Exome Sequencing Reveals Cubilin Mutation as a Single-Gene Cause of Proteinuria


Departments of *Pediatrics and †Human Genetics, §Department of Biological Chemistry and DNA Sequencing Core, and ¶Howard Hughes Medical Institute, University of Michigan, Ann Arbor, Michigan; ‡Department of Medical Biology, Hacettepe University, Ankara, Turkey; and †The Pediatric Nephrology Unit, Alexandria University, Alexandria, Egypt

Protein-encoding exons constitute only 1% of the human genome but harbor 85% of mutations in single-gene disorders. Among about 2800 Mendelian diseases, for which the causative gene has been identified, mutations affect the coding region or canonical splice sites and thereby the function of the encoded proteins. However, PCR amplification of thousands of candidate exons is costly and impractical and has hampered discovery of single-gene disease causes. Recently, re-sequencing of entire coding regions of the human genome, the “exome,” with consecutive massively parallel (MP) re-sequencing has dramatically changed this situation. Exome capture and massively parallel re-sequencing has been successfully applied to identify single-gene causes of rare diseases including Miller syndrome, Kabuki syndrome, Schinzel-Giedion syndrome, Bartter syndrome, and nephronophthisis type 10. Nephrotic syndrome (NS) is a common kidney disease characterized by proteinuria, hypoalbuminemia, generalized edema, and hyperlipidemia. Identification of recessive single-gene causes of NS has provided important insights into the pathogenesis of this enigmatic disorder. Whereas NPHS2 mutations explain 10 to 25% of childhood NS, and two thirds of all NS in the first year of life can be explained by mutations in four genes only (NPHS1, NPHS2, WT1, and PLCE1),7 most other recessive causes of NS are very rare (1 to 3% of cases). Whereas single-gene causes of NS occur in childhood and adolescence, in more than 70% of cases the causative gene mutation is still unknown. This fords the use of cohort studies for gene identification and necessitates the ability to identify disease-causing genes in single families. We therefore combined whole genome homozygosity mapping with consecutive whole human exome capture (WHEC) and massively parallel re-sequencing to overcome this limitation. In this way we here identify a homozygous deleterious frameshift mutation within the cubilin gene (CUBN) as a novel unexpected cause of proteinuria in two siblings with intermittent nephrotic-range proteinuria. This approach will likely be very useful to rapidly solve cases of rare nephropathies.

RESULTS

Two siblings (A 2410-21 and A2410-22) from family A2410 were discovered to have proteinuria on routine examination.

Received April 4, 2011. Accepted May 27, 2011. Published online ahead of print. Publication date available at www.jasn.org.

B.O. and E.A.O. contributed equally to this work.

Correspondence: Dr. Friedhelm Hildebrandt, Howard Hughes Medical Institute, University of Michigan Health System, 8220C MSRB III, 1150 West Medical Center Drive, Ann Arbor, Michigan 48109-5646. Phone: 734-615-7285 (office), 734-615-7895, -7896 (laboratories), Fax: 734-615-1386, -7770, E-mail: fhilde@umich.edu

Copyright © 2011 by the American Society of Nephrology
tion. The amount of proteinuria in both siblings had been fluctuating, sometimes reaching up to 2 g/d and then decreasing without any treatment (see Patients in the Concise Methods section). Given consanguinity of the parents of the two siblings with proteinuria, we hypothesized that its cause was a recessively inherited mutation. Therefore, we performed linkage analysis and homozygosity mapping in both affected siblings (A2410-21 and A2410-22) using Affymetrix 250K SNP/StyI arrays. Nonparametric LOD score analysis was calculated for both siblings together, yielding seven segments of homozygosity by descent on chromosomes 3, 10, 14, 17, 21, and 22 with a total cumulative genomic length of 130 Mb (Figure 1). One segment of homozygosity on chromosome 10 coincided with PLCE1, a known locus for nephrotic syndrome. However, Sanger re-sequencing of all exons and adjacent exon-intron boundaries of PLCE1 yielded no mutation in this gene.

We then performed whole human exome capture in one sibling (A2410-22) using NimbleGen 2.1M Human Exome Array with consecutive massively parallel re-sequencing on two lanes of an Illumina-GAIIx sequencing platform to identify the underlying disease-causing mutation. Sequencing generated 29.5 million single-end reads of 78 bases. Quality-filtered sequence reads were aligned to the human reference genome assembly hg19 (NCBI build 37) using “CLC Genomics Workbench” software (CLC-bio, Aarhus, Denmark). About 20.1 million of these reads (68%) matched the targeted exons contributing to a median coverage of 11-fold (mean: 14-fold). No coverage was obtained for about 2.2% of all 180,000 targeted exons. CLC Genomics Workbench software was further used to call single nucleotide variants or small insertions/deletions. By filtering for homozygous-only variants, we identified 1968 variants from the reference sequence with at least fivefold sequence coverage. Of these, only 48 were not known single nucleotide polymorphisms, either reported in the database dbSNP130 or the 1000 genomes project. When examining the segments of homozygosity (132 Mb total), for which we had initially generated a positional candidate hypothesis by whole genome homozygosity mapping, only 11 non-synonymous exonic variants remained. One of these changes was a novel 1-bp homozygous deletion (c.8355delA) in exon 53 of the gene CUBN (cubilin), resulting in a frameshift and a predicted premature truncation of the encoded protein (p.S2785fsX19) (Figure 2). We confirmed this mutation by exon-PCR and Sanger sequencing homozygously in both affected siblings and heterozygously in both parents (Figure 3). We then searched for additional CUBN mutations in a large worldwide cohort of 540 families with NS in whom we had generated homozygosity mapping data by WGHM. Four affected individuals (A155-21, A849-21, A1605-21, and A2591-21) of these families showed a genomic segment of >2 Mb of homozygosity at the CUBN locus. However, direct Sanger sequencing of all exons of CUBN in these individuals did not yield any additional mutations (data not shown).

DISCUSSION

We here detected a novel 1-bp homozygous deletion of the cubilin (CUBN) gene as the cause of proteinuria in two siblings with NS using a combined strategy of homozygosity mapping and WHEC with massively parallel re-sequencing. The finding represented an unexpected cause of NS because intermittent gross proteinuria is rare in patients with CUBN mutations who are known to have Imerslund-Grasbeck syndrome (IGS; OMIM ID #261100), a form of congenital megaloblastic anemia due to vitamin B12 deficiency caused by a defect in the vitamin B12/intrinsic factor receptor (CUBN; OMIM ID #602997).
The renal effects of genetic ablation of cubilin in a mouse model has been reported previously.\textsuperscript{15} It was observed that proximal tubule cells did not localize the receptor protein “amnionless” to the plasma membrane in the absence of cubilin, indicating a mutual dependency of cubilin and amnionless to form a functional membrane receptor complex. Furthermore, cubilin-deficient mice exhibited markedly decreased uptake of albumin by proximal tubule cells, leading to proteinuria. It was shown that cubilin is essential for albumin re-absorption by proximal tubule cells.\textsuperscript{15}

Recently, an association between a missense variant (I2984V) in the \textit{CUBN} gene and both, urinary albumin-to-creatinin ratio and microalbuminuria has been identified in a genome-wide association study.\textsuperscript{16}

In this study we detected by total human exome capture a novel homozygous deleterious CUBN mutation as the cause of proteinuria in a family in whom absence of megaloblastic anemia did not elicit any suspicion of IGS, most likely because of their young age. IGS represents the full clinical picture of CUBN mutations. This is an example of molecular genetics providing a diagnostic tool that is more sensitive and accurate than clinical diagnosis, a possibility that will be strongly enhanced as WHEC and massively parallel re-sequencing become more established.

Moreover, genetic diagnostics might inform therapy: Recently, another case with IGS and CUBN mutation was reported. The patient was a 15-year-old girl of German descent who presented with megaloblastic anemia, selective proteinuria, and funicular myelosis. After intravenous vitamin B\textsubscript{12} therapy and consecutive intermittent intramuscular injection of vitamin B\textsubscript{12}, the patient’s neurologic symptoms ceased completely after 2 months, hematologic parameters normalized within 5 months, and although proteinuria was persistent, renal function did not deteriorate.\textsuperscript{17} Likewise, the siblings described here might benefit from vitamin B\textsubscript{12} replacement therapy.

At the cellular level, IGS, a monogenic cause of megaloblastic anemia (MGA1: OMIM ID #261100), is due to defective intestinal B\textsubscript{12} malabsorption and/or re-
nal tubular protein reabsorption. IGS occurs worldwide, but its prevalence is higher in several Middle Eastern countries and in Norway, and highest in Finland (0.8 in 100,000 individuals). By genetic mapping, Finnish-type IGS has been mapped to the CUBN gene locus, whereas Norwegian-type IGS has been mapped to the AMN gene locus. In families from the Mediterranean an Israeli Jewish family of Tunisian origin, a Tunisian family, and Turkish families two different AMN mutations and three different CUBN mutations were detected. Outside these geographic regions, only scattered cases have been described, for example, South Africa, France, the United States, and Taiwan. The Scandinavian cases were typical examples of founder effects, whereas in the Mediterranean region, instead of a founder effect, high degrees of consanguinity exposed rare homozygous mutations in both genes.

We here describe a novel homozygous frameshift mutation in two siblings from family A2410. Sibling A2410-21 had intermittent nephrotic range proteinuria and sibling A2410-22 had intermittent nephrotic range proteinuria along with epilepsy as the only clinical presentations. Anemia was not a part of their clinical course. CUBN mutations represent a recessive single-gene cause of proteinuria. Recessive single-gene disease causes convey full penetrance of a disease. They are thereby distinct from genetic variants that are found only to be associated with disease because associated variants usually explain only a low percentage of the phenotypic variance, as is the case for instance in the MYH9 and APOL1 variants that have been found in nephrotic syndrome.

In conclusion, the combination of homozygosity mapping in consanguineous siblings together with WHEC and massively parallel re-sequencing provides a tool for the evaluation of patients with undiagnosed genetic diseases. Our findings also indicate that IGS should be considered when diagnosing individuals with proteinuria to identify potentially treatable variants.

CONCISE METHODS

Patients
We obtained blood samples, pedigrees, and clinical information after receiving informed consent (http://www.renalgenes.org) from all patients. Approval for experiments on humans was obtained from the University of Michigan Institutional Review Board. One of the siblings of a consanguineous family from Egypt (A2410-22) was found to have proteinuria during a routine urine analysis preceding an operation for hypospadias. A2410-22 had proteinuria of 790 mg/d. Serum creatinine was 0.4 mg/dl, serum protein 6.5 g/L, serum albumin 4.2 g/L with normal creatinine clearance, and normal C3 levels at 4 years of age. Then the other sibling of the family (A2410-21) was tested for proteinuria. Sibling A2410-21 had proteinuria of 380 mg/d. Serum creatinine was 0.4 mg/dl, total serum protein 6.8 g/L, and serum albumin 3.5 g/L. Creatinine clearance and complement 3 (C3) level were normal at 5 years of age. Both siblings were treated with ACE inhibitors with no improvement of their condition. The amount of proteinuria in both siblings had been fluctuating, sometimes reaching up to 2 g/d and then decreasing without any treatment. However, no edema developed and no immunosuppressive treatment was instituted. Renal biopsy was not performed. As for megaloblastic anemia, both siblings had normal hemoglobin levels and the red blood cell indices were also within normal range for their ages. A2410-22 has had complex partial seizures since he was 5-months old, requiring treatment with carbamazepine. He has been seizure free on carbamazepine for 1 year now.

Homozygosity Mapping
For genome-wide homozygosity mapping, the GeneChip Human Mapping 250K StyI Array from Affymetrix was used. Nonparametric LOD scores were calculated using a modified version of the program GENEHUNTER 2.26 through stepwise use of a sliding window with sets of 110 SNPs. The program ALLEGRO was employed to identify regions of homozygosity as described, using a disease allele frequency of 0.0001 and CEU marker allele frequencies.

Whole Human Exome Capture
Genomic DNA (10 μg) from affected sibling A2410-22 was captured using the NimbleGen 2.1M Human Exome Array (Roche/NimbleGen) according to protocol. The NimbleGen 2.1M Human Exome Array contains oligonucleotides probes that target approximately 180,000 exons of 18,673 protein-coding genes from the consensus coding sequence (CCDS) set in addition to 551 micro-RNAs. To generate random start positions and to reduce fragment sizes to be appropriate for the Illumina
Massively Parallel Re-sequencing, Sequence Alignment, and Variant Calling

Library construction of the modified captured DNA fragments was performed using the “Genomic DNA Sample Prep Kit” according to the manufacturer’s instructions (Illumina, San Diego). The library was sequenced on two lanes of an Illumina genome analyzer (GAIIL) as 80 base single-end reads. Image analysis and base calling were generated using the Illumina pipeline 1.5.1 using default parameters. Subsequent sequence alignment to the human genome reference genome (hg19) and variant calling was performed using CLC Genomics Workbench software. Only variants with an allele frequency of more than 80% were called to identify homozygous variants.

ACKNOWLEDGMENTS

We thank the patients and their physicians for contribution of blood samples and clinical data. This work was supported by grants to F.H. from the National Institutes of Health (DK076683, RC1-DK086542), from the NephCure Foundation, and from the Thrasher Research Fund. F. Hildebrandt is an Investigator of the Howard Hughes Medical Institute, a Doris Duke Distinguished Clinical Scientist, and the Frederick G.L. Huestew Professor for the Cure and Prevention of Birth Defects.

DISCLOSURES

None.

REFERENCES


