Exome sequencing identified *MYO1E* and *NEIL1* as candidate genes for human autosomal recessive steroid-resistant nephrotic syndrome

Simone Sanna-Cherchi[1], Katelyn E. Burgess[1,16], Shannon N. Nees[1,16], Gianluca Caridi[2], Patricia L. Weng[1,3], Monica Dagnino[2], Monica Bodria[2], Alba Carrea[2], Maddalena A. Allegretta[4,5,6,7], Hyunjae R. Kim[8], Brittany J. Perry[1], Maddalena Gigante[9], Lorraine N. Clark[10], Sergey Kisselev[10], Daniele Cusi[11], Loreto Gesualdo[9], Landino Allegri[12], Francesco Scolari[13], Vivette D'Agati[7], Lawrence S. Shapiro[14], Carmine Pecoraro[15], Teresa Palomero[4,5,6,7], Gian M. Ghiggeri[2] and Ali G. Gharavi[1]

[1]Division of Nephrology, Columbia University College of Physicians and Surgeons, New York, New York, USA; [2]Division of Pediatric Nephrology, G Gaslini Institute, Genoa, Italy; [3]Division of Pediatric Nephrology, Columbia University, New York, New York, USA; [4]NextGen Sequencing Core, Sulzberger Columbia Genome Center, Columbia University, New York, New York, USA; [5]Herbert Irving Comprehensive Cancer Center, Columbia University, New York, New York, USA; [6]Institute for Cancer Genetics, Columbia University, New York, New York, USA; [7]Department of Pathology, Columbia University, New York, New York, USA; [8]Columbia Initiative in Systems Biology, Columbia University, New York, New York, USA; [9]Department of Biomedical Sciences, University of Foggia, Foggia, Italy; [10]Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University, New York, New York, USA; [11]Division of Nephrology, Department of Medicine, Surgery and Dentistry, Chair and Graduate School of Nephrology, University of Milan, San Paolo Hospital Milan and Fondazione Filarete, Milan, Italy; [12]Department of Clinical Medicine, Nephrology and Health Sciences, Section of Nephrology, University of Parma, Parma, Italy; [13]Division of Nephrology, Hospital of Montichiari, Montichiari, Italy; [14]Departments of Biochemistry and Molecular Biophysics and Ophthalmology, Columbia University, New York, New York, USA and [15]Unit of Nephrology and Dialysis, Children's Hospital Santobono, Naples, Italy

To identify gene loci associated with steroid-resistant nephrotic syndrome (SRNS), we utilized homozygosity mapping and exome sequencing in a consanguineous pedigree with three affected siblings. High-density genotyping identified three segments of homozygosity spanning 33.6 Mb on chromosomes 5, 10, and 15 containing 296 candidate genes. Exome sequencing identified two homozygous missense variants within the chromosome 15 segment; an A159P substitution in myosin 1E (MYO1E), encoding a podocyte cytoskeletal protein; and an E181K substitution in nei endonuclease VIII-like 1 (NEIL1), encoding a base-excision DNA repair enzyme. Both variants disrupt highly conserved protein sequences and were absent in public databases, 247 healthy controls, and 286 patients with nephrotic syndrome. The MYO1E A159P variant is noteworthy, as it is expected to impair ligand binding and actin interaction in the MYO1E motor domain. The predicted loss of function is consistent with the previous demonstration that *Myo1e* inactivation produces nephrotic syndrome in mice. Screening 71 additional patients with SRNS, however, did not identify independent *NEIL1* or *MYO1E* mutations, suggesting larger sequencing efforts are needed to uncover which mutation is responsible for the phenotype. Our findings demonstrate the utility of exome sequencing for rapidly identifying candidate genes for human SRNS.

Kidney International (2011) **80,** 389–396; doi:10.1038/ki.2011.148; published online 22 June 2011

KEYWORDS: homozygosity mapping; nephrotic syndrome; next-generation sequencing

Correspondence: *Simone Sanna-Cherchi* or *Ali G. Gharavi, Division of Nephrology, Columbia University College of Physicians and Surgeons, 1150 Street Nicholas Avenue, Russ Berrie Pavilion #412, New York, New York 10032, USA. E-mail: ss2517@columbia.edu* or *ag2239@columbia.edu*

[16]These authors contributed equally to this paper.

Received 6 November 2010; revised 20 April 2011; accepted 26 April 2011; published online 22 June 2011

Nephrotic syndrome, characterized by the triad of heavy proteinuria, hypoalbuminemia, and edema, can lead to progressive damage of kidney glomeruli and end-stage renal disease. Idiopathic nephrotic syndrome has a prevalence of about 16/100,000 children.[1,2] Although the majority of cases (~80%) are steroid responsive and have no long-term renal sequelae,[3,4] steroid-resistant nephrotic syndrome (SRNS) represents one of the most common causes of pediatric end-stage renal disease. The typical histological lesion in SRNS is focal segmental glomerulosclerosis (FSGS). Positional cloning studies have identified several genes mutated in familial, non-syndromic forms of FSGS (*NPHS1*, *NPHS2*, *PLCE1*, *WT1*, *ACTN4*, *TRPC6*, and *INF2*).[5–10] All of these genes encode proteins that are highly expressed in glomerular

podocytes, the terminally differentiated visceral epithelial cells whose highly organized actin–myosin cytoskeleton provides structural support against pulsatile forces at the glomerular filtration barrier. Podocytes also possess elaborate foot processes that contribute to formation of the slit diaphragm, the kidney filter. Although mutations in *NPHS1* and *NPHS2* are a common cause of childhood-onset, autosomal recessive FSGS, mutations in other genes are very rare. Thus, the genetic cause of the vast majority ($>80\%$) of childhood-onset SRNS is still not known,[11–15] indicating that there are more, yet undiscovered, genes responsible for this trait. The ability to identify SRNS by molecular diagnosis has important clinical implications, as this would prevent unnecessary administration of corticosteroids and direct physicians to alternative therapies.

Identification of genes for nephrotic syndrome is complicated by genetic heterogeneity and incomplete penetrance, which limits availability of large pedigrees for traditional positional cloning approaches. However, the recent advent of exome capture, followed by massive parallel sequencing, enables a rapid genome-wide search for rare pathogenic mutations. This approach has successfully identified new genes for rare monogenic diseases for which informative pedigrees were not available and can also provide molecular identification of Mendelian disorders when the clinical diagnosis is uncertain.[16–18] Here, we identified novel candidate genes for human autosomal recessive SRNS by genome-wide homozygosity mapping, followed by exome sequencing.

RESULTS

Patients and families

We established a cohort of Caucasian pediatric patients with nephrotic syndrome, comprising 234 steroid-resistant and 72 steroid-sensitive patients. This cohort includes 20 patients from nine families (Figure 1a and Supplementary Figure S1 online). Mutations in *NPHS2* and *WT1* were excluded by Sanger sequencing in the all nine families and in 219 sporadic patients. In this cohort, one Italian pedigree (K3016, Figure 1a) was ideally suited for genetic studies, as it consisted of three siblings diagnosed with SRNS, born to a consanguineous union between healthy parents, which is strongly indicative of a simple autosomal recessive inheritance (Figure 1a). All three affected individuals had kidney biopsy–documented FSGS that was unresponsive to steroids, with the age of onset ranging between 3 and 10 years. The maximal expected logarithm of odds (LOD) score under a simple autosomal recessive model was 2.7, indicating high power for chromosomal localization of a new disease-causing gene.

Linkage analysis and homozygosity mapping localize a novel gene to 33.6 Mb

A genome-wide analysis of linkage with 393 microsatellite markers (Marshfield Mammalian Genotyping Service, performed in 2006) resulted in three suggestive signals on chromosomes 5 (D5S1725; LOD score = 1.7), 10 (D10S1419; LOD score = 1.4), and 15 (D15S643; LOD score = 2.4) under an autosomal recessive model (Allegro 2.0 statistical package;[18] disease gene frequency = 0.0001 and phenocopy rate = 0.0001). Follow-up studies with more novel high-density genotyping methods (Illumina 610-Quad chip; Illumina, San Diego, CA) and haplotype analysis in the three affected patients precisely localized the segments of homozygosity, defining shared intervals spanning 2.98 Mb on chromosome 5, 0.26 Mb on chromosome 10, and 30.32 Mb on chromosome 15 (Figure 1b). In addition, genome-wide linkage analysis was repeated using a pruned set of 24,441 highly informative single-nucleotide polymorphisms (SNPs) but did not identify additional shared homozygous segments. Thus, altogether, high-density homozygosity mapping localized the causal variant(s) to three regions totaling 33.56 Mb of sequence ($\sim 1\%$ of the human genome), containing 296 positional candidates. The chromosome 15 interval was most likely to harbor the *SRNS* gene(s) as it spanned a large segment of shared homozygosity and reached the maximum expected LOD score of 2.7 between rs9788730 and rs3813573.

Exome sequencing identifies homozygous missense variants in the *MYO1E* and *NEIL1* genes

Given the large number of positional candidates, we performed a comprehensive search for pathogenic mutations by whole-exome sequencing in the index case (Figure 1a). After capturing of 50 Mb of human DNA sequence on an Agilent array (Agilent, Santa Clara, CA), massive parallel sequencing was conducted on a SOLiD 4 instrument (Applied Biosystems, Carlsbad, CA). Nearly 95% of the SureSelect regions were sequenced at an average of $65\times$ depth, and 84% of the bases were covered at $10\times$ depth or more. Analysis restricted to the regions of homozygosity demonstrated that $\sim 96.2\%$ of the exons in these regions were adequately captured or sequenced, indicating near-complete coverage (the 135 exons on chromosome 15 and 3 exons on chromosome 5 that were not captured are listed in Supplementary Table S1 online). This coverage is comparable to published exome-sequencing experiments.[16–18] We next applied a Bayesian SNP calling to evaluate the previous probability of the existence of a heterozygote or a non-reference homozygote at the position.

A Bayesian SNP calling with an error correction algorithm (SOLiD Accuracy Enhancer Tool (SAET), Carlsbad, CA) led to the identification of 14,262 single-nucleotide variants in the sample. After filtering these against the Single Nucleotide Polymorphism Database and the 1000-Genomes Project, we identified 665 unique variants. For quality control, we examined SNP calls within the regions of homozygosity defined by high-density genotyping and identified only four heterozygous calls, indicating a low false-positive rate. There were no homozygous truncating mutations or splice-site variants across the genome. By restricting the analysis to the shared regions of homozygosity on chromosomes 5, 10, and 15, we identified only four missense homozygous variants, all located on chromosome 15: *MYO1E*, p.A159P; *NEIL1*,

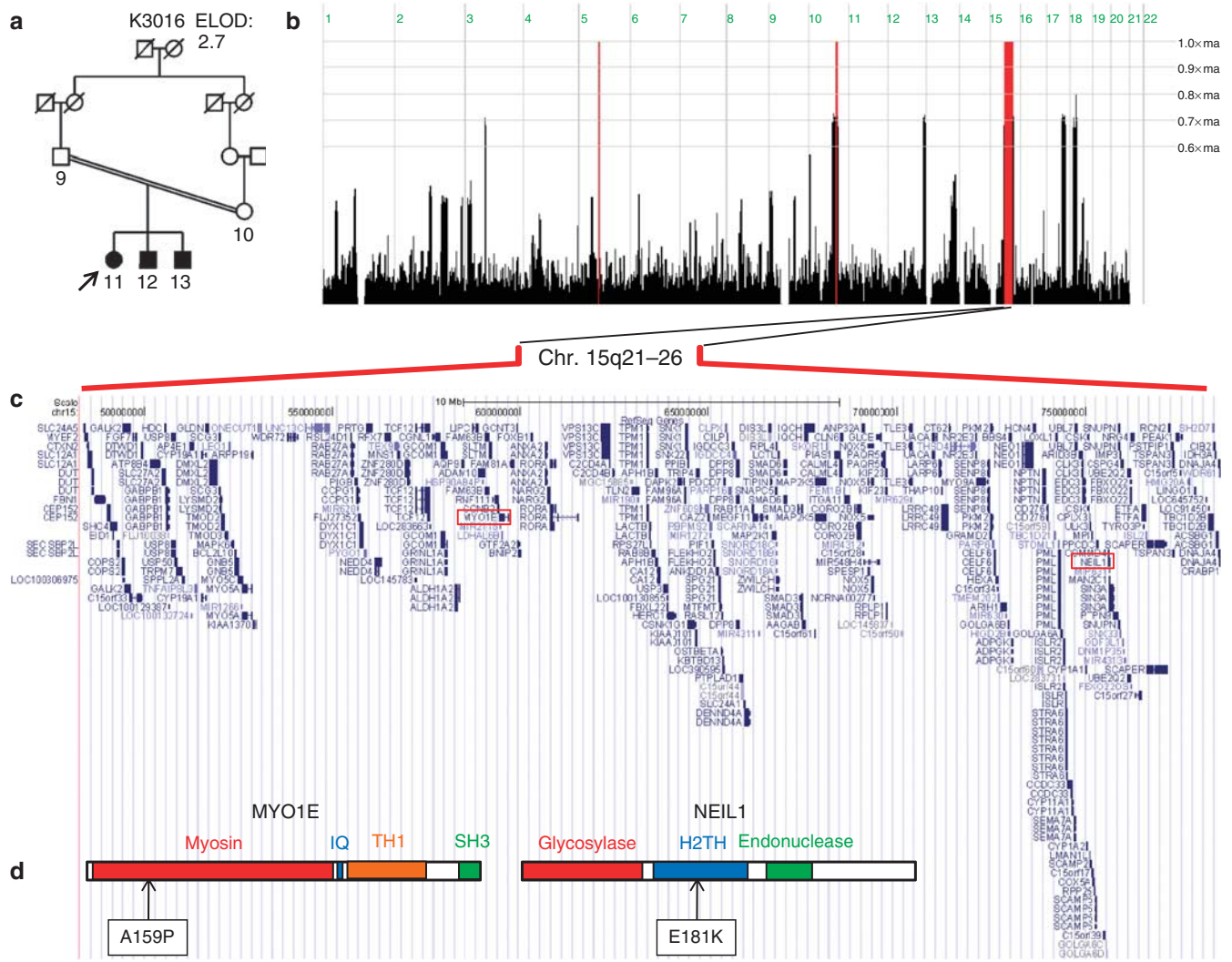
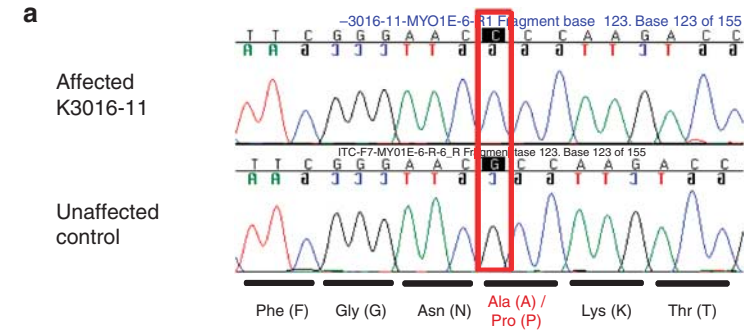
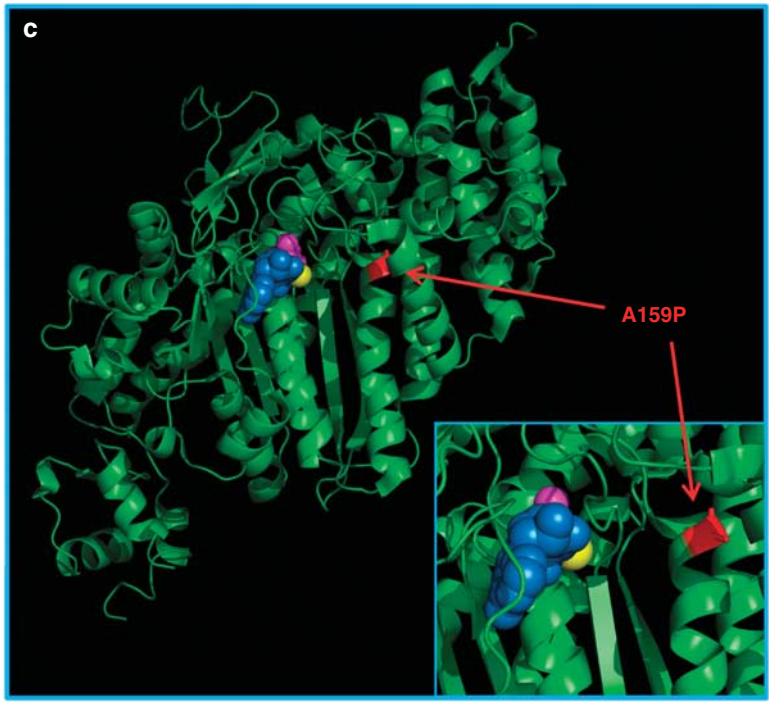


Figure 1 | Mapping a gene for steroid-resistant nephrotic syndrome to Chr. 15q21–26. (**a**) Pedigree structure of family K3016. Individuals 10, 11, 12, and 13 were genotyped for linkage with 393 microsatellite markers across the genome (Marshfield Mammalian Genotyping Service). The three affected siblings were genotyped with high-density single-nucleotide polymorphism arrays (Illumina 610-Quad) for linkage and homozygosity mapping. Individual K3016-11 was subjected to whole-exome sequencing. (**b**) Genome-wide homozygosity plot. In red are indicated the three regions of significant homozygosity in which the haplotypes were identical by descent. (**c**) University of California, Santa Cruz map of the 30.32-Mb segment of homozygosity on chromosome 15 containing 263 positional candidates; *MYO1E* and *NEIL1* are highlighted. (**d**) Graphical representation of the MYO1E and NEIL1 proteins, with the location of the MYO1E A159P and NEIL1 E181K substitutions shown by arrows. The functional domains of each protein are also indicated. chr., chromosome; endonuclease, endonuclease VIII-like 1, DNA-binding domain; glycosylase, N-terminal domain of metazoan Nei-like glycosylase 1; H2TH, formamidopyrimidine–DNA glycosylase helix 2 turn helix domain; IQ, IQ motif; myosin; myosin, motor domain; SH3, Src homology-3 domain; TH1, C-terminal tail homology domain. ma = maximum homozygosity scores (scale 0–1). Scores higher than 0.8 are highlighted in red.

p.E181K; *GOLGA6*, p.R149Q; and *TBD2B*, p.V284L. Sanger sequencing indicated that the *GOLGA6* and *TBD2B* variants were false positives, but the *MYO1E* p.A159P and *NEIL1* p.E181K were present in homozygosity in all three affected children and in heterozygous state in both parents (Figures 2 and 3). Analysis with PolyPhen-2 indicated that the *MYO1E* p.A159P variant was most likely deleterious (score = 0.998) and the *NEIL1* p.E181K (score = 0.836) was possibly damaging. However, *MYO1E* was a highly compelling candidate because a recent study showed that *Myo1e* is expressed in glomerular podocytes and its inactivation in mice results in heavy proteinuria due to podocyte damage, recapitulating human nephrotic syndrome,[19] whereas inactivation of murine *Neil1* on two genetic backgrounds produces only obesity and metabolic defects.[20,21] Nevertheless, both variants were absent in all public databases, 247 ethnically matched healthy controls, and 286 independent pediatric patients with steroid-resistant and -sensitive nephrotic syndrome. We also performed Sanger sequencing in all 28 exons of *MYO1E* and 9 exons of *NEIL1* in affected individuals in K3016 but found no other coding variants.

We next sequenced all *MYO1E* and *NEIL1* exons in the index cases from 8 independent nephrotic syndrome families (Supplementary Figure S1 online) and 63 sporadic cases with

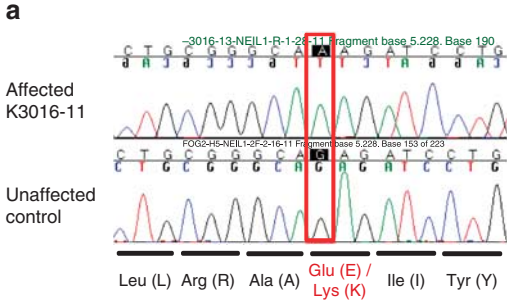




b

Species	Myosin type	Protein sequence
H. sapiens	Myosin 1E	LLEAFGN A KTVRNNN
P. troglodytes	Myosin 1E	LLEAFGN A KTVRNNN
C. familiaris	Myosin 1E	LLEAFGN A KTVRNNN
B. taurus	Myosin 1E	LLEAFGN A KTVRNNN
M. musculus	Myosin 1E	LLEAFGN A KTVRNNN
R. norvegicus	Myosin 1E	LLEAFGN A KTVRNNN
G. gallus	Myosin 1E	LLEAFGN A KTVRNNN
D. rerio	Myosin 1E	LLEAFGN A KTVRNNN
T. nigroviridis	Myosin 1E	LLEAFGN A KTLRNNN
X. laevis	Myosin 1E	LLEAFGN A KTVRNNN
D. discoideum	Myosin 1E	LLEAFGN A KTLRNDN
P. troglodytes	Myosin 1F	LLEAFGN A KTVRNNN
M. mulatia	Myosin 1F	LLEAFGN A KTVRNNN
C. familiaris	Myosin 1F	LLEAFGN A KTVRNNN
M. musculus	Myosin 1F	LLEAFGN A KTVRNNN
M. domestica	Myosin 1F	LLEAFGN A KTVRNNN
G. gallus	Myosin 1F	LLEAFGN A KTVRNNN
X. tropicalis	Myosin 1F	LLEAFGN A KTVRNNN
A. castellani	Myosin 1A	VLEAFGN A KTLRNDN
D. discoideum	Myosin 1A	LLEAFGN A KTLRNNN
A. clavatus	Myosin 1A	LLESFGN A KTLRNNN
C. familiaris	Myosin 1B	VLEAFGN A KTVRNDN
D. discoideum	Myosin 1B	LLEAFGN A KTLRNNN
A. castellani	Myosin 1B	LLEAFGN A KTLRNNN
E. histolytica	Myosin 1B	LLEAFGN A KTLRNNN
T. cruzi	Myosin 1B	LLEAFGN A KTVRNDN
T. brucei	Myosin 1B	VLEAFGN A KTVRNDN
M. domestica	Myosin 1C	VLEAFGN A KTLRNDN
M. mulatta	Myosin 1C	VLEAFGN A KTLRNDN
A. castellani	Myosin 1C	LLEAFGN A KTIRNNN
D. discoideum	Myosin 1C	LLEAFGN A KTLRNNN
D. discoideum	Myosin 1D	LLEAFGN A KTLRNNN
C. immitis	Myosin 1	LLESFGN A KTLRNNN
A. fumigatus	Myosin 1	LLESFGN A KTLRNNN
A. terreus	Myosin 1	LLESFGN A KTLRNNN
C. albicans	Myosin 1	LLESFGC A KTLRNNN
S. pombe	Myosin 1	LLESFGC A KTLRNNN
S. cerevisiae	Myosin 5P	LLESFGC A KTLRNNN
P. stiptis	Myosin 5P	LLESFGC A KTLRNNN
D. melanogaster	Myosin 61F	VLEAFGN A KTNRNDN
S. cerevisiae	Myosin 3	LLESFGC A KTLRNNN

Figure 2 | Description of the MYO1E A159 substitution. (**a**) Chromatogram of the homozygous c.475 G>C variant resulting in the MYO1E alanine-to-proline substitution at position 159 in K3016. The sequence variant is boxed in red. (**b**) Multiple protein sequence alignment shows conservation of the alanine residue at position 159 in the myosin-1E protein, its homologs, and orthologs. (**c**) Ribbon diagram of the crystal structure of the motor domain from MYOE from *Dictyostelium discoideum* (PDB ID 1LKX), the most highly related MYO1E protein of known structure. The position of the proline mutation at A159 (A149 in *D. discoideum*) is colored in red; ADP is shown in blue, Mg^{2+} in yellow, and a bound vanadate ion is shown in magenta. The mutation is positioned at the C-terminal end of helix 7, preceding a loop that is disordered in the structure. This residue is positioned adjacent to the active site and is likely to interfere with ligand binding or catalysis.[32] Ala, alanine; Asn, asparagine phenylalanine; Gly, glycine; Lys, lysine; Phe, phenylalanine; Pro, proline; Thr, threonine.



b

Species	Protein sequence (p.E181K)
H. sapiens	IGNYLRA E ILYRLKI
P. troglodytes	IGNYLRA E ILYRLKI
C. familiaris	IGNYLRA E ILHRLGI
B. taurus	IGNYLRA E ILYRLRI
M. musculus	IGNYLRA E ILYRLKI
R. norvegicus	IGNYLRA E ILYRLKI
M. domestica	IGNYLRA E ILYRLKI
G. gallus	IGNYLRA E ILYRAKI
D. rerio	IGNYLRA E ILFRLQI
X. laevis	IGNYLRA E ILFRLEI

Figure 3 | Description of the *NEIL1* E181K substitution. (**a**) Chromatogram of the homozygous G>A variant resulting in the *NEIL1* E181K substitution in K3016. The sequence variant is boxed in red. (**b**) Multiple protein sequence alignment shows conservation of the glutamic acid residue (E) at position 181 in the NEIL1 protein and in its homologs and orthologs. Ala, alanine; Arg, arginine; Glu, glutamic acid; Ile, isoleucine; Leu, leucine; Lys, lysine; Tyr, tyrosine.

onset of SRNS between 1 and 14 years, and in whom mutations in *NPHS2* and *WT1* gene had been excluded.[12,22] We identified a few rare heterozygous variants (Supplementary Table S2 online), but none of the patients harbored two potentially pathogenic variants for either *MYO1E* or *NEIL1* genes. One affected sib from family K3031 carried two potentially deleterious variants in the *MYO1E* gene (D185G and P1049H), but segregation analysis indicated that both variants segregated in *cis* from the maternal side. Moreover, the affected sib did not inherit this haplotype. Of note, we identified a *NEIL1* variant affecting a canonical splice donor site in intron 1 (c.434 + 2T > C, identified in three cases), which is predicted to result in premature termination signal; however, this variant is present in public databases and in controls (rs5745908, frequency = 0.013 among Europeans), indicating that potentially disruptive mutations in *NEIL1* may be tolerated.

As inactivation of *Myo1e* gene causes nephrotic syndrome in mice, we further analyzed the structure of MYO1E to infer the potential functional consequence of the A159P substitution. Position A159 is located within the myosin motor domain, which mediates ligand binding and interactions with actin, with alanine at this site conserved among virtually all MYO1E orthologs and paralogs (Figure 2b). The crystal structure of the highly conserved MYO1E from *Dictyostelium*[23] shows that the A149 residue, which is equivalent to A159 in the human protein, is located at the C terminus of helix 7, directly adjacent to the switch 1 domain involved in nucleotide binding (Figure 2c). Mutations in this region in MYO1C and myosins II and V affect nucleotide binding or ATP hydrolysis;[24–30] thus, substitution with the larger proline side chain at position 159 is highly likely to interfere with ligand binding or catalysis. Taken together, these data strongly indicate that the A159P substitution is a loss of function mutation.

DISCUSSION

In this study, we used homozygosity mapping and exome sequencing to identify *MYO1E* and *NEIL1* as novel candidate genes for human SRNS. The *MYO1E* and *NEIL1* mutations localized to the only large segment of homozygosity in a consanguineous pedigree; these mutations were absent in 1066 independent chromosomes surveyed by Sanger sequencing; there were no other homozygous deleterious coding mutations elsewhere in the genome, and both variants in MYO1E and NEIL1 affect amino-acid residues that are highly conserved across homologs and orthologs and affect functional domains of the proteins. We screened all exons of *MYO1E* and *NEIL1* genes in 71 additional cases but did not identify independent mutations in either gene, indicating that *MYO1E* and/or *NEIL1* mutations represent a relatively rare cause of recessive SRNS in humans. These findings motivate larger screening efforts to identify the contribution of these genes to nephrotic syndrome.

MYO1E is a ubiquitously expressed member of class I myosins, which are small, monomeric molecules that interact directly with cell membranes via their C-terminal TH1 domain and with actin filaments via their N-terminal motor domain, enabling control of plasma tension and mechanical strain, dynamic gating of ion channels, microvillar vesicle shedding, and endo/exocytosis.[31,32] Cells with specialized cytoskeletal organization, such as glomerular podocytes, auditory hair cells, and enterocytes, are particularly enriched in class I myosins, which enable them to regulate their specialized membranes (for example, podocyte pedicels). Heterozygous mutations in the *MYO1A* gene produce a defect in hair cells, resulting in autosomal dominant hearing loss in humans.[33] Most relevant to the present work, a recent study has shown that *Myo1e* is highly expressed in glomerular podocytes and *Myo1e* null mice develop selective podocyte injury, with foot process effacement, severe proteinuria, and progressive glomerulosclerosis,[19] highlighting the importance of the actin–myosin cytoskeletal integrity for maintenance of the kidney filtration barrier.

NEIL1 gene encodes a ubiquitously expressed DNA glycosylase involved in the base-excision repair of mitochondrial and nuclear DNA, which is important for protection against reactive oxygen species.[34] Inactivation of *Neil1* in mice results in late-onset obesity, extensive in visceral fat deposition, and biochemical features of metabolic syndrome.[20,21] This is accompanied by mitochondrial DNA depletion and impaired oxidative phosphorylation in hepatocytes.[20,21] The renal phenotype of *Neil1* null mice has not been studied in detail, but available data indicate proximal tubule vacuolization without detectable fat deposition.[21] Interestingly, we recently demonstrated that mutations in the DNA repair gene *Prkdc* produce susceptibility to adriamycin nephropathy in mice by impairing mitochondrial DNA maintenance in podocytes.[35] This observation provides a potential mechanism that could link mutations in the *NEIL1* gene to podocyte injury and SRNS.

Although the genetic evidence points to both *NEIL1* and *MYO1E* as the disease gene, and digenic disease cannot be formally excluded in K3016, consideration of the literature favors the MYO1E A159P substitution as the most likely culprit mutation. Mutations in the nucleotide-binding pocket in the myosin superfamily (for example, *MYO1C*, *MYOII*, *MYOV*, and *MYH7*) are associated with many clinical phenotypes, such as sensorineural hearing loss or cardiomyopathy.[24–30] For example, a recently described R156W mutation in the MYO1C switch 1 domain corresponding to R163 in MYO1E, in close proximity to the A159P substitution presently described, severely impairs function by reducing duty ratio and force sensitivity, likely accounting for its association hearing loss.[27,30] Thus, it is likely that the MYO1E substitution from an alanine to a proline at position 159, a virtually invariant site near the ligand-binding site in the myosin motor domain, similarly produces a loss of function by impairing nucleotide binding and altering duty ratio.[24–30] This predicted loss of function mechanism is consistent both with the mouse knockout data and with previous findings, demonstrating that mutations in other components of the

actin–myosin network (for example, *ACTN4* or *INF2*) produce FSGS in humans.[6,8]

In summary, homozygosity mapping and exome sequencing in a consanguineous kindred identified *MYO1E* and *NEIL1* as novel candidate genes for human autosomal recessive SRNS, motivating comprehensive mutational screening of both genes in larger cohorts to distinguish whether mutations in one or both genes produce SRNS. These results highlight the power of next-generation sequencing in defining candidate genes for Mendelian disorders in the setting of high genetic heterogeneity, providing a strong rationale for more systematic whole-exome sequencing in human SRNS.

MATERIALS AND METHODS

Patients and families

Patient recruitment was performed at different research units in Italy (Genova, Brescia, Parma, and Foggia) and New York (Columbia University). Inclusion criteria for enrollment were: nephrotic range proteinuria; age of onset <18 years of age; histological findings of FSGS, diffuse mesangial sclerosis; or minimal change nephropathy. Family K3016 was recruited in 2002. We collected peripheral blood samples for DNA isolation from all affected patients and available family members. The Institutional Review Board for Columbia University, and local ethic review committees in Genova, Brescia, Naples, Parma, and Foggia approved our protocol.

DNA isolation and genotyping

Genomic DNA was purified from peripheral blood cells using standard procedures. The genome-wide microsatellite scan was performed with 393 microsatellites (intermarker distance $\sim$10 cm) genotyped across the genome in the three affected individuals from K3016 and in one parent in 2006 (Marshfield Mammalian Genotyping Service). In 2009, to maximize inheritance information across the genome and to identify areas of shared homozygosity, we also genotyped the three affected individuals using the Illumina 610-Quad gene-chip array, which features over 620,000 markers across the genome. DNA processing and gene-chip hybridization were performed as recommended by the manufacturer. Clustering, normalization, and genotype calls were performed using the dedicated GenomeStudio 2010.3 Genotyping Module (Illumina).

Exclusion of *NPHS2* and *WT1* mutations

Mutational screening of *NPHS2* and *WT1* was performed by bidirectional Sanger sequencing of exons and flanking introns in nine families with autosomal recessive nephrotic syndrome (including K3016) and 219 sporadic patients with steroid-resistant or steroid-dependent nephrotic syndrome, as previously described.[12,22] Mutations in these genes were excluded in pedigree K3016 in 2003. An additional 67 sporadic patients were either steroid sensitive or recently recruited and not yet tested for mutations in known genes.

Linkage analysis and homozygosity mapping

Analysis of linkage was performed using the 393 microsatellite markers (Marshfield Mammalian Genotyping Service) and a pruned dataset of 24,441 SNPs, obtained by removing SNPs with minor allele frequency <0.01, Hardy–Weinberg equilibrium, P-value $<5\times10^{-7}$, and $r^2>0.05$, on a training set of 2716 Italian individuals genotyped with the Illumina 1M-Duo chips. Analyses were conducted using the PLINK software.[36] Parametric linkage analysis was conducted under an autosomal recessive mode of inheritance, with disease gene frequency $=0.0001$ and phenocopy rate $=0.0001$, using Allegro 2.0 statistical package.[18] Homozygosity mapping was conducted on the three affected patients from K3016 genotyped on the Illumina 610-Quad using Homozygosity Mapper (http://www.homozygositymapper.org/).

Whole-exome capture followed by massive parallel sequencing

Next-generation sequencing was performed in 2010 on a SOLiD 4 System (Applied Biosystems) according to the manufacturer's instruction. Briefly, Agilent SureSelect 50 Mb Human All Exon kit for SOLiD was used for the sequence capture. This kit allows to selectively capture 50 Mb of genomic DNA, including over 180,000 exons, 700 human microRNAs from the Sanger v13 database, and over 300 additional human non-coding RNAs, such as snoRNAs and scaRNAs, representing 1.22% of human genomic regions.

Fragment libraries were prepared using 3 μg of genomic DNA as starting material. Upon shearing, the libraries were prepared using the Beckman Coulter SPRIworks automated library preparation kit (Brea, CA) with SureSelect AB adaptors. A total of 500 ng of prepared library was hybridized with the baits for 48 h, washed, and eluted using the protocol provided by Agilent Technologies. The resulting captured DNA was amplified using Platinum PCR SuperMix High Fidelity amplification kit (Invitrogen, Carlsbad, CA). Enrichment in the targeted regions was determined before sequencing by real-time PCR quantification of several genomic loci within the capture design. After quantification by real-time PCR, the amplified captured libraries were subjected to emulsion PCR using the EZ bead system (Applied Biosystems) and sequenced using the SOLiD 4 instrument and paired-end protocols. To minimize the false-positive SNP calls, we applied an error correction algorithm before SNP calling (SAET). We applied a Bayesian SNP calling to evaluate the previous probability of the existence of a heterozygote or a non-reference homozygote at the position. The data were then filtered against the Single Nucleotide Polymorphism Database and the 1000-Genome Project to identify novel variants in the sequenced sample.

Next-generation sequencing data analysis

High-quality sequencing reads were aligned to the human genome reference sequence from University of California, Santa Cruz assembly hg18. For each genome alignment, single-nucleotide discrepancies and small insertion/deletions in comparison with the reference genome were searched. The variations were regarded as reliable SNPs if they achieved minimum quality score. Bioscope 1.2 with progressive mapping option was used to align the pair-end library reads ($50+35$ bp) to human genome hg18. This mapping algorithm chooses the alignment with the highest score and the shortest matches if multiple possibilities have the same score. Bayesian SNP calling was used, which evaluates the posterior probability of the existence of a heterozygote or a non-reference homozygote at the position. On the basis of the probability of SNP evidence being a miscolor call, position error, or probe error, the Bayesian algorithm uses the previous probability of the position being a heterozygote and the probability of observation being correct. The probabilities are calculated from the quality values of color calls, the frequencies of dicolor read mismatches as a function of positions in a read, or from the frequencies of mismatches as the function of 6-mer probe prefix.

To minimize the false-positive SNP calls, we performed the error correction algorithm (SAET) before SNP calling.

Validation and search for independent mutations via traditional Sanger sequencing

All homozygous variants that were identified by exome sequencing and that were included into the pedigree K3016 regions of homozygosity were validated by traditional Sanger sequencing. To identify independent mutations, we conducted Sanger sequencing of all exons and exon–intron boundaries of the *MYO1E* (28 exons) and *NEIL1* (nine exons) in index cases from an additional eight independent families with autosomal recessive nephrotic syndrome and 63 Italian patients with sporadic SRNS with disease onset between 2 and 14 years of age (comparable to family K3016) and no mutations in *NPHS2* or *WT1*. In addition, exon 6 of *MYO1E* and exon 2 of *NEIL1* were sequenced in an additional 211 patients with SRNS (including 192 patients negative for mutations in *NPHS2* and *WT1* gene) and in 67 patients with steroid-sensitive nephrotic syndrome. The pathogenic role of each missense variant was assessed by publicly available prediction programs (PolyPhen-2: http://genetics.bwh.harvard.edu/pph2/) and by protein modeling using the highly related myosin protein from *Dictyostelium* discoideum (PDB ID 1LKX) in the Protein Data Bank (http://www.rcsb.org/pdb/). The frequencies of all rare variants were further determined by direct sequencing of 247 ethnically and geographically matched controls.

DISCLOSURE

All the authors declared no competing interests.

ACKNOWLEDGMENTS

We thank the patients and their families for participating in the study. Genome-wide STR genotyping was performed by the Mammalian Genotyping Service at the Marshfield clinic (NO1-HV-48141). SSC is supported by the American Heart Association Scientist Development Grant 0930151N and by the American Society of Nephrology Carl W Gottschalk Research Scholar Grant. GMG is supported by the European PodoNet research consortium and by the Fondazione Malattie Renali nel Bambino. SNN is supported by the American Society of Nephrology and the Doris Duke Charitable Foundation. We thank all clinicians who referred patients for the present study: L Murer (Padova), R Coppo (Torino), D Somenzi (Parma), C Izzi (Brescia), and F Emma (Roma).
We also thank the investigators of the Hypergenes Consortium (http://www.hypergenes.eu/) for sharing high-density genotyping data for derivation of allele frequencies.

SUPPLEMENTARY MATERIAL

Table S1. List of exons not covered by exome capture experiment.
Table S2. Single-nucleotide variants identified in the *MYO1E* and *NEIL1* genes.
Figure S1. Pedigree structure of the eight additional families with familial autosomal recessive nephrotic syndrome.
Supplementary material is linked to the online version of the paper at http://www.nature.com/ki

REFERENCES

1. International Study of Kidney Disease in Children. Nephrotic syndrome in children: prediction of histopathology from clinical and laboratory characteristics at time of diagnosis. A report of the International Study of Kidney Disease in Children. *Kidney Int* 1978; **13**: 159–165.
2. McKinney PA, Feltbower RG, Brocklebank JT *et al.* Time trends and ethnic patterns of childhood nephrotic syndrome in Yorkshire, UK. *Pediatr Nephrol* 2001; **16**: 1040–1044.
3. International Study of Kidney Disease in Children. The primary nephrotic syndrome in children. Identification of patients with minimal change nephrotic syndrome from initial response to prednisone. A report of the International Study of Kidney Disease in Children. *J Pediatr* 1981; **98**: 561–564.
4. Srivastava T, Simon SD, Alon US. High incidence of focal segmental glomerulosclerosis in nephrotic syndrome of childhood. *Pediatr Nephrol* 1999; **13**: 13–18.
5. Boute N, Gribouval O, Roselli S *et al.* NPHS2, encoding the glomerular protein podocin, is mutated in autosomal recessive steroid-resistant nephrotic syndrome. *Nat Genet* 2000; **24**: 349–354.
6. Brown EJ, Schlöndorff JS, Becker DJ *et al.* Mutations in the formin gene INF2 cause focal segmental glomerulosclerosis. *Nat Genet* 2010; **42**: 72–76.
7. 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 Genet* 2006; **38**: 1397–1405.
8. Kaplan JM, Kim SH, North KN *et al.* Mutations in ACTN4, encoding alpha-actinin-4, cause familial focal segmental glomerulosclerosis. *Nat Genet* 2000; **24**: 251–256.
9. Kestila M, Lenkkeri U, Männikkö M *et al.* Positionally cloned gene for a novel glomerular protein – nephrin – is mutated in congenital nephrotic syndrome. *Mol Cell* 1998; **1**: 575–582.
10. Winn MP, Conlon PJ, Lynn KL *et al.* A mutation in the TRPC6 cation channel causes familial focal segmental glomerulosclerosis. *Science* 2005; **308**: 1801–1804.
11. Boyer O, Benoit G, Gribouval O *et al.* Mutational analysis of the PLCE1 gene in steroid resistant nephrotic syndrome. *J Med Genet* 2010; **47**: 445–452.
12. Caridi G, Bertelli R, Di Duca M *et al.* Broadening the spectrum of diseases related to podocin mutations. *J Am Soc Nephrol* 2003; **14**: 1278–1286.
13. Karle SM, Uetz B, Ronner V *et al.* Novel mutations in NPHS2 detected in both familial and sporadic steroid-resistant nephrotic syndrome. *J Am Soc Nephrol* 2002; **13**: 388–393.
14. Philippe A, Nevo F, Esquivel EL *et al.* Nephrin mutations can cause childhood-onset steroid-resistant nephrotic syndrome. *J Am Soc Nephrol* 2008; **19**: 1871–1878.
15. Weber S, Gribouval O, Esquivel EL *et al.* NPHS2 mutation analysis shows genetic heterogeneity of steroid-resistant nephrotic syndrome and low post-transplant recurrence. *Kidney Int* 2004; **66**: 571–579.
16. Bilguvar K, Oztürk AK, Louvi A *et al.* Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations. *Nature* 2010; **467**: 207–210.
17. Ng SB, Buckingham KJ, Lee C *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* 2010; **42**: 30–35.
18. Gudbjartsson DF, Jonasson K, Frigge ML *et al.* Allegro, a new computer program for multipoint linkage analysis. *Nat Genet* 2000; **25**: 12–13.
19. Krendel M, Kim SV, Willinger T *et al.* Disruption of Myosin 1e promotes podocyte injury. *J Am Soc Nephrol* 2009; **20**: 86–94.
20. Sampath H, Batra AK, Vartanian V *et al.* Variable penetrance of metabolic phenotypes and development of high fat diet-induced adiposity in NEIL1-deficient mice. *Am J Physiol Endocrinol Metab* 2011; **300**: E724–E734.
21. Vartanian V, Lowell B, Minko IG *et al.* The metabolic syndrome resulting from a knockout of the NEIL1 DNA glycosylase. *Proc Natl Acad Sci USA* 2006; **103**: 1864–1869.
22. Aucella F, Bisceglia L, De Bonis P *et al.* WT1 mutations in nephrotic syndrome revisited. High prevalence in young girls, associations and renal phenotypes. *Pediatr Nephrol* 2006; **21**: 1393–1398.
23. Kollmar M, Durrwang U, Kliche W *et al.* Crystal structure of the motor domain of a class-I myosin. *EMBO J* 2002; **21**: 2517–2525.
24. Forgacs E, Sakamoto T, Cartwright S *et al.* Switch 1 mutation S217A converts myosin V into a low duty ratio motor. *J Biol Chem* 2009; **284**: 2138–2149.
25. Kintses B, Gyimesi M, Pearson DS *et al.* Reversible movement of switch 1 loop of myosin determines actin interaction. *EMBO J* 2007; **26**: 265–274.
26. Li XD, Rhodes TE, Ikebe R *et al.* Effects of mutations in the gamma-phosphate binding site of myosin on its motor function. *J Biol Chem* 1998; **273**: 27404–27411.
27. Lin T, Greenberg MJ, Moore JR *et al.* A hearing loss-associated myo1c mutation (R156W) decreases the myosin duty ratio and force sensitivity. *Biochemistry* 2011; **50**: 1831–1838.
28. Rayment I, Holden HM, Sellers JR *et al.* Structural interpretation of the mutations in the beta-cardiac myosin that have been implicated in familial hypertrophic cardiomyopathy. *Proc Natl Acad Sci USA* 1995; **92**: 3864–3868.

29. Shimada T, Sasaki N, Ohkura R *et al.* Alanine scanning mutagenesis of the switch I region in the ATPase site of *Dictyostelium discoideum* myosin II. *Biochemistry* 1997; **36**: 14037–14043.
30. Zadro C, Alemanno MS, Bellacchio E *et al.* Are MYO1C and MYO1F associated with hearing loss? *Biochim Biophys Acta* 2009; **1792**: 27–32.
31. Odronitz F, Kollmar M. Drawing the tree of eukaryotic life based on the analysis of 2269 manually annotated myosins from 328 species. *Genome Biol* 2007; **8**: R196.
32. McConnell RE, Tyska MJ. Leveraging the membrane – cytoskeleton interface with myosin-1. *Trends Cell Biol* 2010; **20**: 418–426.
33. Donaudy F, Ferrara A, Esposito L *et al.* Multiple mutations of MYO1A, a cochlear-expressed gene, in sensorineural hearing loss. *Am J Hum Genet* 2003; **72**: 1571–1577.
34. Dizdaroglu M. Substrate specificities and excision kinetics of DNA glycosylases involved in base-excision repair of oxidative DNA damage. *Mutat Res* 2003; **531**: 109–126.
35. Papeta N, Zheng Z, Schon EA *et al.* Prkdc participates in mitochondrial genome maintenance and prevents Adriamycin-induced nephropathy in mice. *J Clin Invest* 2010; **120**: 4055–4064.
36. Purcell S, Neale B, Todd-Brown K *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–575.