Human C-terminal CUBN variants associate with chronic proteinuria and normal renal function

Mathilda Bedin, … , Corinne Antignac, Matias Simons


**Background:** Proteinuria is considered as an unfavorable clinical condition that accelerates renal and cardiovascular disease. However, it is not clear if all forms of proteinuria are damaging. Mutations in *CUBN* cause Imerslund-Gräsbeck syndrome (IGS) featured 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 are associated with chronic isolated proteinuria with childhood onset. Since the proteinuria displayed 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 proteinuria-lowering treatments. Yet, renal function was normal in all cases. By contrast, we did not find any biallelic pathogenic *CUBN* variants in patients with reduced renal function or focal segmental glomerulosclerosis. Unlike the more N-terminal IGS mutations, 37 out of the […]

Find the latest version:

https://jci.me/129937/pdf
Human C-terminal CUBN variants associate with chronic proteinuria and normal renal function

Mathilda Bedin¹, Olivia Boyer²,³, Aude Servais²,⁴, Yong Li⁵, Laure Villoing-Gaudé¹, Marie-Josephe Tête², Alexandra Cambier⁷, Julien Hogan⁷, Veronique Baudouin⁷, Saoussen Krid³, Albert Bensman³, Florie Lammens⁶, Ferielle Louillet⁹, Bruno Ranchin¹⁰, Cecile Vigneau¹¹, Iseline Bouteau¹², Corinne Isnard-Bagnis¹³, Christoph J. Mache¹⁴, Tobias Schäfer¹⁵, Lars Pape¹⁶, Markus Gödel¹⁷, Tobias B. Huber¹⁷, Marcus Benz¹⁸, Günter Klaus¹⁹, Matthias Hansen²⁰, Kay Latta²⁰, Olivier Gribouval², Vincent Morinière²¹, Carole Tournant²¹, Maik Grohmann²²,²³, Elisa Kuhn²², Timo Wagner²², Christine Bole-Feyssot²⁴, Fabienne Jabot-Hanin²⁴, Patrick Nitschke²⁴, Tarunveer S. Ahluwalia²⁵, Anna Köttgen⁵, Christian Brix Folstedt Andersen²⁶, Carsten Bergmann²²,²³,²⁷, Corinne Antignac²,²¹, Matias Simons¹

¹ Laboratory of Epithelial Biology and Disease, Imagine Institute, INSERM U1163, Université de Paris, Paris, France
² Laboratory of Hereditary Kidney Disease, Imagine Institute, INSERM U1163, Université de Paris, Paris, France
³ Department of Pediatric Nephrology, Necker Hospital, Assistance Publique Hopitaux de Paris (APHP), Paris, France
⁴ Department of Nephrology, Necker Hospital, Assistance Publique Hopitaux de Paris (APHP), Paris, France
⁵ Institute of Genetic Epidemiology, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany
⁶ Department of Pediatric Nephrology and Transplantation, APHP, Robert-Debré Hospital, Paris, France
⁷ Centre Hospitalier Régional Universitaire de Lille, Lille, France
⁸ Centre Hospitalier Universitaire de Rouen, Rouen, France
⁹ Department of Pediatric Nephrology, Hospices Civils de Lyon, Bron, France
¹⁰ Centre Hospitalier Universitaire de Rennes, INSERM, U1085 IRSET-9, Rennes, France
¹¹ Centre Hospitalier Universitaire de Poitiers, Poitiers, France
¹² Children’s Hospital, Medical University Graz, Austria
¹³ Renal Division, University Medical Center Freiburg, Freiburg, Germany
¹⁴ Department of Pediatric Kidney, Liver and Metabolic Disease, Hannover Medical School, Hannover, Germany
¹⁵ III. Department of Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
Correspondance should be addressed to:
Matias SIMONS, INSERM UMR1163, Laboratory of Epithelial Biology and Disease, Imagine Institute, 24 Boulevard du Montparnasse, 75015, Paris, France
e-mail: matias.simons@institutimagine.org
Tel: +33 (0)1 42 75 4331

Running title:
CUBN variants and isolated proteinuria

Conflict of interest statement:
The authors have declared that no conflict of interest exists.
Abstract

Background
Proteinuria is considered as an unfavorable clinical condition that accelerates renal and cardiovascular disease. However, it is not clear if all forms of proteinuria are damaging. Mutations in CUBN cause Imerslund-Gräsbeck syndrome (IGS) featured 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 are associated with chronic isolated proteinuria with an early childhood onset. Since the proteinuria displayed 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 proteinuria-lowering treatments. Yet, renal function was normal in all cases. By contrast, we did not find any biallelic pathogenic CUBN variants in proteinuric patients with reduced renal function or focal segmental glomerulosclerosis. Unlike the more N-terminal IGS mutations, 37 out of the 41 proteinuria-associated CUBN variants led to modifications or truncations after the vitamin B12-binding domain. Finally, we show that four 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.
**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 and diabetic kidney disease (CKD and DKD, respectively), end-stage renal disease (ESRD) and mortality (1-3). Although the reasons for this are not entirely clear, the overload of the renal tubules with proteins and albumin-bound lipids has been proposed to be damaging for tubular epithelial cells (4)(5). Anti-proteinuric therapy, for example through angiotensin-converting-enzyme (ACE) or angiotensin II receptor (AT₁) 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 the proximal tubular protein reabsorption. Tubular proteinuria is of smaller range than glomerular proteinuria because this kind of proteinuria only affects the proteins that are filtered by the glomerulus. Typically, these are proteins, such as β2-microglobulin, with a smaller 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 (gene name: *LRP2*), cubilin (*CUBN*) and amnionless (*AMN*) (10, 11). While mutations in *LRP2* cause Donnai-Barrow syndrome, a multi-system developmental disorder (12), *CUBN* and *AMN* mutations lead to IGS, featured by intestinal vitamin B12 malabsorption and in about half the cases proteinuria (13). The phenotypes reflect the expression patterns of all three proteins: while megalin is expressed more broadly, the expression of cubilin and amnionless is mostly limited to the small intestine and the kidney. Within the kidney, recent single RNA-sequencing studies of mouse and human kidney have suggested exclusive expression in the proximal tubular compartment for all three proteins (14-16).
With regard to the protein structures, megalin and amnionless are type I transmembrane proteins, whereas cubilin 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 three cubilin subunits (17). Each cubilin protomer has 8 EGF domains and 27 CUB domains, some of which are involved in Ca$^{2+}$-dependent ligand binding (17, 18). Most IGS mutations of CUBN are in the N-terminal half of cubilin, either affecting the interactions with amnionless or the vitamin B12/intrinsic factor (IF)-binding CUB domains 5-8 (CUB5-8; CUB stands for Complement C1r/C1s, Uegf (epidermal growth factor-related sea urchin protein) and Bone morphogenic protein-1) (17). Interestingly, one 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 genome-wide association studies (GWAS) (20-22), which is in agreement with the albuminuria observed in cubilin knockout mice (10). Altogether, these findings suggest that cubilin could be necessary for preventing urinary albumin loss in humans. However, it is unclear whether the albuminuria due to cubilin dysfunction impairs renal function.

Here, we identify a large number of patients with isolated proteinuria associated with biallelic variants in the CUBN gene in three different cohorts of individuals with suspected genetic renal disease. Almost all variants are 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 cubilin deficiency leads to proteinuria with a high proportion of urinary albumin without impairing renal filtration function. In addition, we find that four C-terminal missense variants with strong albuminuria associations in GWAS are 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

Next-generation sequencing (NGS) data obtained using a panel containing 309 renal disease genes (Renome panel) on a French cohort of 759 patients were retrospectively analyzed. All cases 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 Text).

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

Based on the recessive nature of IGS, we asked whether there were any patients with biallelic CUBN variants passing established filtering criteria for Mendelian disease in the proteinuric group (see Methods for filtering criteria). We found 14 patients from 11 families of European or African descent with biallelic likely pathogenic variants in the CUBN gene (Table 1; Supplemental Table 1 and 2). By contrast, the non-proteinuric group contained only two cases with biallelic CUBN variants and a control group matched for ethnic groups did not contain any cases (Supplemental Figure 1C). Interestingly, all 14 patients with biallelic CUBN variants showed a very similar phenotype. While no signs of vitamin B12 deficiency, such as megaloblastic anemia, could be detected, the patients shared a chronic proteinuria ranging from 0.5 to 3g/24h. The average age of onset of the proteinuria discovery was 10.9 years (Table 1; Supplemental Table 1). Although not measured in all cases, the proportion of albumin in the urinary protein was higher than 50% and urinary α1- or β2-microglobulin was mostly low or absent (Table 1; Supplemental Table 1) (7, 8), which is in contrast to other forms of tubular proteinuria, including megalin deficiency (12).
In all cases, renal function was normal at an average age of 17 years, as measured by serum creatinine levels and eGFR (Table 1; Supplemental Table 1). For this reason, obtaining follow-up clinical information for these patients was sometimes difficult. Renal biopsies had been performed in 9 cases, and in all of them lesions were minimal, unspecific or not present (Table 1; Supplemental Tables 1 and 2). In 7 patients, treatments with ACE inhibitors had already been started but without any lowering effects on the proteinuria. These findings were confirmed in an independent German and Austrian cohort of 1350 suspected SRNS and AS cases (Genetic kidney disease cohort II), in which we identified 13 additional individuals from 12 families with biallelic CUBN variants (Table 1; Supplemental Tables 1, 2 and 3). Also here, the chronic subnephrotic proteinuria, featured by a high proportion of urinary albumin, was combined with normal renal function in all patients (Table 1; Supplemental Table 1).

Proteinuria-associated CUBN variants localize to C-terminal CUBN domains

Altogether, we found 30 novel variants in these two studies (Figure 2; Supplemental Figure 2) that not only associated with proteinuria as an established cubilin phenotype but also passed rather stringent filtering criteria (see methods). While three variants (c.6125-2A>G, p.R2030* and p.Y3018S) were found in both cohorts, all variants were found with frequencies below 0.1% in our in-house genome database (mostly enriched for European or African ancestries) or in public reference genome databases (e.g. gnomAD (23)) (Supplemental Table 4). 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 3A-D; 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) (24) (Figure 2; Supplemental Figure 2). The only two 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 our in-house genome database or 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; more details in Supplementary Text). Similar to the abovedescribed 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 had normal renal function (serum creatinine <110 µmol/L in men or 90 µmol/L in women or eGFR > 60), whereas 30 patients had ESRD (<10ml/min GFR, transplanted or on hemodialysis) (Table 1; Supplemental Table 1; Figure 4A). Sequencing was performed with a smaller custom-made NGS panel enriched for genes important for proximal tubule function that included for the first time in these patients the gene CUBN. The sequencing revealed that 10.3% of the patients in this cohort have homozygous or two heterozygous filtered variants in CUBN, translating into a mutation rate of 28.2% in individuals with chronic proteinuria and normal renal function and 0% in patients with chronic proteinuria and reduced renal function (Table 1; Supplemental Table 1; Figure 4A).

The 12 patients from this cohort had a very similar phenotype to the abovedescribed patients, including the age of onset, the type and range of proteinuria, the lack of severe lesions in renal biopsies and the lack of proteinuria lowering effects with ACE inhibitors (Table 1; Supplemental Table 1; Figure 4B,C). Most importantly, renal function was also normal, even for the oldest patient with the age of 66 years (Table 1; Supplemental Table 1). Figure 4D depicts a normal age-dependent eGFR decline for the CUBN patients, while the
CUBN-negative cases all show a more rapid decline. Of note, we also identified two patients with hemizygous CLCN5 and OCRL1 variants, respectively, responsible for Dent’s disease 1 and 2 (Supplemental Table 6). However, in these patients serum creatinine levels were elevated, suggesting a fundamental difference between Dent’s disease and cubilin deficiency. Except for one single heterozygous variant localized close to 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 here, structural models showed that all the variants could have effects on the folding and function of the CUB domains (Figure 3C, 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 associate with albuminuria and increased glomerular filtration rate in population-based studies

To more generally evaluate the importance of CUBN variants for kidney function, we turned to large population studies. We focused on four 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 four GWAS variants have the potential to disturb CUB domain stability or ligand binding (Figure 3C; Supplemental Table 5).
To test whether the four GWAS variants affect eGFR, we performed a large meta-analysis of population-based cohorts from the CKDGen Consortium, comprising between 331,340 and 597,710 individuals (28). In all four cases, we found in addition to the association with albuminuria a modest but significant association with higher eGFR for the minor compared to the major allele (p(A1690V)=0.02432; p(N2157D)=0.00854; p(A2914V)=0.0008926; p(I2984V)=0.0005845; Table 2). This was also confirmed in a smaller, independent cohort comprising 13,550 individuals with and without type 2 diabetes (Supplemental Tables 8 and 9). While p.N2157D was only rarely found in this cohort, both eGFR and albuminuria was significantly increased for the other three variants (Supplemental Table 9). While all associations were tested under an additive model, for the most common variant, p.I2984V, also the recessive model could be performed, confirming that homozygotes indeed show the strongest association with albuminuria (Supplemental Tables 10-12). When stratified for diabetes status, the eGFR was significantly increased for all three variants in diabetic people (p(A1690V)=0.04; p(A2914V)=0.03; p(I2984V)=0.02), whereas in the non-diabetic people 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 provide strong support for the benign nature of the albuminuria associated with C-terminal CUBN variants.
Discussion

Altogether, the combined analysis of the three 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 does 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 is often
misinterpreted as glomerular injury in the clinical setting, justifying renal biopsy and ACE or
AT1 inhibition as proteinuria-lowering treatment. Yet, we show that the proteinuria does not
seem to be associated with an unfavorable prognosis for kidney disease in our patients.
Apart from establishing a diagnosis in individuals with isolated subnephrotic proteinuria, the
detection of CUBN variants may therefore avoid 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 tubular
cells. The reabsorption of albumin is particularly affected, supporting mouse studies that
cubilin could be the main albumin receptor in the proximal tubules (10). By contrast, megalin
deficiency as in Donnai-Barrow syndrome typically leads to lower urinary albumin-to-protein
ratios (7, 8, 12, 29). Moreover, renal function has been reported to be reduced in several
families (29-31), which is also the case in diseases like Dent’s disease, in which both
megalin and cubilin are reduced (32). Accordingly, LRP2 variants have been found to be
associated with reduced GFR but not with albuminuria (33), altogether suggesting
fundamental differences for the contribution of megalin and cubilin to renal health.

Another unexpected finding is the clear genotype-phenotype correlation associated with
CUBN variants (11, 34). While all the IGS mutations can exclusively be found before or
within the vitamin B12/IF-binding region (CUB5-8), proteinuria without vitamin B12
malabsorption is 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 26 that possess proteinuria-causing
variants in or close to Ca\textsuperscript{2+}-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 IGS patients that 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. Our findings were supported and extended to the general population by performing a meta-analysis of over half a million participants. By evaluating four \textit{CUBN} variants, previously identified due to their strong association with albuminuria (20-22, 25, 26), we found a moderate but significant association with higher eGFR. While potential correlations of nearby SNPs cannot be excluded for variants identified in GWAS, support for functional effects of these variants comes from our structural modelling and from the observation that compound heterozygosity with a filtered \textit{CUBN} variant can also lead to the combination of chronic proteinuria and normal renal function. As evidence for positive selection of an European haplotype containing one of the GWAS variants (p.I2984V) has already been reported (42), it can be speculated that there may even be some sort of evolutionary advantage in reducing proximal tubular uptake (43), for example in conditions where 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 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, or from mice lacking albumin, that become protected against AS (48). However, as our population-based studies were cross-sectional, it would be interesting to study any protective effects of cubilin deficiency on the progression of specific glomerular diseases in a longitudinal manner. Moreover, given that the observed effects on eGFR are mild and that eGFR itself reflects creatinine clearance with a margin of error, future studies could benefit from more specific tubular damage markers, e.g. urinary epidermal growth factor (49).

In sum, our study proposes a new paradigm for the non-detrimental 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 fully be excluded, we recommend genetic testing for CUBN variants in individuals with chronic subnephrotic proteinuria to avoid unnecessary further medical actions. 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 next-generation sequencing

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 Xpose scanner (Luescher). For Genetic kidney disease cohort I, 759 patients with suspected genetic renal disease were sequenced by the “Renome panel” containing 309 known renal disease genes. Diseases belonging to the proteinuric group are Alport syndrome (AS) and steroid-resistant nephrotic syndrome (SRNS). The non-proteinuric group consists of renal tubular dysgenesis (RTD), renal hypodysplasia (CAKUT), tubulointerstitial nephritis (TIN), nephronophthisis (NPHP) and polycystic kidney disease (PKD). The control group consisted of 694 individuals with matched ethnic backgrounds (527 Caucasians and 157 Maghrebians). All identified CUBN variants were verified by Sanger sequencing. Whenever DNA of the parents was available, segregation was confirmed in the compound heterozygous cases (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 of several SRNS and/or AS genes were sequenced using a custom-made next-generation sequencing panel targeting seventeen genes known to have functions in the proximal tubule (SureSelectXT, Agilent Technologies, France). High-throughput sequencing was carried out using a MiSeq/HiSeq platform (Illumina, San Diego, CA). The selected patients were heterogeneous in ages and were recruited through adult and pediatric nephrology departments from all over France. All the clinical information was provided and collected by clinicians prescribing the genetic testing, including familial information and pedigree. For Genetic kidney disease cohort II, DNA was extracted from human blood samples from a total of 1350 patients with suspected SRNS and AS. All exons and adjacent intronic boundaries of up to 324 glomerular genes (depending on the version of the customized multi-gene panel) known or hypothesized to cause SRNS, FSGS, Alport syndrome and differential
diagnoses were targeted by a custom SeqCap EZ choice sequence capture library (“GLOM panel”; NimbleGen, Madison, Wisconsin, USA) and subsequently sequenced on an Illumina MiSeq or HiSeq platform (2x150 PE) according to the manufacturer’s protocol. DNA samples were analysed with an average coverage of 120-fold (MiSeq) or more than 200-fold (HiSeq), respectively. Bioinformatic analysis was performed using the SeqPilot SeqNext moduleTM (v3.5.2, JSI medical systems, Kippenheim, Germany) 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 provided and collected by 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 two out of three damage prediction algorithms: Mutation Taster (http://www.mutationtaster.org/), PolyPhen-2 (http://genetics.bwh.harvard.edu/php2/) and SIFT (http://sift.jcvi.org/). Only one variant, p.N1303H, was predicted to be damaging only by one 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 Phyre2 server (53). Figures were prepared with PyMOL.

Calculation of the estimated glomerular filtration rate and proteinuria range

The eGFR was calculated with Schwartz formula for pediatric patients (age < 18 years old) and using CKD-EPI formula for adults (54-56). Proteinuria was measured as total amount of protein in 24-hour urine collections. If only spot urine was available, urinary protein-to-creatinine (mostly in mg/mmol) was used as an estimate for the 24-hour urine (57).

Statistics

Continuous values are here reported as means with ± SD (standard deviation). For these values, 95% confidence intervals were calculated using the appropriate Student’s-t probabilities. Dichotomous data are showed as percentages. We applied χ² or Fisher exact tests to dichotomous data in order to compare differences between two groups. For continuous data comparison we used an unpaired t test for the Gaussian sampled data. Two-tailed P values < 0.05 were regarded as statistically significant. Statistical analyses were performed using Rstudio and Prism 4 (GraphPad) software.

Lookup of genetic variants 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). Alleles, effect direction, standard error, p-value 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).
Lookup of genetic variants in a European exome-wide association study (ExWAS) with and without diabetes

GWAS variants were evaluated for association with albuminuria and eGFR using summary statistics (alleles, effect, standard error, p value and sample size) from an 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 I2984V (rs1801239) also with the recessive model. Similarly, association summary for the three variants for eGFR and albuminuria was also reported, based on stratification for diabetes status. Total sample size varied between 13,124 to 13,550 (3837 to 3990 individuals with diabetes and 9251 to 9449 individuals without diabetes) across variants. We used inverse variance fixed effects in the meta-analysis as previously described (20).
Author contributions

M.B. analyzed and compiled patient data with help from O.B., A.S., MJ.T., C.A., M.S.
Y.L., T.S.V., A.K. performed population-based studies.
C.F.B.A performed the structural modeling.
M.S. conceived and supervised the project.
M.S. wrote the paper with help from M.B. and C.A.
All authors critically reviewed the paper.
Funding and Acknowledgements

The work has been supported by the ATIP-Avenir program, the Fondation Bettencourt-Schueller (Liliane Bettencourt Chair of Developmental Biology) as well as State funding by the Agence Nationale de la Recherche (ANR) under the “Investissements d’avenir” program (ANR-10-IAHU-01) and the NEPHROFLY (ANR-14-ACHN-0013) grant to MS. TSA and MS also received funding from the Novo Nordisk Foundation, Steno Collaborative Grant 2018 (NNF18OC0052457). TSA acknowledges all the study participants and collaborations with the Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, 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. The summary statistics from the CKDGen Consortium are available at http://ckdgen.imbi.uni-freiburg.de. 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, German Research Foundation) Collaborative Research Centre (SFB) KIDGEM 1140 (project no. 246781735) and the Federal Ministry of Education and Research (BMBF, 01GM1515C).
References


| Table 1. Clinical and biological data of patients with or without biallelic pathogenic CUBN variants |
|--------------------------------------------------|---------------------------------------|----------------------------------|----------------------------------|----------------------------------|
|                                                  | Genetic kidney disease cohort I       | Genetic kidney disease cohort II  | All patients                     | Chronic PU cohort                |
| Clinical and biological data                     | Patients with biallelic filtered CUBN variants (95% IC or %) | Patients with biallelic filtered CUBN variants (95% IC or %) | Patients with biallelic filtered CUBN variants (95% IC or %) | Patients without biallelic filtered CUBN variants (95% IC or %) |
| N                                                | 14                                   | 13                               | 107                              | 12                               | 95                               |
| Age at first manifestation (yr)                  | 10.87 ± 11.99 (0.36-36.59) [n=14]    | 6.37 ± 5.12 (0.17-14.43) [n=13]  | 24.31 ± 16.64 (0.57-33.3) [n=99] | 7.85 ± 3.11 (1.01-14.69) [n=11]  | 26.37 ± 16.50 (0.80-60.28) [n=88] |
| Proteinuria (g/d)*                               | 0.69 ± 0.45 (0.16-1.65) [n=14]       | 0.76 ± 0.68 (0.0-2.26) [n=11]    | 1.55 ± 1.37 (0.42-2.29) [n=57]   | 1.00 ± 0.61 (0.23-2.35) [n=11]   | 1.68 ± 1.47 (0-20.32) [n=46]      |
| Urinary albumin/protein ratio (%)                | 63.06 ± 8.99 (43.63-82.49) [n=13]    | 64.16 ± 11.76 (37.50-90.76) [n=11] | 61.29 ± 14.00 (19.73-100.00) [n=14] | 60.79 ± 12.06 (29.8-91.78) [n=5] | 61.58 ± 18.00 (15.22-100.00) [n=9] |
| Biopsies showing FSGS                            | 0 [n=9]                              | 1 (25) [n=4]                     | 53 (75.71) [n=70]                | 0 [n=6]                         | 53 (82.81) [n=64]                 |
| Biopsies with no, unspecific or minimal lesions  | 9 (100) [n=9]                        | 3 (75) [n=4]                     | 17 (24.29) [n=70]                | 6 (100) [n=6]                   | 11 (17.2) [n=64]                  |
| Age at last follow-up (yr)                       | 19.87 ± 15.94 (0.54-50.06) [n=14]    | 10.48 ± 6.47 (0.24-24.5) [n=13]  | 36.76 ± 19.84 (0.76-18.6) [n=89] | 26.0 ± 20.27 (0.7-11.17) [n=10]  | 38.13 ± 19.49 (0.77-57) [n=79]    |
| eGFR at last follow-up (mL/min/1.73m2)           | 114.26 ± 18.11 (75.41-153.10) [n=14] | 126.07 ± 20.06 (83.91-172.23) [n=11] | 64.66 ± 50.11 (0-164.22) [n=89]  | 105.64 ± 14.63 (73.04-138.25) [n=10] | 59.47 ± 50.65 (0-160.82) [n=79]    |
| ESRD                                             | 0 [n=14]                             | 0 [n=13]                        | 30 (29.13) [n=103]               | 0 [n=12]                        | 30 (32.97) [n=91]                 |
| Age of ESRD (yr)                                 | NA                                   | NA                              | 41.90 ± 19.79 (1.49-82.31) [n=30] | NA                              | 41.9 ± 19.79 (1.94-81.68) [n=30]  |

(a) Two-sample t-test for significance between CUBN mutated and non-mutated patients from the chronic PU cohort
(b) Fisher exact test for significance between CUBN mutated and non-mutated patients from the chronic PU cohort
[n] indicates the total number of patients with the given information; NA, not available; yr, years; FSGS, focal segmental glomerulosclerosis; ACE, Angiotensin-conversion enzyme; ESRD, end-stage renal disease.

* Proteinuria in g/24h were measured in most patients. For children and adults where only urinary protein to creatinine ratios in spot urine were reported, conversion into g/24h 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.
### Table 2. Albuminuria and eGFR meta-analysis of CKDGen cohorts for CUBN variants

<table>
<thead>
<tr>
<th>Variant</th>
<th>Protein</th>
<th>f</th>
<th>Effect on log(UACR)</th>
<th>SE</th>
<th>P</th>
<th>N</th>
<th>Effect on log(eGFR)</th>
<th>SE</th>
<th>P</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs14164097</td>
<td>A1690V</td>
<td>0.0041</td>
<td>0.4258</td>
<td>0.0207</td>
<td>3.52E-94</td>
<td>481598</td>
<td>0.0105</td>
<td>0.0046</td>
<td>0.02432</td>
<td>569696</td>
</tr>
<tr>
<td>rs14436024</td>
<td>N2157D</td>
<td>0.0083</td>
<td>0.1152</td>
<td>0.0159</td>
<td>3.92E-13</td>
<td>463842</td>
<td>0.008</td>
<td>0.0031</td>
<td>0.00854</td>
<td>331340</td>
</tr>
<tr>
<td>rs45551835</td>
<td>A2914V</td>
<td>0.0152</td>
<td>0.2008</td>
<td>0.0084</td>
<td>2.77E-126</td>
<td>545479</td>
<td>0.0049</td>
<td>0.0015</td>
<td>0.000892</td>
<td>345028</td>
</tr>
<tr>
<td>rs1801239</td>
<td>I2984V</td>
<td>0.1056</td>
<td>0.0615</td>
<td>0.0032</td>
<td>4.77E-81</td>
<td>557672</td>
<td>0.0019</td>
<td>0.0006</td>
<td>0.000584</td>
<td>597710</td>
</tr>
</tbody>
</table>

Effect on estimated glomerular filtration (eGFR) in the CKDgen cohort of the four GWAS CUBN variants A1690V, N2157D, A2914V and I2984V. f is the frequency; Effect on log(UACR) and on log(eGFR) is the effect size or “beta” (positive value means positive directionality and therefore an increased value of albuminuria and eGFR); SE is the standard error; P is the P-value; and N is the sample size. The unit for UACR is mg/g and for eGFR is ml min⁻¹ per 1.73 m².

The table shows the meta-analysis results for albuminuria and eGFR for CUBN variants in the CKDGen cohort. The variants rs14164097, rs14436024, rs45551835, and rs1801239 are associated with changes in albuminuria and eGFR. The effect size ("beta") is calculated as the change in log-transformed values for both albuminuria and eGFR. The standard error (SE) and p-value (P) provide measures of the variability and statistical significance of the effect, respectively. The sample size (N) indicates the number of participants in the analysis.
Figure 1. Flow chart for cohort genotyping.
The three different cohorts consisted of individuals with suspected genetic diseases (Alport syndrome (AS), steroid-resistant nephrotic syndrome (SRNS), nephronophthisis (NPHP), congenital anomalies of the kidney and urinary tract (CAKUT), tubulointerstitial nephritis (TIN), polycystic kidney disease (PKD) and renal tubular dysgenesis (RTD)). Altogether, 39 individuals with biallelic filtered CUBN variants (biallelic) were identified. In the chronic PU cohort, 11 individuals with biallelic filtered CUBN variants were identified plus 1 additional individual carrying one biallelic filtered CUBN variant and the GWAS variant p.N2157D (see Supplemental Table 7).
Figure 2. Frequency, density and position of CUBN variants along the cubilin protein. Cubilin protein structure summarized with the 8 EGF-like (lighter blue) and 27 CUB domains (brighter blue); the red dots correspond to the theoretical Ca2+-binding sites; arrows on top indicate the variants found in the three cohorts of this study; the green arrows represent the filtered CUBN variants, and the red arrows indicate the four GWAS missense variants A1690V, N2157D, A2914V and I2984V. Gaussian kernel density curves (with 0.2 width parameter) as well as the position of HGMD variants (red), the variants from this study (green), PTVs from the in-house (yellow) and gnomAD (blue) are shown as curves as vertical lines, respectively. The minor allele frequencies (MAF) of HGMD and gnomAD variants are shown as red and blue dots, respectively.
Figure 3. Structural modeling of CUBN missense variants.  
(A) Location of the cubilin T55M missense variant in RCSB entry 6GJE. The variant is located in the hydrophobic core of the cubilin b-helix that interacts with amnionless. (B) Location of the cubilin N1303H missense variant in RCSB entry 3KQ4. The variant is located in CUB domain 8 close to the interface with IF/B12. Red spheres represent Ca\(^{2+}\) ions binding to the CUB domains. (C) Location of cubilin missense plus the four GWAS variants (in red) in in silico structural models of domains CUB 11 to 27. Red spheres represent Ca\(^{2+}\) ions predicted to bind domains CUB 11, 13 and 26. Variants S1947Y and D3492Y are notably close to these Ca\(^{2+}\)-binding motifs. (D) Example of one CUB domain with b-sheets (b2-10) and loops (L2-9).
Figure 4. Clinical and biological profiles of patients carrying biallelic filtered CUBN variants.

(A) Proportions of biallelic filtered CUBN variants carriers in the Chronic proteinuria (PU) cohort (N=107), when considering only patients with normal renal function (eGFR > 60 ml/min per 1.73m2, N=39) or only the patients in end-stage renal disease (ESRD; eGFR < 15 ml/min per 1.73m2, N=30).

(B-D) Patients with biallelic filtered CUBN variants from the three cohorts (genetic kidney disease cohorts I and II, chronic PU cohort) are merged as “Patients with biallelic filtered CUBN variants” group. Urinary albumin to protein ratio (UAPR) is plotted in patients with (n=27) and without (n=9) biallelic filtered CUBN variants. As a general rule, glomerular proteinuria is characterized by a UAPR above 50%, while tubular proteinuria is below 50 % (B) (7, 43). Age of first manifestation of proteinuria for patients with (n=38) and without (n=88) biallelic filtered CUBN. Data represent mean values ± 95% confidence intervals (C). t test: *p < 0.01; **p < 0.001; ***p < 0.0001; NS, not significant. Renal function of patients with biallelic filtered CUBN variants (n=35) is declining according to normal age decline, whereas in patients without biallelic filtered CUBN variants (n=79) the decline is more rapid (D). Renal function (eGFR, estimated glomerular filtration rate) was calculated using Schwartz formula in children and CKD-EPI formula for adults. The blue and red dotted lines represent the logarithmic trend curves for both groups.