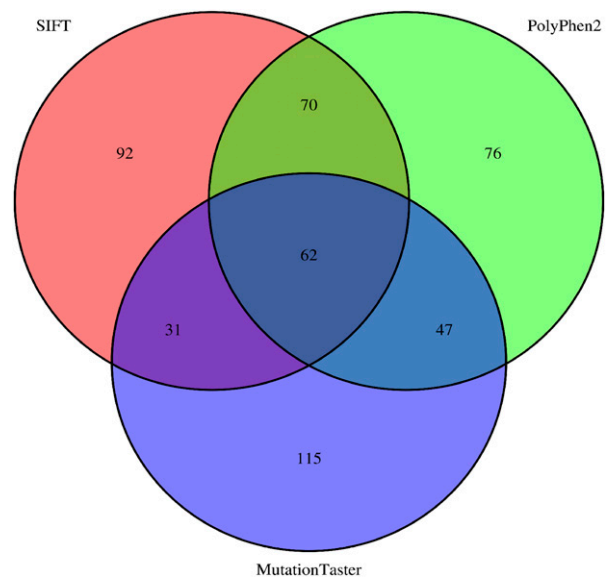


Figure 1. Substantial discordance between variant level functional prediction methods. From 1220 variants in 21 SRNS genes identified in the 1000G, the top 25% scored as most deleterious (n=255) by each of Polyphen2, SIFT, and MutationTaster were studied for concordance of prediction. Requiring one of three methods to classify variants as damaging implicates 493 variants as causal. Two of three implicate 210 variants as causal. Requiring all three programs to be in concordance reduced the number of causal variants to 62.

reproduce. The degree of negative selection acting on a gene can differ by transcript. Variants found in transcripts under the most negative selection would be expected to be the most deleterious, whereas healthy individuals may be expected to have variants in less selected transcripts.²³ It was hypothesized that background prevalence could be reduced by considering as causal only variants occurring in the transcript of a gene under the highest degree of negative selection. An existing metric was used (TIMS score) to identify the transcript in each gene under the most negative selection.²⁴

Among 21 genes, only *INF2* and *LAMB2* had a transcript under substantially greater negative selection. When considering *INF2*'s longest transcript (Figure 2) and holding all other filtering criteria constant, the odds of a putative pathogenic variant being present in a patients versus a control are 1.3 (5 of 312 patients; 32 of 2535 subjects in the 1000G). However, when using the most negatively selected transcript of *INF2* (ENST00000398337), the odds increased to 24.6 (3 of 312 patients; 1 of 2535 subjects in the 1000G). This is consistent with recent studies reporting *INF2* mutations in NS.^{25,26} There were no patients or controls with two qualifying *LAMB2* variants.

Number of Genes Sequenced and Estimated Prevalence of True Monogenic NS

The list of genes implicated in monogenic NS continues to grow. Additionally, patients with a lower likelihood of truly

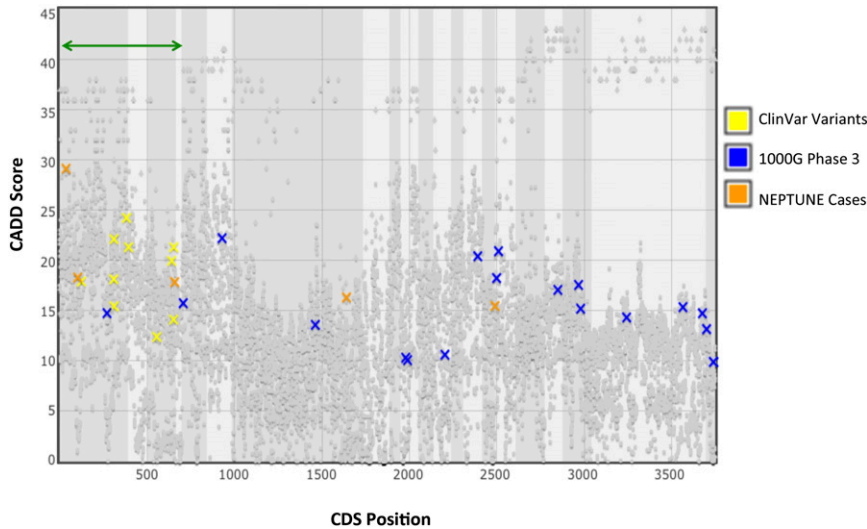


Figure 2. Transcript-specific filtering influences the accuracy of monogenic diagnosis. The horizontal axis represents the base position in the coding sequence of the *INF2* gene on the basis of the longest transcript. The vertical axis is the combined annotated-dependent depletion (CADD) score of each potential variant. All possible missense and nonsense variants in this region are plotted in gray. Blue indicates variants observed in the 1000G, and orange indicates variants observed in 311 subjects with NS. Yellow indicates known pathogenic variants reported in ClinVar 138. The green interval represents the exons for the most selected transcript with the highest TMS score (ENST00000398337). It is apparent that rare variants found in patients with NS are enriched in this transcript, whereas rare variants found in the 1000G are not. CDS, coding DNA sequence.

having monogenic NS can now be sequenced. By using the background prevalence in 21 SRNS genes and making assumptions regarding false positive and false negative rates, simulations were performed to quantify the effect on the accuracy of monogenic classification as a function of (1) the number of genes sequenced or (2) the estimated true prevalence of monogenic NS in the population under study.

Varying Numbers of Genes Analyzed

The default pathogenicity filter was applied to 21 SRNS genes in a theoretical cohort of patients with NS whose predicted prevalence of true monogenic NS was 15% (Figure 3). In this way, 16.5% of participants were classified with monogenic NS. Although 11.7% would have true monogenic NS, 4.9% would be falsely classified as such because of the background prevalence.

The next simulation considered a theoretical scenario in which an additional 21 newly discovered monogenic NS genes were added to a diagnostic panel, with similar background prevalence rates per

gene as the initial panel. Now, with this 42-gene panel, 22.9% of subjects were classified with monogenic NS. Although 13.1% would be true positives, 9.8% of subjects would be falsely classified. Thus, because of the rarity of monogenic NS, doubling the number of candidate genes analyzed results in a doubling of false positive classification without significant changes in true positives.

Varying Underlying Prevalence of True Monogenic NS

Diagnostic genetic screening of subjects with a higher pretest probability of monogenic NS improves the accuracy of pathogenic classification. Screening the SRNS gene set in a population with a 2% estimated true prevalence of monogenic NS was simulated. In this scenario, the relative risk of putative mutations being in a patient versus a subject in the 1000G was 1.25 (Figure 4). However, if the prevalence of true NS in the cohort under study was increased to 20%, the relative risk was increased to 3.5.

Systematic Assessment of Parameters to Create a Stringent Pipeline

Ultimately, insights derived from these analyses led to the creation of a stringent filtering strategy (Table 4). Using this “stringent filter” reduced the background prevalence in the 1000G to 1.5%, a 74% decrease compared

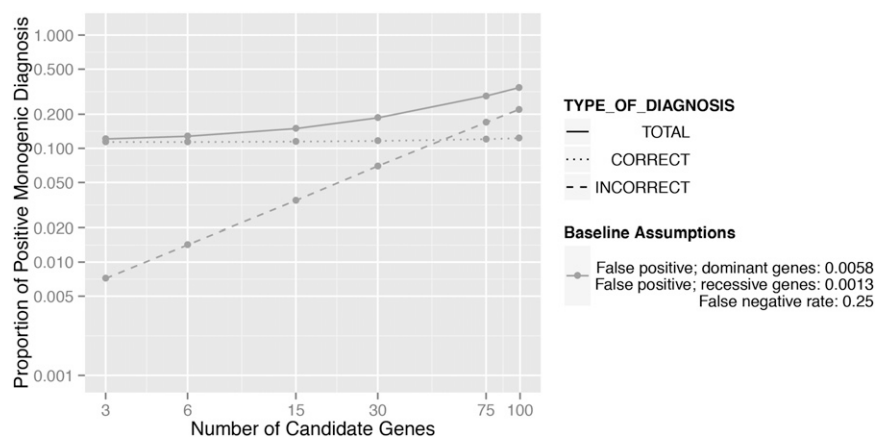


Figure 3. Increasing the number of genes sequenced increased the prevalence of incorrectly identified variants. The horizontal axis is the number of candidate genes sequenced assuming median gene length (1271 bp), and the vertical axis is the proportion of individuals classified with a monogenic diagnosis. Under assumptions of a fixed background prevalence and false negative rate, curves represent the total proportions of monogenic diagnosis, those expected to be correct, and those expected to be incorrect.

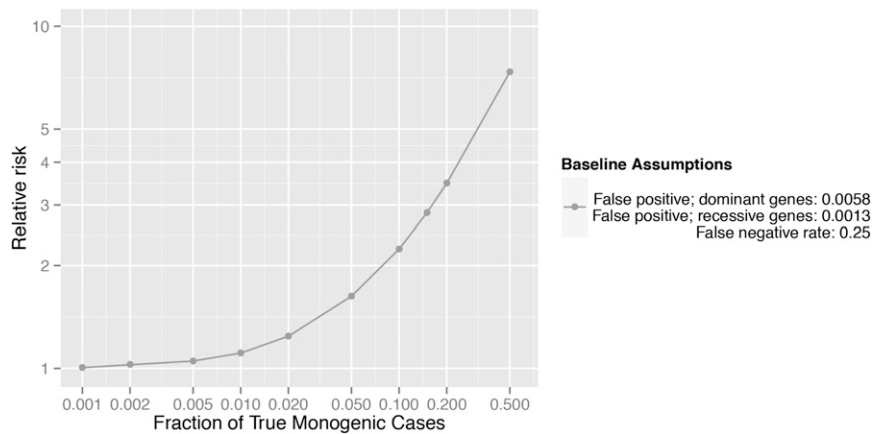


Figure 4. Effect of prevalence of true monogenic NS on the relative risk of a variant found in the 21-gene SRNS gene set being in a patient who is monogenic versus population control. The horizontal axis is the prevalence of monogenic NS in a population. As the prevalence of monogenic NS increases, the relative risk of the variant being in a patient versus a subject in the 1000G increases as well.

with the 5.8% from the default filter. The stringent filter was then applied to three kindred with familial forms of NS sequenced by the investigators and a curated list of 70 mutations reported in the initial gene discovery papers, which were assumed to be established as causal for NS.^{27–45} This filter correctly identified the causal mutation in these three families, while rejecting common polymorphisms found in each gene. Of 70 established mutations from familial NS, the filter classified 90% as causal as well (63 of 70). Importantly, two of seven of the putative mutations that the filter did not classify as causal, on manual review, were convincingly unlikely to represent causal mutations. Thus, the stringent filter had 93% sensitivity overall when using these mutations as a gold standard. Stratified by inheritance, the filter was 96% and 83% sensitive for published mutations in recessive and dominant SRNS genes, respectively.

Putative Monogenic NS in Patients under the Stringent Filter

Applying the stringent filter, 9 of 312 (2.9%) patients with NS in NEPTUNE were classified as monogenic (Table 5). Putative mutations were present in 7 of 21 of the genes sequenced. Seven participants had mutations in dominantly inherited genes. The remaining two were homozygous for *NPHS1* and *NPHS2* variants; 4 of 10 distinct mutations had been previously reported in the Human Gene Mutation Database (HGMD).

Sixty-seven percent (six of nine) of patients who were monogenic had pediatric onset; 22% (two of nine) had a likely family history of NS. Four participants had a histologic diagnosis of FSGS, and three had minimal change disease. The remaining two in the other glomerulopathy group were diagnosed with IgM nephropathy and immune complex GN. Additional characteristics of participants with putative

monogenic versus nonmonogenic NS are presented in Supplemental Table 2. Prevalence estimates of monogenic NS in NEPTUNE stratified by clinical criteria of interest are presented in Table 6.

On the basis of proteinuria measurements obtained longitudinally, 56% of patients with a monogenic diagnosis (five of nine) attained complete remission (CR). This included four of six pediatric-onset patients and one of three adult-onset patients. Additional details in Table 5 show that remission was achieved with a variety of therapeutic strategies, including steroid monotherapy and renin-angiotensin-aldosterone system blockade monotherapy as well as regimens of multiple immunosuppressants. In survival analyses, the chance of achieving CR in those classified with monogenic NS was not significantly different than in those without NS in

both unadjusted (hazard ratio, 1.4; *P*=0.44) (Figure 5) and adjusted (hazard ratio, 1.5; *P*=0.35) models.

Genome-Wide Prevalence of Potential False Monogenic Diagnosis

The effect of background prevalence on potential false diagnosis extends beyond known NS genes. In subjects of the 1000G, the background prevalence of putative pathogenic mutations was calculated genome-wide⁴⁶ using the default pipeline and under both dominant and recessive models. A subset analysis of a gene set known to cause Mendelian diseases under dominant inheritance (hOMIM dominant)⁴⁷ was also performed.

Under the dominant model, 83.1% of all genes and 87.7% of hOMIM-dominant genes had putative pathogenic variants in >0.1% of the population (Figure 6). This persisted even after normalizing all genes to a typical gene length. By contrast, only 7.3% of genes under the recessive model had >0.1% of

Table 4. Characteristics of the stringent filter devised to classify variants as causal for NS

Parameter	Default
No. of genes	21
AF threshold in sequenced samples	<5%
AF threshold in the EVS	<0.1% (dominant gene); <0.1% (recessive gene)
AF threshold in the 1000G	<1% in any of 26 populations
Functional prediction threshold	Predicted damaging by two of three of PolyPhen2, SIFT, and MutationTaster
Transcript specificity	Most negatively selected
Conservation score (GERP)	>4

GERP, genome evolutionary rate profiling.

Table 5. Characteristics of patients with NS classified with monogenic NS

Subject	Ancestry	Histology	Gene	Genotype	Amino Acid Change	HGMD	Pediatric	Remission	Immunosuppressive Agent	RAAS Agent
1	AFR	MCD	<i>NPHS1</i>	HOM_ALT	p.A765V	No	Yes	Complete	None	Yes
2	AMR	FSGS	<i>TRPC6</i>	HET	p.N765S	No	Yes	Complete	Steroid	Yes
3	AFR	MCD	<i>CD2AP</i>	HET	p.M496I	Yes	Yes	Complete	Steroid	No
4	EUR	IgM nephropathy	<i>WT1</i>	HET	p.C350R	No	Yes	Complete	Steroid, tacrolimus, MMF	No
5	EUR	ICGN	<i>INF2</i>	HET	p.E220K	Yes	Yes	None	None	Yes
6	EUR	FSGS	<i>NPHS2</i>	HOM_ALT	p.A284V	Yes	Yes	None	None	Yes
7	EUR	MCD	<i>INF2</i>	HET	p.A13T	Yes	No	Complete	Steroid	Yes
8	EUR	FSGS	<i>TRPC6</i>	HET	p.D798H	No	No	Partial	None	Yes
			<i>LMX1B</i>	HET	p.D374N	No				
9	AMR	FSGS	<i>INF2</i>	HET	p.P35R	No	No	Partial	Steroid, CNI	Yes

AFR, African 1000G continental ancestry; AMR, American 1000G continental ancestry; EUR, European continental ancestry; MCD, minimal change disease; ICGN, immune complex GN; HOM_ALT, homozygous for the alternative allele; HET, heterozygous variant; MMF, mycophenolate mofetil; CNI, calcineurin inhibitor; RAAS, renin-angiotensin-aldosterone system.

subjects predicted to be monogenic (91% fewer than the dominant model). Thus, especially under a dominant mode of transmission, there is substantial exome-wide background prevalence of predicted harmful, protein-altering variants. Perhaps unexpectedly, although these genes are implicated in Mendelian disease, they do not have fewer pathogenic variants compared with all genes.

DISCUSSION

The growing efforts to perform genetic screening in patients with NS are motivated by reports of the clinical significance of deriving a diagnosis of monogenic NS.¹⁶ Many of the studies that have shaped the current understanding of monogenic NS were performed in a limited number of genes in patients who, in various ways, were selected for a higher prevalence of monogenic NS. This includes those with steroid resistance, those with congenital or infantile onset, those with familial disease, and those from regions in which there was an elevated degree of consanguinity.^{48,49} Now, in this genomic era, increasingly diverse populations of affected patients (both demographically and clinically) are undergoing genetic screening of an expanded number of previously implicated NS genes. Given the potential influence on clinical decision making of making this diagnosis, there was motivation to discover whether previous findings about monogenic NS were generalizable in a clinically familiar scenario, namely sporadically affected North American patients with family data that are absent.

This study was designed to determine the spectrum of genetic variation independently in patients and controls under the same filtering strategies rather than solely screening control databases to detect the presence of putative mutations found in patients. In doing so, it was appreciated that, under a typical pathogenicity filter, there was a substantial background prevalence of protein-altering variants in the population

(e.g., 5.8% under the default filter). The majority of the background prevalence in the NS gene set resulted from putative pathogenic variants in genes with a dominant inheritance model. The lower background contributed by recessive genes under stringent filtering conditions suggests that finding two predicted harmful variants in the same recessive NS gene in a patient is less likely to be caused by chance. Thus, without accounting for the background prevalence and seeking filtering strategies to minimize it, sequencing studies for rare diseases are at risk for inaccurately classifying variants as causal mutations.

Table 6. Prevalence of monogenic NS by selected subsets

Description of Subsets	Classified Monogenic (%; Fraction)
Overall	2.9 (9/312)
Pediatric	6.3 (6/95)
Pediatric + FSGS	6.7 (2/30)
Pediatric + FSGS + no CR	6.3 (1/16)
Pediatric + FSGS + CR	7.1 (1/14)
Pediatric + FSGS or MCD	5.5 (4/73)
Pediatric + no CR	5.6 (2/36)
Pediatric + CR	6.8 (4/59)
Adult	1.4 (3/217)
Adult + FSGS	2.7 (2/75)
Adult + FSGS + no CR	3.8 (2/53)
Adult + FSGS + CR	0.0 (0/22)
Adult + FSGS or MCD	3.2 (3/95)
Adult + no CR	1.5 (2/133)
Adult + CR	1.2 (1/84)
All + FSGS	3.8 (4/105)
All + FSGS + no CR	4.3 (3/69)
All + FSGS + CR	2.8 (1/36)
All + FSGS or MCD	4.2 (7/168)
All + not FSGS	2.4 (5/207)
All + no CR	2.4 (4/169)
All + CR	3.5 (5/143)

MCD, minimal change disease.

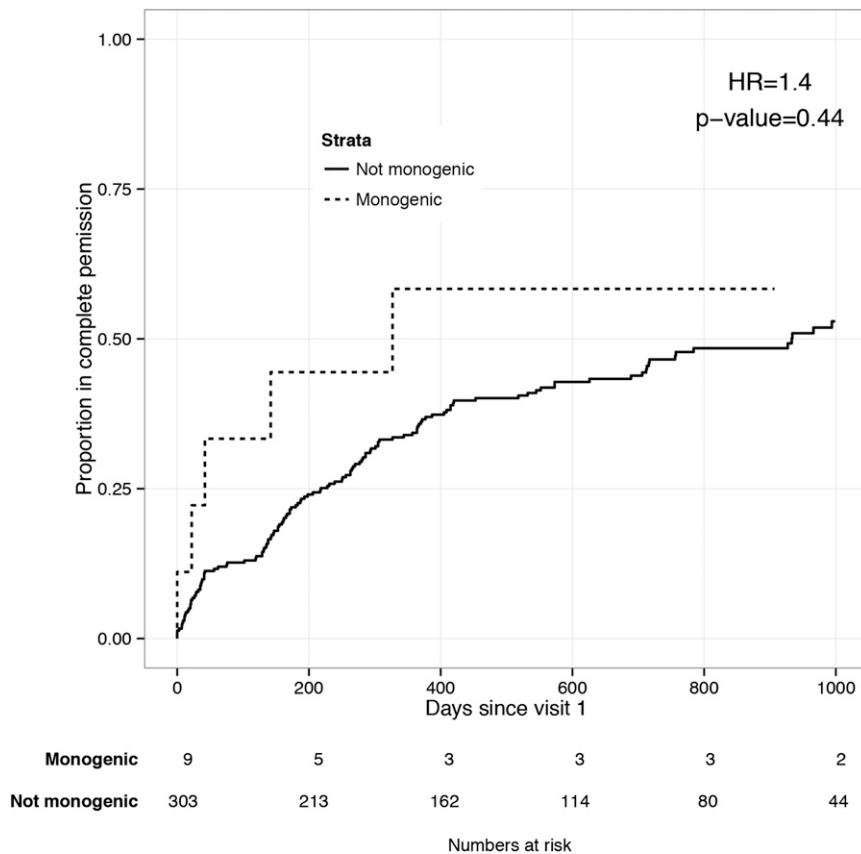


Figure 5. Achievement of complete remission does not significantly differ by monogenic status. Unadjusted Kaplan–Meier plot of the proportion of CR in patients with NS stratified by monogenic status. Plot is truncated at day 1000. P value was determined with an unadjusted Cox proportional hazards model of CR and monogenic status. HR, hazard ratio.

This background prevalence was useful to aid in simulations, showing for one, that increasing the pretest probability (*i.e.*, the estimated prevalence) of true monogenic NS in the population under study substantially improves the accuracy of classification. This supports the classic approach of incorporating clinical, laboratory, and histologic data to enrich the true prevalence of monogenic disease. Also, increasing the number of NS genes screened in patients (particularly those with dominant transmission) also increases the false discovery rate. Recognizing that sequencing increased numbers of genes in patients with a low pretest probability of monogenic NS increases inaccurate monogenic diagnosis seems particularly relevant at this time, because expanded genetic screening for NS in the clinical and research arenas may enrich for both of these factors.

Computing background prevalence also provided an outcome metric to calibrate a number of different filtering strategies on prevalence of monogenic diagnosis in subjects in the 1000G and patients. For instance, filters that achieved a background prevalence matching prevalence estimates of NS in the population resulted in an absence of patients with NS being classified with monogenic NS. However, only requiring one

method to predict a variant as pathogenic or using AF filters of 1% substantially boosted background prevalence, including for recessive genes. The stringent filter devised reduced background prevalence but also, had face validity. Although the 1.5% background prevalence for NS under the stringent filter is far greater than would be predicted under a Mendelian paradigm, it is consistent with reports of incomplete penetrance for other conditions with monogenic causes.^{50,51}

Recently, the Exome Aggregation Consortium (ExAC) released the ExAC Browser (<http://exac.broadinstitute.org>). ExAC provides summary variant calls of whole-exome sequencing data from 60,076 unrelated individuals worldwide. ExAC provides an increasingly accurate estimate of allele frequencies worldwide. In this study, our inferences of putative causal variants were comparable when using ExAC-based AF thresholds versus our existing strategy of combining an EVS AF of <0.1% with the 1000G local population AF threshold of <1%. Moving forward, ExAC should be a powerful resource for the NS community to identify optimal filtering strategies for determining causality of variants for NS. In particular, its increased power to detect rare variants may contribute to our understanding of the prevalence of putative disease-causing mutations in the general population.

Applying the stringent filter to patients in NEPTUNE resulted in a number of notable observations; 2.9% (9 of 312) of patients in NEPTUNE were classified with monogenic NS. However, in NEPTUNE, as in many clinical care settings, parental and sibling sequence data were unavailable to perform segregation analysis, raising the potential that some of these putative causal variants are present in healthy family members.

There was a question of whether the NEPTUNE cohort, reflecting a patient population of all comers *de novo* diagnosed in North American academic medical centers, fully accounts for some of the differences in prevalence estimates compared with previous studies.^{2,4,25,52,53} Thus, subset analyses were also performed, stratifying patients by more traditional classifications, such as pediatric onset, FSGS histology, or SRNS. Still, the group that one may predict to have the highest prevalence of monogenic diagnosis, children with FSGS who have not achieved CR, had a monogenic prevalence of only 6.3%.

Of equal interest to the prevalence of monogenic NS in the cohort was the observation that 56% (five of nine) of participants classified with monogenic NS achieved CR. This includes one participant with FSGS and two participants whose

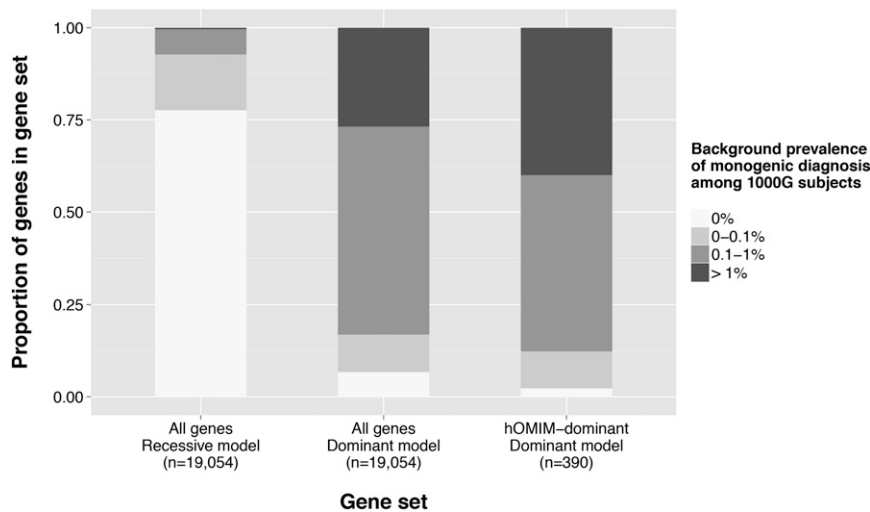


Figure 6. Genome-wide background prevalence of rare and deleterious variants. All coding variants observed in the 1000G were classified as pathogenic or benign on the basis of stringent filtering criteria (maximum ancestry-specific EVS $MAF < 0.1\%$, 1000G $MAF < 5\%$, and loss of function or nonsynonymous and passing the two of three functional filter). Then, for each gene, the resulting background prevalence of predicted pathogenic mutations is calculated. A background prevalence of 1% for a given gene indicates that 1% of subjects in the 1000G carry a rare and deleterious variant in this gene. Subsets of genes (Concise Methods) and inheritance models considered are represented on the horizontal axis. Bars depict the proportions of genes within a gene set having background prevalence of predicted pathogenic variation within specific ranges. hOMIM, hand curated Online Mendelian Inheritance in Man; MAF, minor allele frequency.

causal variants were previously reported in the HGMD. In regard to specific therapies used to achieve CR, it was notable that there was no specific strategy used. CR was achieved with monotherapy with ACE inhibitors or steroids, ACE inhibitors and steroids, and in one participant, steroids, calcineurin inhibitors, and mycophenolate. Furthermore, there was no significant association between monogenic status and failure to achieve CR.

A potential reason for this observation is that the harmful effects of these rare variants are potentiated or protected against by environmental or genetic backgrounds that differ in this population versus those previously reported. Another possibility is that NEPTUNE's prospective, longitudinal design may have captured CR events that are not ascertained or not appreciated in a cross-sectional or retrospective design. Finally, it is possible that the filtering parameters were still not stringent enough and that functional studies and family segregation analysis would have eliminated all participants who achieved CR. For a number of reasons, including having a child with a homozygous variant in *NPHS1* (p.A765V) who achieved CR and a participant with an *INF2* variant previously reported in a family study with SRNS,³⁶ this explanation seems less likely.

Although variable expressivity of the SRNS phenotype in patients with monogenic NS has certainly been reported,^{4,29,54–57} the immunosuppressive-resistant phenotype in monogenic NS has been of greater focus.^{2,5,21,22,25,58–60} Arguably, the

strongest motivation for clinicians and families to perform genetic screening is the belief that this diagnosis would allow avoidance of immunosuppressant agents that would only harm with no chance of benefit. Altogether, this study's findings suggest a previously unrecognized degree of variable expressivity of putatively fully penetrant variants in monogenic NS genes.

Thus, at least in this population of sporadically affected North American patients, making a monogenic diagnosis of NS is challenged by background prevalence, particularly in dominant genes, and incomplete penetrance of the SRNS phenotype. Like any screening test, variant prediction filters have imperfect sensitivity and specificity. Under the stringent filter, there was 1.5% background prevalence in the 1000G, suggesting that some variants, even if found in patients, may be false positives. In regard to false negatives, when using 92 published mutations as a gold standard, the sensitivity was not 100%. This was more of a factor in genes with a dominant mode of transmission. Thus, care must be taken to consider the output

of this filter in the context of other complementary information about both the variants and affected patients/research participants under consideration. Future sequencing studies of hundreds to thousands of longitudinally phenotyped patients with sporadic and familial NS and controls with proteinuria and renal functional data should increase our accuracy of these prediction filters.

Monogenic prediction may also be challenged by the existence of more complex modes of inheritance in NS, such as alleles that modify the effect's causal variants, triallelism or bigenic heterozygosity,^{61,62} or risk alleles that exist independently of rare coding variants in known monogenic NS genes.^{63–66} These complex forms of genetic architecture should continue to be the focus of study as well. Altogether, these factors seemingly challenge the ability to make treatment decisions solely on the basis of a subject's genetic sequence. Thus, it may be more accurate to reconsider predicted monogenic mutations as probabilistic risk factors.

There are a number of advantages of a probabilistic model. First, quantifying the pathogenicity of variants discovered along a continuous spectrum may more accurately reflect their effect on disease. Second, consideration of factors additional to variant prediction could allow more informed adjustment of the probability of a patient having monogenic disease. Although more complex than a deterministic, dichotomous diagnosis, quantifying relative risk of a variant being pathogenic for NS may improve the precision of clinical care.

For rare diseases in general, whether for family counseling, clinical management, or study design, there is strong motivation to perform genetic screening and optimize the accuracy and interpretation of a monogenic diagnosis. This study showed that systematically devising a variant filtering strategy calibrated to genetic variation in the population can improve the accuracy and potential clinical use of genetic screening for monogenic forms of NS in a group increasingly undergoing diagnostic sequencing. This study also showed that the background prevalence of rare, protein-altering variants exists genome wide, implying that benefits of accounting for its effect could extend beyond NS.

To this end, the web application GeneVetter (<http://genevetter.org>) was created to allow investigators of any phenotype to quantify the background prevalence of predicted casual variants in the 1000G within their genes of interest under default or custom filtering strategies.⁶⁷ In addition, GeneVetter also allows users to analyze sequence data from their own subjects without transmitting genotype information off their servers. Overall, GeneVetter should aid users in their own sequencing studies by improving accuracy in implicating variants as causal as well as interpreting the results published from other reports.

CONCISE METHODS

Study Participants

Participants were enrolled in NEPTUNE at their first clinically indicated biopsy for suspicion of having minimal change disease, FSGS, or membranous nephropathy.¹⁷ From each enrolled participant, blood, renal biopsy tissue, and urine along with highly detailed demographic and clinical data were collected. Patients are followed four times in the first year and biannually thereafter. Unlike other studies, sporadic patients without ascertainment by family history or age of onset were collected; 325 participants with available DNA for sequencing were included in this study. Of the participants, 13 were removed, because they failed quality control for a total of 312 participants. Per the NEPTUNE protocol, CR was defined as a reduction of the urine protein-to-creatinine ratio to <0.3 . Partial remission was defined as a reduction of the urine protein-to-creatinine ratio to between 0.3 and 3.5 and a 50% reduction compared with measurement at the eligibility visit.

Matching 1000G Controls

The first 256 subjects with NS were also genotyped using Illumina Exome Chip. Genotypes were called by GenomeStudio software with default cluster. After excluding single-nucleotide polymorphisms with $<95\%$ call rates, all samples had $>98\%$ call rates. These genotypes were combined with the Exome Chip genotypes of 2089 samples from the 1000G across all overlapping markers and produced principal components using the EFACTS software package (<http://genome.sph.umich.edu/wiki/EFACTS>). The five closest remaining participants with NS and their computed centroid were then iteratively identified. Next, the closest subjects in the 1000G to this

centroid were identified by Euclidian distance in the space of the first four principal components, where each principal component was weighted by its eigenvalue. When the number of remaining subjects was <32 , centroids were computed using two samples instead of five, and the final three participants with NS were matched directly to the closest remaining subjects in the 1000G. This was repeated until 61 controls were selected. The number of participants with NS per centroid was reduced as the number of remaining subjects became small, because the remaining participants were relatively far from each other. One control was removed after sequencing because of poor sequencing quality.

Targeted Sequencing of 21 Genes

Microfluidic PCR paired with next generation sequencing of all exons of 21 monogenic NS genes was performed using methods previously described.^{53,68,69}

Sequence Alignment and Variant Calling

Genome Analysis Toolkit (GATK)⁷⁰ best practices (<https://www.broadinstitute.org/gatk/guide/best-practices>; accessed 7/13/2015) were followed for preprocessing, sequence alignment, and variant calling. First adapter sequences were trimmed using cutadapt 1.8,⁷¹ BWA mem⁷² (0.7.12) software was used to align the raw sequence data. Duplicated reads were not removed or marked, because it is impossible to distinguish PCR duplicates from independent segments with the microfluidic PCR data. GATK (3.4) was applied to recalibrate base quality scores, perform indel realignment, and detect single-nucleotide polymorphisms and short indels. GATK's Unified Genotyper was used for variant calling instead of Haplotype Caller, because some known variants were missed with Haplotype Caller. The default settings for the Unified Genotyper were used, except that the down sampling was turned off. GATK best practices recommend using hard variant filters for low-coverage data. However, greater sensitivity and specificity in terms of Sanger-sequenced confirmed variants was achieved using a Support Vector Machine (SVM)⁷³ with a radial basis function kernel. The SVM was trained using the R package e1071 (<http://cran.r-project.org/web/packages/e1071/>; R interface to libsvm)⁷⁴ with the following features: read position rank sum, call rate, quality by depth, mean allele balance for heterozygous subjects, strand odds ratio, and inbreeding coefficient. The impute R package (<http://www.bioconductor.org/packages/release/bioc/html/impute.html>), which uses the k nearest neighbor algorithm, was used to impute missing values. A site was labeled a negative training example if it failed at least two criteria of the default GATK best practices hard filter or was in the 1000G Phase 3 or ExAC v0.3 and not labeled PASS. A site was labeled a positive training example if it was labeled PASS in the 1000G or ExAC. All remaining sites were labeled as positive or negative training examples on the basis of their presence or absence in dbSNP b138. Sites passing the SVM filter were further refined at the genotype level. For heterozygous genotypes, at least five alternative alleles, an allele balance of $\geq 10\%$, and a genotype quality of 40 were required. For nonreference homozygous genotypes, the same parameters were required, with the exception of allele balance.

Functional Annotation of Variants

The functional consequences of exonic variants were annotated using the SnpEff, version 3.5 software package⁷⁵ with the GenCODE v9 database.⁴⁶ Each missense and nonsense variant was annotated using 19 functional scores annotated in the dbNSFP 2.5.⁷⁶ Variants were annotated as being in the HGMD⁷⁷ if they matched the chromosome, position, and amino acid change for some HGMD entry.

Variant Filtering Parameters Considered

The following criteria for each variant or individual were considered when devising variant filtering criteria: (1) maximum AF in the EVS across European and black samples; (2) functional prediction results from SIFT, PolyPhen2, and MutationTaster as annotated in dbNSFP; (3) maximum continental AF and population AF calculated from the 1000G after leaving one sample out, which was calculated by $[\text{allele count} - 1]/[\text{total allele count} - 2]$ within each continental or population group; (4) minimum genomic evolutionary rate profiling++ score⁷⁸ threshold; and (5) the longest versus the most negatively selected transcript. Variants with $\text{AF} \geq 5\%$ in the sequenced pool of samples were excluded in all of the analyses. The individual thresholds are as stated in each section or are not used if not mentioned.

Concordance Estimated between Functional Prediction Methods

In total, 1220 variants in 21 SRNS genes identified were identified in the subjects of the 1000G. The pathogenicity score for each of these variants was computed by each of Polyphen2, SIFT, and MutationTaster. The top 25% most damaging from each method were then brought forward for evaluation of concordance of classification. This represented 225 variants per program, which because of overlap, comprised 493 distinct variants.

Population Stratification

For each individual, the number of rare (EVS $\text{AF} < 0.5\%$ or $< 0.1\%$) and deleterious (deleterious in two of three functional prediction) variants in 21 SRNS genes is calculated. The average number of rare and deleterious variants per person is then calculated and stratified by continental ancestry. Variants were further filtered by continental AF ($\text{AF} < 1\%$) or population AF ($\text{AF} < 1\%$) in the 1000G, leaving one sample out in AF calculations to avoid bias. Average numbers of predicted pathogenic variants per each of these populations are then recalculated.

Transcript Filtering

Analyses focused on the longest transcript per gene, except where specified otherwise. Only transcripts that were labeled protein coding from protein-coding genes for autosomal chromosomes were considered. To ensure fair comparison between transcripts, 1718 (2.3%) of 74,569 transcripts that had entries in dbNSFP were removed.

Simulation Methods

The ratio of patients who are true dominant monogenic dominant to patients who are true dominant recessive was assumed to be three. This ratio is the ratio observed in our cohort. The fraction of subjects who are true monogenic was assumed to be 0.15. Subjects who are true

monogenic were generated as $B(1, 0.116) + B(1, 0.039)$, where $B(1, p)$ is a binomial distribution with one trial and probability p . Patients who were false monogenic were generated by $B(1, \text{no. of dominant genes} \times \text{observed background rate per dominant gene}) + B(1, \text{no. of recessive genes} \times \text{observed background rate per recessive gene})$. The observed background rate for dominant genes was $103/2535/7 = 0.0058$. The observed background rate for recessive genes was $45/2535/14 = 0.0013$. The subjects predicted to be monogenic had a false negative rate of 0.25; therefore, 0.75 of the subjects who are true monogenic were predicted to be monogenic, and all of the subjects who are false monogenic were predicted to be monogenic. For the calculation where we doubled the genes from 21 to 42, we decreased the false negative rate to 15% to account for subjects who were previously classified as a negative simply because their disease gene was not sequenced. The fractions of true positives, false positives, true negatives, and false negatives were calculated using 1 million simulated subjects.

Genome-Wide Background Prevalence

All coding variants in 19,054 genes observed in the 1000G were classified as pathogenic or benign on the basis of stringent filtering criteria (maximum ancestry-specific EVS minor allele frequency $< 0.1\%$; 1000G $\text{MAF} < 5\%$; loss of function or nonsynonymous and passing the two of three functional filter). For the entire gene set, the background prevalence of predicted pathogenic mutations in the 1000G was calculated under a dominant and recessive model (homozygous or two heterozygous variants). This analysis was repeated in a gene set of 390 genes implicated in dominant Mendelian diseases (hOMIM dominant).⁴⁷ A background prevalence of 1% for a given gene indicates that 1% of subjects of the 1000G carry a rare and deleterious variant in this gene.

Time to Event Analyses of CR

A Cox proportional hazards model was used to determine an association between time to first CR of proteinuria and the status of a subject being classified as monogenic. One model was estimated without adjustment, and the other was adjusted for cohort, baseline age, and sex.

ACKNOWLEDGMENTS

M.G.S. is a Carl Gottschalk Research Scholar of the American Society of Nephrology and is supported by the National Institute of Diabetes and Digestive and Kidney Diseases Grant 1K08-DK100662-01. M.K. is supported by Nephrotic Syndrome Study Network Consortium (NEPTUNE) Grant U54DK083912 and George M. O'Brien Kidney Research Core Center at the University of Michigan Grant P30-DK081943. The NEPTUNE is a part of the National Center for Advancing Translational Sciences (NCATS) Rare Disease Clinical Research Network (RDCRN), which is supported through a collaboration between the Office of Rare Diseases Research (ORDR), the NCATS, and the National Institute of Diabetes, Digestive, and Kidney Diseases. The RDCRN is an initiative of the ORDR and the NCATS.

Additional funding and/or programmatic support for this project have also been provided by the University of Michigan, NephCure Kidney International, and the Halpin Foundation.

DISCLOSURES

None.

REFERENCES

- Saleem MA: New developments in steroid-resistant nephrotic syndrome. *Pediatr Nephrol* 28: 699–709, 2013
- Sadowski CE, Lovric S, Ashraf S, Pabst WL, Gee HY, Kohl S, Engelmann S, Vega-Warner V, Fang H, Halbritter J, Somers MJ, Tan W, Shril S, Fessi I, Lifton RP, Bockenbauer D, El-Desoky S, Kari JA, Zenker M, Kemper MJ, Mueller D, Fathy HM, Soliman NA, Hildebrandt F; SRNS Study Group: A single-gene cause in 29.5% of cases of steroid-resistant nephrotic syndrome. *J Am Soc Nephrol* 26: 1279–1289, 2015
- McCarthy HJ, Bierzynska A, Wherlock M, Ognjanovic M, Kerecuk L, Hegde S, Feather S, Gilbert RD, Krischock L, Jones C, Sinha MD, Webb NJ, Christian M, Williams MM, Marks S, Koziell A, Welsh GI, Saleem MA; RADAR the UK SRNS Study Group: Simultaneous sequencing of 24 genes associated with steroid-resistant nephrotic syndrome. *Clin J Am Soc Nephrol* 8: 637–648, 2013
- Lipska BS, Iatropoulos P, Maranta R, Caridi G, Ozaltin F, Anarat A, Balat A, Gellermann J, Trautmann A, Erdogan O, Saeed B, Emre S, Bogdanovic R, Azocar M, Balasz-Chmielewska I, Benetti E, Caliskan S, Mir S, Melk A, Ertan P, Baskin E, Jardim H, Davitaia T, Wasilewska A, Drozd D, Szczepanska M, Jankauskiene A, Higuaita LM, Ardissino G, Ozkaya O, Kuzma-Mroczkowska E, Soylemezoglu O, Ranchin B, Medynska A, Tkaczyk M, Peco-Antic A, Akil I, Jarmolinski T, Firszt-Adamczyk A, Dusek J, Simonetti GD, Gok F, Gheissari A, Emma F, Krmar RT, Fischbach M, Printza N, Simkova E, Mele C, Ghiggeri GM, Schaefer F; PodoNet Consortium: Genetic screening in adolescents with steroid-resistant nephrotic syndrome. *Kidney Int* 84: 206–213, 2013
- Giglio S, Provenzano A, Mazzinghi B, Becherucci F, Giunti L, Sansavini G, Ravaglia F, Roperto RM, Farsetti S, Benetti E, Rotondi M, Murer L, Lazzeri E, Lasagni L, Materassi M, Romagnani P: Heterogeneous genetic alterations in sporadic nephrotic syndrome associate with resistance to immunosuppression. *J Am Soc Nephrol* 26: 230–236, 2015
- Bulich G, Trujillano D, Santin S, Ossowski S, Mendizabal S, Fraga G, Madrid A, Ariceta G, Ballarin J, Torra R, Estivill X, Ars E: Targeted next-generation sequencing in steroid-resistant nephrotic syndrome: Mutations in multiple glomerular genes may influence disease severity. *Eur J Hum Genet* 23: 1192–1199, 2015
- MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, Jostins L, Habegger L, Pickrell JK, Montgomery SB, Albers CA, Zhang ZD, Conrad DF, Lunter G, Zheng H, Ayub Q, DePristo MA, Banks E, Hu M, Handsaker RE, Rosenfeld JA, Fromer M, Jin M, Mu XJ, Khurana E, Ye K, Kay M, Saunders GI, Suner MM, Hunt T, Barnes IH, Amid C, Carvalho-Silva DR, Bignell AH, Snow C, Yngvadottir B, Bumpstead S, Cooper DN, Xue Y, Romero IG, Wang J, Li Y, Gibbs RA, McCarroll SA, Dermitzakis ET, Pritchard JK, Barrett JC, Harrow J, Hurles ME, Gerstein MB, Tyler-Smith C; 1000 Genomes Project Consortium: A systematic survey of loss-of-function variants in human protein-coding genes. *Science* 335: 823–828, 2012
- Xue Y, Chen Y, Ayub Q, Huang N, Ball EV, Mort M, Phillips AD, Shaw K, Stenson PD, Cooper DN, Tyler-Smith C; 1000 Genomes Project Consortium: Deleterious- and disease-allele prevalence in healthy individuals: Insights from current predictions, mutation databases, and population-scale resequencing. *Am J Hum Genet* 91: 1022–1032, 2012
- MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, Adams DR, Altman RB, Antonarakis SE, Ashley EA, Barrett JC, Biesecker LG, Conrad DF, Cooper GM, Cox NJ, Daly MJ, Gerstein MB, Goldstein DB, Hirschhorn JN, Leal SM, Pennacchio LA, Stamatoyannopoulos JA, Sunyaev SR, Valle D, Voight BF, Winckler W, Gunter C: Guidelines for investigating causality of sequence variants in human disease. *Nature* 508: 469–476, 2014
- Tennesen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, McGee S, Do R, Liu X, Jun G, Kang HM, Jordan D, Leal SM, Gabriel S, Rieder MJ, Abecasis G, Altshuler D, Nickerson DA, Boerwinkle E, Sunyaev S, Bustamante CD, Bamshad MJ, Akey JM; Broad GO; Seattle GO; NHLBI Exome Sequencing Project: Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* 337: 64–69, 2012
- Norton N, Robertson PD, Rieder MJ, Züchner S, Rampersaud E, Martin E, Li D, Nickerson DA, Hershberger RE; National Heart, Lung and Blood Institute GO Exome Sequencing Project: Evaluating pathogenicity of rare variants from dilated cardiomyopathy in the exome era. *Circ Cardiovasc Genet* 5: 167–174, 2012
- Weeke P, Denny JC, Basterache L, Shaffer C, Bowton E, Ingram C, Darbar D, Roden DM: Examining rare and low-frequency genetic variants previously associated with lone or familial forms of atrial fibrillation in an electronic medical record system: A cautionary note. *Circ Cardiovasc Genet* 8: 58–63, 2015
- Yang RQ, Jabbari J, Cheng XS, Jabbari R, Nielsen JB, Risgaard B, Chen X, Sajadieh A, Haunsø S, Svendsen JH, Olesen MS, Tfelt-Hansen J: New population-based exome data question the pathogenicity of some genetic variants previously associated with Marfan syndrome. *BMC Genet* 15: 74, 2014
- Jabbari J, Jabbari R, Nielsen MW, Holst AG, Nielsen JB, Haunsø S, Tfelt-Hansen J, Svendsen JH, Olesen MS: New exome data question the pathogenicity of genetic variants previously associated with catecholaminergic polymorphic ventricular tachycardia. *Circ Cardiovasc Genet* 6: 481–489, 2013
- Flannick J, Beer NL, Bick AG, Agarwala V, Molnes J, Gupta N, Burt NP, Florez JC, Meigs JB, Taylor H, Lyssenko V, Irgens H, Fox E, Burslem F, Johansson S, Brosnan MJ, Trimmer JK, Newton-Cheh C, Tuomi T, Molven A, Wilson JG, O'Donnell CJ, Kathiresan S, Hirschhorn JN, Njølstad PR, Rolph T, Seidman JG, Gabriel S, Cox DR, Seidman CE, Groop L, Altshuler D: Assessing the phenotypic effects in the general population of rare variants in genes for a dominant Mendelian form of diabetes. *Nat Genet* 45: 1380–1385, 2013
- Sampson MG, Pollak MR: Opportunities and challenges of genotyping patients with nephrotic syndrome in the genomic era. *Semin Nephrol* 35: 212–221, 2015
- Gadegbeku CA, Gipson DS, Holzman LB, Ojo AO, Song PX, Barisoni L, Sampson MG, Kopp JB, Lemley KV, Nelson PJ, Lienczewski CC, Adler SG, Appel GB, Cattran DC, Choi MJ, Contreras G, Dell KM, Fervenza FC, Gibson KL, Greenbaum LA, Hernandez JD, Hewitt SM, Hingorani SR, Hladunewich M, Hogan MC, Hogan SL, Kaskel FJ, Lieske JC, Meyers KE, Nachman PH, Nast CC, Neu AM, Reich HN, Sedor JR, Sethna CB, Trachtman H, Tuttle KR, Zhdanova O, Zilleruelo GE, Kretzler M: Design of the Nephrotic Syndrome Study Network (NEPTUNE) to evaluate primary glomerular nephropathy by a multidisciplinary approach. *Kidney Int* 83: 749–756, 2013
- 1000 Genomes Project Consortium: Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA: An integrated map of genetic variation from 1,092 human genomes. *Nature* 491: 56–65, 2012
- 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR: A global reference for human genetic variation. *Nature* 526(7571): 68–74, 2015
- Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR: A method and server for

- predicting damaging missense mutations. *Nat Methods* 7: 248–249, 2010
21. Kumar P, Henikoff S, Ng PC: Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 4: 1073–1081, 2009
 22. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D: MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* 7: 575–576, 2010
 23. Sampson MG, Gillies CE, Ju W, Kretzler M, Kang HM: Gene-level integrated metric of negative selection (GIMS) prioritizes candidate genes for nephrotic syndrome. *PLoS One* 8: e81062, 2013
 24. Sampson MG, Kretzler M, Shaw AS, Kang HM: Negative selection of deleterious variants in genes associated with Mendelian forms of nephrotic syndrome. Presented at the American Society of Nephrology Kidney Week, San Diego, CA, 2012
 25. Barua M, Brown EJ, Charoonratana VT, Genovese G, Sun H, Pollak MR: Mutations in the INF2 gene account for a significant proportion of familial but not sporadic focal and segmental glomerulosclerosis. *Kidney Int* 83: 316–322, 2013
 26. Gbadegesin RA, Lavin PJ, Hall G, Bartkowiak B, Homstad A, Jiang R, Wu G, Byrd A, Lynn K, Wolfish N, Ottati C, Stevens P, Howell D, Conlon P, Winn MP: Inverted formin 2 mutations with variable expression in patients with sporadic and hereditary focal and segmental glomerulosclerosis. *Kidney Int* 81: 94–99, 2012
 27. Has C, Spartà G, Kiritsi D, Weibel L, Moeller A, Vega-Warner V, Waters A, He Y, Anikster Y, Esser P, Straub BK, Hausser I, Bockenbauer D, Dekel B, Hildebrandt F, Bruckner-Tuderman L, Laube GF: Integrin $\alpha 3$ mutations with kidney, lung, and skin disease. *N Engl J Med* 366: 1508–1514, 2012
 28. Heeringa SF, Chernin G, Chaki M, Zhou W, Sloan AJ, Ji Z, Xie LX, Salviati L, Hurd TW, Vega-Warner V, Killen PD, Raphael Y, Ashraf S, Ovunc B, Schoeb DS, McLaughlin HM, Airik R, Vlangos CN, Gbadegesin R, Hinkes B, Saisawat P, Trevisson E, Doimo M, Casarin A, Pertegato V, Giorgi G, Prokisch H, Rötig A, Nürnberg G, Becker C, Wang S, Ozaltin F, Topaloglu R, Bakkaloglu A, Bakkaloglu SA, Müller D, Beissert A, Mir S, Berdeli A, Varpizen S, Zenker M, Matejas V, Santos-Ocaña C, Navas P, Kusakabe T, Kispert A, Akman S, Soliman NA, Krick S, Mundel P, Reiser J, Nürnberg P, Clarke CF, Wiggins RC, Faul C, Hildebrandt F: COQ6 mutations in human patients produce nephrotic syndrome with sensorineural deafness. *J Clin Invest* 121: 2013–2024, 2011
 29. Hinkes B, Wiggins RC, Gbadegesin R, Vlangos CN, Seelow D, Nürnberg G, Garg P, Verma R, Chaib H, Hoskins BE, Ashraf S, Becker C, Hennies HC, Goyal M, Wharram BL, Schachter AD, Mudumana S, Drummond I, Kerjaschi D, Waldherr R, Dietrich A, Ozaltin F, Bakkaloglu A, Cleper R, Basel-Vanagaite L, Pohl M, Griebel M, Tsygin AN, Soyulu A, Müller D, Sorli CS, Bunney TD, Katan M, Liu J, Attanasio M, O'toole JF, Hasselbacher K, Mucha B, Otto EA, Airik R, Kispert A, Kelley GG, Smrcka AV, Gudermann T, Holzman LB, Nürnberg P, Hildebrandt F: Positional cloning uncovers mutations in PLCE1 responsible for a nephrotic syndrome variant that may be reversible. *Nat Genet* 38: 1397–1405, 2006
 30. Winn MP, Conlon PJ, Lynn KL, Farrington MK, Creazzo T, Hawkins AF, Daskalakis N, Kwan SY, Ebersviller S, Burchette JL, Pericak-Vance MA, Howell DN, Vance JM, Rosenberg PB: A mutation in the TRPC6 cation channel causes familial focal segmental glomerulosclerosis. *Science* 308: 1801–1804, 2005
 31. Boute N, Gribouval O, Roselli S, Benessy F, Lee H, Fuchshuber A, Dahan K, Gubler MC, Niaudet P, Antignac C: NPHS2, encoding the glomerular protein podocin, is mutated in autosomal recessive steroid-resistant nephrotic syndrome. *Nat Genet* 24: 349–354, 2000
 32. Dreyer SD, Zhou G, Baldini A, Winterpacht A, Zabel B, Cole W, Johnson RL, Lee B: Mutations in LMX1B cause abnormal skeletal patterning and renal dysplasia in nail patella syndrome. *Nat Genet* 19: 47–50, 1998
 33. Pelletier J, Bruening W, Kashtan CE, Mauer SM, Manivel JC, Striegel JE, Houghton DC, Junien C, Habib R, Fouser L, Fine RN, Silverman BL, Haber DA, Housman D: Germline mutations in the Wilms' tumor suppressor gene are associated with abnormal urogenital development in Denys-Drash syndrome. *Cell* 67: 437–447, 1991
 34. Sanna-Cherchi S, Burgess KE, Nees SN, Caridi G, Weng PL, Dagnino M, Bodria M, Carrea A, Allegretta MA, Kim HR, Perry BJ, Gigante M, Clark LN, Kisselev S, Cusi D, Gesualdo L, Allegri L, Scolari F, D'Agati V, Shapiro LS, Pecoraro C, Palomero T, Ghiggeri GM, Gharavi AG: Exome sequencing identified MYO1E and NEIL1 as candidate genes for human autosomal recessive steroid-resistant nephrotic syndrome. *Kidney Int* 80: 389–396, 2011
 35. Mele C, Iatropoulos P, Donadelli R, Calabria A, Maranta R, Cassis P, Buelli S, Tomasoni S, Piras R, Krendel M, Bettoni S, Morigi M, Delledonne M, Pecoraro C, Abbate I, Capobianchi MR, Hildebrandt F, Otto E, Schaefer F, Macciardi F, Ozaltin F, Emre S, Ibsirlioglu T, Benigni A, Remuzzi G, Noris M; PodoNet Consortium: MYO1E mutations and childhood familial focal segmental glomerulosclerosis. *N Engl J Med* 365: 295–306, 2011
 36. Brown EJ, Schlöndorff JS, Becker DJ, Tsukaguchi H, Tonna SJ, Uscinski AL, Higgs HN, Henderson JM, Pollak MR: Mutations in the formin gene INF2 cause focal segmental glomerulosclerosis. *Nat Genet* 42: 72–76, 2010
 37. Kaplan JM, Kim SH, North KN, Rennke H, Correia LA, Tong HQ, Mathis BJ, Rodríguez-Pérez JC, Allen PG, Beggs AH, Pollak MR: Mutations in ACTN4, encoding alpha-actinin-4, cause familial focal segmental glomerulosclerosis. *Nat Genet* 24: 251–256, 2000
 38. Kestilä M, Lenkkeri U, Männikkö M, Lamerdin J, McCready P, Putaala H, Ruotsalainen V, Morita T, Nissinen M, Herva R, Kashtan CE, Peltonen L, Holmberg C, Olsen A, Tryggvason K: Positionally cloned gene for a novel glomerular protein—nephrin—is mutated in congenital nephrotic syndrome. *Mol Cell* 1: 575–582, 1998
 39. López LC, Schuelke M, Quinzii CM, Kanki T, Rodenburg RJ, Naini A, Dimauro S, Hirano M: Leigh syndrome with nephropathy and CoQ10 deficiency due to decaprenyl diphosphate synthase subunit 2 (PDSS2) mutations. *Am J Hum Genet* 79: 1125–1129, 2006
 40. Zenker M, Aigner T, Wendler O, Tralau T, Müntefering H, Fenski R, Pitz S, Schumacher V, Royer-Pokora B, Wühl E, Cochat P, Bouvier R, Kraus C, Mark K, Madlon H, Dötsch J, Rascher W, Maruniak-Chudek I, Lennert T, Neumann LM, Reis A: Human laminin beta2 deficiency causes congenital nephrosis with mesangial sclerosis and distinct eye abnormalities. *Hum Mol Genet* 13: 2625–2632, 2004
 41. Boerkoel CF, Takashima H, John J, Yan J, Stankiewicz P, Rosenbarker L, André JL, Bogdanovic R, Burguet A, Cockfield S, Cordeiro I, Fründ S, Illies F, Joseph M, Kaitila I, Lama G, Loirat C, McLeod DR, Milford DV, Petty EM, Rodrigo F, Saraiva JM, Schmidt B, Smith GC, Spranger J, Stein A, Thiele H, Tizard J, Weksberg R, Lupski JR, Stockton DW: Mutant chromatin remodeling protein SMARCAL1 causes Schimke immuno-osseous dysplasia. *Nat Genet* 30: 215–220, 2002
 42. Ovunc B, Otto EA, Vega-Warner V, Saisawat P, Ashraf S, Ramaswami G, Fathy HM, Schoeb D, Chernin G, Lyons RH, Yilmaz E, Hildebrandt F: Exome sequencing reveals cubilin mutation as a single-gene cause of proteinuria. *J Am Soc Nephrol* 22: 1815–1820, 2011
 43. Ozaltin F, Ibsirlioglu T, Taskiran EZ, Baydar DE, Kaymaz F, Buyukcelik M, Kilic BD, Balat A, Iatropoulos P, Asan E, Akarsu NA, Schaefer F, Yilmaz E, Bakkaloglu A; PodoNet Consortium: Disruption of PTPRO causes childhood-onset nephrotic syndrome. *Am J Hum Genet* 89: 139–147, 2011
 44. Balreira A, Gaspar P, Caiola D, Chaves J, Beirão I, Lima JL, Azevedo JE, Miranda MC: A nonsense mutation in the LIMP-2 gene associated with progressive myoclonic epilepsy and nephrotic syndrome. *Hum Mol Genet* 17: 2238–2243, 2008
 45. Diomedi-Camassei F, Di Giandomenico S, Santorelli FM, Caridi G, Piemonte F, Montini G, Ghiggeri GM, Murer L, Barisoni L, Pastore A,

- Muda AO, Valente ML, Bertini E, Emma F: COQ2 nephropathy: A newly described inherited mitochondriopathy with primary renal involvement. *J Am Soc Nephrol* 18: 2773–2780, 2007
46. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadzisa A, Searle S, Barnes I, Bignell A, Boychenko V, Hunt T, Kay M, Mukherjee G, Rajan J, Despacio-Reyes G, Saunders G, Steward C, Harte R, Lin M, Howald C, Tanzer A, Derrien T, Chrast J, Walters N, Balasubramanian S, Pei B, Tress M, Rodriguez JM, Ezkurdia I, van Baren J, Brent M, Haussler D, Kellis M, Valencia A, Reymond A, Gerstein M, Guigó R, Hubbard TJ: GENCODE: The reference human genome annotation for The ENCODE Project. *Genome Res* 22: 1760–1774, 2012
 47. Blekhman R, Man O, Herrmann L, Boyko AR, Indap A, Kosiol C, Bustamante CD, Teshima KM, Przeworski M: Natural selection on genes that underlie human disease susceptibility. *Curr Biol* 18: 883–889, 2008
 48. Al-Hamed MH, Al-Sabban E, Al-Mojalli H, Al-Harbi N, Faqeih E, Al Shaya H, Alhasan K, Al-Hissi S, Rajab M, Edwards N, Al-Abbad A, Al-Hassoun I, Sayer JA, Meyer BF: A molecular genetic analysis of childhood nephrotic syndrome in a cohort of Saudi Arabian families. *J Hum Genet* 58: 480–489, 2013
 49. Hinkes BG, Mucha B, Vlangos CN, Gbadegesin R, Liu J, Hasselbacher K, Hangan D, Ozaltin F, Zenker M, Hildebrandt F; Arbeitsgemeinschaft für Paediatrische Nephrologie Study Group: Nephrotic syndrome in the first year of life: Two thirds of cases are caused by mutations in 4 genes (NPHS1, NPHS2, WT1, and LAMB2). *Pediatrics* 119: e907–e919, 2007
 50. Begg CB, Haile RW, Borg A, Malone KE, Concannon P, Thomas DC, Langholz B, Bernstein L, Olsen JH, Lynch CF, Anton-Culver H, Capanu M, Liang X, Hummer AJ, Sима C, Bernstein JL: Variation of breast cancer risk among BRCA1/2 carriers. *JAMA* 299: 194–201, 2008
 51. Beutler E, Felitti VJ, Koziol JA, Ho NJ, Gelbart T: Penetrance of 845G→A (C282Y) HFE hereditary haemochromatosis mutation in the USA. *Lancet* 359: 211–218, 2002
 52. Santín S, Bullich G, Tazón-Vega B, García-Maset R, Giménez I, Silva I, Ruiz P, Ballarín J, Torra R, Ars E: Clinical utility of genetic testing in children and adults with steroid-resistant nephrotic syndrome. *Clin J Am Soc Nephrol* 6: 1139–1148, 2011
 53. Lovric S, Fang H, Vega-Warner V, Sadowski CE, Gee HY, Halbritter J, Ashraf S, Saisawat P, Soliman NA, Kari JA, Otto EA, Hildebrandt F; Nephrotic Syndrome Study Group: Rapid detection of monogenic causes of childhood-onset steroid-resistant nephrotic syndrome. *Clin J Am Soc Nephrol* 9: 1109–1116, 2014
 54. Heeringa SF, Vlangos CN, Chernin G, Hinkes B, Gbadegesin R, Liu J, Hoskins BE, Ozaltin F, Hildebrandt F; Members of the APN Study Group: Thirteen novel NPHS1 mutations in a large cohort of children with congenital nephrotic syndrome. *Nephrol Dial Transplant* 23: 3527–3533, 2008
 55. Ovunc B, Ashraf S, Vega-Warner V, Bockenbauer D, Elshakhs NA, Joseph M, Hildebrandt F; Gesellschaft für Pädiatrische Nephrologie (GPN) Study Group: Mutation analysis of NPHS1 in a worldwide cohort of congenital nephrotic syndrome patients. *Nephron Clin Pract* 120: c139–c146, 2012
 56. Wong W, Morris MC, Kara T: Congenital nephrotic syndrome with prolonged renal survival without renal replacement therapy. *Pediatr Nephrol* 28: 2313–2321, 2013
 57. Matejas V, Hinkes B, Alkandari F, Al-Gazali L, Annexstad E, Aytac MB, Barrow M, Bláhová K, Bockenbauer D, Cheong HI, Maruniak-Chudek I, Cochat P, Dötsch J, Gajjar P, Hennekam RC, Janssen F, Kagan M, Kariminejad A, Kemper MJ, Koenig J, Kogan J, Kroes HY, Kuwertz-Bröking E, Lewanda AF, Medeira A, Muscheites J, Niaudet P, Pierson M, Saggat A, Seaver L, Suri M, Tsygin A, Wühl E, Zurowska A, Uebe S, Hildebrandt F, Antignac C, Zenker M: Mutations in the human laminin beta2 (LAMB2) gene and the associated phenotypic spectrum. *Hum Mutat* 31: 992–1002, 2010
 58. Benoit G, Machuca E, Antignac C: Hereditary nephrotic syndrome: A systematic approach for genetic testing and a review of associated podocyte gene mutations. *Pediatr Nephrol* 25: 1621–1632, 2010
 59. Büscher AK, Kranz B, Büscher R, Hildebrandt F, Dworniczak B, Pennekamp P, Kuwertz-Bröking E, Wingen AM, John U, Kemper M, Monnens L, Hoyer PF, Weber S, Konrad M: Immunosuppression and renal outcome in congenital and pediatric steroid-resistant nephrotic syndrome. *Clin J Am Soc Nephrol* 5: 2075–2084, 2010
 60. Gbadegesin R, Hinkes B, Vlangos C, Mucha B, Liu J, Hopcian J, Hildebrandt F: Mutational analysis of NPHS2 and WT1 in frequently relapsing and steroid-dependent nephrotic syndrome. *Pediatr Nephrol* 22: 509–513, 2007
 61. Löwik M, Levtschenko E, Westra D, Groenen P, Steenberg E, Weening J, Lilien M, Monnens L, van den Heuvel L: Bigenic heterozygosity and the development of steroid-resistant focal segmental glomerulosclerosis. *Nephrol Dial Transplant* 23: 3146–3151, 2008
 62. Koziell A, Grech V, Hussain S, Lee G, Lenkkeri U, Tryggvason K, Scambler P: Genotype/phenotype correlations of NPHS1 and NPHS2 mutations in nephrotic syndrome advocate a functional inter-relationship in glomerular filtration. *Hum Mol Genet* 11: 379–388, 2002
 63. Gbadegesin RA, Adeyemo A, Webb NJ, Greenbaum LA, Abeyagunawardena A, Thalagahagoda S, Kale A, Gipson D, Srivastava T, Lin JJ, Chand D, Hunley TE, Brophy PD, Bagga A, Sinha A, Rheault MN, Ghali J, Nicholls K, Abraham E, Janjua HS, Omolaja A, Barletta GM, Cai Y, Milford DD, O'Brien C, Awan A, Belostotsky V, Smoyer WE, Homstad A, Hall G, Wu G, Nagaraj S, Wigfall D, Foreman J, Winn MP; Mid-West Pediatric Nephrology Consortium: HLA-DQA1 and PLCG2 are candidate risk loci for childhood-onset steroid-sensitive nephrotic syndrome. *J Am Soc Nephrol* 26: 1701–1710, 2015
 64. Stanescu HC, Arcos-Burgos M, Medlar A, Bockenbauer D, Kottgen A, Dragomirescu L, Voinescu C, Patel N, Pearce K, Hubank M, Stephens HA, Laundy V, Padmanabhan S, Zawadzka A, Hofstra JM, Coenen MJ, den Heijer M, Kiemeny LA, Bacq-Daia D, Stengel B, Powis SH, Brenchley P, Feehally J, Rees AJ, Debiec H, Wetzels JF, Ronco P, Mathieson PW, Kleta R: Risk HLA-DQA1 and PLA(2)R1 alleles in idiopathic membranous nephropathy. *N Engl J Med* 364: 616–626, 2011
 65. Gharavi AG, Kiryluk K, Choi M, Li Y, Hou P, Xie J, Sanna-Cherchi S, Men CJ, Julian BA, Wyatt RJ, Novak J, He JC, Wang H, Lv J, Zhu L, Wang W, Wang Z, Yasuno K, Gunel M, Mane S, Umlauf S, Tikhonova I, Beerman I, Savoldi S, Magistroni R, Ghiggeri GM, Bodria M, Lugani F, Ravani P, Ponticelli C, Allegri L, Boscutti G, Frasca G, Amore A, Peruzzi L, Coppo R, Izzi C, Viola BF, Prati E, Salvadori M, Mignani R, Gesualdo L, Bertinetto F, Mesiano P, Amoroso A, Scolari F, Chen N, Zhang H, Lifton RP: Genome-wide association study identifies susceptibility loci for IgA nephropathy. *Nat Genet* 43: 321–327, 2011
 66. Genovese G, Friedman DJ, Ross MD, Lecordier L, Uzureau P, Freedman BI, Bowden DW, Langefeld CD, Oleksyk TK, Uscinski Knob AL, Bernhardt AJ, Hicks PJ, Nelson GW, Vanhollebeke B, Winkler CA, Kopp JB, Pays E, Pollak MR: Association of trypanolytic ApoL1 variants with kidney disease in African Americans. *Science* 329: 841–845, 2010
 67. Gillies CE, Robertson CC, Sampson MG, Kang H: GeneVetter: A web tool for quantitative monogenic assessment of rare diseases [published online ahead of print July 23, 2015]. *Bioinformatics* doi:bvt432
 68. Halbritter J, Porath JD, Diaz KA, Braun DA, Kohl S, Chaki M, Allen SJ, Soliman NA, Hildebrandt F, Otto EA; GPN Study Group: Identification of 99 novel mutations in a worldwide cohort of 1,056 patients with a nephronophthisis-related ciliopathy. *Hum Genet* 132: 865–884, 2013
 69. Halbritter J, Diaz K, Chaki M, Porath JD, Tarrier B, Fu C, Innis JL, Allen SJ, Lyons RH, Stefanidis CJ, Omran H, Soliman NA, Otto EA: High-throughput mutation analysis in patients with a nephronophthisis-associated ciliopathy applying multiplexed barcoded array-based PCR amplification and next-generation sequencing. *J Med Genet* 49: 756–767, 2012

70. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA: The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20: 1297–1303, 2010
71. Martin M: Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* 17: 10–12, 2011
72. Li H, Durbin R: Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–1760, 2009
73. Cortes C, Vapnik V: Support-vector networks. *Mach Learn* 20: 273–297, 1995
74. Chang CC, Lin CJ: LIBSVM: A library for support vector machines. *ACM Trans Intell Syst Technol* 2: 1–27, 2011
75. Cingolani P, Platts A, Wang L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM: A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 6: 80–92, 2012
76. Liu X, Jian X, Boerwinkle E: dbNSFP v2.0: A database of human non-synonymous SNVs and their functional predictions and annotations. *Hum Mutat* 34: E2393–E2402, 2013
77. Stenson PD, Mort M, Ball EV, Shaw K, Phillips A, Cooper DN: The Human Gene Mutation Database: Building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet* 133: 1–9, 2014
78. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S: Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLOS Comput Biol* 6: e1001025, 2010

This article contains supplemental material online at <http://jasn.asnjournals.org/lookup/suppl/doi:10.1681/ASN.2015050504/-/DCSupplemental>.