Meta-GWAS Reveals Novel Genetic Variants Associated with Urinary Excretion of Uromodulin

Supplemental Material

Suppl. Appendix S1. Summary characteristics of the study cohorts.

Suppl. Figure S1. Genetic loci associated with uUMOD and uUCR.

Suppl. Figure S2. Effect size of rs12934455 and regional association plot of *UMOD-PDILT* locus from raw uromodulin levels.

Suppl. Figure S3. Effect size of rs9672398 and regional association plot of *WDR72* locus for uUCR meta-analysis.

Suppl. Figure S4. Effect size of rs13335818 and regional association plot of *UMOD-PDILT* locus from uUCR meta-analysis.

Supp. Figure S5. Manhattan plot of meta-GWAS of uUMOD and uUCR using sample size and *P* values for analysis of the 13 study cohorts.

Supp. Figure S6. Forest plot showing effect sizes of rs8067385 (*KRT40* locus) on uUMOD and uUCR meta-analysis in the 13 cohorts.

Suppl. Figure S7. Effect of *UMOD* genotype on urinary uromodulin (uUMOD and uUCR) levels.

Suppl. Figure S8. Manhattan plot showing GWAS results for uUMOD and uUMOD conditioned for rs12934455 or for rs11864909 using GS:SFHS.

Suppl. Figure S9. Candidate genes influencing the urinary excretion of uromodulin.

Suppl. Figure S10. *In situ* hybridization for *Umod*, *Krt40* and *Krt39* on mouse kidney.

Suppl. Figure S11. Immunofluorescence staining for AQP2 and KRT40 or WDR72 on mouse kidney.

Suppl. Figure S12. eQTL data for the *KRT40* variant rs8067385.

Suppl. Figure S13. Uromodulin (Z-stack) and ROMK distribution in mTAL cells following KRT40 knock-down.

Suppl. Methods

Suppl. References

Suppl. Appendix S1. Summary characteristics of the study cohorts.

The CARTaGENE study is a population-based study with over 20,000 individuals, aged 40-69 years, recruited from Quebec. A subset of 675 individuals of European descent with genotype, urinary uromodulin and creatinine measurements were used in this study.[1]

The CoLaus study is a population-based study involving more than 6,000 people of European descent aged 35–75 years from the city of Lausanne, Switzerland. Individuals were recruited between 2003 and 2006.[2]

CROATIA-Korcula is a family-based, cross-sectional study of the isolate population in the island of Korcula, Croatia that included 1687 individuals aged 18 years or over with urine samples collected.[3]

CROATIA-Split is a population-based, cross-sectional study in the Dalmatian City of Split, Croatia, that included 500 individuals aged 18 years or over with urine samples collected.[4]

CROATIA-Vis is a family-based, cross-sectional study of the isolate population in the island of Vis, Croatia that included 200 individuals aged 18 years or over with urine samples collected.[5]

The Framingham Heart Study (FHS) is a community-based family study involving three generations (1971, original cohort; 1984, offspring cohort; 2002, third generation). A subset of 2,640 participants from the offspring cohort with urinary uromodulin and eGFR levels measured were used in this analysis.[6]

The genetic and phenotypic determinants of blood pressure and other cardiovascular risk factors (GAPP) study is a population-based cohort study comprising of healthy individuals from the Principality of Liechtenstein. Individuals with any cardiovascular disease, diabetes, BMI > 35 kg/m^2 and on anti-inflammatory medications were excluded from the study. Genotype, urinary uromodulin and creatinine measurements are available for 1,518 of the participants.[7]

The German Chronic Kidney Disease (GCKD) cohort is a national cohort study. Between 2010 and 2012, 5,217 patients with chronic kidney disease under regular care of nephrologists were recruited and are since followed. At the time of recruitment, patients had an eGFR of 30-60 mL/min/1.73 m^2 or increased proteinuria (UACR >300 mg/g or UPCR >500 mg/g) with an eGFR above 60 mL/min /1.73 m^2. Urinary uromodulin measurements, coavariables and genotypes were available for 4716 individuals.[8]

Generation Scotland: Scottish Family Health Sudy (GS:SFHS): is a family-based and population-based study of individuals aged 18 years over from across Scotland with European ancestry of which 87% were born in Scotland. 7,660 volunteers had morning spot urine collected along with clinical and biochemical measures and lifestyle and health questionnaires. The participants also consented for their data to be linkable to their NHS electronic health records using the CHI number.[9]

INGI-Carlantino (INGI-CARL): Carlantino is a small village in the Province of Foggia in southern Italy. Genetic analyses of chromosome Y haplotypes as well as mitochondrial DNA show that Carlantino is a genetically homogeneous population and not only a geographically isolated village. Participant were randomly selected in a range of 15 – 90 years of age.

Subjects gave their written informed consent for participating in these studies. The project was approved by the local administration of Carlantino, the Health Service of Foggia Province, Italy, and ethical committee of the IRCCS Burlo-Garofolo of Trieste.[10]

The INGI-Val Borbera (INGI-VB) cohort is a population-based study involving individuals from the geographically isolated Borbera Valley of Northwest Italy, in Piedmont. The study was initiated in 2005 and biological samples and phenotype information were obtained from 1803 inhabitants between the ages of 18 and 102 years.[11] Subjects gave their written informed consent for participating in these studies. The project was approved by the ethical committee of IRCCS San Raffaele Hospital of Milan.

The Lothian Birth Cohort 1936 (LBC1936) mostly comprises surviving participants of the Scottish Mental Survey 1947 (SMS1947), most of whom lived in the Edinburgh City or wider Lothian area of Scotland when recruited. 1091 SMS1947 survivors were recruited into the study between 2004 and 2007, when they were approximately 70 years old. At this time they underwent a series of cognitive and physical tests. A second wave of cognitive and physical testing occurred at approximately 73 years of age at which time a urine sample was collected.[12,13]

The Viking Health Study-Shetland (VIKING) is a family-based, cross-sectional study of the isolate population in the islands of Shetland, Scotland, that included 2,089 individuals aged 18 years or over with urine samples collected along with other biochemical measurements taken and completing a health survey questionnaire.[14]
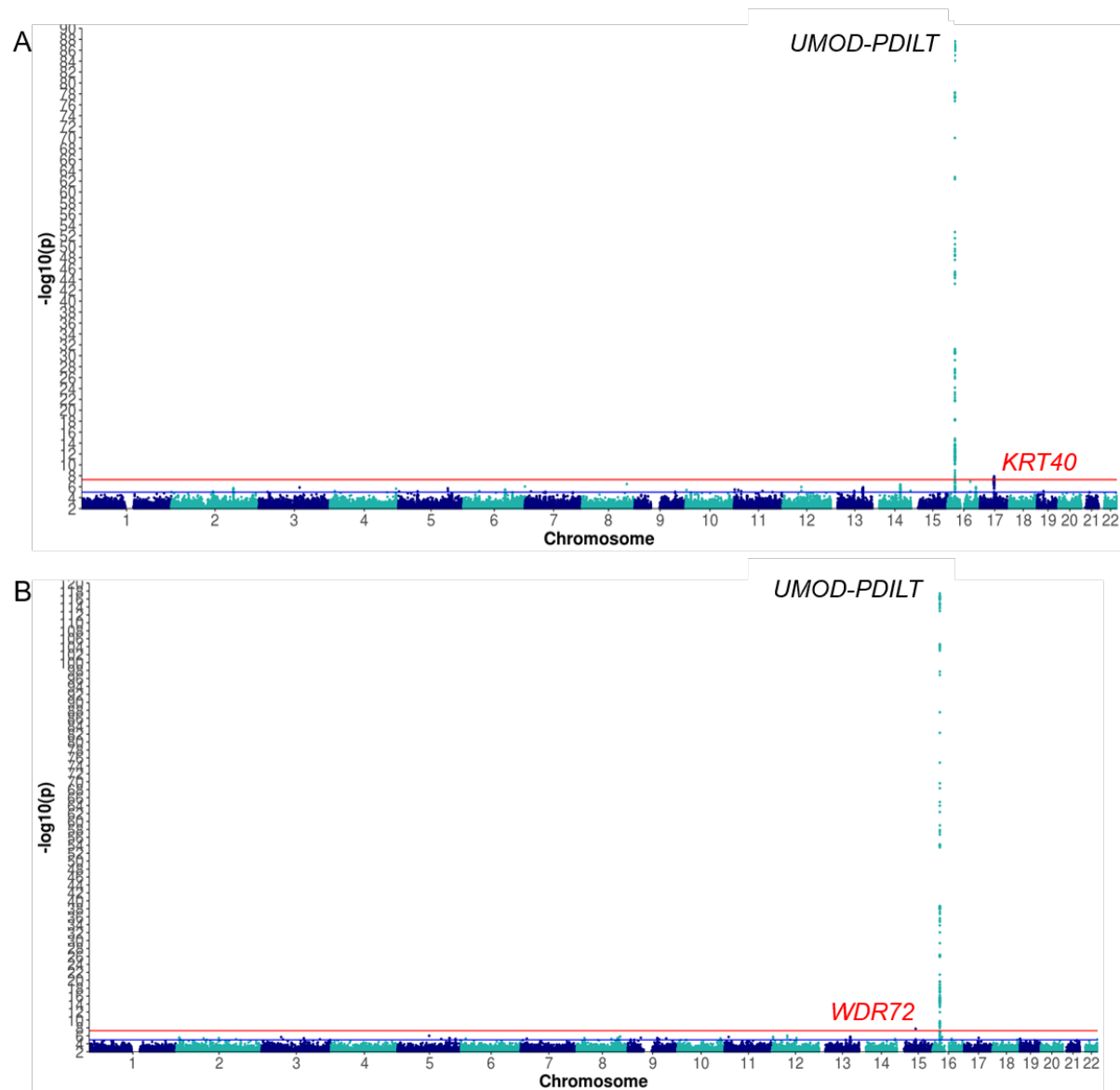
Figure S1. Genetic loci associated with uUMOD and uUCR.
Manhattan plot of meta-GWAS for **(A)** raw uromodulin (uUMOD) and **(B)** uromodulin indexed to creatinine (uUCR) in the 13 cohorts. The blue line is at the 1E-05 'suggestive' level and the red line is at the commonly used 5E-08 threshold for significance in GWAS. The two novel genome-wide significant loci are indicated in red.
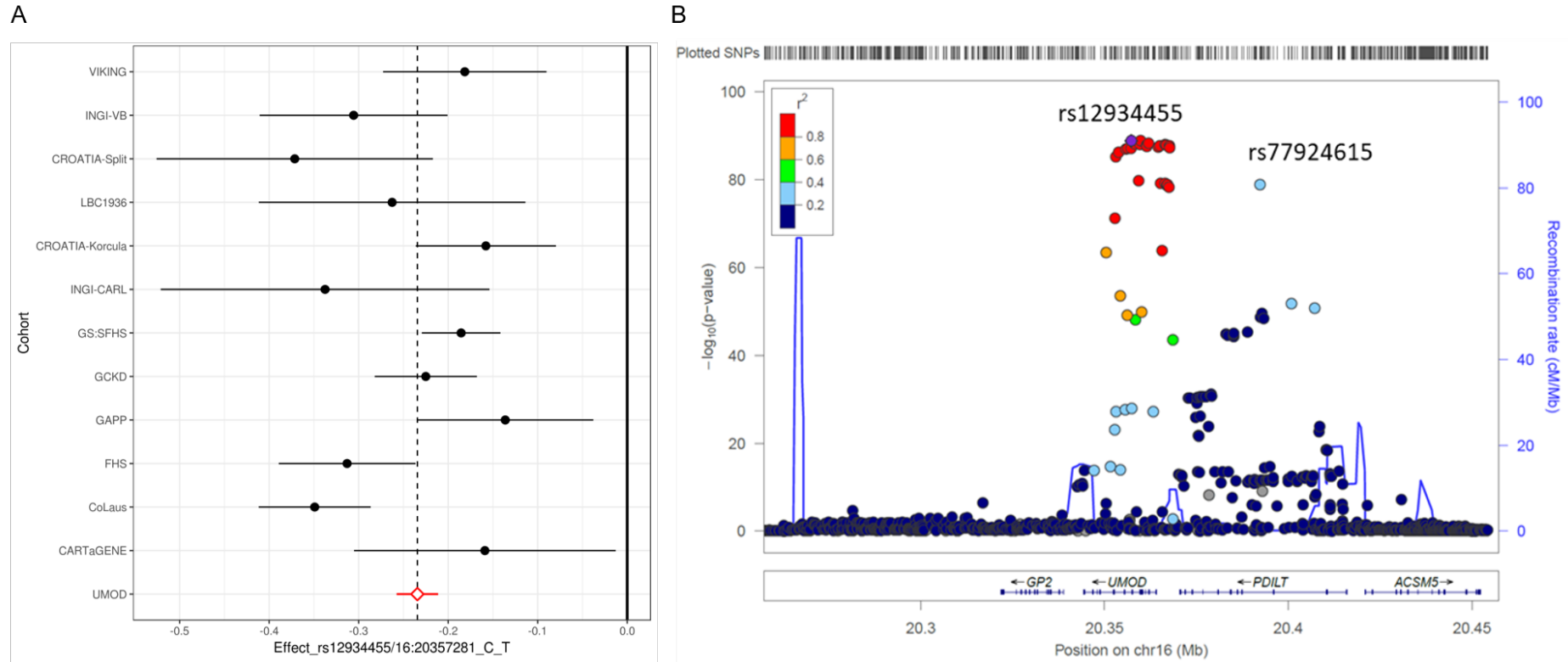
4

Figure S2. Effect size of rs12934455 and regional association plot of *UMOD-PDILT* locus from raw uromodulin levels.
(A) Forest plot showing effect sizes of rs12934455 (top SNP in *UMOD-PDILT* locus) on uUMOD meta-analyses in the 13 cohorts. The red diamond represents the average effect size of -0.2344 with a standard error of 0.0118 of the minor, T allele of rs12934455 in association with uUMOD. Effect sizes are shown for cohorts with at least 10 individuals for each of the genotypes of rs8067385. **(B)** Regional association plot of the *UMOD-PDILT* locus for uUMOD meta-analysis in 13 cohorts. The genome-wide significant signal includes two independent sets of SNPs: the top rs12934455 (P value 2.17E-88; purple diamond) is located on *UMOD*, whereas an independent set of SNPs (top rs77924615, P value 5.33E-79) is also present on *PDILT*. Each dot represents a SNP; the colour code refers to the LD toward the top SNP: red dot represents high LD with the top SNP.
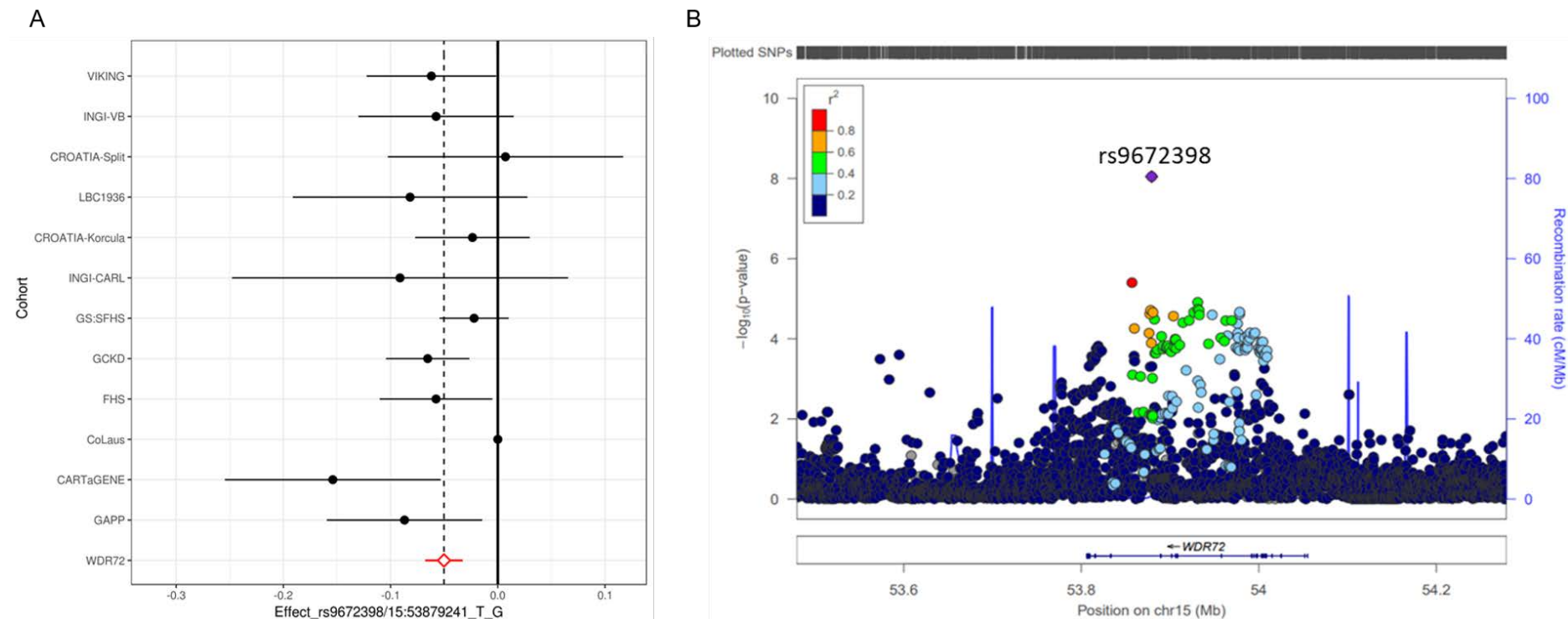
Figure S3. Effect size of rs9672398 and regional association plot of *WDR72* locus from uUCR meta-analysis.
(A) Forest plot showing effect sizes of rs9672398 in uUCR meta-analyses in 12 cohorts. The red diamond represents the average effect size of -0.0502 and a standard error of 0.0089 of the minor allele, G, of the SNP with lowest *P* value (rs9672398) in *WDR72* gene in association with uUCR. Information on this SNP was not available in the GWAS for the CoLaus cohort. **(B)** Locus zoom into the top SNP shows that the genome-wide significant locus spans the *WDR72* gene. Each dot represents a SNP; the colour code refers to the linkage disequilibrium (LD) toward the top SNP (purple diamond). Red dot represents high LD with the top SNP.
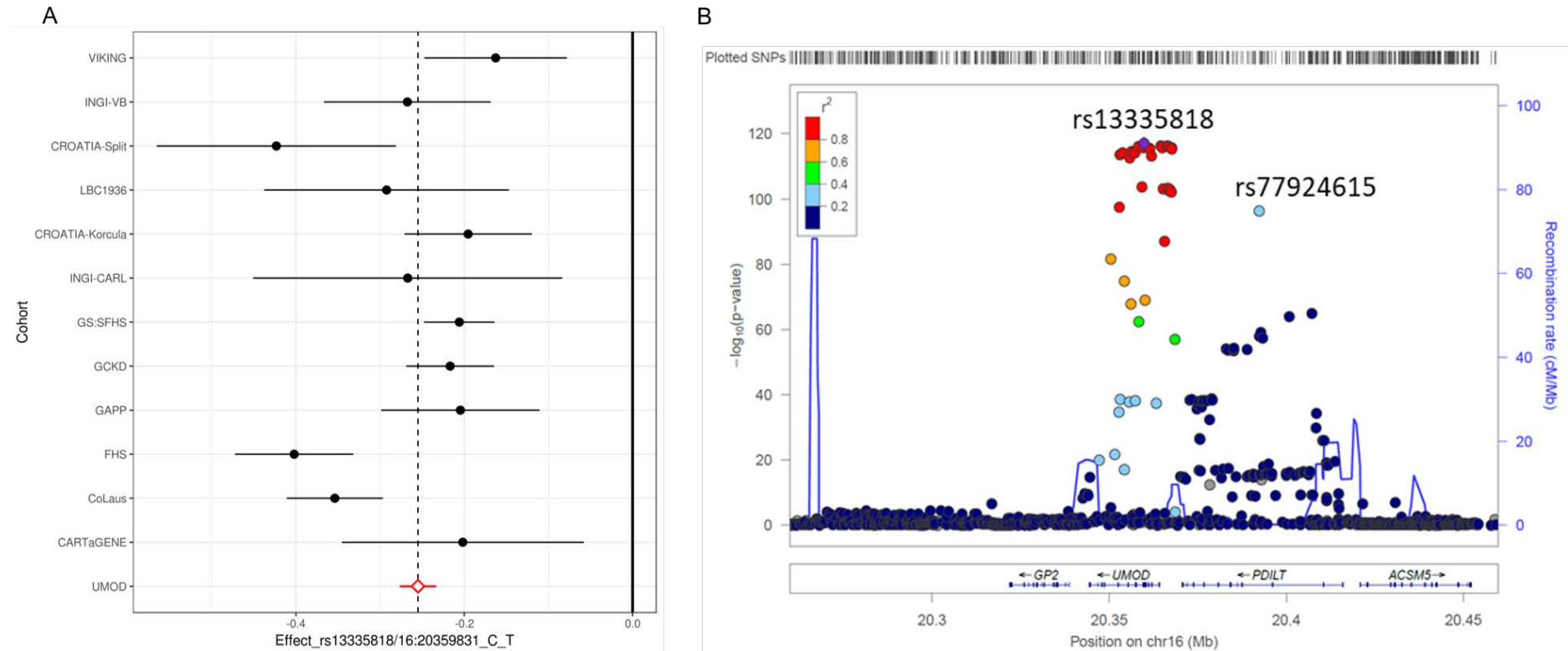
Figure S4. Effect size of rs13335818 and regional association plot of *UMOD-PDILT* locus from uUCR meta-analysis.
(A) Forest plot showing effect sizes for the minor allele, T, of rs13335818 in association with uUCR in 13 cohorts. The red diamond represents the average effect size of -0.255 with a standard error of 0.011. Variant rs13335818 is the SNP with the lowest *P* value within the *UMOD-PDILT* locus in association with uUCR. The minor allele of this SNP is associated with lower levels of uUCR in all of the 13 cohorts. **(B)** Locus zoom into the top SNP, rs13335818 (*P* value 3.86E-118), shows that the genome-wide significant locus spans over *UMOD* and *PDILT* genes. Each dot represents a SNP; the colour code refers to the LD toward the top SNP (purple diamond). Red dot represents high LD with the top SNP.
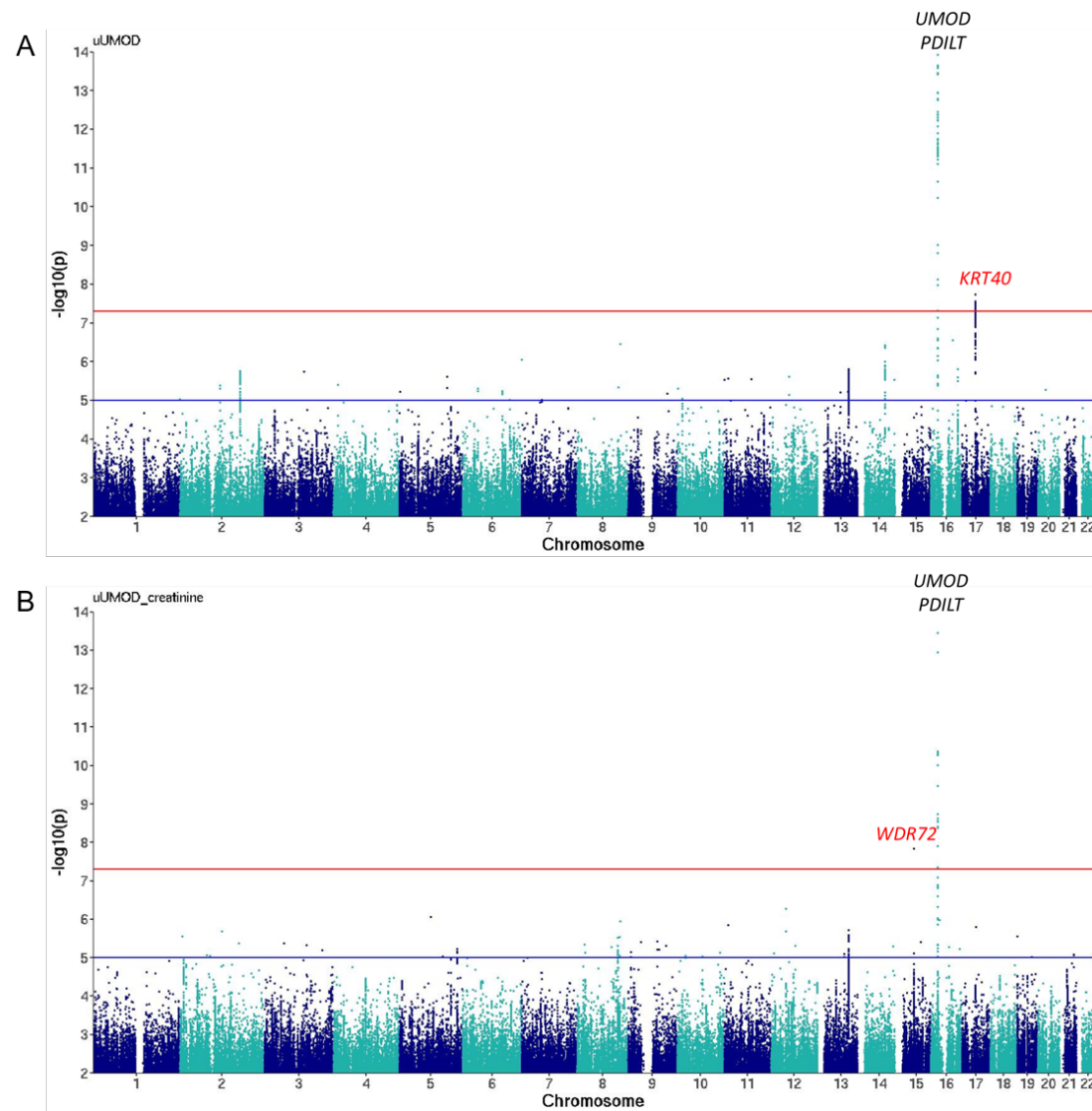
Figure S5. Manhattan plot of meta-GWAS of uUMOD and uUCR using sample size and *P* values for analysis of the 13 study cohorts. The blue line is at the 1E-05 'suggestive' level and the red line is at the commonly used 5E-08 threshold for significance in GWAS. **(A)** The two genome-wide significant loci, *UMOD/PDILT* (rs12934455 with *P* value 5.01E-90) and *KRT40* (rs8067385 with *P* value 1.82E-08), were consistent with the findings from meta-analysis using the effect size and standard error for analysis. **(B)** The two genome-wide significant loci, *UMOD/PDILT* (rs13335818 with *P* value 7.99E-119) and *WDR72* (rs9672398 with *P* value 1.45E-08), were consistent with the findings from meta-analysis using the effect size and standard error for analysis. The y-axis is cut-off at 1 x 10-14.
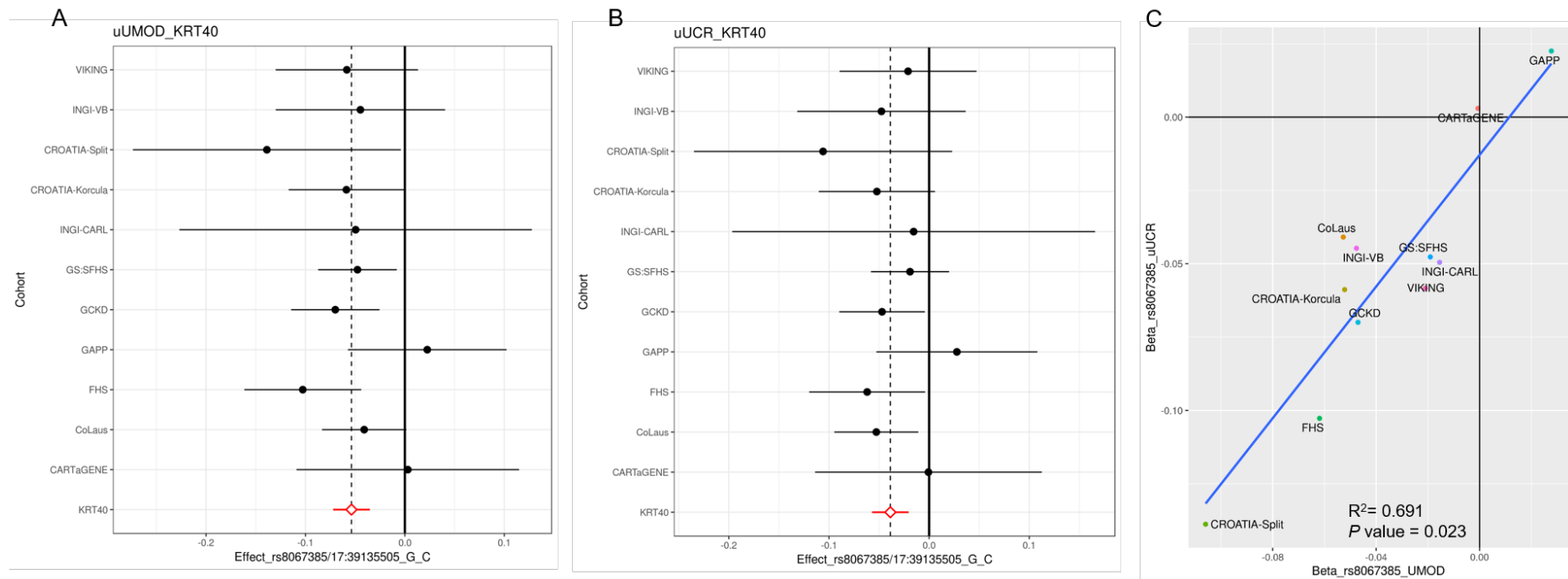
Suppl. Figure S6. Forest plot showing effect sizes of rs8067385 (KRT40 locus) on uUMOD and uUCR meta-analysis in the 13 cohorts.
(A) The red diamond represents the average effect size of -0.0537 and a standard error of 0.0094 of the minor, C allele of rs8067385 in association with uUMOD. **(B)** For uUCR, the average effect size is -0.0387 and the standard error is 0.0093. Information on this SNP was not available in the GWAS for the LBC1936 cohort. **(C)** Scatter plot showing effect size of rs8067385 from GWAS of uUMOD (x-axis) plotted against uUCR (y-axis). The horizontal and vertical blue lines represent zero. Correlation coefficient is 0.691 and Spearman's rank correlation of the effect sizes generated *P* value of 0.023. Effect sizes are shown for cohorts with at least 10 individuals for the genotypes of rs8067385.
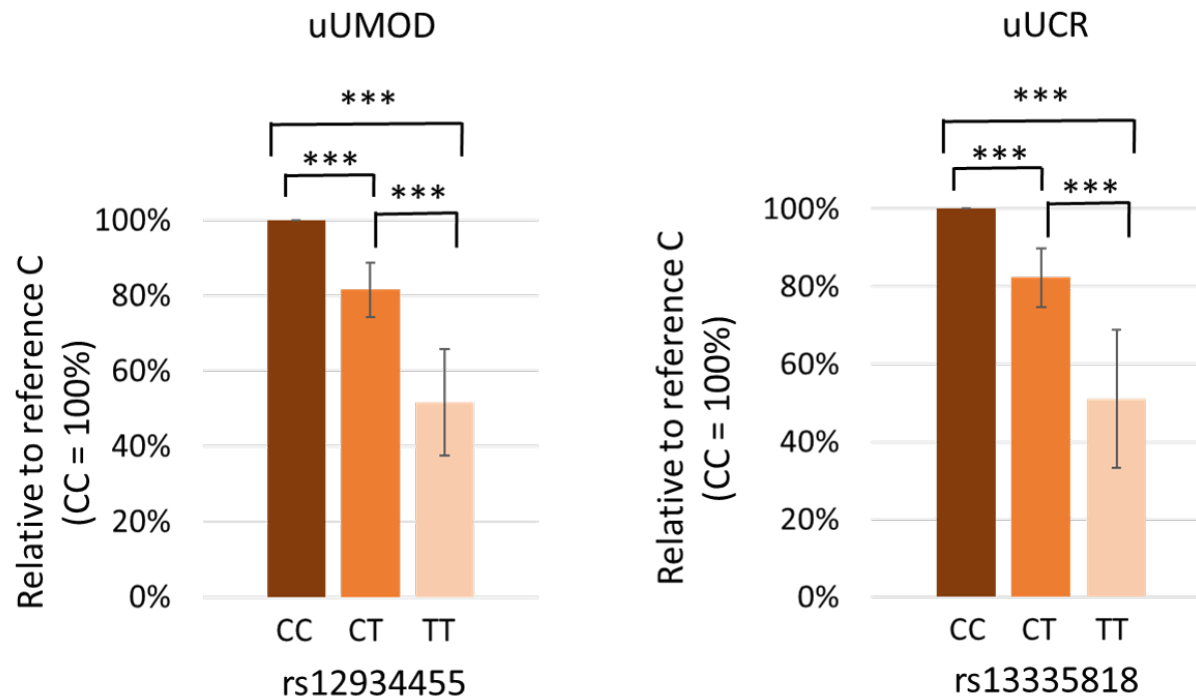
Suppl. Figure S7. Effect of *UMOD* genotype on urinary uromodulin (uUMOD and uUCR) levels.
The minor allele of the top variants rs12934455 (raw uromodulin: uUMOD) and rs13335818 (uromodulin indexed to creatinine: uUCR) are associated with lower levels of urinary uromodulin compared to the homozygous carriers of the reference allele (CC for both variants, taken as 100%). ANOVA analysis *** $P < 0.0001$.
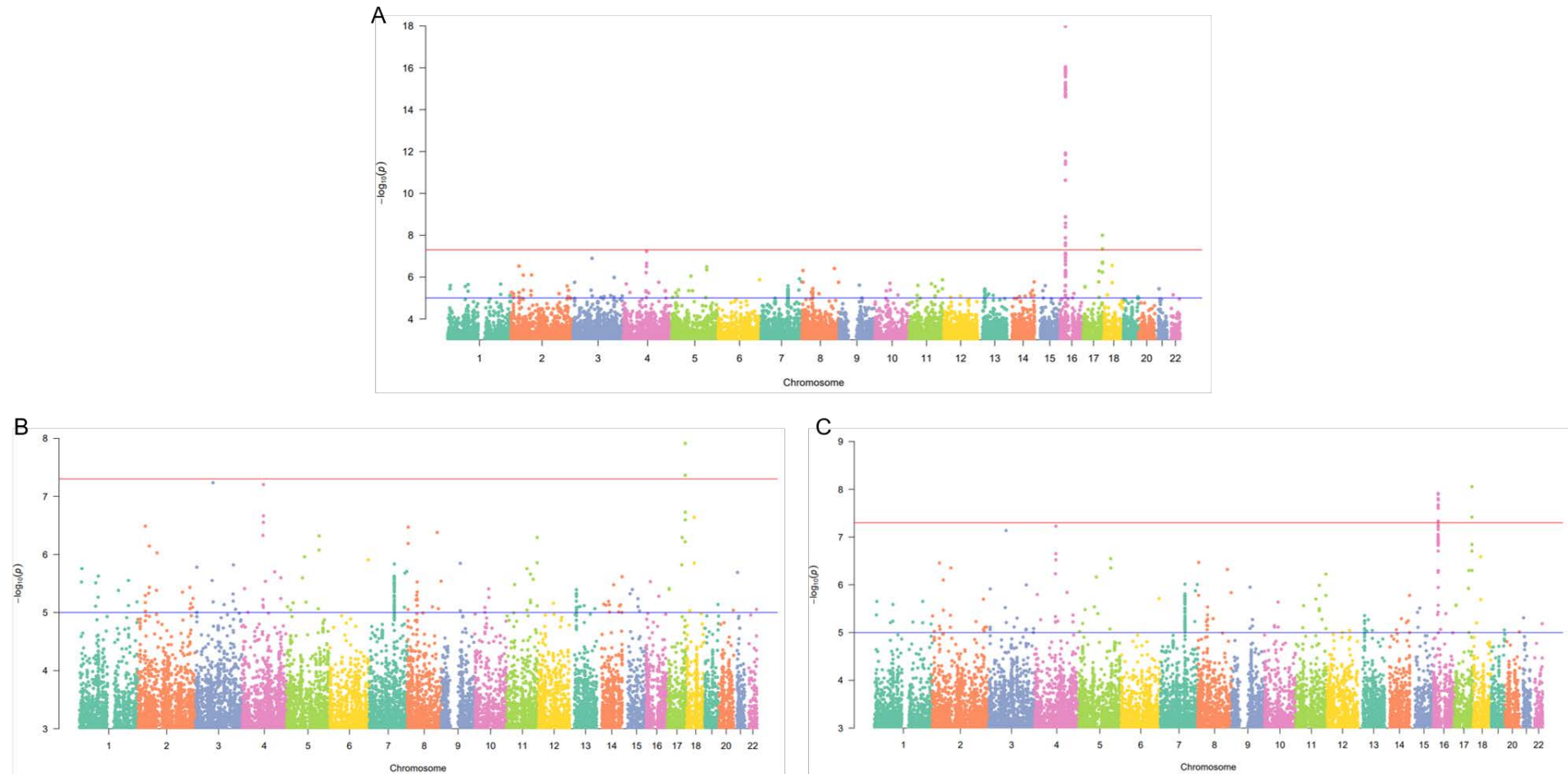
Suppl. Figure S8. Manhattan plot showing GWAS results for uUMOD and uUMOD conditioned for rs12934455 or for rs11864909 using GS:SFHS.
The blue line is at the 1E-05 'suggestive' level and the red line is at the commonly used 5E-08 threshold for significance in GWAS. The genome-wide significant locus within *UMOD* is observed in both uUMOD (**A**). The genome-wide significant locus within the *UMOD/PDILT* locus was not observed after conditional analysis for rs12934455 (**B**) but the *UMOD/PDILT* signal remained when conditioned for rs11864909 (**C**). Similar results were observed for uUCR.

A
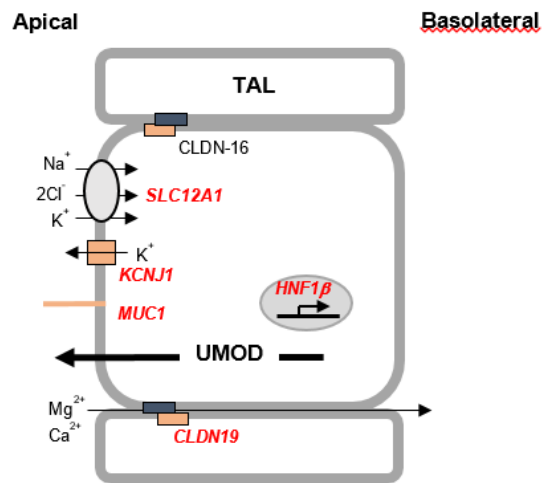
B

| Gene	Protein	SNP ID with Lowest P Value	SNP P value	Trait
SLC12A1	NKCC2	rs34685202	0.009	uUCR
KCNJ1	ROMK	rs190478015	0.006	uUCR
CLDN19	Claudin-19	rs41269513	0.006	uUCR
HNF1B	HNF1β	rs1058166	0.003	uUMOD
MUC1	Mucin-1	rs149945265	0.021	uUCR
MUC1	Mucin-1	rs149945265	0.024	uUMOD

Suppl. Figure S9. Candidate genes influencing the urinary excretion of uromodulin.
The *SLC12A1*, *KCNJ1*, *CLDN19*, *HNF1B* and *MUC1* genes, involved in rare inherited disorders affecting the thick ascending limb (TAL), contain at least one SNP with a *P* value below the gene-specific threshold associated with the raw (uUMOD) and/or normalized (uUCR) urinary levels of uromodulin. **(A)** Subellular localization of these genes and their function in the cells lining the TAL. **(B)** SNPs with the lowest *P* value in each gene, for uUMOD and uUCR.
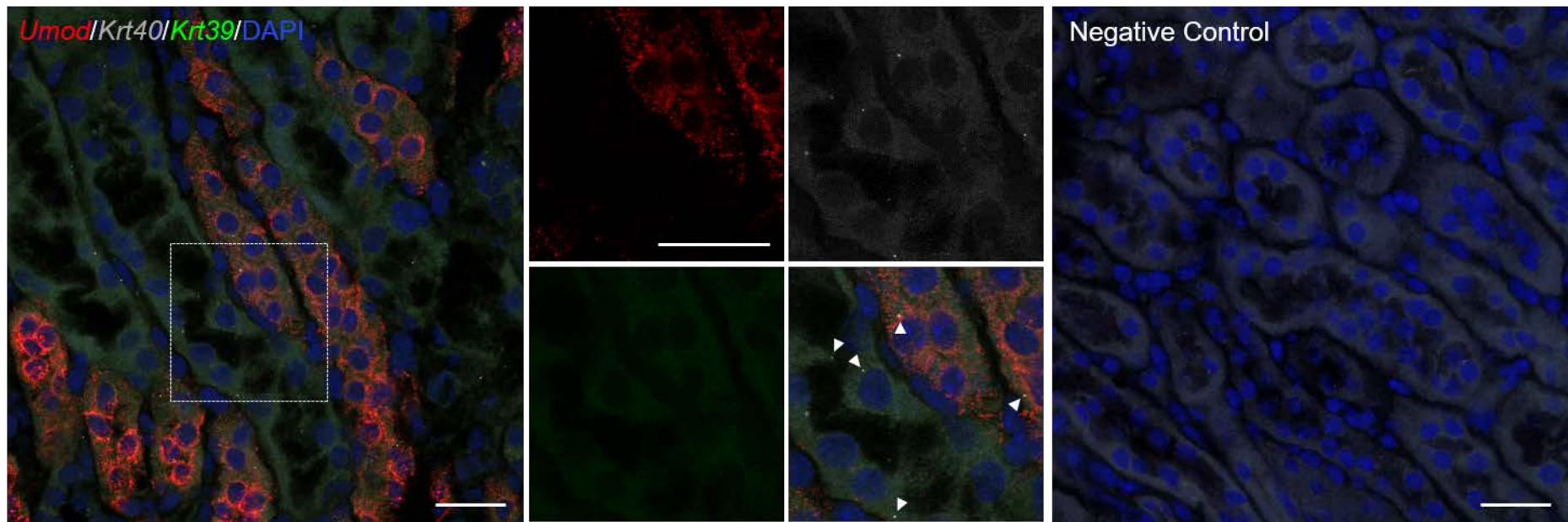
Suppl. Figure S10. *In situ* **hybridization for** *Umod*, *Krt40* **and** *Krt39* **on mouse kidney.**

Representative pictures of fluorescent multiplex *in situ* hybridization (RNAscope) on 10 μm cryo-sections from wild-type mouse kidney. A weak signal for *Krt40* was detected in both *Umod*-positive and negative tubules, while no signal was detected for *Krt39*. Left panel: RNAscope for *Umod* (red), *Krt40* (gray), and *Krt39* (green). Right panel: RNAscope 3-plex negative control for channels Alexa 488, Atto 550, Atto 647N. Nuclei are counterstained with DAPI (blue). Scale bar: 25μm.
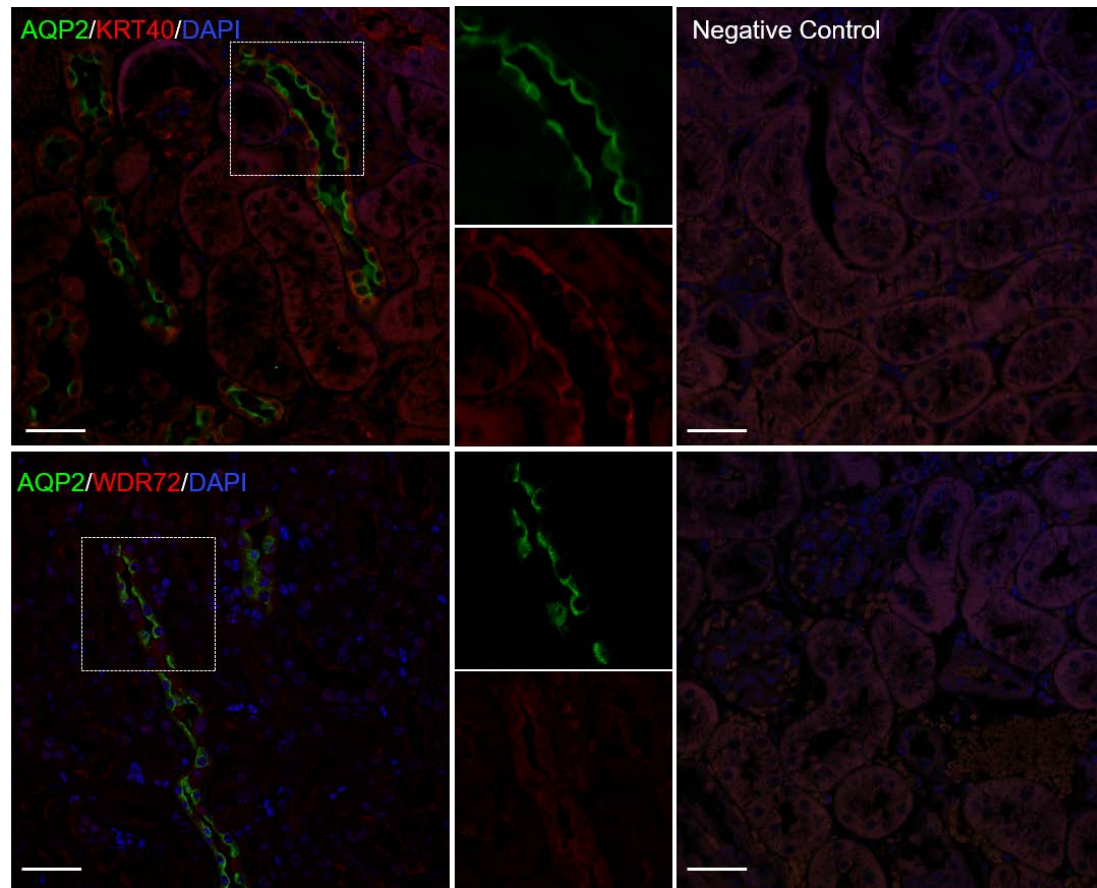
Suppl. Figure S11. Immunofluorescence staining for AQP2 and KRT40 or WDR72 on mouse kidney.
Representative immunofluorescence staining for AQP2 (green) and KRT40 or WDR72 (red) on paraffin-embedded kidney sections from wild-type mice, showing localization of both KRT40 and WDR72 in the collecting duct. Nuclei are counterstained with DAPI (blue). Negative control (right panel) was probed only with secondary antibodies. Scale bar: 25 μm.
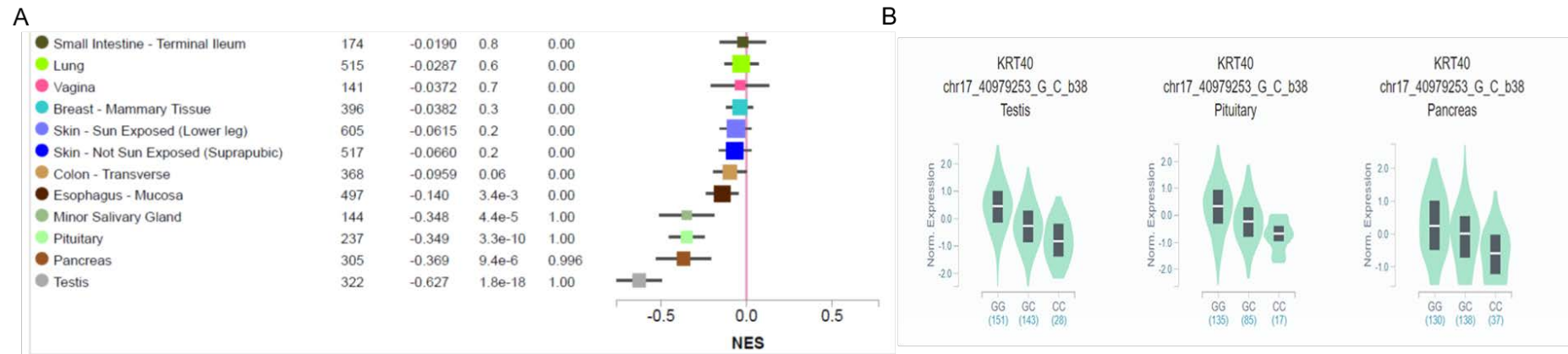
**Suppl. Figure S12. eQTL data for the *KRT40* variant rs8067385.**

(A) Box plot showing rank normalised expression level of *KRT40* associated with the different genotypes of rs8067385 in different epithelial tissues from the GTEx database V8. **(B)** Individuals homozygous for the minor, C allele of rs8067385 show lower levels of *KRT40* expression in testis, pituitary and pancreas. No significant eQTLs were found for *KRT40* gene in kidney cortex tissue and no precomputed eQTL data is available for kidney medulla tissue in GTEx. Source: https://gtexportal.org/home/snp/rs8067385; last accessed on December 22, 2020.

A
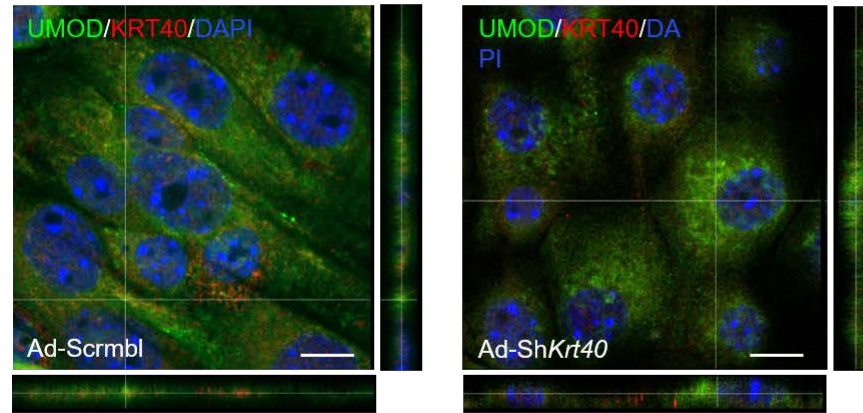
B
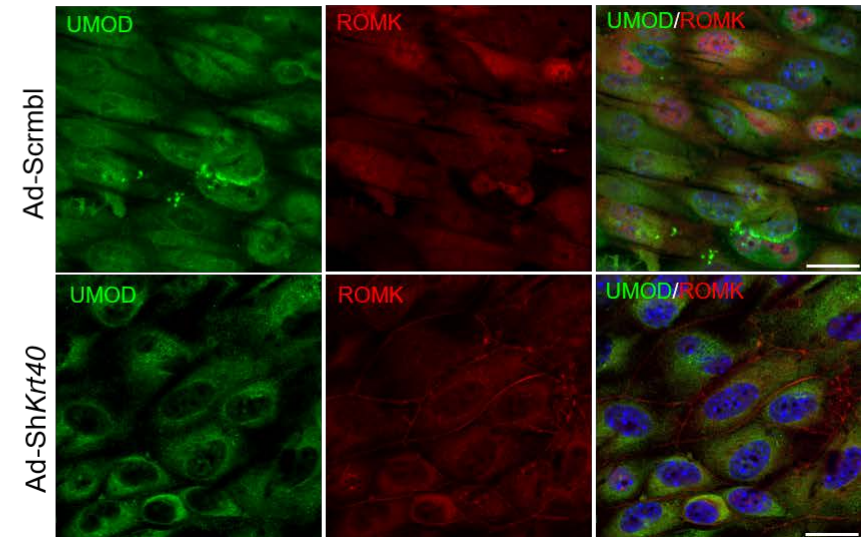
Suppl. Figure S13. Uromodulin (Z-stack) and ROMK distribution in mTAL cells following KRT40 knock-down.
(A) Representative immunofluorescence staining for uromodulin (UMOD, green) and KRT40 (red) on mTAL cells following transduction with Ad-shKrt40. Reconstructions of the z-plane are shown, showing perinuclear uromodulin localization in Ad-ShKrt40 treated cells. Nuclei are counterstained with DAPI (blue). Scale bar: 15 µm. **(B)** Representative immunofluorescence staining for uromodulin (UMOD, green) and ROMK (red) on mTAL cells following transduction with Ad-shKrt40, showing basolateral localization of ROMK in Ad-ShKrt40 cells. Nuclei are counterstained with DAPI (blue). Scale bar: 25 µm.

Supplementary Methods

In situ **hybridization:** Fluorescent multiplex *in situ* hybridization (RNAscope) assays (Advanced Cell Diagnostics, Hayward, CA, USA) were used to visualize single RNA molecules per cell in 10-µm cryosections of C57BL/6 wild-type mouse kidney fixed with 10% neutral buffered formalin, as previously described.[15] Kidney sections were incubated with probes for mouse *Krt40* (Mm-Krt40; #553001), *Krt39* (Mm-Krt39-C2; #553011-C2) and *Umod* (Mm-Umod-C3; #476301-C3). As controls, 3 plex negative control probe (#320871) and 3 plex positive control probe (#320881) were used. Images were obtained with an SP8 confocal microscope (Leica Microsystems, Wetzlar, Germany).

Electrophysiology: Confluent TAL (mTAL) monolayer on filters from both Ad-Scrmbl and Ad-Sh*Krt40* treated cells were subjected to simultaneous transepithelial voltage (V_{te}) and resistance (R_{te}) measurements using an EVOM-G potentiometer (WPI, USA) and Endohm 6 electrodes (WPI) as described previously.[16] (V_{te}) and (R_{te}) were recorded daily during the 96h following *Krt40* knockdown to assess the induction of mTAL differentiation.

Supplementary References

1. Awadalla P, Boileau C, Payette Y, et al.: Cohort profile of the CARTaGENE study: Quebec's population-based biobank for public health and personalized genomics. Int J Epidemiol. 42:1285-99, 2013.

2. Firmann M, Mayor V, Vidal PM, et al.: The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. BMC Cardiovasc Disord. 8:6, 2008.

3. Zemunik T, Boban M, Lauc G, et al.: Genome-wide association study of biochemical traits in Korcula Island, Croatia. Croat Med J. 50:23-33, 2009.

4. Rudan I, Marusić A, Janković S, et al.: "10001 Dalmatians:" Croatia launches its national biobank. Croat Med J. 50:4-6, 2009.

5. Polasek O, Hayward C, Bellenguez C, et al.: Comparative assessment of methods for estimating individual genome-wide homozygosity-by-descent from human genomic data. BMC Genomics.11:139, 2010.

6. Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham Offspring Study. Design and preliminary data. Prev Med. 4:518-25, 1975.

7. Conen D, Schön T, Aeschbacher S, et al.: Genetic and phenotypic determinants of blood pressure and other cardiovascular risk factors (GAPP). Swiss Med Wkly. 143:w13728, 2013.

8. Eckardt, K. U., Bärthlein, B., Baid-Agrawal, S., Beck, A., Busch, M., Eitner, F., The German Chronic Kidney Disease (GCKD) study: design and methods. Nephrol Dial Transplant. 27: 1454–1460, 2012.

9. Smith BH, Campbell A, Linksted P, et al.: Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. Int J Epidemiol. 42: 689-700, 2013.

10. Cocca M, Barbieri C, Concas MP, et al.: A bird's-eye view of Italian genomic variation through whole-genome sequencing. Eur J Hum Genet. 28:435-44, 2020.

11. Traglia M, Sala C, Masciullo C, et al. : Heritability and demographic analyses in the large isolated population of Val Borbera suggest advantages in mapping complex traits genes. Plos One. 4:e7554, 2009.

12. Deary IJ, Gow AJ, Taylor MD, et al.: The Lothian Birth Cohort 1936: a study to examine influences on cognitive ageing from age 11 to age 70 and beyond. BMC Geriatr 7:28, 2007.

13. Taylor AM, Pattie A and Deary IJ: Cohort Profile Update: The Lothian Birth Cohorts of 1921 and 1936. Int J Epidemiol. 47:1042, 2018.

14. Kerr SM, Klaric L, Halachev M, et al.: An actionable KCNH2 Long QT Syndrome variant detected by sequence and haplotype analysis in a population research cohort. Sci Rep. 9:10964, 2019.

15. Tokonami N, Takata T, Beyeler J, et al. Uromodulin is expressed in the distal convoluted tubule, where it is critical for regulation of the sodium chloride cotransporter NCC. Kidney Int. 94:701-715, 2018.

16. Glaudemans B, Terryn S, Gölz N, et al. A primary culture system of mouse thick ascending limb cells with preserved function and uromodulin processing. Pflugers Arch. 466:343-56, 2014.