Proteinuria is considered an unfavorable clinical condition that accelerates renal and cardiovascular disease. However, it is not clear whether all forms of proteinuria are damaging. Mutations in \( CUBN \) cause Imerslund-Gräsbeck syndrome (IGS), which is characterized by intestinal malabsorption of vitamin B12 and in some cases proteinuria. \( CUBN \) encodes for cubilin, an intestinal and proximal tubular uptake receptor containing 27 CUB domains for ligand binding.

We used next-generation sequencing for renal disease genes to genotype cohorts of patients with suspected hereditary renal disease and chronic proteinuria. \( CUBN \) variants were analyzed using bioinformatics, structural modeling, and epidemiological methods.

We identified 39 patients, in whom biallelic pathogenic variants in the \( CUBN \) gene were associated with chronic isolated proteinuria and early childhood onset. Since the proteinuria in these patients had a high proportion of albuminuria, glomerular diseases such as steroid-resistant nephrotic syndrome or Alport syndrome were often the primary clinical diagnosis, motivating renal biopsies and the use of proteinuria-lowering treatments. However, renal function was normal in all cases. By contrast, we did not found any biallelic \( CUBN \) variants in proteinuric patients with reduced renal function or focal segmental glomerulosclerosis. Unlike the more N-terminal […]
Human C-terminal CUBN variants associate with chronic proteinuria and normal renal function


1Laboratory of Epithelial Biology and Disease and 2Laboratory of Hereditary Kidney Disease, Imagine Institute, INSERM U1163, Université de Paris, Paris, France. 3Department of Pediatric Nephrology and 4Division, Necker Hospital, Assistance Publique Hôpitaux de Paris (APHP), Paris, France. 5Institute of Genetic Epidemiology, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany. 6Department of Pediatric Nephrology and Transplantation, Robert-Debré Hospital, APHP, Paris, France. 7Centre Hospitalier Régional Universitaire de Lille, Lille, France. 8Department of Pediatric Kidney, Liver and Metabolic Disease, Hannover Medical School, Hannover, Germany. 9Department of Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. 10Kindernephrologie Dachau, Dachau, Germany. 11Children's Hospital, Medical University Graz, Graz, Austria. 12Department of Pediatric Kidney, Liver and Metabolic Disease, Hannover Medical School, Hannover, Germany. 13Renal Division, University Medical Center Freiburg, Freiburg, Germany. 14Renal Division, University Medical Center Freiburg, Freiburg, Germany. 15Department of Pediatric Kidney, Liver and Metabolic Disease, Hannover Medical School, Hannover, Germany. 16Department of Pediatrics, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. 17Department of Child and Adolescent Medicine, University Medical Center Marburg-Giessen, Marburg, Germany. 18KfH-Nierenzentrum für Kinder und Jugendliche und Clementine-Kinderhospital, Frankfurt, Germany. 19Department of Genetics, Necker Hospital, APHP, Paris, France. 20Center for Human Genetics, Bioscientia, Ingelheim, Germany. 21Center for Human Genetics, Mainz, Germany. 22Bioinformatics Platform, INSERM U1163, Institut Imagine, Paris, France. 23Bioinformatics Core Facility, Structure Fedérative de Recherche Necker, INSERM US24/CNRS UMS3633, Université de Paris, Paris, France. 24Steno Diabetes Center Copenhagen, Gentofte, Denmark. 25Department of Biomedicine, Aarhus University, Aarhus, Denmark. 26Renal Division, Department of Medicine, University Hospital Freiburg, Freiburg, Germany.

BACKGROUND. Proteinuria is considered an unfavorable clinical condition that accelerates renal and cardiovascular disease. However, it is not clear whether all forms of proteinuria are damaging. Mutations in CUBN cause Imerslund-Gräsbeck syndrome (IGS), which is characterized by intestinal malabsorption of vitamin B12 and in some cases proteinuria. CUBN encodes for cubilin, an intestinal and proximal tubular uptake receptor containing 27 CUB domains for ligand binding.

METHODS. We used next-generation sequencing for renal disease genes to genotype cohorts of patients with suspected hereditary renal disease and chronic proteinuria. CUBN variants were analyzed using bioinformatics, structural modeling, and epidemiological methods.

RESULTS. We identified 39 patients, in whom biallelic pathogenic variants in the CUBN gene were associated with chronic isolated proteinuria and early childhood onset. Since the proteinuria in these patients had a high proportion of albuminuria, glomerular diseases such as steroid-resistant nephrotic syndrome or Alport syndrome were often the primary clinical diagnosis, motivating renal biopsies and the use of proteinuria-lowering treatments. However, renal function was normal in all cases. By contrast, we did not find any biallelic CUBN variants in proteinuric patients with reduced renal function or focal segmental glomerulosclerosis. Unlike the more N-terminal IGS mutations, 37 of the 41 proteinuria-associated CUBN variants led to modifications or truncations after the vitamin B12-binding domain. Finally, we show that 4 C-terminal CUBN variants are associated with albuminuria and slightly increased GFR in meta-analyses of large population-based cohorts.

CONCLUSIONS. Collectively, our data suggest an important role for the C-terminal half of cubilin in renal albumin reabsorption. Albuminuria due to reduced cubilin function could be an unexpectedly common benign condition in humans that may not require any proteinuria-lowering treatment or renal biopsy.

FUNDING. ATIP-Avenir program, Fondation Bettencourt-Schueller (Liliane Bettencourt Chair of Developmental Biology), Agence Nationale de la Recherche (ANR) Investissements d’avenir program (ANR-10-IAHU-01) and NEPHROFLY (ANR-14-ACHN-0013, to MS), Steno Collaborative Grant 2018 (NNF18OC0052457, to TSA and MS), Heisenberg Professorship of the German Research Foundation (KO 3598/5-1, to AK), Deutsche Forschungsgemeinschaft (DFG) Collaborative Research Centre (SFB) KIDGEM 1140 (project 246781735, to CB), and Federal Ministry of Education and Research (BMB) (01GM1515C, to CB).
Introduction

The loss of proteins into the urine (or proteinuria) is an important risk factor for renal and cardiovascular disease. In particular, albuminuria is associated with an increased risk for chronic kidney disease (CKD) and diabetic kidney disease (DKD), end-stage renal disease (ESRD), and mortality (1–3). Although the reasons are not entirely clear, an overload of the renal tubules with proteins and albumin-bound lipids has been proposed to be damaging for tubular epithelial cells (4, 5). Antiproteinuric therapy, for example through angiotensin-converting enzyme (ACE) or angiotensin II receptor (AT,) inhibition, is therefore an important renoprotective therapy (6).

The main cause of proteinuria is the dysfunction of the glomerular filtration barrier, which leads to symptoms like edema related to the massive albumin loss into the urine. Genetic forms of glomerular proteinuria are steroid-resistant nephrotic syndrome (SRNS) or Alport syndrome (AS). Another form of proteinuria is caused by defects in proximal tubular protein reabsorption. Tubular proteinuria has a smaller range than glomerular proteinuria, because the former only affects proteins that are filtered by the glomerulus. Typically, these are proteins such as β2-microglobulin, which is smaller in size than albumin (7). Albumin itself is filtered to a limited extent and usually accounts for less than half of the urinary protein in tubular proteinuria (8). One example is Dent’s disease, in which mutations in CLCN5 and OCRL1 lead to defective trafficking of the uptake receptor complex (9), consisting of megalin (LRP2), cubulin (CUBN), and amnionless (AMN) (10, 11). Whereas mutations in LRP2 cause Donnai-Barrow syndrome, a multisystem developmental disorder (12), CUBN and AMN mutations lead to Imerslund-Gräsbeck syndrome (IGS), which is characterized by intestinal vitamin B12 malabsorption and, in about half the cases, proteinuria (13). The phenotypes reflect the expression patterns of all 3 proteins: megalin is expressed more broadly, whereas the expression of cubulin and amnionless is mostly limited to the small intestines and kidneys. Within the kidney, recent single RNA-Seq studies of mouse and human kidney have indicated exclusive expression in the proximal tubular compartment for all 3 proteins (14–16).

With regard to the protein structures, megalin and amnionless are type I transmembrane proteins, whereas cubulin is a peripheral protein that requires amnionless for anchoring to the membrane. Anchoring occurs via β-helix–β-helix association between amnionless and the N-terminal hydrophobic stretches of 3 cubulin subunits (17). Each cubulin protomer has 8 EGF domains and 27 CUB (complement C1r/C1s, UEGF [EGF-related sea urchin protein] and bone morphogenic protein 1) domains, some of which are involved in Ca2+-dependent ligand binding (17, 18). Most IGS mutations of CUBN are in the N-terminal half of cubulin, either affecting the interactions with amnionless or the vitamin B12/intrinsic factor–binding (IF-binding) CUB domains 5–8 (CUB5–8) (17). Interestingly, 1 individual with a homozygous deletion of exon 53 harboring CUB20 was shown to have proteinuria without vitamin B12 malabsorption (19). Several CUBN variants have also shown strong associations with albuminuria in recent GWAS (20–22), which is in agreement with the albuminuria observed in cubulin-KO mice (10). Altogether, these findings suggest that cubulin could be necessary for preventing urinary albumin loss in humans. However, it is unclear whether the albuminuria due to cubulin dysfunction impairs renal function.

Here, we identified a large number of patients with isolated proteinuria associated with biallelic variants in the CUBN gene in 3 different cohorts of individuals with suspected genetic renal disease. Almost all variants were located after the vitamin B12/IF-binding domain, suggesting that more C-terminal CUB domains are crucial for renal protein reabsorption. Importantly, we show that cubulin deficiency led to proteinuria with a high proportion of urinary albumin, without impairing renal filtration function. In addition, we found that 4 C-terminal missense variants with strong albuminuria associations in GWAS were related to slightly higher estimated glomerular filtration rate (eGFR) levels in a large meta-analysis of individuals from the general population.

Results

Biallelic CUBN variants cause isolated proteinuria with normal renal function. We performed a retrospective analysis of next-generation sequencing (NGS) data obtained using a panel containing 309 renal disease genes (Renome panel) in a French cohort of 759 patients. All patients had been referred to our reference center for genetic testing because of suspected genetic kidney disease (genetic kidney disease cohort I; Figure 1 and Supplemental materials; supplemental material available online with this article; https://doi.org/10.1172/JCI29937DS1).

We grouped the patients into nonproteinuric and proteinuric groups, with the latter consisting of suspected SRNS or AS (Supplemental Figure 1). Although synonymous and nonsynonymous missense variants were equally distributed between the 2 groups, all protein-truncating variants (PTVs), including splicing, frameshift, and stop-gain variants, were strongly enriched in the proteinuric group (Supplemental Figure 1, A and B), confirming the association of CUBN variants with proteinuria.

Given the recessive nature of IGS, we asked whether there were any patients with biallelic CUBN variants who passed established filtering criteria for Mendelian disease in the proteinuric group (see Methods for the filtering criteria). We identified 14 patients from 11 families of European or African descent who had biallelic likely pathogenic variants in the CUBN gene (Table 1 and Supplemental Tables 1 and 2). By contrast, the nonproteinuric group had only 2 individuals with biallelic CUBN variants, and a control group matched for ethnicity did not include any such cases (Supplemental Figure 1C). Interestingly, all 14 patients with biallelic CUBN variants showed a very similar phenotype. Although no signs of vitamin B12 deficiency such as megaloblastic anemia could be detected, the patients shared the characteristic of chronic proteinuria ranging from 0.5 to 3 g/24 h. The average age of onset of discovery of the proteinuria was 10.9 years (Table 1 and Supple-
In the crystal structure of cubilin (18) or in silico models of individual CUB domains, all missense variants were predicted to have different detrimental effects on cubilin function, ranging from effects on amnionless binding (p.T55M) to stability and ligand binding of CUB domains (all other variants) (Figure 3, A–D, and Supplemental Table 5). Most variants affected residues or led to truncations after CUB8 (aa 1487–3618), which is in contrast to the previously described IGS mutations that are all before CUB8 (aa 66–1390) (ref. 24, Figure 2, and Supplemental Figure 2). The only 2 variants before CUB8, p.T55M and p.W1158*, were found to be in trans with variants after CUB8, consistent with the association of C-terminal CUBN variants with isolated proteinuria. As most PTVs present in the general population, such as those identified in our in-house genome database and gnomAD, lead to truncations after CUB8 (Figure 2), it may even be concluded that the loss of C-terminal CUB domains is more tolerated than the loss of those related to vitamin B12 absorption.

Biallelic CUBN variants are enriched in patients with normal renal function in a chronic proteinuria replication cohort. To investigate whether CUBN variants that lead to proteinuria are always associated with normal renal function, we next assembled an additional cohort of 107 patients with chronic subnephrotic proteinuria (chronic PU cohort; see the Supplemental material for details). Similar to the above-described patients, all patients had been referred to our reference center because of a suspected diagnosis SRNS or AS. However, prior efforts using Sanger sequencing or smaller NGS panels than the Renome panel had failed to identify the molecular cause. In this study population, 39 individuals with normal renal function (serum creatinine <110 μmol/L in men and 90 μmol/L in women or an eGFR > 60 ml/min with a minimum of 85 ml/min at the last follow-up in this series), whereas 30

![Flow chart for cohort genotyping](image-url)

Figure 1. Flow chart for cohort genotyping. The 3 different cohorts consisted of individuals with suspected genetic diseases (see Methods for more details). Altogether, 39 individuals with biallelic filtered CUBN variants were identified. In the chronic PU cohort, 11 individuals with biallelic filtered CUBN variants were identified, plus 1 additional individual carrying 1 biallelic filtered CUBN variant and the GWAS variant p.N2157D (see Supplemental Table 7). HGMD, Human Gene Mutation Database; PT, proximal tubule.
Table 1. Clinical and biological data of patients with or without biallelic filtered CUBN variants

<table>
<thead>
<tr>
<th>Clinical and biological data</th>
<th>Genetic kidney disease cohort I</th>
<th>Genetic kidney disease cohort II</th>
<th>All patients (95% IC or %)</th>
<th>Chronic PU cohort</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>14</td>
<td>13</td>
<td>107</td>
<td>12</td>
<td>95</td>
</tr>
<tr>
<td>Age at first manifestation (yr)</td>
<td>10.87 ± 11.99 (0–36.59) (n = 14)</td>
<td>6.37 ± 5.12 (0–17.43) (n = 13)</td>
<td>24.31 ± 16.64 (0–57.33) (n = 99)</td>
<td>7.85 ± 3.11 (1.01–14.69) (n = 11)</td>
<td>26.37 ± 16.50 (0–60.28) (n = 88)</td>
</tr>
<tr>
<td>Proteinuria (g/d)</td>
<td>0.69 ± 0.45 (0–1.65) (n = 14)</td>
<td>0.76 ± 0.68 (0–2.26) (n = 14)</td>
<td>1.55 ± 1.37 (0–4.29) (n = 57)</td>
<td>1.00 ± 0.61 (0–2.35) (n = 11)</td>
<td>1.68 ± 1.47 (0–20.32) (n = 46)</td>
</tr>
<tr>
<td>UAPR (%)</td>
<td>63.06 ± 8.99 (43.63–82.49) (n = 13)</td>
<td>64.16 ± 11.76 (37.55–90.76) (n = 9)</td>
<td>61.29 ± 14.00 (19.73–100.00) (n = 14)</td>
<td>60.79 ± 12.06 (29.8–91.78) (n = 5)</td>
<td>61.58 ± 18.00 (15.22–100.00) (n = 9)</td>
</tr>
<tr>
<td>Biopsies showing FSGS</td>
<td>0 (n = 9)</td>
<td>1 (25) (n = 4)</td>
<td>53 (75.7) (n = 70)</td>
<td>0 (n = 6)</td>
<td>53 (82.81) (n = 64)</td>
</tr>
<tr>
<td>Biopsies with no, unspecific, or minimal lesions</td>
<td>9 (100) (n = 9)</td>
<td>3 (75) (n = 4)</td>
<td>17 (24.29) (n = 70)</td>
<td>6 (100) (n = 6)</td>
<td>11 (17.2) (n = 64)</td>
</tr>
<tr>
<td>Age at last follow-up (yr)</td>
<td>19.87 ± 15.94 (0–54.06) (n = 14)</td>
<td>10.48 ± 6.47 (0–24.45) (n = 13)</td>
<td>36.76 ± 19.84 (0–76.18) (n = 10)</td>
<td>26.0 ± 20.27 (0–71.17) (n = 10)</td>
<td>38.13 ± 19.49 (0–77.58) (n = 79)</td>
</tr>
<tr>
<td>eGFR at last follow-up (mL/min/1.73 m²)</td>
<td>114.26 ± 28.11 (75.41–153.10) (n = 14)</td>
<td>128.07 ± 20.06 (83.91–172.23) (n = 11)</td>
<td>64.66 ± 50.11 (0–164.22) (n = 89)</td>
<td>105.64 ± 14.63 (73.04–138.25) (n = 10)</td>
<td>59.47 ± 50.65 (0–160.82) (n = 79)</td>
</tr>
<tr>
<td>ESRD</td>
<td>0 (n = 14)</td>
<td>0 (n = 13)</td>
<td>30 (29.13) (n = 103)</td>
<td>0 (n = 12)</td>
<td>30 (32.97) (n = 91)</td>
</tr>
<tr>
<td>Age of ESRD (yr)</td>
<td>NA</td>
<td>NA</td>
<td>41.90 ± 19.79 (1.49–82.31) (n = 30)</td>
<td>NA</td>
<td>41.9 ± 19.79 (1.94–81.86) (n = 30)</td>
</tr>
</tbody>
</table>

P values: *Two-sample t test for significance between CUBN mutated and nonmutated patients from the chronic PU cohort; *Fisher’s exact test for significance between CUBN mutated and nonmutated patients from the chronic PU cohort; *Proteinuria in g/24 h was measured in most patients. For children and adults where only urinary protein to creatinine ratios in spot urine were reported, conversion into g/24 h was performed (56). Patients not included here are patients where only albuminuria or protein dipstick was available or patients without further information on the amount of proteinuria. n, total number of patients with the given information; NA, not available.

Table 2. Albuminuria and eGFR meta-analysis of CKDGen cohorts with regard to CUBN variants

<table>
<thead>
<tr>
<th>Variant</th>
<th>Protein</th>
<th>f</th>
<th>Albuminuria (Effect on log(UACR))</th>
<th>eGFR (Effect on log(eGFR))</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs141640975</td>
<td>A1690V</td>
<td>0.0041</td>
<td>0.4258</td>
<td>0.0207</td>
</tr>
<tr>
<td>rs144360241</td>
<td>N2157D</td>
<td>0.0083</td>
<td>0.1152</td>
<td>0.0159</td>
</tr>
<tr>
<td>rs455581835</td>
<td>A2914V</td>
<td>0.0152</td>
<td>0.2008</td>
<td>0.0084</td>
</tr>
<tr>
<td>rs1801239</td>
<td>I2984V</td>
<td>0.1056</td>
<td>0.0615</td>
<td>0.0032</td>
</tr>
</tbody>
</table>

Effect on eGFR in the CKDGen cohort of the 4 GWAS CUBN variants A1690V, N2157D, A2914V, and I2984V. The effect on log(UACR) and log(eGFR) is the effect size or “beta” (positive value means positive directionality and therefore an increased value for albuminuria and eGFR). The units of measure are mg/g for UACR and mL/min per 1.73 m² for eGFR. f, frequency; RS, reference SNP.
identified in the chronic PU cohort in 
trans, with a low-frequency variant passing our filtering criteria (p.S1947Y) in a child with chronic proteinuria and normal renal function (Supplemental Table 7). As homozygous p.S1947Y has previously been shown to cause proteinuria in a child of similar age (27), p.N2157D seems to be a variant affecting cubilin function. Furthermore, according to the structural modeling, all 4 GWAS variants have the potential to disturb CUB domain stability or ligand binding (Figure 3C and Supplemental Table 5).

To test whether the 4 GWAS variants affect eGFR, we performed a large meta-analysis of the CKDGen Consortium’s population-based cohorts consisting of 331,340 to 597,710 individuals (28). In all 4 cases, we found, in addition to the association with albuminuria, a modest but significant association with a higher eGFR for the minor versus the major allele 

\[
P(A1690V) = 0.02432; \quad P(N2157D) = 0.00854; \quad P(A2914V) = 0.0008926; \quad P(I2984V) = 0.0005845; \quad \text{Table 2}
\]

This was also confirmed in a smaller, independent cohort consisting of 13,550 individuals with and without type 2 diabetes (Supplemental Tables 8 and 9). Although p.N2157D was only rarely found in this cohort, both eGFR and albuminuria were significantly increased for the other 3 variants localized close to the vitamin B12/IF–binding region (p.N1303H; Figure 3B), all 14 likely functional CUBN variants from this cohort were in the C-terminal CUB11–27 (aa 1928–3618; Figure 2). Also, structural models showed that all the variants could have effects on the folding and function of the CUB domains (Figure 3C and Supplemental Table 5). Altogether, these data confirm that biallelic C-terminal CUBN variants are associated with chronic proteinuria and normal renal function.

Four C-terminal missense variants are associated with albuminuria and an increased GFR in population-based studies. For a more general evaluation of the importance of CUBN variants in kidney function, we turned to large population studies. We focused on 4 C-terminal variants (p.A1690V, p.N2157D, p.A2914V, and p.I2984V) that previously had shown strong associations with albuminuria in GWAS (20–22, 25, 26). The frequencies of these GWAS variants were higher than the 0.1% threshold, which is why they were not included in the initial analysis [in-house genome database: f(A1690V) = 0.00109, f(N2157D) = 0.00667, f(A2914V) = 0.00945, f(I2984V) = 0.0932; gnomAD: f(A1690V) = 0.00173, f(N2157D) = 0.00565, f(A2914V) = 0.0122, f(I2984V) = 0.0875; Supplemental Table 3]. However, one of them (p.N2157D) was identified in the chronic PU cohort in 
trans, with a low-frequency variant passing our filtering criteria (p.S1947Y) in a child with chronic proteinuria and normal renal function (Supplemental Table 7). As homozygous p.S1947Y has previously been shown to cause proteinuria in a child of similar age (27), p.N2157D seems to be a variant affecting cubilin function. Furthermore, according to the structural modeling, all 4 GWAS variants have the potential to disturb CUB domain stability or ligand binding (Figure 3C and Supplemental Table 5).
Although we tested all associations under an additive model, for the most common variant, p.I2984V, the recessive model could also be performed, confirming that homozygotes indeed had the strongest association with albuminuria (Supplemental Tables 10–12). When stratified for diabetes status, we found that eGFR was significantly increased for all 3 variants in individuals with diabetes ($P_{(A1690V)} = 0.04; P_{(A2914V)} = 0.03; P_{(I2984V)} = 0.02$), whereas in the nondiabetics, this was only the case for p.A2914V ($P_{(A1690V)} = 0.34; P_{(A2914V)} = 0.01; P_{(I2984V)} = 0.16$; Supplemental Table 9). Together, these data strongly indicate the benign nature of the albuminuria associated with C-terminal CUBN variants.

**Discussion**

Altogether, the combined analysis of the 3 different cohorts with suspected glomerular disease identified 39 patients with isolated proteinuria and normal renal function due to biallelic filtered CUBN variants. However, despite the early onset, the proteinuria did not seem to be associated with an unfavorable prognosis for kidney disease in our patients. We further show that the high percentage of urinary albumin in these patients was often misinterpreted as glomerular injury in the clinical setting, justifying renal biopsy and ACE or AT1 inhibition as a proteinuria-lowering treatment. Apart from establishing a diagnosis in individuals with isolated subnephrotic proteinuria, the detection of CUBN variants may therefore avoid the use of inefficient therapies aimed at reducing glomerular proteinuria.

Assuming that cubilin mostly functions in the proximal tubules (14–16), the proteinuria is explained by the reduced protein reabsorption on the luminal surface of proximal tubule cells. The reabsorption of albumin was particularly affected, supporting the findings from mouse studies showing that cubilin could be the main albumin receptor in the proximal tubules (10). By contrast, megalin deficiency as seen in Donnai-Barrow syndrome typically leads to lower urinary albumin–to–protein ratios (UAPRs) (7, 8, 12, 29). Moreover, renal function has been reported to be...
reduced in several families (29–31), which is typically also the case in Dent’s disease, in which both megalin and cubilin are reduced (32). Accordingly, LRP2 variants have been found to be associated with a reduced GFR but not albuminuria (33). Altogether, these findings indicate fundamental differences in the contribution of megalin and cubilin to renal health.

Another unexpected finding is the clear genotype-phenotype correlation associated with CUBN variants (11, 34). Whereas all the IGS mutations can exclusively be found before or within the vitamin B12/IF–binding region (CUB5-8), proteinuria without vitamin B12 absorption was caused by variants located after this region. In contrast to previous in vitro studies (35), this means that the renal ligands, most notably albumin, should bind to more C-terminal CUB domains. In particular, CUB13 and CUB26, which possess proteinuria-causing variants in or close to Ca²⁺-binding motifs (p.S1947Y and p.D3492Y, respectively), are strong candidate domains for albumin binding. Although this has not yet been studied systematically, it seems that missense variants that only affect vitamin B12/IF binding, such as the Finnish variant p.P1297L, are not associated with proteinuria (36). The reverse conclusion is therefore that patients with IGS who have proteinuria should have mutations that either affect general expression or the interaction with amnionless or lead to cubilin truncation. According to the literature, such patients do not seem to present with renal insufficiency (13, 37–39), which is in agreement with the isolated proteinuria cases described here. Functional studies have also shown that vitamin B12 uptake is maintained when the receptor is truncated after CUB8 (40), consistent with the presence of a putative intestinal transcript truncated directly after CUB8 in the Genotype-Tissue Expression (GTEx) database (41). Combined with our finding that premature truncation of cubilin is more likely to happen after CUB8 in the normal population, it can thus be concluded that in humans, vitamin B12 malabsorption is less tolerated than albuminuria.

We confirmed our findings and extended them to the general population by performing a meta-analysis of over half a million participants. By evaluating 4 CUBN variants, previously identi-
fied due to their strong association with albuminuria (20–22, 25, 26), we found a moderate but significant association with higher eGFR. Although potential correlations of nearby SNPs cannot be excluded for variants identified in GWAS, support for the functional effects of these variants comes from our structural modeling and from the observation that compound heterozygosity with a filtered CUBN variant can also lead to the combination of chronic proteinuria and normal renal function. As evidence for positive selection of a European haplotype containing one of the GWAS variants (p.I2984V) has already been reported (42), it can be speculated that there is some sort of evolutionary advantage in reducing proximal tubule uptake (43), for example in conditions in which the tubules are overloaded with proteins and lipids. Support for this view comes from the clinical observation that tubular damage is key for the progression of DKD and primary focal segmental glomerulosclerosis (FSGS) (44, 45), from mouse models of proteinuria and hyperlipidemia (46, 47), in which the reduction of tubular uptake was shown to prevent early injury, and from mice lacking albumin, which become protected against AS (48). However, as our population-based studies were cross-sectional and also included a substantial amount of healthy individuals, it would be interesting to study any protective effects of cubulin deficiency on the progression of specific glomerular diseases in a longitudinal manner. Moreover, given that the observed effects on eGFR were mild and that eGFR itself reflected creatinine clearance with a margin of error, future studies could benefit from more specific tubular damage markers, such as urinary EGF (49).

In summary, our study proposes a new paradigm for the nondetrimental effects of tubular proteinuria, which contrasts with the general dogma that proteinuria is always damaging. Although adverse long-term effects associated with CUBN deficiency cannot be fully excluded, we recommend genetic testing for CUBN variants in individuals with chronic subnephrotic proteinuria to avoid unnecessary further medical treatments. Analogous to the familial renal glucosuria caused by SGLT2 mutations (50), benign Mendelian traits such as the one presented here may have the potential to define safe drug targets, especially if protective effects for a specific disease can be demonstrated for the genetic variants (51).

**Methods**

**DNA extraction and preparation for NGS.** Genomic DNA was extracted from blood samples and washed in Amicon columns (Merck Millipore). DNA quality was evaluated by agarose gel electrophoresis. The concentration of DNA and presence of impurities were calculated using the Xpose scanner (Luescher). For genetic kidney disease cohort I, a total of 759 patients with suspected genetic renal disease were sequenced by the "Renome panel" containing 309 known renal disease genes. The proteinuric group consisted of patients with AS or SRNS. Patients in the nonproteinuric group had renal tubular dysgenesis (RTD), renal hypoplasia (CAKUT), tubulointerstitial nephritis (TIN), nephronophthisis (NPHP), or polycystic kidney disease (PKD). The control group consisted of 694 individuals with matched ethnic backgrounds (527 whites and 157 Maghrebians). All identified CUBN variants were verified by Sanger sequencing. Whenever DNA of the parents was available, segregation was confirmed in the individuals with compound heterozygosity (Supplemental Table 2). For the chronic PU cohort, 107 selected individuals exhibiting isolated chronic proteinuria (between 0.5 and 3 g/d) but no genetic diagnosis despite the previous testing for several SRNS and/or AS genes were sequenced using a custom-made, NGS panel targeting 17 genes known to have functions in the proximal tubule (SureSelectXT, Agilent Technologies). High-throughput sequencing was carried out using the MiSeq/HiSeq platform (Illumina). The selected patients were heterogeneous in terms of age and were recruited through adult and pediatric nephrology departments throughout France. All the clinical information, including family information and pedigree, was collected and provided by clinicians who prescribed the genetic testing.

For genetic kidney disease cohort II, DNA was extracted from blood samples from a total of 1350 patients with suspected SRNS or AS. All exons and adjacent intronic boundaries of up to 324 glomerular genes (depending on the version of the customized multigene panel) known or hypothesized to cause SRNS, FSGS, or AS and differential diagnoses were targeted by a custom SeqCap EZ Choice Sequence Capture Library (GLOM panel, Roche NimbleGen) and subsequently sequenced on an Illumina MiSeq or HiSeq platform (2 × 150 paired-end reads [PE]) according to the manufacturer’s protocol. DNA samples were analyzed with an average coverage of 120-fold (MiSeq) or more than 200-fold (HiSeq), respectively. Bioinformatic analysis was performed using the SeqPilot SeqNext Module, version 3.5.2 (JSI Medical Systems) as well as an in-house bioinformatic pipeline. For all approaches, potential pathogenic variants were confirmed by Sanger sequencing. For compound heterozygous individuals, segregation was confirmed by sequencing the parents. All the clinical information was collected and provided by the clinicians prescribing the genetic testing.

**Variant filtering.** In order to evaluate the likelihood of pathogenicity for the identified variants, we used the guidelines of the American College of Medical Genetics (52). As a general approach, we combined a defined frequency cutoff, bioinformatic damage prediction, and structural modeling for the filtering of variants. For the filtering of splicing variants, only variants in coding regions or essential splice sites were considered in our study. All missense variants were predicted to be damaging with at least 2 of 3 damage prediction algorithms: Mutation-Taster (http://www.mutationtaster.org/), PolyPhen-2 (http://genetics.bwh.harvard.edu/php2/), and Sorting Intolerant from Tolerant (SIFT) (http://sift.bii.a-star.edu.sg). Only 1 variant, p.N1303H, was predicted to be damaging only by 1 algorithm. However, structural modeling showed that this variant is located directly in the CUB8–vitamin B12/IF interface (Supplemental Table 5). All variants were either absent from reference populations (e.g., gnomAD) (23) or rare, with global allele frequencies below 0.001. Only p.S1947Y has previously been reported in the Human Gene Mutation Database (24). In CUBN-positive cases, no additional gene variant with pathogenic relevance for the disease phenotype was present among the patients described in this manuscript.

**Structural analysis of variants.** Structural models of individual cubilin CUB domains were generated from previously determined structures using the Pyre2 server (53). Figures were prepared using PyMOL software (Schrödinger).

**Calculation of the eGFR and proteinuria range.** The eGFR was calculated with the Schwartz formula for pediatric patients (<18 years of age) and the CKD Epidemiology Collaboration (CKD-EPI) formula for adults (54–56). Proteinuria was measured as the total amount of protein in 24-hour urine collections. If only spot urine was available, the urinary protein–to–creatinine ratio (mostly in mg/mmol) was used as an estimate for the 24-hour urine measurement (57).
Statistical analyses were performed using Rstudio and GraphPad Prism 4 (GraphPad Software).

**CUBN variant lookup in the CKDGen Consortium.** Four low-frequency or common missense variants known to be associated with albuminuria from GWAS of population-based cohorts (rs141640975, rs144360241, rs45551835, and rs1801239) (20–22, 25, 26) were evaluated for association with albuminuria and eGFR using summary statistics from a large-scale meta-analysis of mostly population-based studies within the CKDGen Consortium (28, 58). The summary statistics from the CKDGen Consortium are available at http://ckdgen.imbi.uni-freiburg.de. Alleles, effect direction, SEM, P values, as well as sample size were extracted from genome-wide results using an additive model. Sample size varied between 331,340 and 597,710 across variants. We used fixed effects in the meta-analysis as previously described (28).

**Study of individuals with or without diabetes.** GWAS variants were evaluated for association with albuminuria and eGFR using summary statistics (alleles, effect, SEM, P value, and sample size) from a European exome-wide association study (ExWAS) discovery meta-analysis comprising 5 studies (3 population-based and 2 type 2 diabetes studies) from Denmark (20) under an additive model and for 12984V (rs1801239) also with the recessive model. Similarly, an association summary for the 3 variants for eGFR and albuminuria was also reported and based on stratification for diabetes status. Total sample size varied between 13,124 to 13,550 (5837–3990 individuals with diabetes and 9251–9449 individuals without diabetes across variants). We used inverse variance fixed effects in the meta-analysis as previously described (20).

**Author contributions**
M. Bedin analyzed and compiled patient data with help from OB, AS, MJT, CA, and MS. MB, LVG, MS, OG, CBF, VM, CT, CA, M. Grohmann, EK, TW, and CB performed genotyping and analyzed the sequencing results. M. Bedin, FJH, and PN performed statistical analysis. YL, TSA, and AK performed population-based studies. AC, JH, VB, SK, AB, F. Lammens, F. Louillet, BR, CV, IB, CIB, CJM, TS, LP, M. Gödel, TBH, M. Benz, GK, MH, KL, OB, and AS recruited patients and gathered the clinical data for the study. CBFA performed the structural modeling. MS conceived and supervised the project. MS wrote the manuscript with help from MB and CA. All authors critically reviewed the manuscript.

**Acknowledgments**
This work was supported by the ATIP-Avenir program, the Foundation Bettencourt-Schueller (Liliane Bettencourt Chair of Developmental Biology), and state funding by the Agence Nationale de la Recherche (ANR) under the Investissements d’avenir program (ANR-10-IAHU-01) and NEPHROFLY (ANR-14-ACHN-0013, to MS). TSA and MS also received funding from the Novo Nordisk Foundation and a Steno Collaborative Grant 2018 (NNF18OC0052457). TSA thanks all the study participants and acknowledges collaborations with the Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics at the University of Copenhagen (Copenhagen, Denmark) and the Center for Clinical Research and Prevention, Bispebjerg and Frederiksberg Hospital (Capital Region, Copenhagen, Denmark), where the genotyping and phenotyping for the Danish studies were performed. We thank the CKDGen Consortium for their collaboration in providing early access to data. The work of AK was supported by a Heisenberg Professorship of the German Research Foundation (KO 3598/5-1). CB received support from the Deutsche Forschungsgemeinschaft (DFG) Collaborative Research Centre (DFG) KIDGEM 1140 (project 246781735) and the Federal Ministry of Education and Research (BMB) (01GM1515C).

Address correspondence to: Matias Simons, INSERM UMR1163, Laboratory of Epithelial Biology and Disease, Imagine Institute, 24 Boulevard du Montparnasse, 75015, Paris, France. Tel: 33.0.1.42.75.4331; Email: matias.simons@institutimagine.org.

**CLINICAL MEDICINE**