JCI insight

Genome-wide studies reveal factors associated with circulating uromodulin and its relations with complex diseases

Yong Li, … , Luca Rampoldi, Anna Köttgen

JCI Insight. 2022. https://doi.org/10.1172/jci.insight.157035.

Research In-Press Preview Genetics Nephrology

Graphical abstract

Genome-wide studies reveal factors associated with circulating uromodulin and its relations with complex diseases

Yong Li[*1], Yurong Cheng[*1,2], Francesco Consolato[*3], Guglielmo Schiano[4], Michael R. Chong[5,6], Maik Pietzner[7,8], Ngoc Quynh H. Nguyen[9], Nora Scherer[1,10], Mary L. Biggs[11,12], Marcus E. Kleber[13,14], Stefan Haug[1], Burulça Göçmen[1], Marie Pigeyre[15,16], Peggy Sekula[1], Inga Steinbrenner[1], Pascal Schlosser[1,17], Christina B. Joseph[18], Jennifer A. Brody[11], Morgan E. Grams[17,19], Caroline Hayward[18], Ulla T. Schultheiss[1,20], Bernhard K. Krämer[14], Florian Kronenberg[21], Annette Peters[22,23], Jochen Seissler[24], Dominik Steubl[25,26,27], Cornelia Then[24], Matthias Wuttke[1], Winfried März[14,28,29], Kai-Uwe Eckardt[30,31], Christian Gieger[22,32,33], Eric Boerwinkle[34,35], Bruce M. Psaty[11,36,37], Josef Coresh[17], Peter J. Oefner[38], Guillaume Pare[16,39], Claudia Langenberg[7,8], Jürgen E. Scherberich[40], Bing Yu[9], Shreeram Akilesh[41], Olivier Devuyst[4], Luca Rampoldi[**3], Anna Köttgen[**1,17,42]

* These individuals contributed equally
** These individuals jointly supervised the work

1 Institute of Genetic Epidemiology, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany
2 Faculty of Biology, University of Freiburg, Freiburg, Germany
3 Molecular Genetics of Renal Disorders, Division of Genetics and Cell Biology, IRCCS Ospedale San Raffaele, Milan, Italy
4 Institute of Physiology, University of Zurich, Zurich, Switzerland
5 Population Health Research Institute, David Braley Cardiac, Vascular and Stroke Research Institute, Thrombosis and Atherosclerosis Research Institute, Hamilton Health Sciences, Hamilton, Canada
6 Department of Biochemistry and Biomedical Sciences, Department of Pathology and Molecular Medicine, McMaster University, Hamilton, Canada
7 MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge School of Clinical Medicine, Cambridge, UK
8 Computational Medicine, Berlin Institute of Health (BIH) at Charité – Universitätsmedizin Berlin, Berlin, Germany
9 Department of Epidemiology, Environmental Sciences and Human Genetics, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA
10 Spemann Graduate School of Biology and Medicine (SGBM), University of Freiburg, Freiburg, Germany
11 Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA
12 Department of Biostatistics, University of Washington, Seattle, WA, USA
13 SYNLAB MVZ Humangenetik Mannheim GmbH, Mannheim, Germany
14 Vth Department of Medicine, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany
15 Department of Medicine, McMaster University, Michael G. DeGroote School of Medicine, Hamilton, Canada
16 Population Health Research Institute, Thrombosis and Atherosclerosis Research Institute, Hamilton Health Sciences, Hamilton, Canada
17 Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health,

Baltimore, MD, USA
18 MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Western General Hospital, Edinburgh, UK
19 Johns Hopkins University, School of Medicine, Division of Nephrology, Baltimore MD, USA
20 Department of Medicine IV, Nephrology and Primary Care, Faculty of Medicine and Medical Center -University of Freiburg, Freiburg, Germany
21 Institute of Genetic Epidemiology, Medical University of Innsbruck, Innsbruck, Austria
22 Institute of Epidemiology, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany
23 Chair of Epidemiology, Institute of Medical Information Sciences, Biometry and Epidemiology, Ludwig-Maximilians University, Munich, Germany
24 Medizinische Klinik und Poliklinik IV, Klinikum der Universität München, LMU München, Germany
25 Division of Nephrology, Tufts Medical Center, Boston, MA, USA
26 Department of Nephrology, Klinikum rechts der Isar, Technical University Munich, Munich, Germany
27 Boehringer Ingelheim International GmbH, Ingelheim, Germany
28 Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Graz, Austria
29 SYNLAB Academy, SYNLAB Holding Deutschland GmbH, Augsburg and Mannheim, Germany
30 Department of Nephrology and Medical Intensive Care, Charité-Universitätsmedizin Berlin, Berlin, Germany
31 Department of Nephrology and Hypertension, University Hospital Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
32 Research Unit of Molecular Epidemiology, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany
33 German Center for Diabetes Research (DZD), Partner München, Neuherberg, Germany
34 Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA
35 Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA
36 Department of Epidemiology, University of Washington, Seattle, WA, USA
37 Department of Health Systems and Population Health, University of Washington, Seattle, WA, USA
38 Institute of Functional Genomics, University of Regensburg, Regensburg, Germany
39 Department of Pathology and Molecular Medicine, McMaster University, Hamilton, Canada
40 München Klinik Harlaching, teaching hospital of the Ludwig-Maximilians-Universität München, München, Germany
41 Department of Laboratory Medicine and Pathology, University of Washington, Seattle, WA, USA
42 CIBSS – Centre for Integrative Biological Signalling Studies, Albert-Ludwigs-Universität Freiburg, Freiburg, Germany

Correspondence to:

Anna Köttgen, MD MPH
Institute of Genetic Epidemiology
University Medical Center Freiburg
Hugstetter Str. 49
79106 Freiburg
Germany
Tel: +49 (0)761 270-78050
anna.koettgen@uniklinik-freiburg.de

Luca Rampoldi, PhD
Molecular Genetics of Renal Disorders Unit
Division of Genetics and Cell Biology
San Raffaele Scientific Institute
Via Olgettina 58
20132 Milan
Italy
Tel: +39 0226434947
rampoldi.luca@hsr.it

Abstract

UMOD is a major risk gene for monogenic and complex forms of kidney disease. The encoded kidney-specific protein uromodulin is highly abundant in urine and related to chronic kidney disease, hypertension, and pathogen defense. To gain insights into potential systemic roles, we performed genome-wide screens of circulating uromodulin using complementary antibody-based (N=13,985) and aptamer-based (N=18,070) assays. We detected 3 and 10 distinct significant (p<5e-8) loci, respectively. Integration of antibody-based results at the *UMOD* locus with functional genomics data (RNA-seq, ATAC-seq, Hi-C) of primary human kidney tissue highlights an upstream variant with differential accessibility and transcription in uromodulin-synthesizing kidney cells as underlying the observed *cis* effect. Shared association patterns with complex traits, including chronic kidney disease and blood pressure, place the *PRKAG2* locus in the same pathway as *UMOD*. Experimental validation of the third antibody-based locus, *B4GALNT2*, shows that the p.Cys466Arg variant of the encoded N-acetylgalactosaminyltransferase has a loss-of-function effect leading to higher serum uromodulin levels. Aptamer-based results point to enzymes writing glycan marks present on uromodulin and to their receptors in the circulation, suggesting that this assay permits investigating uromodulin's complex glycosylation rather than its quantitative levels. Overall, our study provides new insights into circulating uromodulin and its emerging functions.

Introduction

Chronic kidney disease (CKD) can progress to kidney failure, is a major risk factor for cardiovascular morbidity and mortality, and a leading cause of death (1-3). CKD affects approximately 10% of adults (1). Genome-wide association studies (GWAS) of kidney function, CKD, and CKD progression in population-based studies have consistently identified the largest effect for common variants at the *UMOD* locus (4-7). The encoded protein uromodulin, previously named Tamm-Horsfall protein, is the most abundant protein in the urine of healthy individuals (8). It is exclusively synthesized in the kidney's thick ascending limb (TAL) of the loop of Henle (LOH) and the distal convoluted tubule (DCT) (9). Urinary uromodulin has important roles in protecting against urinary tract infections (10). Glycosylation accounts for approximately 30% of the mature protein's molecular weight in urine and may be important for some of the protein's functions, including an emerging immuno-modulatory role (8).

Common CKD risk variants in *UMOD* are also associated with higher risk of hypertension, hyperuricemia and gout, and lower risk of kidney stone disease (4, 11-14). Their association with higher uromodulin transcript levels in kidney (7, 15) and higher uromodulin levels in urine (7, 16) directly implicates a pathophysiologic role of uromodulin. Rare mutations in *UMOD* cause one of the most common monogenic kidney diseases, autosomal-dominant tubulo-interstitial kidney disease (17, 18).

UMOD is hence a main driver of genetic kidney disease, and genetic studies of the kidney-specific protein uromodulin may yield insights not only into kidney disease but also into the protein's other diverse functions and associated diseases. Such studies can also reveal regulators and interaction partners that can help to understand potential consequences of therapeutic manipulation and may reveal new entry points to do so, with the final goal to

reach pharmacological intervention (19). Previous studies of uromodulin have almost exclusively focused on urine. The protein is, however, also released from the basolateral membrane of renal TAL and DCT cells and reaches the blood, where its concentration is about 100-fold lower than in urine (8). In a previous study, urine and plasma uromodulin levels were moderately correlated (20), although they are both associated with the kidney function measure estimated glomerular filtration rate (eGFR). The mechanisms influencing circulating uromodulin, whether circulating and urine uromodulin share association patterns with complex diseases, and any factors related to the glycans carried by uromodulin are unknown. Quantification of circulating uromodulin on a population scale has recently become feasible (21-24). A small GWAS of serum uromodulin levels reported only an association with the known CKD-associated *UMOD* variants in *cis* (23).

Here, we perform meta-analyses of GWAS of circulating uromodulin to obtain insights into factors that may be relevant to CKD pathophysiology and into any systemic functions of this kidney-specific protein. Using an antibody-based assay, we (i) identify an upstream variant at the *UMOD* locus with differential accessibility and transcription in human uromodulin-synthesizing kidney cell types and compartments that is strongly associated with circulating and urine uromodulin, CKD and hypertension, (ii) place the *PRKAG2* locus in the same pathway as *UMOD* with respect to its disease associations, and (iii) show that p.Cys466Arg in the uromodulin-glycosylating enzyme B4GALNT2 is a loss-of-function allele leading to higher serum uromodulin levels. Using an aptamer-based assay, we identify non-overlapping loci that point to enzymes writing glycan marks present on uromodulin and to their receptors in the circulation. Together, our study based on human genetic evidence provides new insights into circulating uromodulin and its emerging functions.

Results

GWAS meta-analysis identifies 13 genetic loci associated with circulating uromodulin

Characteristics of the 32,055 individuals from seven participating studies (ARIC, CHS, Fenland, GCKD, KORA, LURIC, ORIGIN), including distributions of age, sex, and the estimated glomerular filtration rate (eGFR), are shown **Supplementary Table 1**. There were 29,439 participants of European ancestry (EA), 400 African Americans (AA), and 2,216 Hispanics (HIS). GWAS of age-, sex-, and eGFR-adjusted and inverse normal rank transformed circulating uromodulin measurements were carried out in each of the seven studies using densely imputed genotypes (25, 26) (**Supplementary Table 2**) and combined via meta-analysis (Methods).

Trans-ethnic meta-analysis of 10,735,251 genetic variants of minor allele frequency (MAF) >1% across five studies with antibody-based uromodulin quantification (CHS, GCKD, KORA, LURIC, ORIGIN; N=13,985) revealed three genomic loci with at least one significantly associated (p<5e-8) genetic variant (**Figure 1A**; **Supplementary Table 3**): *UMOD*/*PDILT* (index SNP rs77924615, p=6.4e-577), *B4GALNT2* (rs7224888, p=1.8e-32), and *PRKAG2* (rs55791829, p=2.9e-09). The genomic control parameter was 0.99, consistent with the absence of undetected population stratification (**Supplementary Figure 1A**). The estimated SNP-based heritability of uromodulin was 0.135 (95% confidence interval [CI] 0.010-0.259, Methods). Except for the *UMOD* locus, there was little heterogeneity of genetic effects in the five contributing studies (**Supplementary Figure 2**). The index variant rs77924615 at the locus with the strongest association, *UMOD*/*PDILT*, explained an estimated 18% of the serum uromodulin variance (**Table 1**, Methods).

The GWAS meta-analysis of 8,815,558 genetic variants across two studies with plasma aptamer uromodulin-readout (ARIC and Fenland; N=18,070) showed no evidence of inflation

(λ=0.99; **Supplementary Figure 1B**) and revealed 10 genome-wide significant loci (**Figure 1B**; **Supplementary Table 3**), with the statistically strongest association observed at rs34211178 upstream of *ST3GAL6* on chromosome 3 (p=6.9e-442). *ST3GAL6* encodes for ST3 Beta-Galactoside Alpha-2,3-Sialyltransferase (27), an enzyme with alpha-2,3-sialyltransferase activity toward Gal-beta1,4-GlcNAc structures that are present on the glycoprotein uromodulin (28). The largest effect size was observed for a low frequency variant in *ASGR2*, with each minor allele associated with one standard deviation higher age- and sex-adjusted plasma aptamer uromodulin-readout. The estimated SNP-based heritability was 0.177 (95% CI -0.032-0.386). For each of the two assays, regional association plots for all 13 loci that achieved genome-wide significance are shown in **Supplementary Figure 3**, and association statistics in **Supplementary Table 3**.

The two meta-analyses of antibody- and aptamer-based uromodulin measurements identified different genetic loci. A *cis*-association between SNPs in the *UMOD* gene and levels of the encoded protein uromodulin were only observed with antibody quantification, supporting that this assay measures the amount of protein. Within each of the two meta-analyses, the association results showed consistent effect sizes and directions in all contributing studies, and both assays had low coefficients of variation (Methods). Moreover, genes identified with both assays can be connected to uromodulin through different sources of external evidence (see below). This indicates that the antibody- and aptamer-based assays for circulating uromodulin deliver reproducible measurements, but assess different properties of their respective targets such as protein amount and glycosylation pattern, respectively.

Secondary analyses, sex-specific effects, and association with urine uromodulin levels

Genome-wide discovery screens without adjustment for eGFR showed virtually identical results (**Supplementary Figure 4A** and **B**), indicating that kidney function did not confound genetic associations with uromodulin. A secondary analysis restricting to 11,369 EA participants with antibody-based measurements yielded very similar results as the primary trans-ethnic meta-analysis (**Supplementary Table 3**).

We next evaluated the presence of sex-specific genetic effects, motivated by the observation that women have higher serum uromodulin levels than men (22, 29, 30). Higher circulating uromodulin in women as compared to men was observed for both the antibody- (mean of the mean uromodulin 103.62 ng/ml in women vs. 92.76 ng/ml in men) and the aptamer-based assay (10329 vs. 9813 relative fluorescence units). Sex-specific analyses identified several genome-wide significant loci for both assays (**Supplementary Figure 5**), all of which were also identified in the primary combined analyses. The index SNPs at the 13 significant loci did not show evidence for sex-specific differences (**Supplementary Figure 6**), nor did GWAS of the X-chromosome or a genome-wide test for differences of SNP effects on uromodulin between men and women yield significant findings (**Supplementary Table 4**).

Given that a previous GWAS meta-analysis of urine uromodulin reported significant associations at the *UMOD* locus (16), we queried the association between the 13 index SNPs identified in this study and urine uromodulin levels among 29,262 EA individuals (Methods). Except for rs77924615 at *UMOD*/*PDILT* (p=5.3e-97), which explained 1.4% of the urine uromodulin variance, none of the other SNPs showed significant (p<3.8e-03 = 0.05/13) associations (**Supplementary Table 5**).

Prioritization of causal variants in uromodulin-associated loci

Statistical fine-mapping was carried out to identify the most likely causal variants in uromodulin-associated loci (Methods). Conditional analyses supported the presence of more than one independent signal at *UMOD*/*PDILT* (n=2), *B4GALNT2* (n=3), *ST3GAL6* (n=4) and *ASGR1*/*ASGR2* (n=2; **Supplementary Table 6**). For each of 20 independent, uromodulin-associated signals within the 13 identified loci, we calculated a SNP set that contains the variant driving the respective association signal with 99% posterior probability. There were 12 sets with <20 variants, five of which had <5 variants (**Supplementary Table 6**).

Credible set variant annotation showed several noteworthy findings (**Supplementary Table 7**). The antibody-based association on chromosome 16 could be mapped to a single intronic variant, rs77924615, in *PDILT*, the gene upstream of *UMOD*. Another independent set of 10 variants in the locus mapped to *UMOD* (**Figure 2**), with the lead SNP rs4293393 experimentally shown to affect *UMOD* transcription (15). To study whether rs77924615 may be an upstream variant regulating *UMOD* transcription, we generated functional genomic annotation data of chromatin accessibility (ATAC-seq) and gene expression (RNA-seq) from cortex and medulla of native human kidney tissue (Methods) that showed transcription of *UMOD*, more strongly in medulla than in cortex, but not of *PDILT*. Both independent variants, rs77924615 in *PDILT* and rs4293393 in the *UMOD* promoter, mapped into regions of open chromatin in medulla, where *UMOD* transcript levels in human kidney are highest (GTEx Project V8) (31). These regions aligned with open chromatin in LOH and DCT kidney cells from single-nucleus ATAC-seq data (Methods), that were not observed in several other kidney and immune cell types (**Figure 2**). Both variants mapped into the same topological associated domain, with predicted contacts based on chromatin conformation capture (Hi-C, Methods). Thus, the identified SNPs likely are regulatory variants in kidney cell types producing uromodulin.

The lead SNP at the *B4GALNT2* locus, rs7224888, was identified with the antibody assay and is a missense variant. Its minor C allele encodes a cysteine-to-arginine substitution (p.Cys466Arg; NP_703147.2) in the encoded enzyme beta-1,4-N-acetyl-galactosaminyltransferase 2. The C allele was associated with higher circulating uromodulin in our study, and has been linked to the absence of the Sda antigen (32), a blood group antigen synthesized by B4GALNT2 that is present on uromodulin (33). The *B4GALNT2* locus also contained an independent small credible set of three variants. The most likely causal rs72835417 maps into a splice region; its minor allele was associated with higher uromodulin in our study, and with lower *B4GALNT2* expression (p=1.8e-7) in micro-dissected kidney tubules (34), further supporting that reduced B4GALNT2 function relates to higher circulating uromodulin.

At *SIGLEC9*, the major G allele at the most likely causal variant rs2075803 leads to a lysine-to-glutamine substitution in SIGLEC9 (p.Lys100Glu; NP_001185487.1). It was associated with lower aptamer signal in our study (p=3.8e-100), and with lower circulating SIGLEC9 protein (p=6e-2142) (35) and serum C-reactive protein (p=5e-10) (36) in previous studies. The encoded sialic acid-binding Ig-like lectin-9 is an inhibitory receptor mainly present on neutrophils and monocytes. It has been experimentally shown to interact with urinary uromodulin (37), indicating that genetically encoded variation in *SIGLEC9* levels relates to differences in the aptamer readout of circulating uromodulin.

Uromodulin-associated loci are associated with distinct sets of biomarkers and diseases

Genetic studies have linked variation in *UMOD* to monogenic autosomal-dominant tubulointerstitial kidney disease (17, 38), to complex kidney function traits and CKD (4), blood pressure and hypertension (11, 12), uric acid levels and gout (13) as well as to kidney stone

disease (14). In order to investigate whether any of the 13 significant loci share phenotype association patterns, and to detect additional disease associations that may be mediated by altered uromodulin levels or properties, we performed colocalization with (i) levels of 30 biomarkers and (ii) 1,404 complex traits and diseases based on data from the UK Biobank study, as well as (iii) additional traits previously linked to common *UMOD* variants – namely eGFR, CKD, systolic (SBP), diastolic blood pressure (DBP) and uromodulin levels in urine (Methods). Interestingly, genetic associations with antibody-based circulating uromodulin at the *PRKAG2* and *UMOD*/*PDILT* loci shared a very similar pattern of colocalization with numerous kidney-related traits (creatinine, cystatin C, urea, eGFR, CKD, urinary calculus, DBP, SBP, hypertension; **Supplementary Table 8**, **Figure 3**). The directions of association of colocalizing traits were consistent with biological knowledge based on studies of uromodulin levels in urine, for example higher serum uromodulin and higher risk of CKD (**Figure 3**). These observations are consistent with a common biological context of the *PRKAG2* and *UMOD*/*PDILT* loci in the pathophysiology of CKD, hypertension, and kidney stone disease, and with the earlier identification of the *PRKAG2* locus in GWAS of CKD (39). Conditional co-localization of two independent SNP sets at the *UMOD* locus further supported a shared genetic cause between the levels of circulating and urine uromodulin levels (**Figure 3**, **Supplementary Table 9**).

There were several other examples of positive colocalizations supported by biological knowledge: first, genetic associations at the *B4GALNT2* locus with antibody-based circulating uromodulin colocalized with the odds of multiple gestation. This is consistent with a role of the B4GALNT2-mediated Sda antigen in embryo implantation in mice (40) (**Figure 3**). Second, SNPs at *SIGLEC9* associated with plasma aptamer uromodulin-readout colocalized with levels of alkaline phosphatase in blood, which is in line with altered bone turnover described in

recent knockout mouse model of the homologous gene (41). These observations suggest that the aptamer-based assay is particularly well suited to generate insights into the generation of glycosylation residues that are present on uromodulin, and how such glycosylation residues are recognized in the circulation. Tests of pairwise interactions of the 13 index SNPs, which could point towards non-additive effects when the same pathway is affected, showed a significant interaction between genotype at the lead SNPs in *ST3GAL6* and *SIGLEC9* on plasma aptamer uromodulin-readout (interaction p-value=3e-07; Methods). These two genes are indeed functionally related, as *ST3GAL6* is involved in the synthesis of sialic acid moiety that is bound by SIGLEC9.

Additionally, we tested the aggregate effect of rare (MAF <0.1%), potentially deleterious variants in the genes prioritized at each of the 13 uromodulin-associated genetic loci on 770 complex diseases (Methods). Using data from whole exome sequencing of 173,688 UK Biobank participants, significant associations (p<4.9e-06) were identified between carrier status of rare *UMOD* variants and anemia of chronic disease (OR=2.8, p=6.1e-09), hypertensive CKD (OR=2.6, p=1.8e-06), and CKD (OR=1.89, p=2.0e-09; **Supplementary Table 10**).

Prioritization of causal genes in uromodulin-associated loci

Colocalization analyses of the uromodulin association signals at all 13 significant loci were also performed with the expression of genes in *cis* based on transcriptome-wide RNA-sequencing of 36 non-brain tissues (GTEx Project V8) (31), tubulo-interstitial and glomerular kidney tissue portions (42), as well as with circulating plasma proteins (35) in order to prioritize the most likely causal genes (Methods). Plasma proteomics captures information about the abundance, structure, and context of circulating proteins and, when integrated

with genomics, can reveal new insights into proteins that mediate genetic associations with complex traits and diseases (35, 43, 44). At least one positive colocalization with gene expression or plasma proteins was observed for most loci (**Supplementary Table 11, Supplementary Figure 7**). For example, the association of genetic variants on chromosome 19 with uromodulin colocalized with their association with SIGLEC9 protein, further supporting that genetic variation in *SIGLEC9* relates to changes in the aptamer readout of circulating uromodulin. We confirmed that the aptamer readouts of plasma uromodulin and SIGLEC9 were correlated in the ARIC study (Spearman coefficient 0.53, p<2.2e-16).

Integration of colocalization evidence with additional sources of annotation (Methods) implicated *PRKAG2*, *B4GALNT2*, and *UMOD*/*PDILT* as the genes most likely causing the association with antibody-based uromodulin, and *CFH*, *MGAT5*, *ST3GAL6*, *HLA-DRB1*, *B4GALT1*, *ABO*, *DPP7*/*MAN1B1*, *ST3GAL4*, *ASGR1*/*ASGR2*, and *SIGLEC9* for the aptamer-based readout (**Supplementary Table 12**).

Antibody- and aptamer-based measurement of circulating uromodulin differ in their approach (**Figure 4A**). Whereas the antibody-based methods quantify abundance of the circulating protein as evidenced by the *cis*-association at *UMOD/PDILT*, the aptamer-based assay may rather identify differences related to glycan marks known to be present on uromodulin and their receptors, for example because such modifications may lead to differential aptamer binding. In order to detect any shared functions, processes and pathways among the respective genes identified by each assay, enrichment analyses were performed (Methods). Terms and pathways related to protein glycosylation were highly enriched for genes identified via aptamer plasma uromodulin readout (**Figure 4B**), with the genes that drove the enrichment encoding for enzymes and receptors involved in the biosynthesis and recognition of glycans, respectively (**Figure 4C**, **Supplementary Table 13**).

B4GALNT p.Cys466Arg causes reduced enzyme function and processing

To study the possible biological effect of rs7224888, we first generated a homology-based model of B4GALNT2. The arginine insertion at p.Cys466Arg was predicted to reduce protein structural stability by three different programs (Pymol, Site Direct Mutation, Missense3D), likely as a consequence of the higher steric hindrance of arginine (**Figure 5A** and data not shown). To validate the *in silico* findings, we transfected Madin-Darby Canine Kidney (MDCK) cells with expression vectors for the two B4GALNT2 allelic variants. Two isoforms were reported for B4GALNT2, long and short (45). For this project, we employed an already described expression vector for B4GALNT2 short isoform, where cysteine 466 corresponds to residue 406. A previous study (32) suggested that the p.Cys466Arg variant affects a region that glycosyltransferases typically use to interact with their substrate, impairing B4GALNT2 activity. Thus, we tested the activity of the enzyme isoforms (wildtype [wt] and Arg406) by taking advantage of the well-established interaction between the *Dolichos biflorus* agglutinin (DBA) and the Sda antigen (46). By using a rhodamine labelled version of DBA we observed a clear signal, mostly localized on the plasma membrane, in MDCK cells expressing wt B4GALNT2, while virtually no signal could be detected for Arg406 expressing cells (**Figure 5B**), confirming absent activity. Double staining of wt B4GALNT2 and DBA confirmed the specific presence of Sda antigen only in cells expressing the enzyme (**Figure 5C**).

Western blot analysis on cells lysates showed that Arg406 B4GALNT2 had a slightly reduced molecular weight compared with wt protein. Such difference is related to different glycosylation, due to retention of Arg406 B4GALNT2 in the endoplasmic reticulum (ER). Indeed, the lower band observed for the Arg406 isoform was fully sensitive to treatment with Endo H, a deglycosylating enzyme that is specific for high-mannose, ER-type N-glycans, while

only a minor fraction of wt B4GALNT2 was cleaved by Endo H (**Figure 6A**). To substantiate this finding, we analyzed the intracellular localization of B4GALNT2 by immunofluorescence. While the wt isoform showed the expected predominant localization in the Golgi compartment, the Arg406 isoform fully colocalized with the ER marker KDEL, confirming its ER retention (**Figure 6B**, **C**). These results demonstrate that the B4GALNT2 variant p.Cys466Arg is functional and leads to loss of B4GALNT2 function and ER retention, likely due to protein misfolding.

Previous in vitro studies demonstrated that a N-acetyl-β-D-galactosaminyltransferase activity present in microsomal preparations of guinea-pig kidney transfers *N*-[^{14}C]-acetylgalactosamine to N-linked glycans of UMOD for the synthesis of Sda antigen (47). To assess whether B4GALNT2 directly acts on uromodulin under physiological conditions, we first verified their co-expression in kidney cells using real-time RT-qPCR on RNA extracted from micro-dissected mouse nephron segments and immunofluorescence on mouse and human kidney tissue (Methods). Real-time RT-qPCR demonstrated the presence of *B4GALNT2* transcript in TAL and DCT segments where uromodulin is expressed (**Figure 7A**). These data were confirmed by immunofluorescence analysis that showed a strong B4GALNT2 signal in collecting ducts (AQP2+), and a low but consistent signal in UMOD-positive cells (**Figure 7B**). B4GALNT2 expression in UMOD-positive cells was also confirmed in human kidney tissue (**Figure 7C**). To demonstrate the activity of B4GALNT2 on uromodulin glycosylation, we generated MDCK clones stably expressing uromodulin +/- B4GALNT2. Western blot analysis showed that uromodulin has a slightly increased molecular weight in lysates of B4GALNT2-positive cells that is due to different protein glycosylation, as demonstrated by removal of N-glycans through PNGase F treatment (**Figure 7D**).

Finally, we excluded that the association of B4GALNT2 loss-of-function with higher circulating uromodulin can be ascribed to altered immunoreactivity due to absence of the Sda antigen based on two observations: first, the quantitative, additive effect of *UMOD* variants on serum uromodulin was clearly detected regardless of the *B4GALNT2* genotype (**Figure 8A**). Second, the immunoreactivity of both ELISA antibodies did not differ from a reference antibody in detecting increasing amounts of uromodulin produced by cells expressing or not expressing B4GALNT2, hence carrying or not the Sda antigen glycan moiety (**Figure 8B**).

Discussion

This GWAS meta-analysis of circulating uromodulin using complementary antibody-based (N=13,985) and aptamer-based (N=18,070) assays has four principal findings: first, it identifies an upstream variant at the *UMOD*/*PDILT* locus for which integration with functional genomics data from primary human kidney tissue supports differential chromatin accessibility and transcription in cells synthesizing uromodulin as underlying its strong association with circulating and urine uromodulin, as well as CKD and hypertension. Second, shared association patterns of uromodulin-associated genes with complex traits and diseases are plentiful and place the *PRKAG2* and *UMOD* loci into the same context with respect to their associations with CKD, hypertension, and kidney stone disease. Third, the missense variant p.Cys466Arg in the uromodulin-glycosylating enzyme B4GALNT2 is a loss-of-function allele leading to higher levels of circulating uromodulin. Fourth, our study reveals enzymes that write glycan marks found on uromodulin and their receptors that may be related to uromodulin's complex glycosylation pattern, function and clearance.

Previous GWAS of circulating uromodulin quantified with antibody assays only identified the *UMOD* locus (16, 23). At the *UMOD* locus, findings were consistent with those

from urine (16) and a previous small GWAS of serum uromodulin (23), e.g., the major allele at the index SNP was associated with higher uromodulin levels. The generation of functional genomic data from kidney tissue now allowed for new insights at this locus, by providing a plausible mechanism by which an intronic variant rs77924615 in the upstream gene *PDILT* is associated with uromodulin levels (16), despite the absence of *PDILT* transcription in kidney. Evidence for accessible chromatin at the SNP's position solely in target kidney cell types for uromodulin synthesis, TAL and DCT cells, and mapping of rs77924615 and the functional *UMOD* promoter index SNP rs4293393 within the same topological associated domain, are indicative of a regulatory effect of this upstream variant on uromodulin transcription. This mechanism is also likely to underlie the reported associations of rs77924615 with urine uromodulin levels, kidney function and CKD (7).

There are several potential explanations why a previous meta-analysis of urine uromodulin levels did not detect any of the other loci identified here despite a similar sample size (16). First, uromodulin occurs as a polymer in urine but is present as a monomer in blood. In addition, the 100-fold higher levels in urine as well as the high biological variability of urine concentration and composition may preclude the detection of slight variations that can be observed in plasma. Our study supports a shared genetic basis for urine and circulating uromodulin levels, but the index variant rs77924615 explained more than 10 times as much of the uromodulin variance in the circulation compared to the urine. Thus, circulating uromodulin may be a more attractive biomarker to estimate uromodulin production in the kidney. Second, it is conceivable that receptors recognizing glycan marks present on uromodulin, potentially affecting its stability or clearance, differ between urine and the circulation. Third, previous urine studies did not use aptamer-based assays. Our results suggest that the aptamer detects genetic loci related to the writing of glycans that are present

on many secreted glycoproteins, including uromodulin, and their recognition in the circulation, rather than representing the abundance of the intact protein in blood. Future validation of the uromodulin aptamer, and investigations how the amount of uromodulin protein relates to its glycosylation patterns, are of interest. Regardless, the aptamer readout carries complementary information by delivering insights into the glycan component of this important glycoprotein. Despite the distinct set of loci detected with the two assays, there are also connections: the *ABO* encoded glycosyltransferase acts on precursor chains that are also substrates of the enzymes encoded by *ST3GAL4* and *ST3GAL6* to synthesize type 2 monosialyl-galactosylgloboside that is used for the synthesis of sialyl Lewis X antigen, and by B4GALNT2 for the synthesis of the Sda antigen (48, 49).

Concerning B4GALNT2, our functional studies demonstrate that the p.Cys466Arg variant leads to retention of the Arg466 isoform in the ER and its absence in the Golgi compartment. This is likely due to protein misfolding, as suggested by prediction analysis of the effect of the missense change on B4GALNT2 structural stability. The consequence, loss of protein function, is demonstrated by virtual absence of DBA-positivity. These data are consistent with and provide a mechanistic explanation for previous results showing that the Arg466 variant is statistically correlated with absence of the Sda antigen (32).

The role of B4GALNT2 in uromodulin glycosylation is supported by the proteins' demonstrated co-expression in TAL and DCT segments in mouse and human kidney. Moreover, we show that expression of B4GALNT2 in cells expressing uromodulin leads to addition of a glycan moiety, presumably the Sda antigen. The functional role of the Sda antigen on uromodulin is not known. Through the addition of β1,4-linked GalNAc it may hinder binding of bacterial adhesins and, hence, have a role in pathogen resistance. There are several potential explanations for how B4GALNT2 loss-of-function, i.e. absence of the Sda

antigen, may lead to higher serum levels of uromodulin as quantified by antibody. Neither our experimental nor our population study data provides evidence for an altered immunoreactivity due to absence of the Sda antigen, making this an unlikely option. It is conceivable that the presence of the Sda antigen is associated with lower stability of circulating uromodulin, as observed for von Willebrand factor (vWF) through a mechanism that depends on ASPGR activity (50). Alternatively, the absence of the Sda antigen may influence uromodulin polarized trafficking, partly redirecting the protein towards the basolateral membrane and from there to the circulation.

Common variants at the *PRKAG2* locus showed a striking similarity to those at *UMOD* with respect to shared disease association patterns. Considering that we tested for colocalization with hundreds of human traits and diseases, these mirroring patterns are extremely unlikely to result from chance. Our results imply that genetic variants at *PRKAG2* are associated with higher risk for CKD and hypertension, as well as lower risk for kidney stone disease, through the same biological context as *UMOD*. *PRKAG2* encodes for the regulatory gamma subunit of the AMP-activated protein kinase (AMPK), an enzyme with a key role in regulating multiple processes related to cellular energy metabolism. AMPK has been described to phosphorylate the kidney-specific Na^+-K^+-$2Cl^-$ cotransporter (51), the molecular link between CKD-associated *UMOD* variants and hypertension (15). However, although this connection is biologically plausible, we cannot exclude that *GALNT11*, *GALNT5* or other genes or elements in the *PRKAG2* locus represent the causal link.

The loci identified through aptamer-readout point towards the importance of uromodulin glycosylation in general and sialylation in particular. SIGLEC9 receptor binding to uromodulin depends on the presence of terminal sialic acid on uromodulin N-glycans (37). ASGPR, mainly expressed in liver hepatocytes and to lower extent in several cell types

including monocytes (52), mediates binding, endocytosis and degradation of glycoproteins with decreased sialylation and exposed terminal galactose or N-acetylgalactosamine (GalNAc) residues (52-55). An observed four-fold increase of circulating uromodulin in mice upon ablation of Asgr2 and the mannose receptor (56) as well as the association between a rare *ASGR2* variant and the aptamer uromodulin-readout in our study raise the possibility that circulating uromodulin may be another glycoprotein recognized by the ASGPR. The function of SIGLEC9 and ASGPR intersects with that of the ST3Gal family of sialyltransferases, which add sialic acid in alpha2,3 linkage to terminal galactose, thereby generating potential SIGLEC9 ligands, while masking ASPGR ligands (53, 57-59). Overall, these data point at a clear functional relationship between SIGLEC9 and ASGPR, potentially involved in uromodulin-mediated immunomodulatory signaling and clearance, and ST3GAL4 and 6, modulating such activities through sialylation of glycan moieties.

Our study also provides insights of potential clinical relevance. It indicates that a common genetic basis of urine uromodulin levels with higher risk of CKD and hypertension extends to circulating uromodulin levels, and identifies kidney cell-type specific regulation of uromodulin expression as a mechanism. Interventions aimed at reducing uromodulin synthesis can therefore be expected to have concomitant effects on both urine and serum levels of the protein, which may be of importance given its emerging systemic relevance. The association of genetic variants at *PRKAG2* with higher risk for CKD and hypertension as well as lower risk for kidney stone disease suggests a biological link between PRKAG2 and UMOD and suggest that PRKAG2 represents another target to modulate uromodulin-mediated risk of CKD. Finally, therapeutic targeting of (a)sialoprotein receptors may impact on circulating uromodulin, possibly modulating cross-talk between the kidney and the innate and adaptive immune system.

In conclusion, our study provides human genetic evidence of new pathway members of uromodulin and delivers novel insights into its determinants and systemic role in the circulation.

Methods

Study design and participants

Seven prospective studies participated in the genome-wide analyses of serum/plasma uromodulin levels (**Supplementary Table 1**): the Cardiovascular Health Study (CHS) (60), German Chronic Kidney Disease (GCKD) study (61), the Cooperative Health Research in the Region Augsburg (KORA) study (62), the LUdwigshafen RIsk and Cardiovascular Health (LURIC) study (63), the Outcome Reduction with an Initial Glargine Intervention (ORIGIN) trial (64), the Atherosclerosis Risk in Communities (ARIC) (65) study and the Fenland study (66). Each study contributed data from EA participants. Data from AA and HIS participants were contributed by the CHS study (AA) and the ORIGIN trial (AA, HIS). The ARIC, CHS, Fenland and KORA studies have a population-based design, the GCKD study recruited patients with chronic kidney disease, the LURIC study recruited patients with cardiovascular disease, and the ORIGIN trial patients with impaired glucose tolerance or early type 2 diabetes. Demographic information including age and sex was collected using standardized procedures. The estimated glomerular filtration rate (eGFR) was calculated from IDMS-traceable serum creatinine measurements using the four-variable CKDEpi equation (7).

Genotyping and imputation

Details about genotyping and imputation in each of the six studies are provided in **Supplementary Table 2**. In brief, all samples were genotyped for genome-wide single nucleotide polymorphisms (SNPs) using Illumina or Affymetrix arrays and called using commercial software. Variant-level quality control and cleaning included removal for low call rate and deviation from Hardy-Weinberg equilibrium. Genotype imputation was then performed using phasing and imputation software, based on the Trans-Omics for Precision

Medicine (TOPMED) haplotypes version r2 (ARIC) or the Haplotype Reference Consortium (HRC) haplotypes version r1.1 (all other studies) reference panels.

Uromodulin quantification

The CHS, GCKD, KORA, and LURIC studies quantified uromodulin from serum using a commercial ELISA (23, 30, 67, 68) (Euroimmun, Medizinische Labordiagnostika AG). The assay is based on a colorimetric sandwich immunoassay, in which the capture antibody was a mouse monoclonal antibody against human uromodulin, and the detection antibody was a biotinylated mouse monoclonal antibody against human uromodulin. The intra-assay precision of the assay at 30–214 ng/mL was 1.8–3.2%, and the inter-assay precision at 35–228 ng/mL 6.6% to 7.8% (21). The ORIGIN study measured serum uromodulin using an immunoassay that is part of the Human DiscoveryMAP panel (Myriad RBM Inc.) with a biotinylated polyclonal detection antibody against human uromodulin (69). In the ARIC and Fenland study, uromodulin was quantified as part of plasma proteome profiling using the SOMAscan assay (Seq-ID 9451-20), a multiplexed modified DNA-based aptamer technology by SomaLogic as described previously (43). The aptamer was raised against amino acids 24-611 of human uromodulin (NP_001008390.1). Its signal-to-noise ratio was 114, and its intra- and inter-assay coefficients of variation were 3.2% and 3.6%, respectively.

GWAS of serum uromodulin levels and meta-analyses

Each study performed four sets of GWAS according to a pre-specified analysis plan: a primary analysis, in which inverse normal rank transformed age-, sex-, and eGFR-adjusted residuals of uromodulin levels were used as the dependent variable and regressed on genotypes as the predictor, controlling for principal components. In addition, three secondary analyses were

carried out: the primary analysis was repeated separately for men and women (without adjustment for sex and only when at least 50 participants were available), and a sex-combined analysis was performed in which the residuals were not adjusted for eGFR. Studies with participants of different ancestries performed separate GWAS for each ancestry group. Sex-stratified GWAS of chromosome X markers, assuming X inactivation and an additive model, were only available from the CHS and GCKD studies. Genome-wide summary statistics were collected in a prespecified format and uploaded to a central server for meta-analysis.

Prior to meta-analysis, GWAS summary statistics from individual studies were subjected to thorough quality control using GWAtoolbox (70). The variant identifiers in each GWAS file were harmonized to the format chromosome:position:ref:alt, where ref and alt are the REF and ALT alleles in the HRC r1.1 reference site file. Genome-wide summary statistics were combined for studies with antibody-based uromodulin quantification using inverse-variance weighted meta-analysis of effect estimates. Previous meta-analyses of uromodulin levels in urine quantified from ELISA and the RBM immunoassay showed little heterogeneity and comparable results when using inverse variance weighted and sample size weighted meta-analysis (16). The primary meta-analysis was a trans-ethnic analysis, combining data from EA, AA, and HIS participants, and performed using metal (71). Sex-specific summary statistics of chromosome X were combined via meta-analysis. Genomic control was applied to individual GWAS files when the inflation parameter was >1. Genome-wide significance was defined as p<5e-08. EA-specific meta-analyses were also performed, as EA participants were the largest subsample. SNP-based heritability was estimated using LDSC v1.0.1 with the option --h2. Pre-computed LD scores from 1000 Genomes European data were used as reference. Input files for LDSC were GWAS summary statistics from the primary association analyses filtered for MAF >0.01 (antibody-based assay) or MAF >0.005 (aptamer-based assay).

Downstream characterization of GWAS meta-analysis results

Several complementary approaches to characterize genetic loci identified through genome-wide screens of circulating uromodulin were employed, with detailed Methods described in the **Supplementary Material**. These include (i) associations with urine uromodulin levels, (ii) annotation, enrichment analyses and functional genomics, (iii) independent SNP selection, statistical fine mapping and credible set annotation, (iv) colocalization with gene expression, plasma protein levels, biomarkers and diseases, (v) PheWAS of serum uromodulin levels, as well as (vi) gene-by-gene interaction analyses.

3D modelling of B4GALNT2 and prediction of the effect of p.Cys466Arg

The sequence of B4GALNT2 isoform 2 (Uniprot Q8NHY0-2) was analyzed in PFAM (http://pfam.xfam.org) to map functional domains. The region 254-464 containing the glycosyltransferase domain was analyzed in iTasser (72) to generate a homology-based 3D model. The effect of the p.Cys466Arg substitution was analyzed in Pymol (version 2.3.4) (Schrödinger) with the Mutagenesis Wizard function, in Site Direct Mutator (73) and in Missense3D (74).

Experimental studies of B4GALNT2

Detailed information regarding functional studies of B4GALNT2 (constructs, cell lines and culture conditions, protein extracts and Western blot analysis, RNA isolation and quantitative RT-PCR, immunofluorescence analysis, antibodies) are reported in Methods section in the **Supplementary Material**. Protein and RNA extracts, Western blot, quantitative RT-PCR and immunofluorescence analyses were carried out as previously described (9, 75).

Study approval

All participants of CHS, GCKD, KORA, LURIC, ORIGIN, ARIC and Fenland studies provided written informed consent, and the studies were approved by their local ethics committees as outlined in their respective design publications (60-66). The use of human kidney biopsies in experimental studies of B4GALNT2 has been approved by the UCLouvain Ethical Review Board. Human kidney tissues used in functional genomics were collected in deidentified fashion through Northwest Biotrust at the University of Washington Medical Center (Seattle, WA) with local IRB approval (Study 1297).

All mouse experiments were performed in accordance with the ethical guidelines at University of Zurich (Zurich, Switzerland) and the legislation of animal care and experimentation of Canton Zurich (Kanton Zürich Gesundheitsdirektion Veterinäramt; protocol ZH049/17).

Acknowledgements

We thank Rory Wilson for sharing code to generate a heatmap figure. We are grateful for the willingness of the patients to participate in the GCKD study. The enormous effort of the study personnel of the various regional centers is highly appreciated. We thank the large number of nephrologists who provide routine care for the patients and collaborate with the GCKD study. We thank the Fenland Study Investigators, Fenland Study Co-ordination team and the Epidemiology Field, Data and Laboratory teams. We are grateful to Prof. Fabio Dall'Olio (University of Bologna) for providing the B4GALNT2 wt expression construct and useful suggestions, to Dr. Céline Schaeffer (IRCCS Ospedale San Raffaele) for generating transduced MDCK cells expressing HA-tagged uromodulin. We thank ALEMBIC, the advanced microscopy laboratory of IRCCS Ospedale San Raffaele and Università Vita-Salute San Raffaele, for excellent technical assistance. We thank Victor Herbst, Matthias Block and Wolfgang Schlumberger, Institute of Experimental Immunology, Euroimmun for providing the uromodulin assay and quality control. We thank Marta Mariniello for help with the studies of micro-dissected mouse nephron sections. We thank Sibylle Hess (Sanofi), who suggested to measure uromodulin in the biomarker sub-study of the Origin trial and who also coordinated the biomarker screen at Myriad RBM Inc.

M.C. is supported by a Canadian Institutes of Health Research (CIHR) Frederick Banting and Charles Best Canada Graduate Scholarships Doctoral Award. Y.C was supported by the Chinese Scholarship Council. C.H. is supported by an MRC Human Genetics Unit programme grant 'Quantitative traits in health and disease' (U. MC_UU_00007/10). The work of A.K. was supported by German Research Foundation (DFG) grant KO 3598/5-1, DFG - Project-ID 431984000 - SFB 1453, and DFG CIBSS – EXC-2189 – Project ID 390939984. C.L. and M.P. are funded by the Medical Research Council (MC_UU_00006/1 - Aetiology and Mechanisms). L.R.

is supported by the Italian Society of Nephrology (SIN) under the “Adotta un progetto di ricerca” program, and the Italian Ministry of Health (grant RF-2016-02362623). The work of Pa.S. was funded by the EQUIP Program for Medical Scientists, Faculty of Medicine, University of Freiburg. C.T. obtained further support from the Deutsche Diabetes Gesellschaft (DDG). The work of M.W. was supported by DFG Project-ID 431984000 - SFB 1453. The work of U.T.S. was supported by the German Federal Ministry of Education and Research (BMBF) within the framework of the e:Med research and funding concept (grant 01ZX1912B). The work of O.D. and G.S. is supported by the Swiss National Science Foundation (project grant 310030_189044) and the Swiss National Centre of Competence in Research (NCCR), Kidney Control of Homeostasis (Kidney.CH).

The ARIC study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services, under Contract nos. (HHSN268201700001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700004I, HHSN268201700005I). The authors thank the staff and participants of the ARIC study for their important contributions. Funding was also supported by R01HL087641, R01HL059367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. SomaLogic Inc. conducted the SomaScan assays in exchange for use of ARIC data. This work was supported in part by NIH/NHLBI grant R01 HL134320. This CHS research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086, 75N92021D00006; and NHLBI grants U01HL080295, R01HL087652, R01HL105756,

R01HL103612, R01HL120393, and U01HL130114 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through R01AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The Fenland Study (10.22025/2017.10.101.00001) is funded by the Medical Research Council (MC_UU_12015/1). We further acknowledge support for genomics from the Medical Research Council (MC_PC_13046). Proteomic measurements in Fenland Study were supported and governed by a collaboration agreement between the University of Cambridge and Somalogic. The GCKD study was funded by the German Ministry of Research and Education (Bundesminsterium für Bildung und Forschung, BMBF, grant number 01ER0804, K.U.E.) and by the Foundation KfH Stiftung Präventivmedizin. Unregistered grants to support the study were provided by Bayer, Fresenius Medical Care and Amgen. Genotyping was supported by Bayer Pharma AG. Uromodulin measurements in GCKD were supported by the Swiss National Centre of Competence in Research Kidney Control of Homeostasis program and the Swiss National Science Foundation grant 31003A_169850. The KORA study was supported by the German Federal Ministry of Education and Research to the German Center for Diabetes Research (DZD), and by the Bavarian State Ministry of Health and Care through the research project DigiMed Bayern (www.digimed-bayern.de). The KORA study was initiated and financed by the Helmholtz Zentrum München – German Research Center for

Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC-Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ. We thank the LURIC study team who were either temporarily or permanently involved in patient recruitment as well as sample and data handling, in addition to the laboratory staff at the Ludwigshafen General Hospital and the Universities of Freiburg, Ulm, and Heidelberg, Germany. LURIC was supported by the 7th Framework Programs Atheroremo (grant agreement number 201668), RiskyCAD (grant agreement number 305739) and the H2020 Program TO_AITION (grant agreement number 848146) of the European Union. Genetic analysis of ORIGIN were sponsored by Sanofi and the Canadian Institutes of Health Research (award 125794 to G.P.)

Author contributions

Conception and design of the study: Y.L., Y.C., F.C., L.R., A.K.; Statistical analysis: Y.L., Y.C., M.R.C., M.P., N.Q.H.N., N.S., M.L.B., M.E.K., B.G., Pe.S., I.S., Pa.S., M.W., S.A., L.R., A.K.; Management of an individual study: M.E.K., M.P., J.A.B., M.E.G., C.H., U.T.S., B.K.K., F.K., A.P., J.S., D.S., C.T., KU.E., C.G., E.B., B.M.P., J.C., C.L., B.Y.; Measurement of uromodulin: J.S., D.S., C.T., W.M., G.P., J.E.S.; Bioinformatic data integration: Y.L., Y.C., M.E.K., N.S., C.B.J., M.W.; Interpretation of results: Y.L., Y.C., F.C., G.S., P.J.O., S.A., O.D., L.R., A.K.; Experimental studies: F.C., G.S., S.A., O.D., L.R.; Initial manuscript draft: Y.L., Y.C., F.C., G.S., S.A., L.R., A.K.; Critical revision of the manuscript: all authors. The order of the co-first authors was determined based on the overall analytical and conceptual contribution to the project.

Competing interest statement

Marcus E. Kleber is employed by SYNLAB Holding Deutschland GmbH. Winfried März reports grants from Siemens Healthineers, grants and personal fees from Aegerion Pharmaceuticals, grants and personal fees from AMGEN, grants from AstraZeneca, grants and personal fees from Sanofi, grants and personal fees from Amryt Pharmaceuticals, grants and personal fees from BASF, grants and personal fees from Abbott Diagnostics, grants and personal fees from Numares AG, grants and personal fees from Berlin-Chemie, grants and personal fees from Akzea Therapeutics, grants from Bayer Vital GmbH, grants from bestbion dx GmbH, grants from Boehringer Ingelheim Pharma GmbH Co KG, grants from Immundiagnostik GmbH, grants from Merck Chemicals GmbH, grants from Novartis Pharma GmbH, grants from Olink Proteomics, grants and personal fees from AMGEN, personal fees from Novartis Pharma, personal fees from Vifor Pharma, other from SYNLAB Holding Deutschland GmbH, outside the submitted work. Bruce M. Psaty serves on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. Jürgen E. Scherberich has a patent (1/3) at the Charité-Universitätsmedizin Berlin. Dominik Steubl is currently employed by Boehringer Ingelheim International GmbH.

References

1. Eckardt KU, Coresh J, Devuyst O, Johnson RJ, Kottgen A, Levey AS, et al. Evolving importance of kidney disease: from subspecialty to global health burden. *Lancet.* 2013;382(9887):158-69.
2. Matsushita K, Coresh J, Sang Y, Chalmers J, Fox C, Guallar E, et al. Estimated glomerular filtration rate and albuminuria for prediction of cardiovascular outcomes: a collaborative meta-analysis of individual participant data. *Lancet Diabetes Endocrinol.* 2015;3(7):514-25.
3. GBD 2016 Causes of Death Collaborators. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet.* 2017;390(10100):1151-210.
4. Kottgen A, Glazer NL, Dehghan A, Hwang SJ, Katz R, Li M, et al. Multiple loci associated with indices of renal function and chronic kidney disease. *Nat Genet.* 2009;41(6):712-7.
5. Kanai M, Akiyama M, Takahashi A, Matoba N, Momozawa Y, Ikeda M, et al. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat Genet.* 2018;50(3):390-400.
6. Morris AP, Le TH, Wu H, Akbarov A, van der Most PJ, Hemani G, et al. Trans-ethnic kidney function association study reveals putative causal genes and effects on kidney-specific disease aetiologies. *Nat Commun.* 2019;10(1):29.
7. Wuttke M, Li Y, Li M, Sieber KB, Feitosa MF, Gorski M, et al. A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat Genet.* 2019;51(6):957-72.
8. Schaeffer C, Devuyst O, and Rampoldi L. Uromodulin: Roles in Health and Disease. *Annu Rev Physiol.* 2021;83:477-501.
9. Tokonami N, Takata T, Beyeler J, Ehrbar I, Yoshifuji A, Christensen EI, et al. Uromodulin is expressed in the distal convoluted tubule, where it is critical for regulation of the sodium chloride cotransporter NCC. *Kidney Int.* 2018;94(4):701-15.
10. Weiss GL, Stanisich JJ, Sauer MM, Lin CW, Eras J, Zyla DS, et al. Architecture and function of human uromodulin filaments in urinary tract infections. *Science.* 2020;369(6506):1005-10.
11. Padmanabhan S, Melander O, Johnson T, Di Blasio AM, Lee WK, Gentilini D, et al. Genome-wide association study of blood pressure extremes identifies variant near UMOD associated with hypertension. *PLoS Genet.* 2010;6(10):e1001177.
12. Hoffmann TJ, Ehret GB, Nandakumar P, Ranatunga D, Schaefer C, Kwok PY, et al. Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nat Genet.* 2017;49(1):54-64.
13. Tin A, Marten J, Halperin Kuhns VL, Li Y, Wuttke M, Kirsten H, et al. Target genes, variants, tissues and transcriptional pathways influencing human serum urate levels. *Nat Genet.* 2019;51(10):1459-74.
14. Gudbjartsson DF, Holm H, Indridason OS, Thorleifsson G, Edvardsson V, Sulem P, et al. Association of variants at UMOD with chronic kidney disease and kidney stones-role of age and comorbid diseases. *PLoS Genet.* 2010;6(7):e1001039.

15. Trudu M, Janas S, Lanzani C, Debaix H, Schaeffer C, Ikehata M, et al. Common noncoding UMOD gene variants induce salt-sensitive hypertension and kidney damage by increasing uromodulin expression. *Nat Med.* 2013;19(12):1655-60.
16. Olden M, Corre T, Hayward C, Toniolo D, Ulivi S, Gasparini P, et al. Common variants in UMOD associate with urinary uromodulin levels: a meta-analysis. *J Am Soc Nephrol.* 2014;25(8):1869-82.
17. Hart TC, Gorry MC, Hart PS, Woodard AS, Shihabi Z, Sandhu J, et al. Mutations of the UMOD gene are responsible for medullary cystic kidney disease 2 and familial juvenile hyperuricaemic nephropathy. *J Med Genet.* 2002;39(12):882-92.
18. Devuyst O, Olinger E, Weber S, Eckardt KU, Kmoch S, Rampoldi L, et al. Autosomal dominant tubulointerstitial kidney disease. *Nat Rev Dis Primers.* 2019;5(1):60.
19. Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, et al. A comprehensive map of molecular drug targets. *Nat Rev Drug Discov.* 2017;16(1):19-34.
20. Steubl D, Buzkova P, Ix JH, Devarajan P, Bennett MR, Chaves PHM, et al. Association of serum and urinary uromodulin and their correlates in older adults-The Cardiovascular Health Study. *Nephrology (Carlton).* 2020;25(7):522-6.
21. Steubl D, Block M, Herbst V, Nockher WA, Schlumberger W, Satanovskij R, et al. Plasma Uromodulin Correlates With Kidney Function and Identifies Early Stages in Chronic Kidney Disease Patients. *Medicine (Baltimore).* 2016;95(10):e3011.
22. Scherberich JE, Gruber R, Nockher WA, Christensen EI, Schmitt H, Herbst V, et al. Serum uromodulin-a marker of kidney function and renal parenchymal integrity. *Nephrol Dial Transplant.* 2018;33(2):284-95.
23. Delgado GE, Kleber ME, Scharnagl H, Kramer BK, Marz W, and Scherberich JE. Serum Uromodulin and Mortality Risk in Patients Undergoing Coronary Angiography. *J Am Soc Nephrol.* 2017;28(7):2201-10.
24. Then C, Then H, Meisinger C, Heier M, Peters A, Koenig W, et al. Serum uromodulin is associated with but does not predict type 2 diabetes in elderly KORA F4/FF4 study participants. *J Clin Endocrinol Metab.* 2019.
25. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet.* 2016;48(10):1279-83.
26. Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature.* 2021;590(7845):290-9.
27. Okajima T, Fukumoto S, Miyazaki H, Ishida H, Kiso M, Furukawa K, et al. Molecular cloning of a novel alpha2,3-sialyltransferase (ST3Gal VI) that sialylates type II lactosamine structures on glycoproteins and glycolipids. *J Biol Chem.* 1999;274(17):11479-86.
28. van Rooijen JJ, Voskamp AF, Kamerling JP, and Vliegenthart JF. Glycosylation sites and site-specific glycosylation in human Tamm-Horsfall glycoprotein. *Glycobiology.* 1999;9(1):21-30.
29. Then C, Then HL, Lechner A, Thorand B, Meisinger C, Heier M, et al. Serum uromodulin and risk for cardiovascular morbidity and mortality in the community-based KORA F4 study. *Atherosclerosis.* 2020;297:1-7.
30. Steubl D, Schneider MP, Meiselbach H, Nadal J, Schmid MC, Saritas T, et al. Association of Serum Uromodulin with Death, Cardiovascular Events, and Kidney Failure in CKD. *Clin J Am Soc Nephrol.* 2020;15(5):616-24.

31. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science.* 2020;369(6509):1318-30.
32. Stenfelt L, Hellberg A, Moller M, Thornton N, Larson G, and Olsson ML. Missense mutations in the C-terminal portion of the B4GALNT2-encoded glycosyltransferase underlying the Sd(a-) phenotype. *Biochem Biophys Rep.* 2019;19:100659.
33. Dall'Olio F, Malagolini N, Chiricolo M, Trinchera M, and Harduin-Lepers A. The expanding roles of the Sd(a)/Cad carbohydrate antigen and its cognate glycosyltransferase B4GALNT2. *Biochim Biophys Acta.* 2014;1840(1):443-53.
34. Qiu C, Huang S, Park J, Park Y, Ko YA, Seasock MJ, et al. Renal compartment-specific genetic variation analyses identify new pathways in chronic kidney disease. *Nat Med.* 2018;24(11):1721-31.
35. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic atlas of the human plasma proteome. *Nature.* 2018;558(7708):73-9.
36. Han X, Ong JS, An J, Hewitt AW, Gharahkhani P, and MacGregor S. Using Mendelian randomization to evaluate the causal relationship between serum C-reactive protein levels and age-related macular degeneration. *Eur J Epidemiol.* 2020;35(2):139-46.
37. Patras KA, Coady A, Olson J, Ali SR, RamachandraRao SP, Kumar S, et al. Tamm-Horsfall glycoprotein engages human Siglec-9 to modulate neutrophil activation in the urinary tract. *Immunol Cell Biol.* 2017;95(10):960-5.
38. Rampoldi L, Caridi G, Santon D, Boaretto F, Bernascone I, Lamorte G, et al. Allelism of MCKD, FJHN and GCKD caused by impairment of uromodulin export dynamics. *Hum Mol Genet.* 2003;12(24):3369-84.
39. Kottgen A, Pattaro C, Boger CA, Fuchsberger C, Olden M, Glazer NL, et al. New loci associated with kidney function and chronic kidney disease. *Nat Genet.* 2010;42(5):376-84.
40. Li PT, Liao CJ, Yu LC, Wu WG, and Chu ST. Localization of B4GALNT2 and its role in mouse embryo attachment. *Fertil Steril.* 2012;97(5):1206-12 e1-3.
41. Andes FT, Adam S, Hahn M, Aust O, Frey S, Grueneboom A, et al. The human sialic acid-binding immunoglobulin-like lectin Siglec-9 and its murine homolog Siglec-E control osteoclast activity and bone resorption. *Bone.* 2021;143:115665.
42. Gillies CE, Putler R, Menon R, Otto E, Yasutake K, Nair V, et al. An eQTL Landscape of Kidney Tissue in Human Nephrotic Syndrome. *Am J Hum Genet.* 2018;103(2):232-44.
43. Williams SA, Kivimaki M, Langenberg C, Hingorani AD, Casas JP, Bouchard C, et al. Plasma protein patterns as comprehensive indicators of health. *Nat Med.* 2019;25(12):1851-7.
44. Pietzner M, Wheeler E, Carrasco-Zanini J, Cortes A, Koprulu M, Worheide MA, et al. Mapping the proteo-genomic convergence of human diseases. *Science.* 2021:eabj1541.
45. Malagolini N, Santini D, Chiricolo M, and Dall'Olio F. Biosynthesis and expression of the Sda and sialyl Lewis x antigens in normal and cancer colon. *Glycobiology.* 2007;17(7):688-97.
46. Sanger R, Gavin J, Tippett P, Teesdale P, and Eldon K. Plant agglutinin for another human blood-group. *Lancet.* 1971;1(7709):1130.
47. Serafini-Cessi F, and Dall'Olio F. Guinea-pig kidney beta-N-acetylgalactosaminyltransferase towards Tamm-Horsfall glycoprotein. Requirement of sialic acid in the acceptor for transferase activity. *Biochem J.* 1983;215(3):483-9.

48. Serafini-Cessi F, Dall'Olio F, and Malagolini N. Characterization of N-acetyl-beta-D-galactosaminyltransferase from guinea-pig kidney involved in the biosynthesis of Sda antigen associated with Tamm-Horsfall glycoprotein. *Carbohydr Res.* 1986;151:65-76.
49. Montiel MD, Krzewinski-Recchi MA, Delannoy P, and Harduin-Lepers A. Molecular cloning, gene organization and expression of the human UDP-GalNAc:Neu5Acalpha2-3Galbeta-R beta1,4-N-acetylgalactosaminyltransferase responsible for the biosynthesis of the blood group Sda/Cad antigen: evidence for an unusual extended cytoplasmic domain. *Biochem J.* 2003;373(Pt 2):369-79.
50. Mohlke KL, Purkayastha AA, Westrick RJ, Smith PL, Petryniak B, Lowe JB, et al. Mvwf, a dominant modifier of murine von Willebrand factor, results from altered lineage-specific expression of a glycosyltransferase. *Cell.* 1999;96(1):111-20.
51. Fraser SA, Gimenez I, Cook N, Jennings I, Katerelos M, Katsis F, et al. Regulation of the renal-specific Na+-K+-2Cl- co-transporter NKCC2 by AMP-activated protein kinase (AMPK). *Biochem J.* 2007;405(1):85-93.
52. Harris RL, van den Berg CW, and Bowen DJ. ASGR1 and ASGR2, the Genes that Encode the Asialoglycoprotein Receptor (Ashwell Receptor), Are Expressed in Peripheral Blood Monocytes and Show Interindividual Differences in Transcript Profile. *Mol Biol Int.* 2012;2012:283974.
53. Grewal PK, Uchiyama S, Ditto D, Varki N, Le DT, Nizet V, et al. The Ashwell receptor mitigates the lethal coagulopathy of sepsis. *Nat Med.* 2008;14(6):648-55.
54. Nioi P, Sigurdsson A, Thorleifsson G, Helgason H, Agustsdottir AB, Norddahl GL, et al. Variant ASGR1 Associated with a Reduced Risk of Coronary Artery Disease. *N Engl J Med.* 2016;374(22):2131-41.
55. Weigel PH, and Yik JH. Glycans as endocytosis signals: the cases of the asialoglycoprotein and hyaluronan/chondroitin sulfate receptors. *Biochim Biophys Acta.* 2002;1572(2-3):341-63.
56. Mi Y, Coonce M, Fiete D, Steirer L, Dveksler G, Townsend RR, et al. Functional Consequences of Mannose and Asialoglycoprotein Receptor Ablation. *J Biol Chem.* 2016;291(36):18700-17.
57. Ellies LG, Ditto D, Levy GG, Wahrenbrock M, Ginsburg D, Varki A, et al. Sialyltransferase ST3Gal-IV operates as a dominant modifier of hemostasis by concealing asialoglycoprotein receptor ligands. *Proc Natl Acad Sci U S A.* 2002;99(15):10042-7.
58. Rodriguez E, Boelaars K, Brown K, Eveline Li RJ, Kruijssen L, Bruijns SCM, et al. Sialic acids in pancreatic cancer cells drive tumour-associated macrophage differentiation via the Siglec receptors Siglec-7 and Siglec-9. *Nat Commun.* 2021;12(1):1270.
59. Wisnovsky S, Mockl L, Malaker SA, Pedram K, Hess GT, Riley NM, et al. Genome-wide CRISPR screens reveal a specific ligand for the glycan-binding immune checkpoint receptor Siglec-7. *Proc Natl Acad Sci U S A.* 2021;118(5):e2015024118.
60. Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA, et al. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol.* 1991;1(3):263-76.
61. Eckardt KU, Barthlein B, Baid-Agrawal S, Beck A, Busch M, Eitner F, et al. The German Chronic Kidney Disease (GCKD) study: design and methods. *Nephrol Dial Transplant.* 2012;27(4):1454-60.
62. Holle R, Happich M, Lowel H, Wichmann HE, and Group MKS. KORA--a research platform for population based health research. *Gesundheitswesen.* 2005;67 Suppl 1:S19-25.

63. Winkelmann BR, Marz W, Boehm BO, Zotz R, Hager J, Hellstern P, et al. Rationale and design of the LURIC study--a resource for functional genomics, pharmacogenomics and long-term prognosis of cardiovascular disease. *Pharmacogenomics.* 2001;2(1 Suppl 1):S1-73.
64. Origin Trial Investigators, Gerstein H, Yusuf S, Riddle MC, Ryden L, and Bosch J. Rationale, design, and baseline characteristics for a large international trial of cardiovascular disease prevention in people with dysglycemia: the ORIGIN Trial (Outcome Reduction with an Initial Glargine Intervention). *Am Heart J.* 2008;155(1):26-32, e1-6.
65. Wright JD, Folsom AR, Coresh J, Sharrett AR, Couper D, Wagenknecht LE, et al. The ARIC (Atherosclerosis Risk In Communities) Study: JACC Focus Seminar 3/8. *J Am Coll Cardiol.* 2021;77(23):2939-59.
66. Lindsay T, Westgate K, Wijndaele K, Hollidge S, Kerrison N, Forouhi N, et al. Descriptive epidemiology of physical activity energy expenditure in UK adults (The Fenland study). *Int J Behav Nutr Phys Act.* 2019;16(1):126.
67. Steubl D, Buzkova P, Garimella PS, Ix JH, Devarajan P, Bennett MR, et al. Association of Serum Uromodulin With ESKD and Kidney Function Decline in the Elderly: The Cardiovascular Health Study. *Am J Kidney Dis.* 2019;74(4):501-9.
68. Then C, Then H, Lechner A, Huth C, Meisinger C, Heier M, et al. Serum uromodulin is inversely associated with the metabolic syndrome in the KORA F4 study. *Endocr Connect.* 2019;8(10):1363-71.
69. Sjaarda J, Gerstein HC, Yusuf S, Treleaven D, Walsh M, Mann JFE, et al. Blood HER2 and Uromodulin as Causal Mediators of CKD. *J Am Soc Nephrol.* 2018;29(4):1326-35.
70. Fuchsberger C, Taliun D, Pramstaller PP, Pattaro C, and consortium CK. GWAtoolbox: an R package for fast quality control and handling of genome-wide association studies meta-analysis data. *Bioinformatics.* 2012;28(3):444-5.
71. Willer CJ, Li Y, and Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010;26(17):2190-1.
72. Roy A, Kucukural A, and Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc.* 2010;5(4):725-38.
73. Pandurangan AP, Ochoa-Montano B, Ascher DB, and Blundell TL. SDM: a server for predicting effects of mutations on protein stability. *Nucleic Acids Res.* 2017;45(W1):W229-W35.
74. Ittisoponpisan S, Islam SA, Khanna T, Alhuzimi E, David A, and Sternberg MJE. Can Predicted Protein 3D Structures Provide Reliable Insights into whether Missense Variants Are Disease Associated? *J Mol Biol.* 2019;431(11):2197-212.
75. Schaeffer C, Santambrogio S, Perucca S, Casari G, and Rampoldi L. Analysis of uromodulin polymerization provides new insights into the mechanisms regulating ZP domain-mediated protein assembly. *Mol Biol Cell.* 2009;20(2):589-99.

Figures

Figure 1: Manhattan plot of GWAS of antibody-based and aptamer-based circulating uromodulin. The plots represent, for each SNP, the p-value from meta-analyses of GWAS of antibody-based (N=13,985, dark blue; panel **A** and of aptamer-based circulating uromodulin (N=18,070, light blue, panel **B**. The x-axis shows chromosomal location and the y-axis the -$\log_{10}$(p-value) of SNP associations with circulating uromodulin. The plots were generated using the R package EasyStrata v8.6. Meta-analyses of X-chromosomal markers did not yield any significant findings.

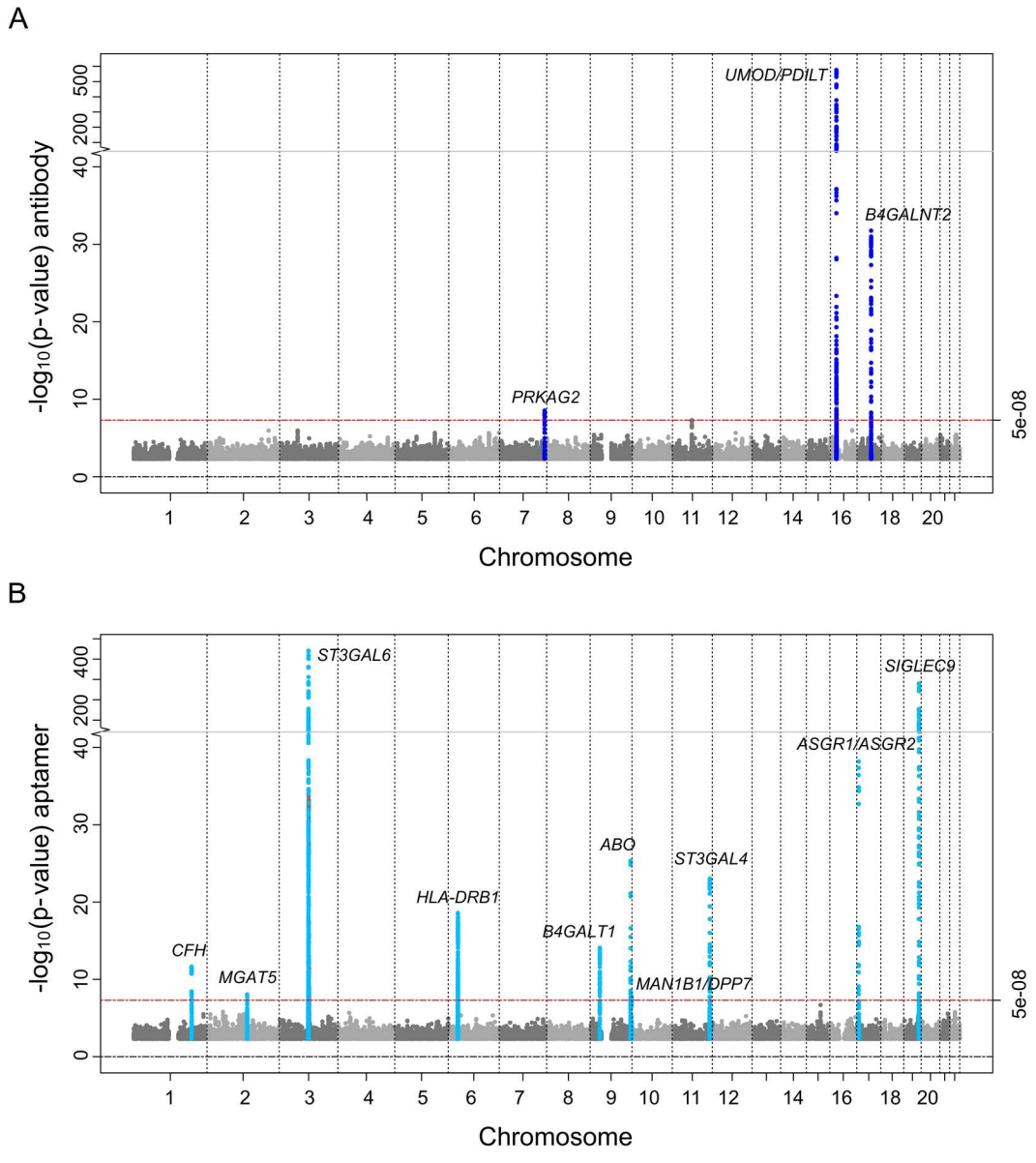


Figure 2: Functional genomic annotation of significantly associated independent variants at the *UMOD*/*PDILT* locus using gene expression and chromatin accessibility data from primary human kidney. The upper part shows the regional association plot of the *UMOD/PDILT* locus, using the two independent variants as reference SNPs. For non-reference SNPs, the extent of LD with the reference SNP with higher correlation is shown by color gradients. Genetic positions (x-axis) represent GRCh38 coordinates. Open chromatin peaks in different kidney cell type tracks based on single nuclear (sn)ATAC-seq are shown underneath the regional association plot. Gene expression and open chromatin tracks of cortex (light blue tracks) and medulla (dark green tracks) based on bulk RNA-seq and ATAC-seq are shown in the lower part as density peaks. SNPs in the two independent credible sets are marked by ticks (purple) and the 10 kb windows encompassing them indicated by the black horizontal bars. Hi-C data generated from kidney cortex was analyzed for contacts (orange arcs) between the 10kb windows encompassing the indicated SNPs with contacts closest to the causal SNPs arbitrarily shown in bold. Intervals for DomainCaller computed topology associated domains (TADs) are shown as black bars below contact arcs. A heatmap of all Hi-C contacts encompassing this region is shown at the bottom (purple). Podo: podocyte, PT: proximal tubule, LOH: loop of Henle, DCT: distal convoluted tubule, CDPC: collecting duct principal cells, CDIC: collecting duct intercalated cells.

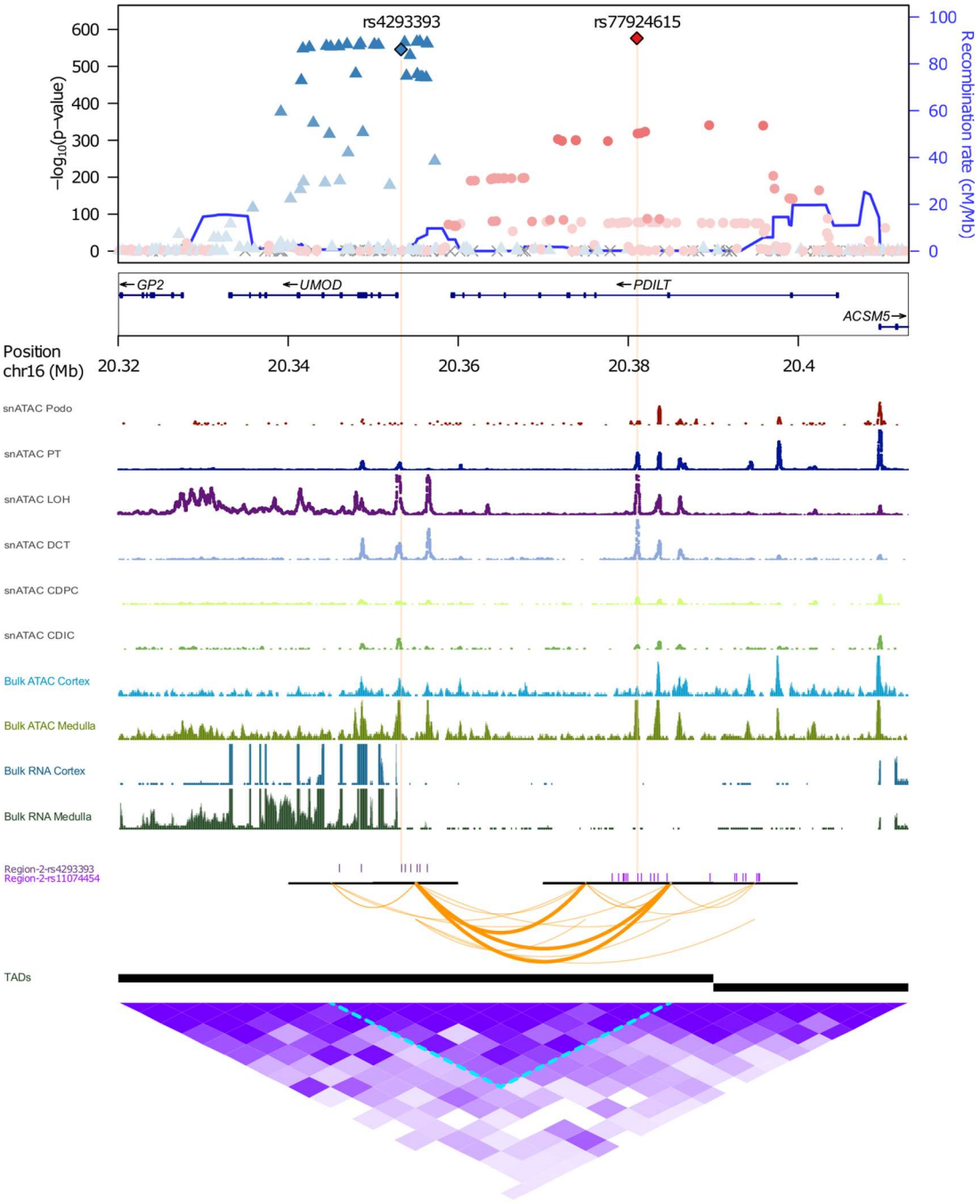

rs4293393
rs77924615
−log10(p−value)
Recombination rate (cM/Mb)
GP2
UMOD
PDILT
ACSM5
Position
chr16 (Mb)
20.32
20.34
20.36
20.38
20.4
snATAC Podo
snATAC PT
snATAC LOH
snATAC DCT
snATAC CDPC
snATAC CDIC
Bulk ATAC Cortex
Bulk ATAC Medulla
Bulk RNA Cortex
Bulk RNA Medulla
Region-2-rs4293393
Region-2-rs11074454
TADs

Figure 3: Summary of findings from colocalization of uromodulin signals with associations from GWAS of biomarkers and diseases. The colocalization analyses findings are shown in two categories, biomarkers and diseases. The x-axis indicates the index SNPs with the likely causal genes. The y-axis shows the traits for biomarkers and diseases, and only top level UKB PheCodes are shown. Within each category, horizontal lines separate different data sources (Methods). Included traits had at least one positive colocalization signal (H4>0.8, Methods). Dots are black when H4>0.8 and grey otherwise, and scaled in size to reflect the different ranges of H4. The trait-to-uromodulin effect size ratios are shown as gradient background for positive colocalization signals, with red indicating positive and blue negative changes per unit higher uromodulin levels. The colocalization with urine uromodulin (uUMOD) was based on conditional association statistics (Methods). H4: hypothesis that one shared SNP underlies the association with two traits; PP: posterior probability.

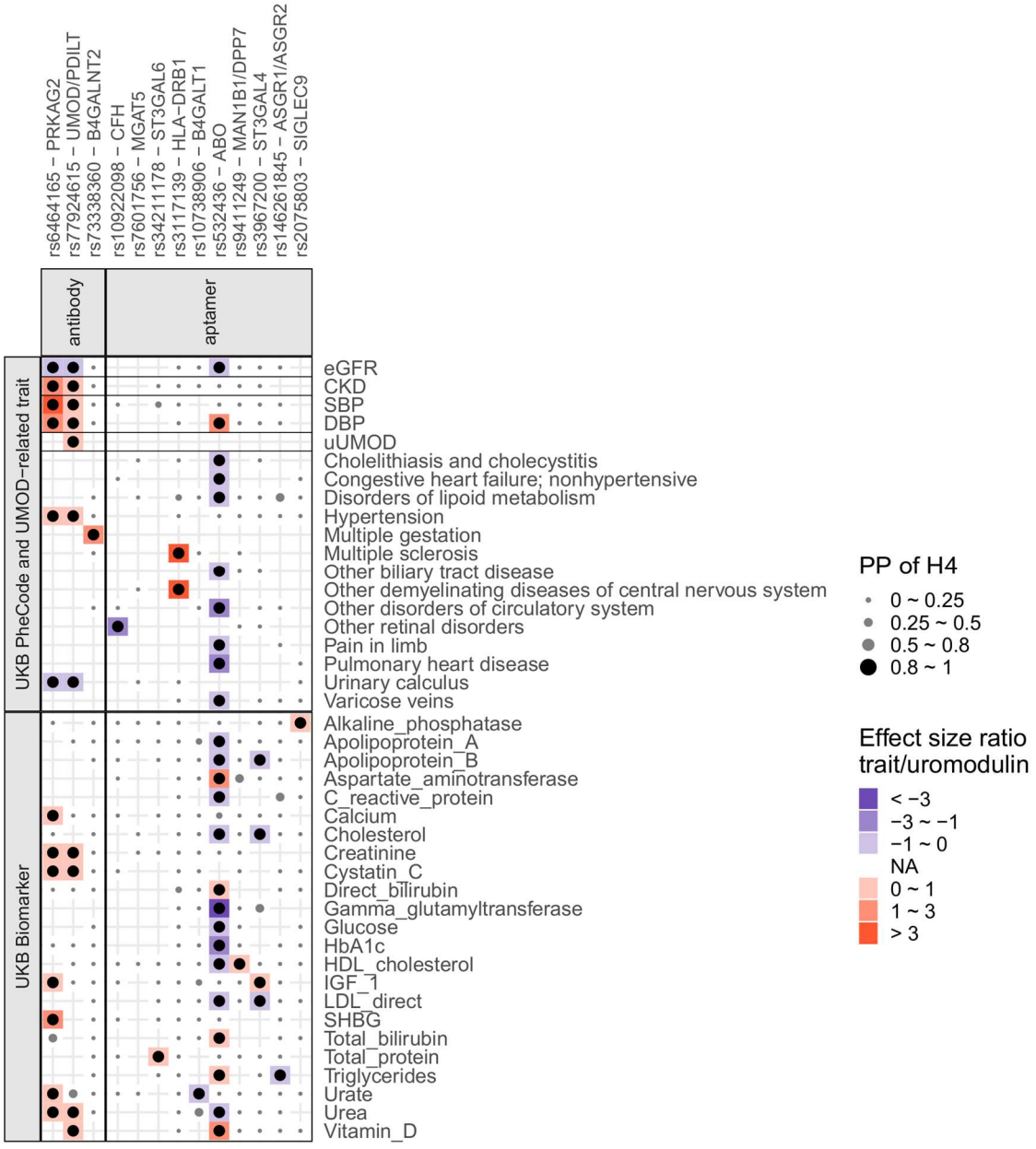


Figure 4: Biological context of genes associated with circulating uromodulin and conceptual model. A, Schematic of antibody- and aptamer-based measurement of circulating uromodulin. **B**, Dot plot shows Gene Ontology (GO) terms - grouped into three categories (BP: biological process, MF: molecular function, CC: cellular component) - and KEGG pathways enriched for uromodulin-associated genes from the aptamer assay on the y-axis. X-axis shows the proportion of the genes in the corresponding GO term or KEGG pathway. Only terms and pathways with more than two uromodulin associated-genes are displayed. The color intensity of the dots' scales with the -$\log_{10}$(Benjamini Hochberg-adjusted p-value). **C**, Conceptual model placing the most likely causal genes associated with circulating uromodulin into their biological context. Loci detected with the aptamer assay predominantly affect differential synthesis or recognition of glycan marks present on uromodulin.

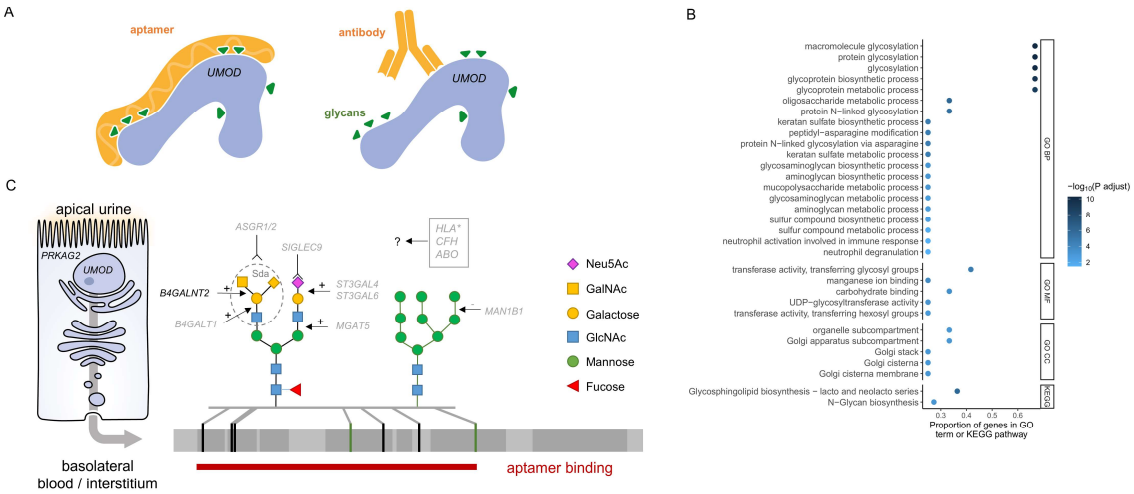


Figure 5: B4GALNT2 p.Cys466Arg is a functional allele. A, Homology-based model of the tridimensional structure of B4GALNT2 enzyme. The partial sequence of B4GALNT2 isoform 2 (Uniprot Q8NHY0-2) containing the glycosyltransferase domain (residues 254-464) was analyzed with iTasser. The top scoring model is shown. The position of cysteine 406 (corresponding to position 466 in B4GALNT2 isoform 1, Uniprot Q8NHY0-1) in the reference sequence is shown. The effect of the p.Cys406Arg substitution was analyzed in Pymol with the Mutagenesis Wizard function. For each isoform, the most likely stereoisomer is shown. Visible red disks indicate significant contacts and bumps. The arginine substitution at position 406 is predicted to increase steric clashes, destabilizing protein structure. **B**, Representative immunofluorescence analysis showing DBA signal (red) on the plasma membrane of unpermeabilized MDCK cells, transiently transfected with wt or p.Cys406Arg human B4GALNT2 (N=3). **C**, Representative immunofluorescence analysis showing wt B4GALNT2 (green) and DBA (red) in stably transfected MDCK cells. DBA signal is mostly evident in B4GALNT2-expressing cells (N=3).

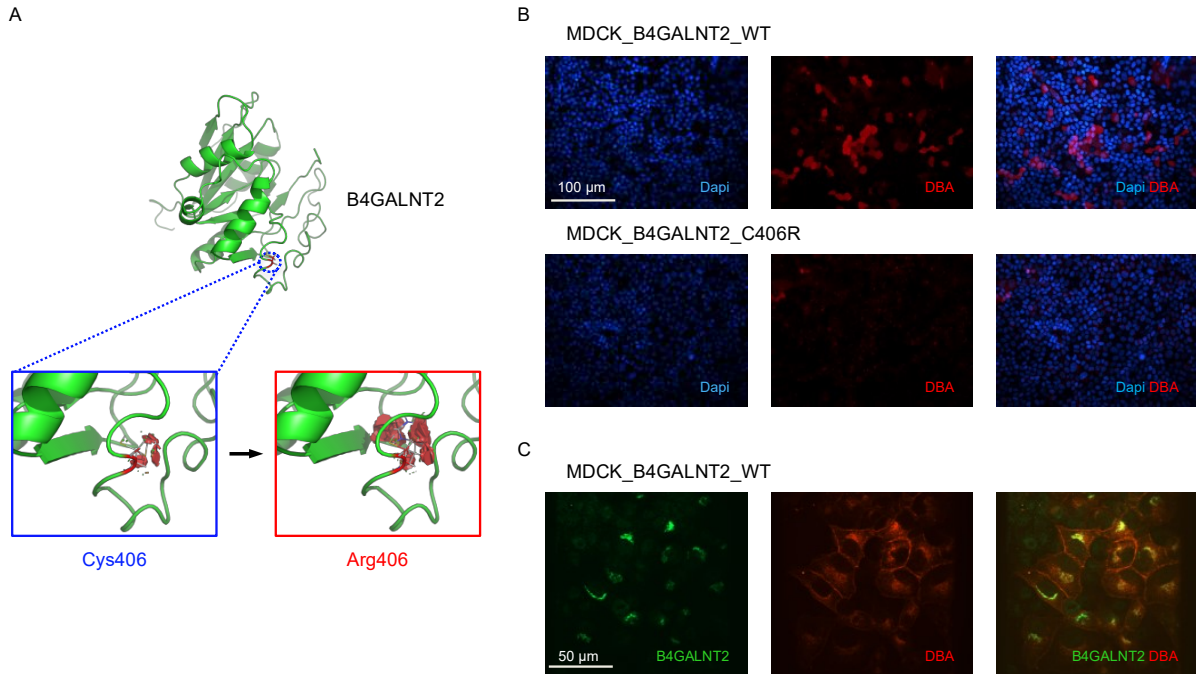


Figure 6: B4GALNT2 p.Cys466Arg is retained in the ER. A, Representative western blot analysis showing B4GALNT2 wt or p.Cys406Arg in stably transfected MDCK cells lysates, untreated or after deglycosylation with Endo H or PNGase F (N=3). **B,C**, Representative immunofluorescence analysis showing intracellular signal of wt (**B**) or p.Cys406Arg (**C**) B4GALNT2 (red) and GM130 (Golgi marker, green) or KDEL (ER marker, green) and merged pictures (N=3).

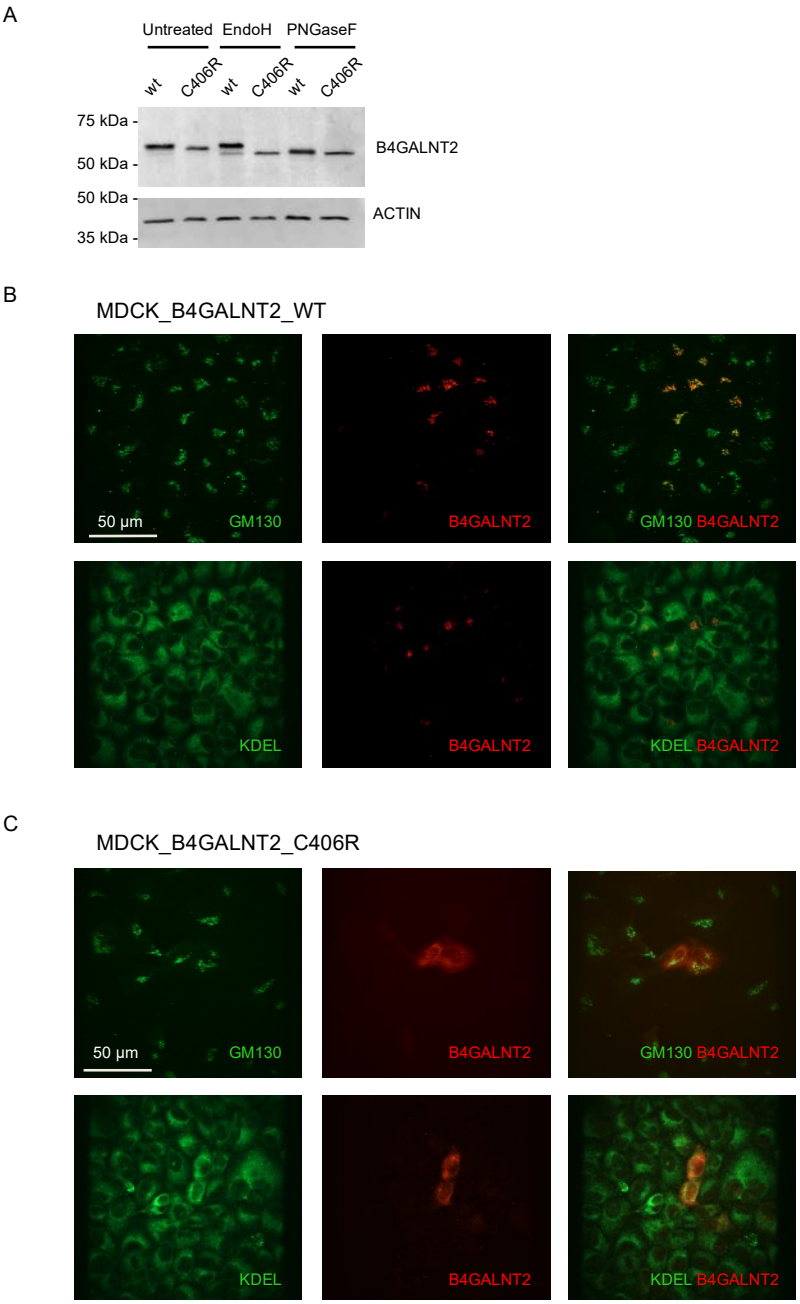


Figure 7: B4GALNT2 and uromodulin expression analysis. A, RT-qPCR analysis of *B4galnt2* expression in isolated mouse nephron segments. Glom, glomerulus; PCT, proximal convoluted tubule; PST, proximal straight tubule; TAL, thick ascending limb; DCT, distal convoluted tubule; CD, collecting duct. Bars indicate mean ± SEM. n ≥ 3 fractions. **B**, Upper panels: Immunofluorescence analysis showing UMOD (green) and B4GALNT2 (red) on paraffin-embedded kidney sections from wild-type mouse (N=2). Right panels show high magnification pictures of UMOD-positive and UMOD-negative tubules. Nuclei are counterstained with DAPI. Lower panels: Immunofluorescence analysis showing AQP2 (green) and B4GALNT2 (red) on mouse kidney, showing a strong signal in the intercalated cells of collecting ducts. Right panels show high magnification pictures of AQP2- and B4GALNT2-positive tubules. Nuclei are counterstained with DAPI. **C**, Immunofluorescence analysis showing UMOD (green) and B4GALNT2 (red) on paraffin-embedded kidney sections from a normal human kidney. Right panels show high magnification pictures of UMOD-positive and UMOD-negative tubules. Nuclei are counterstained with DAPI. **D**, Western blot analysis showing uromodulin (UMOD) and B4GALNT2 in lysates of MDCK cell clones expressing UMOD with or without B4GALNT2 (see Methods), untreated or after deglycosylation with PNGase F. Actin is shown as a normalizer.

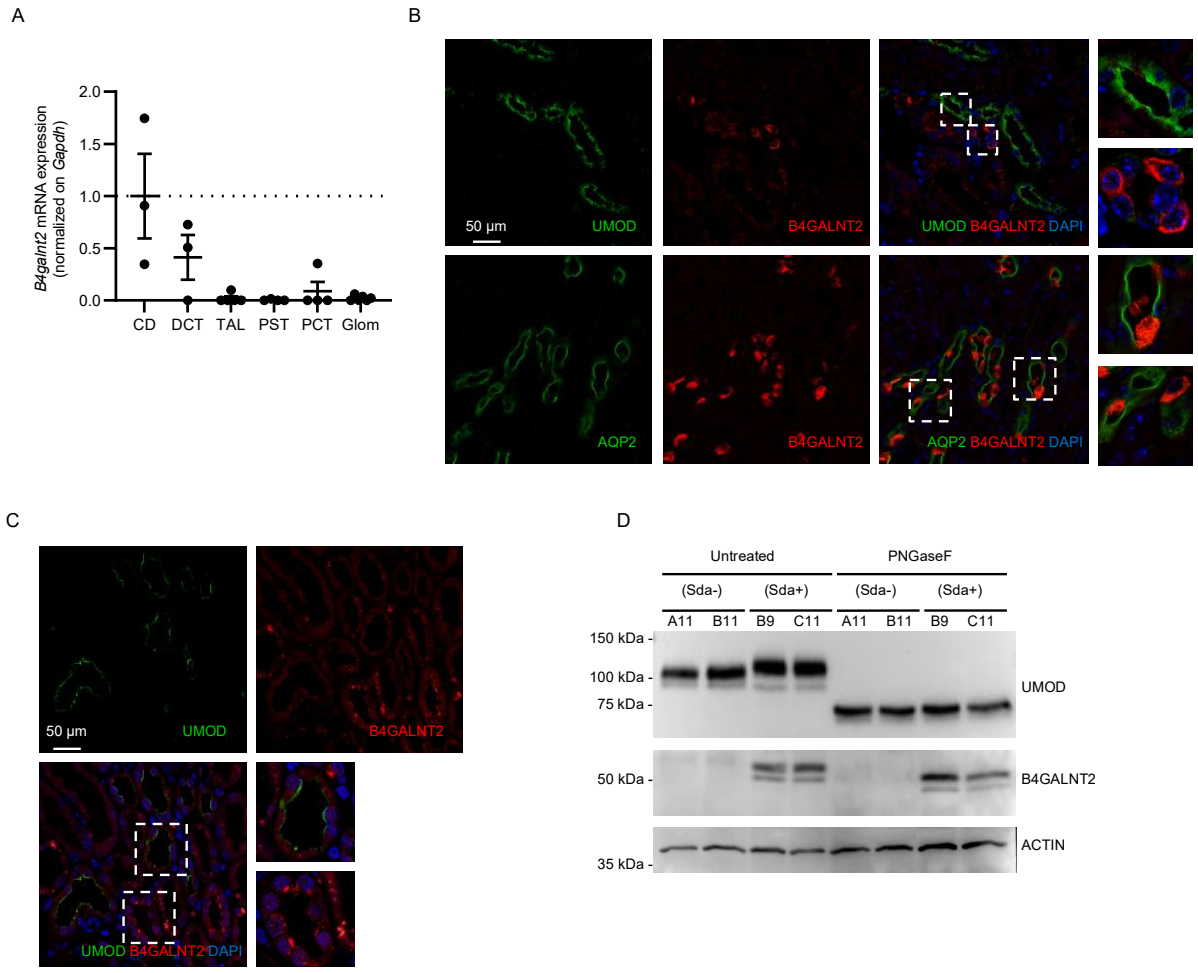
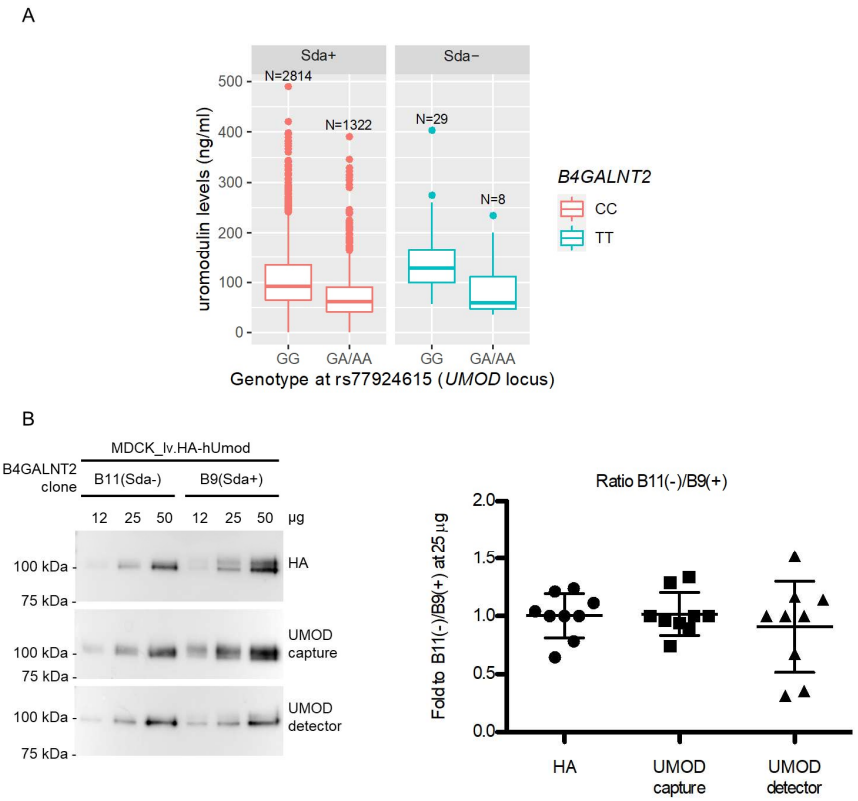


Figure 8: **ELISA-based uromodulin quantification is not affected by presence of Sda antigen.** **A**, Uromodulin serum levels in individuals carrying GG or GA/AA genotype at *UMOD* variant rs77924615, stratified for their genotype at *B4GALNT2* variant rs7224888 (CC, Sda+; TT, Sda-). The expected differences in uromodulin levels are detected regardless of the presence/absence of Sda antigen. The start and end of boxes represent the 25th and 75th percentiles of the uromodulin distribution. The line inside the box represents the median, and the dots indicate outliers above the 75% + 1.5 × (interquartile range) of uromodulin values. **B**, Representative western blot analysis (left) and relative quantification (right) of uromodulin in lysates of MDCK cells transduced with lentiviral vector expressing HA-tagged uromodulin (lv.HA-hUMOD) and stably expressing B4GALNT2 (Sda+) or not (Sda-). The immunoreactivity of 3 different antibodies (anti-HA, and the two antibodies of the Euroimmun ELISA anti-UMOD capture and anti-UMOD detector) was assessed by loading and quantifying increasing amount of cell lysate. Each value represents the ratio between B11 (negative clone) and B9 (positive clone) expressed as fold relative to the ratio measured with 25 µg of cell lysate (n=3 independent experiments). Data are represented as vertical scatter plot expressed as mean ± S.D. (one way ANOVA; p=0.66). The ratios obtained for the different antibodies are comparable, suggesting similar immunoreactivity that is not modified by the presence of the Sda antigen.



Tables

Table 1: Summary of genomic loci with genetic variants significantly associated with uromodulin levels. The SNP with the lowest association p-value (index SNP) at each of the genomic loci is shown.

Uromodulin assay	SNP	chr	Position (b37)	Locus	Effect allele / non-effect allele	Effect allele frequency	Effect	StdErr	P-value	Sample size	Proportion of explained variance
antibody	rs55791829	7	151407429	*PRKAG2*	C/G	0.72	-0.08	0.014	2.89E-09	13956	0.003
antibody	rs77924615	16	20392332	*UMOD/PDILT*	A/G	0.19	-0.76	0.015	6.39e-577	13956	0.180
antibody	rs7224888	17	47246163	*B4GALNT2*	T/C	0.90	-0.25	0.021	1.77E-32	13956	0.011
aptamer	rs10922098	1	196664651	*CFH*	T/C	0.61	0.08	0.011	2.30E-12	18070	0.003
aptamer	rs7601756	2	134893447	*MGAT5*	A/G	0.28	-0.07	0.012	8.81E-09	18070	0.002
aptamer	rs34211178	3	98383562	*ST3GAL6*	A/G	0.45	-0.46	0.010	6.88e-442	18070	0.105
aptamer	rs3117139	6	32364667	*HLA-DRB1*	A/G	0.13	0.14	0.016	2.60E-19	18070	0.005
aptamer	rs10738906	9	33135634	*B4GALT1*	T/C	0.69	-0.09	0.011	8.44E-15	18070	0.003
aptamer	rs532436	9	136149830	*ABO*	A/G	0.20	-0.14	0.013	4.74E-26	18070	0.006
aptamer	rs9411249	9	139992907	*MAN1B1/DPP7*	T/G	0.36	0.06	0.011	1.77E-08	18070	0.002
aptamer	rs3967200	11	126232385	*ST3GAL4*	T/C	0.14	-0.15	0.015	9.24E-24	18070	0.006
aptamer	rs146261845	17	7012254	*ASGR1/ASGR2*	T/C	0.01	1.05	0.080	6.74E-39	16741	0.015
aptamer	rs2075803	19	51628529	*SIGLEC9*	A/G	0.44	0.37	0.011	2.40E-280	18070	0.069

Data availability

Genome-wide summary statistics of the meta-analyses of circulating uromodulin are available at web page https://nxc-1453.imbi.uni-freiburg.de/s/gReQNkMJtkYYLxa.