

RESEARCH ARTICLE

Development of next generation sequencing panel for *UMOD* and association with kidney disease

Caitlin Bailie, Jill Kilner, Alexander P. Maxwell, Amy Jayne McKnight*

Nephrology Research, Centre for Public Health, Queen's University of Belfast, Belfast City Hospital, Belfast, Northern Ireland

* a.j.mcknight@qub.ac.uk



OPEN ACCESS

Citation: Bailie C, Kilner J, Maxwell AP, McKnight AJ (2017) Development of next generation sequencing panel for *UMOD* and association with kidney disease. PLoS ONE 12(6): e0178321. <https://doi.org/10.1371/journal.pone.0178321>

Editor: Tatsuo Shimosawa, The University of Tokyo, JAPAN

Received: July 3, 2016

Accepted: May 11, 2017

Published: June 13, 2017

Copyright: © 2017 Bailie et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The majority of data reported in this manuscript is provided within the main text and supplementary material. The microarray data is held in dbGAP (Study Accession: phs000389.v1.p1) and sequence files were deposited in our institutional repository (Research Portal, <http://libguides.qub.ac.uk/c.php?g=322847&p=2164812>) in accordance with our open access guidelines.

Funding: The Warren 3/UK GoKinD Study Group (UK-ROI collection) was jointly funded by Diabetes UK and the JDRF. Genome-wide genotyping from

Abstract

Chronic kidney disease (CKD) has a prevalence of approximately 10% in adult populations. CKD can progress to end-stage renal disease (ESRD) and this is usually fatal unless some form of renal replacement therapy (chronic dialysis or renal transplantation) is provided. There is an inherited predisposition to CKD with several genetic risk markers now identified. The *UMOD* gene has been associated with CKD of varying aetiologies. An AmpliSeq next generation sequencing panel was developed to facilitate comprehensive sequencing of the *UMOD* gene, covering exonic and regulatory regions. SNPs and CpG sites in the genomic region encompassing *UMOD* were evaluated for association with CKD in two studies; the UK Wellcome Trust Case-Control 3 Renal Transplant Dysfunction Study (n = 1088) and UK-ROI GENIE GWAS (n = 1726). A technological comparison of two Ion Torrent machines revealed 100% allele call concordance between S5 XL™ and PGM™ machines. One SNP (rs183962941), located in a non-coding region of *UMOD*, was nominally associated with ESRD (p = 0.008). No association was identified between *UMOD* variants and estimated glomerular filtration rate. Analysis of methylation data for over 480,000 CpG sites revealed differential methylation patterns within *UMOD*, the most significant of these was cg03140788 p = 3.7 x 10⁻¹⁰.

Introduction

Chronic kidney disease (CKD) is defined by sustained and often progressive decrease in renal function (over months to years) which is generally irreversible. Renal function is most commonly measured by analysis of the serum creatinine concentration and then calculating an estimated glomerular filtration rate (eGFR) using formulae based on the individual's age, gender, ethnicity and serum creatinine value [1]. Initially, CKD is typically asymptomatic and this can mean that diagnosis is delayed and CKD has progressed further before it is recognised [2]. With earlier diagnosis and effective management, CKD progression may be slowed or even halted [3]. CKD affects approximately 10% of adults however the prevalence in older age groups is much higher [4], and CKD is strongly associated with cardiovascular disease

the GENIE consortium was supported by a US Ireland R&D partnership award funded by the Science Foundation Ireland under Grant No. SFI/08/US/B1517, the Northern Ireland Research Development Office, and NIH NIDDK R01 DK081923. This work was also supported by the SFI_DfE Investigators Programme (15/IA/3152) to AJM and a PhD studentship from the Department of Employment and Learning to CB. The funders had no role in study design, data collection and analysis, interpretation of data, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

(cardiovascular events and hospitalisations) and premature mortality [5]. Individuals with CKD can progress to end-stage renal disease (ESRD), prompting the need for renal replacement therapies, such as chronic dialysis or renal transplantation. Treatments for CKD and ESRD represent a substantial socio-economic burden [6,7]. It is therefore important to identify earlier biological risk markers for CKD prediction as well as to highlight potential therapeutic targets.

Although there is evidence for a genetic predisposition to CKD, the risk alleles identified by genome-wide association studies (GWAS) are considered to contribute only a small proportion of overall complex disease risk [8,9,10,11,12]. CKD may be secondary to disorders such as diabetes, hypertension and glomerulonephritis [13,14,15,16]. Of interest, not all persons with diabetes, hypertension or glomerulonephritis will develop progressive CKD prompting investigation of genetic variants and epigenetic features which are associated with susceptibility to kidney disease or “protection” from developing this complication [17,18].

The *UMOD* gene encodes uromodulin protein (UMOD) that is synthesized solely in the epithelial cells of the thick ascending loop of Henle within the kidney. The role of UMOD in normal kidney function is incompletely understood, however, mutations of the *UMOD* gene are associated with autosomal dominant tubulointerstitial kidney disease (ADTKD-*UMOD*) which as previously been known as familial juvenile hyperuricaemic nephropathy (FJHN) and medullary cystic kidney disease (MCKD2), also known as *UMOD*-associated kidney diseases [19,20,21,22,23]. *UMOD* variants have also been identified as possible risk markers for more complex renal phenotypes such as CKD [24,25,26,27,28,29]. *UMOD* risk variants are associated with increased *UMOD* expression *in vitro* and *in vivo* [25]. For example, Olden and colleagues conducted meta-analysis of uromodulin levels in urine, highlighting SNPs in an LD block around the *UMOD* promoter, where the G allele of rs12917707 was consistently associated with decreased eGFR and up to a 2-fold increase in urinary excretion of uromodulin in a dose-dependent manner [28]. However, despite adequate power to identify a statistically significant association with variants in the *UMOD* gene, only a few genome-wide association studies for common renal phenotypes report *UMOD* as a top-ranked loci, suggesting associations may be restricted to particular population-specific phenotypes. Associations with ageing and age-related diseases such as diabetes and hypertension have also been suggested [30,31].

UMOD may have several “reno-protective” properties. The protein can form a gel-like permeability barrier in the nephron, possibly aiding with chemical concentration gradient maintenance [32]. The non-specific, carbohydrate structure of UMOD also allows it to bind and neutralize many molecules including urinary tract infection (UTI) causing bacteria and kidney stone forming calcium crystals [33,34]. Furthermore, UMOD may have a role in triggering an inflammatory response when the nephron has been damaged, thus recruiting immune cells [35] whilst, on the other hand, basal levels of UMOD may confer a protective effect in the event of acute ischaemic renal injury by decreasing inflammation [36]. Prolonged inflammation can trigger the development of progressive tissue fibrosis. This is in part mediated by up-regulation of transforming growth factor-beta (TGF- β) and recruitment of fibroblasts with resulting expansion of extracellular matrix (ECM) [37]. This progressive fibrotic process within the kidney is associated with pathological features such as glomerular sclerosis and tubulo-interstitial fibrosis which are correlated with clinical indicators of progressive CKD e.g. increased proteinuria and decreased eGFR.

UMOD is therefore an attractive biological candidate gene for both the development and progression of kidney disease. Previous studies on *UMOD* have examined selected exonic and upstream regions for variants that might be associated with CKD and eGFR. We have developed a next generation sequencing (NGS) panel that covers all coding and regulatory regions, within 2 kb upstream and downstream of *UMOD*. Published data was also reviewed and

association evaluated between *UMOD* SNPs and ESRD. Additionally, DNA methylation data for the *UMOD* gene, were analysed for association with CKD.

Methods

This study had full ethical approval from Queens University Belfast and the Office of Research Ethics Northern Ireland. PubMed was searched (search terms: *UMOD*, Uromodulin, Tamm-Horsfall protein, kidney disease, renal disease, Chronic Kidney Disease) to retrieve all publications relevant to *UMOD* SNPs and methylation data for any form of kidney disease (*last accessed [10 September 2015]*).

Sequencing samples

Existing high quality DNA was used for this study; DNA was originally extracted using the salting-out approach. For discovery of tag SNPs, 45 samples (23 transplant recipients with ESRD and 22 matched kidney donors with no known renal disease) were sequenced on an Ion Torrent Personal Genome Machine (PGM™) and 46 samples (23 transplant recipients with ESRD and 23 matched kidney donors with no known renal disease, including the 45 samples previously sequenced on the PGM) independently sequenced on the Ion Torrent 5S XL™ ([Fig 1](#)). This provided >95% power to detect all polymorphisms with minor allele frequency greater than 5%. All individuals were Caucasian and from the UK.

AmpliSeq next generation sequencing panel design

Using <http://www.ampliseq.com>, a custom next generation sequencing panel was designed to sequence 23,928 bp, including 2 kb both upstream and downstream, of *UMOD* on chromosome 16p12.3. The design went through several iterations before achieving maximal coverage that was cost-effective.

Ampliseq library preparation and NGS of *UMOD* gene

DNA libraries were prepared according to the Ampliseq Library Preparation Kit 2.0—96Lv (Thermo Fisher, San Diego, USA, MAN0006735) standard protocol using 20 ng of input DNA, with samples barcoded using Ion Xpress Barcodes 1–96 (Thermo Fisher). Prepared libraries were diluted to 26 pM using MilliQ water and combined into four samples pools, for sequencing on four 318 chips. Twenty-five microliters of each sample pool was enriched using a One-Touch Two (OT2) machine and automated enrichment system (Thermo Fisher) with the Ion PGM™ Template OT2 400 Kit (Thermo Fisher) to prepare Template-positive Ion Sphere Particles. Enriched products were prepared for sequencing using the Ion Personal Genome Machine™ (PGM) Hi-Q Sequencing Kit and run on the Personal Genome Machine using the Ion 318 chip v2 (Thermo Fisher, MAN0009816). In parallel, libraries for each sample were diluted to 70 pM and loaded onto an Ion Chef for sequencing on an Ion Torrent S5 XL™ using an Ion 530 chip (Thermo Fisher, MAN0010851) ([Fig 1](#)).

Bioinformatic analysis for SNPs

Aligned BAM files from the PGM server was reformatted with SNP files generated and validated using PLINK [38], before using HaploView [39] software to assess the level of linkage disequilibrium to discern tag SNPs which could be used for optimised association analysis. Ion Torrent suite software version 4.6 aligned reads against the designed. BED file associated with the hg19 reference genome; data freely available from the authors on request. Using PARTEK genomics suite, data was mapped to hg19 and dbSNP138, filtered for log odds ratio ≥ 300 and

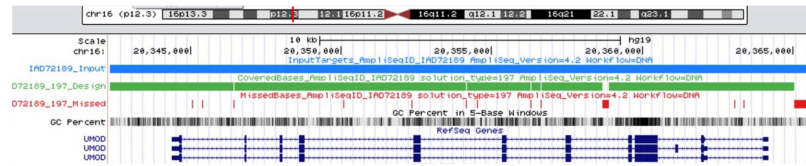


Fig 1. Ampliseq design. The *UMOD* gene region plus 2 kb upstream and downstream is highlighted by the blue line, coverage by Ampliseq amplicons is shown in green and regions not covered by the optimised assay design is shown in red. The structure of *UMOD* from RefSeq genes is represented in dark blue.

<https://doi.org/10.1371/journal.pone.0178321.g001>

total coverage at locus ≥ 20 to ensure reliable results, then annotated with rs# unique identifiers (if available) and functional effects. For the association analysis we used genome-wide genotyping information from two studies; the UK Wellcome Trust Case-Control 3 Renal Transplant Dysfunction Study (n = 1088; Affymetrix 6.0 array) comparing renal transplant recipients with ESRD to their matched donors with no kidney disease, and UK-ROI GENIE GWAS (n = 1726; Illumina OmniQuad array with data imputed to 1000 genomes) for controls with type 1 diabetes and no kidney disease compared to individuals with type 1 diabetes and kidney disease progressed to ESRD. PLINK [38] whole genome data analysis toolset was utilised to extract existing GWAS data for all SNPs between the first and last flanking SNPs from PARTEK analysis (rs9935655 to rs13333226 for Wellcome Trust data and rs4238595 and rs4293393 for GENIE data), within the GWAS datasets. We also conducted new association analysis using eGFR for non-ESRD samples in the UK-ROI GENIE collection. Characteristics of samples analysed in this study are presented in S1 Table. Nominal significance for association was set at p-value <0.05.

450K human methylation data analysis

Quantitative methylation data was previously generated using the Infinium Human Methylation 450K BeadChip (Illumina Inc.) according to manufacturer's instructions for 255 cases with CKD and compared to 152 controls without kidney disease [17]. Stringent quality control was followed and adjustment for multiple testing applied to determine association between methylation patterns and CKD and methylation data extracted for *UMOD*, *UMODL1* and *UMODL1-AS1* genes.

Results

Next generation sequencing

Variants within the *UMOD* gene were determined for a total for 45 individuals; 23 kidney transplant recipients with ESRD, and 22 matching donors with no known kidney disease. Donor age ranged from 6–65 yrs and recipient age ranged from 11–65 yrs. Slightly more females were present in the recipient group compared to the donor group (seven and 12 respectively).

The final AmpliSeq panel design employed 89 amplicons (125–375 bp) to cover 96.4% of the entire *UMOD* gene, starting at chromosome 16 position 20342322 and finishing at position 20366250 and including 100% coverage of exons. This panel covers all mutations currently screened by the UK Genetic Testing Network (*ukgtn.nhs.uk*) or published (Fig 1). An average of 445,650 mapped reads per sample was observed using the PGM™ (range 237,240–912,642 reads) of which >97% were on target, providing a mean depth of 2,052. Two primer pools were used to ensure that overlapping amplicons would optimally multiplex (45 amplicons in pool 1 and 44 amplicons in pool 2) (Fig 2).

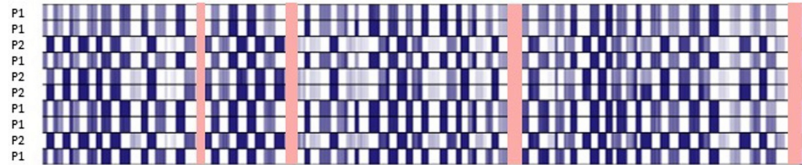


Fig 2. Coverage by amplicons. Sequence heat map from Partek Genomics Suite 6.6 showing complementary sequence coverage by pool 1 (P1) and pool 2 (P2) amplicons for pooled DNA. Areas not covered by either pool are highlighted in red.

<https://doi.org/10.1371/journal.pone.0178321.g002>

During quality control, results from two amplicons were removed as they did not meet the minimum required reads (AMPL7158733025 failed with an average of 42 reads per sample; AMPL7158732941 performed poorly with an average of 92 reads per sample) from the PGM data; following removal, an average of 2,498 reads per amplicon was observed.

In total we found 119 genetic variants within our population of 23 cases with ESRD and 22 controls without renal disease; 29 of these were insertions, deletions or multiallelic polymorphisms (S2 Table). Genotype completion obtained for most variants was 100%; one variant (20363262) showed 91% genotype completion. Of the remaining 90 SNPs, 60 were observed with minor allele frequency greater than 5% (S1 Fig). Using the tagger option in HaploView revealed that 20 SNPs, in 21 tests, captured 100% of the alleles at $r^2 > 0.8$ (mean $r^2 = 0.97$) and are thus sufficient for association analysis in larger cohorts (Figs 3, 4 and 5).

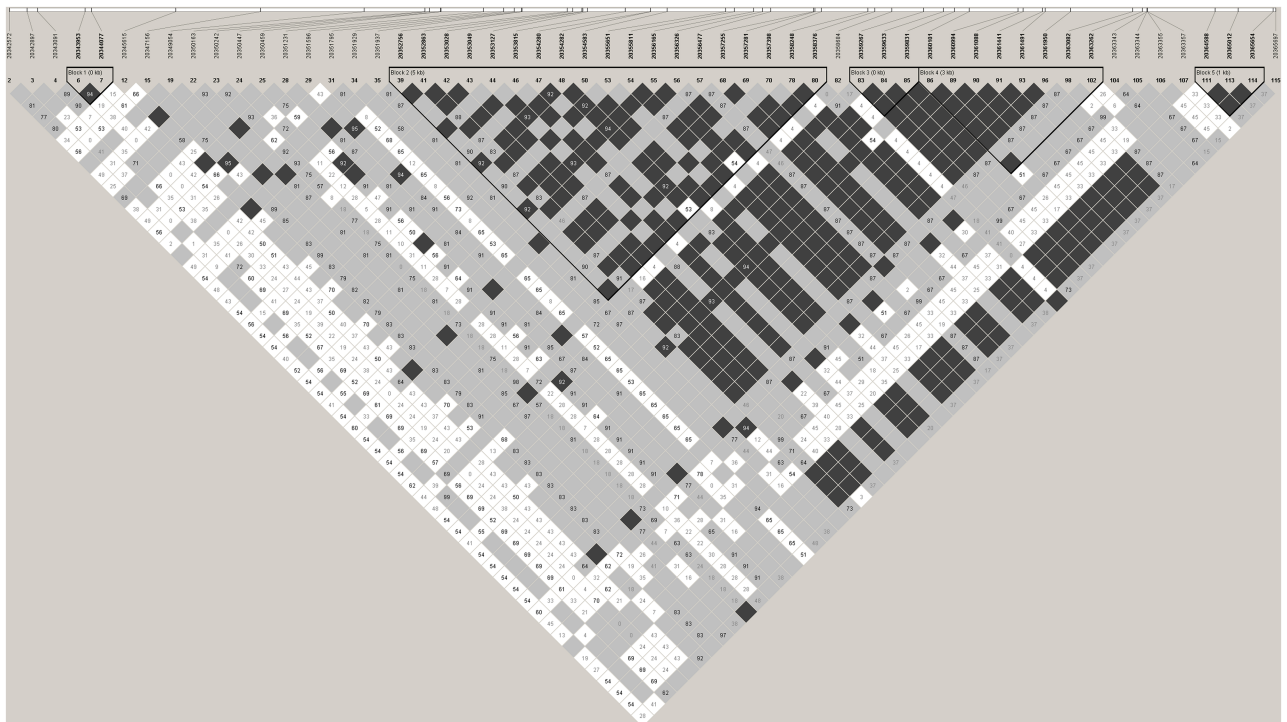


Fig 3. Linkage disequilibrium plot using D' confidence intervals. Linkage disequilibrium plot for all SNPs with minor allele frequency $>5\%$ and in Hardy-Weinberg Equilibrium. Colour scheme for D' confidence interval plot: Strong evidence of LD (dark grey), Strong evidence of recombination (white), Uninformative (light grey).

<https://doi.org/10.1371/journal.pone.0178321.g003>

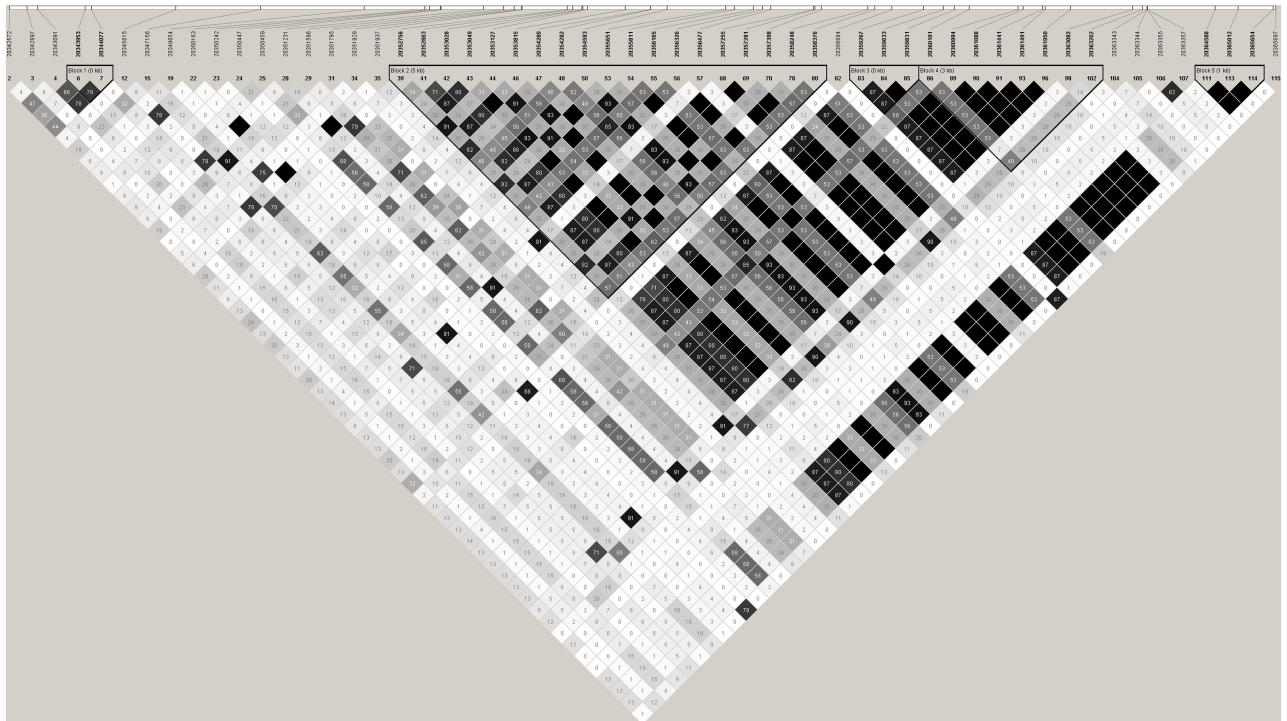


Fig 4. Linkage disequilibrium plot using r^2 . Linkage disequilibrium plot for all SNPs with minor allele frequency >5% and in Hardy-Weinberg Equilibrium. Colour scheme for r^2 plot as follows: $r^2 = 0$ (white), $0 < r^2 < 1$ (shades of grey), $r^2 = 1$ (black).

<https://doi.org/10.1371/journal.pone.0178321.g004>

Initial PARTEK analysis revealed a total of 3,317 differences from the reference sequence across all samples ($n = 45$) where log odds ratio ≥ 5 . 116 unique variants were discovered in the *UMOD* region with log odds ratio ≥ 300 and at least 20 reads of which 78 had rs identifiers in dbSNP build 138. The functional effect of SNPs differs depending on which *UMOD* transcript is considered; for example, one variant at 20352618 (rs55772253) in our discovery cohort exerted a non-synonymous effect on different exons depending on the transcript (Table 1; Fig 6).

An average of 195,205 mapped reads per sample was observed using the S5 XL™ (range 100,081–258,626 reads) of which >98% were on target, providing a mean depth of 965. Comparative results obtained from 318 and 530 chips (Fig 7) confirm both PGM and S5 XL are suitable for this analysis, with the choice of machine optimally guided by the number of samples to be sequenced.

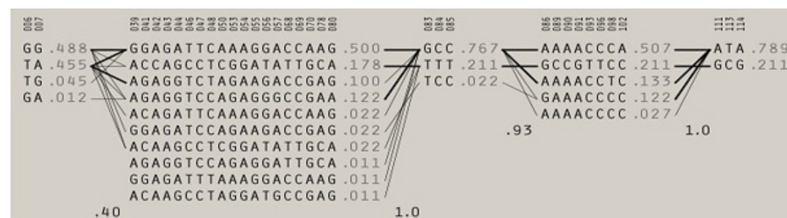


Fig 5. Haplotype blocks. Haplotype blocks with population frequencies displayed next to each haplotype. Lines represent the most common crossings between block (thicker lines are more common) beneath which the value of multiallelic D' is shown, measuring the level of linkage disequilibrium (LD) between blocks. Greater recombination is denoted by a value closer to zero.

<https://doi.org/10.1371/journal.pone.0178321.g005>

Table 1. Chromosome location and alleles associated with multiple transcripts and associated gene locations for non-synonymous SNP rs55772253.

Chromosome	Position	Reference allele	Alternate allele	Transcript	Gene Section	Functional Effect	Nucleotide Change	Amino Acid change
Chr 16	20352618	C	A	NM_001008389 NM_003361	Exon 7 Exon 7	Missense	c.1372G>T	V458L
Chr 16	20352618	C	A	NM_001278614	Exon 8	Missense	c.1471G>T	V491L

<https://doi.org/10.1371/journal.pone.0178321.t001>



Fig 6. UMOD transcripts. Three UMOD transcripts (3'-5') from NCBI gene website [last accessed: 17/10/2015]. Green blocks indicate the position of exons on the sequence with arrows denoting the direction of expression.

<https://doi.org/10.1371/journal.pone.0178321.g006>

Association analysis

Ninety-two SNPs were extracted from the UK/ROI Genetics of Nephropathy—an International Effort GWAS study (GENIE) and seven SNPs were extracted from the Wellcome Trust Case-Control 3 Renal Transplant Dysfunction Study (WTCCC3-RTD) study (two of these

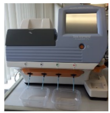

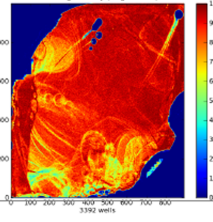


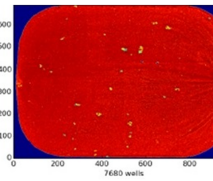
Sequencing Machine	Chip	Libraries sequenced for <i>UMOD</i> (2 pools per sample)	Hands on time for enrichment and sequencing	Chip loading (ISP density)	Addressable Wells	Library ISPs	Reads	Mean Raw accuracy	Output
Ion Torrent Personal Genome Machine 	318 	90 libraries (45 samples x 2 pools)	4-5 hrs	83% 	~11 million	~ 9.3 million	~ 6 million	99.3	~1.2 Gb
Ion Torrent S5 XL 	530 	92 libraries (46 samples x 2 pools)	20 min	91% 	~38 million	~34 million	~21 million	99.4	~4 Gb

Fig 7. Comparison of suitability of PGM and S5XL next generation sequencers for analysis of *UMOD*.

<https://doi.org/10.1371/journal.pone.0178321.g007>

Table 2. Showing sites within *UMOD*, *UMODL1* and *UMODL1-AS1* that are significantly associated with CKD.

Symbol	Description	CpG site	Features	Adj. P-value
<i>UMOD</i>	Uromodulin	cg03140788	Body; CpG Island	3.7×10^{-10}
		cg07817806	3'UTR	4.4×10^{-6}
		cg06861044	Body; CpG Island	6.6×10^{-5}
<i>UMODL1</i>	Uromodulin-Like 1	cg16624482	Body; S_Shore	2.90×10^{-32}
		cg21935742	TSS200	8.10×10^{-31}
		cg09727148	3'UTR	2.10×10^{-28}
		cg23931796	TSS200	4.70×10^{-27}
		cg00785029	Body	3.20×10^{-8}
		cg03643948	Body; Island	8.90×10^{-12}
		cg01542693	Body; S_Shore	2.40×10^{-9}
		cg03240473	Body; Island	1.50×10^{-12}
<i>UMODL1-AS1</i>	Uromodulin-Like 1- Anti-Sense 1	cg01542693	TSS200	2.40×10^{-9}
		cg03240473	Body; Island	1.50×10^{-12}

<https://doi.org/10.1371/journal.pone.0178321.t002>

SNPs were not found in GENIE) (S3 Table). Only one SNP extracted from the GENIE study, was nominally significantly associated with ESRD (rs183962941: $p = 0.008$). This SNP was not in linkage disequilibrium (LD) with any previously reported associated variants. Four other variants, extracted from the GENIE study, showed a trend towards significance, (rs191101580, rs187087030, rs8060932, rs8062123), and rs8060932 is in LD with another SNP (rs4238595) associated with FJHN [21]. Our study did not reveal any association with eGFR for non-ESRD samples with or without diabetes. No SNPs in the WTCCC3-RTD study revealed any significant association with ESRD or eGFR.

Methylation data analysis

Methylation status was determined for 485,577 CpG sites in a total of 407 individuals. Significant cg sites were extracted from the large dataset for *UMOD*, *UMODL1* and *UMODL1-AS1*. Following stringent quality control and analysis differential methylation of three cg sites was observed in *UMOD* revealing cg03140788 as the most significant with a p-value of 3.7×10^{-10} (Table 2). Two related genes that are also expressed in the kidney; *UMODL1* and *UMODL1-AS1*, were explored, revealing eight associated CpG sites, with genome wide significance; $P_{\max} = 2.9 \times 10^{-32}$ (cg16624482) was the most significant for non-diabetic ESRD. A list of SNPs that could influence probe binding and thus affect DNA methylation is provided in S4 Table.

Discussion

This study aimed to develop a next generation sequencing (NGS) assay for the detection of variants in the *UMOD* gene, covering both coding and non-coding regions. A cost-effective approach for targeted sequencing or comprehensive genotyping of individuals and larger population-based cohorts was developed using ≤ 20 ng of input DNA (Fig 8).

We observed one SNP, rs183962941 nominally associated with ESRD ($p = 0.008$), which has not been previously reported and was not shown to be in LD with any other variant using the SNP Annotation and Proxy Search (SNAP—Broad Institute [last accessed: 14 September 2015] [40]). Located in the intronic region between exons 6 and 7 of *UMOD*, the potential function of rs183962941 remains to be determined; searching the Blood eQTL browser [41], the GTEx Portal [42], or HaploReg v4 [43] does not reveal functional associations with this SNP, however it should be noted this SNP is not present on common Illumina or Affymetrix

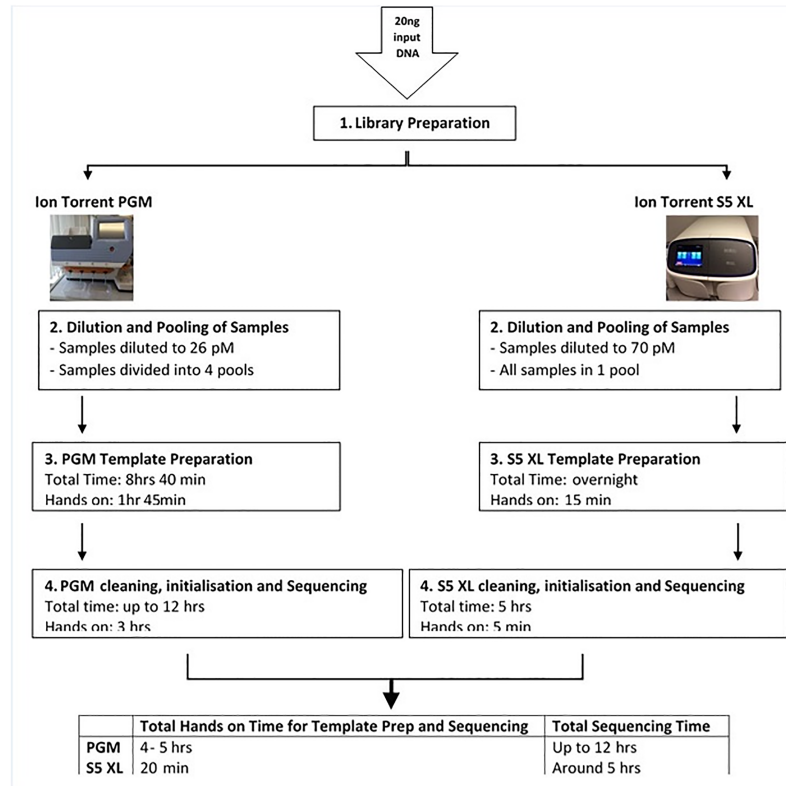


Fig 8. Sequencing workflow. Workflow for Ion Torrent PGM and S5 XL showing hands on time and total time for template preparation and sequencing.

<https://doi.org/10.1371/journal.pone.0178321.g008>

genotyping arrays, HapMap CEU release 22, or 1000 genomes pilot 1 according to SNAP [40]. Of the 116 SNPs annotated in PARTEK analysis, 91 were identified within intronic and / or potential regulatory regions, which may be important to help create a better risk profile for kidney disease.

Four other variants were extracted from the GENIE study, which were trending towards significance at the five percent level (rs191101580, rs187087030, rs8060932, rs8062123); one of these (rs8060932; $p = 0.08$) was shown to be in LD with another SNP, rs4238595 ($r^2 = 0.89$), associated with FJHN [21]. rs13333226 has been previously reported to be associated with reduced urinary uromodulin excretion as well as lower risk of hypertension and cardiovascular disease [13], however there was no evidence of association in this cohort. rs13333226 is also in LD with the most common SNP for *UMOD* (rs12917707), which has been reported to be associated with CKD, eGFR and ESRD [8,9,11,26,28].

Two non-synonymous SNPs were identified from a variant at position 20352618 (rs55772253), resulting in an amino acid change V458L or V491L depending on the *UMOD* transcript. V458L was previously reported by Kottgen *et al.* [24] who suggested it was associated with decreased eGFR and concentration of uromodulin; this SNP is ‘probably damaging’ (score 0.998) according to prediction software SIFT [44] and PolyPhen-2 [45].

The significant differential methylation pattern in *UMOD* (cg03140788 $p = 3.7 \times 10^{-10}$) was detected in a gene body CpG island (CGI). Jones *et al.* [46] suggest that methylation of a gene body CGI may be a method of controlling transcription of a gene that has two promoters (alternative promoter usage) by stopping transcription from this site while still allowing transcription and elongation from the earlier promoter.

Association was also identified in *UMODL1* and *UMODL1-AS1*. *UMODL1* was first characterised in 2004, but there has been little work published on this gene, with only 11 articles in PubMed returned from a search of '*UMODL1*', and one returned from a search of '*UMODL1 and kidney*' (February 2017). This gene is expressed in the kidney [47] and genome-wide association analysis of 3,851 individuals revealed association with *UMODL1* and CKD [48]. *UMODL1-AS1* is a long non-coding RNA, which has been little studied to date, but has potential to provide prognostic information for CKD.

We observed 100% SNP concordance between the S5 XL™ and PGM™ in SNPs with a quality score greater than 100. A visual review of highlighted differences between TorrentSuite-based SNP calling from the two sequencers revealed a microsatellite and homopolymer region (eight A/T) which generated experimental artefacts. Both PGM™ and S5 XL™ are ideally suited for targeted SNP and indel detection, while S5 XL™ represents a significant cost saving in terms of labour and consumables compared to using PGM™.

This study combines genetic and epigenetic investigations to improve understanding of the genetic architecture of the *UMOD* gene region. Strengths of the study included its significant power to identify 95% of variants with MAF >5% in 45 carefully phenotyped individuals, as well as the replication on the S5 XL™ sequencer and 100% concordance of SNPs between both next generation sequencers. However, tens of thousands of individuals will be required to evaluate rare variants in this gene to provide adequate power to identify significant associations for rare SNPs. A potential limitation to this study was that all samples were Caucasian and so variants that might be present within other ethnicities were not assessed.

In summary, there is sufficient evidence to suggest variants at the *UMOD* locus are associated with Mendelian and more common complex CKD aetiologies; larger, more comprehensive studies are warranted. Many SNPs that have been significantly associated with CKD are only nominally associated with ESRD if at all, perhaps suggesting different roles of action or the need for more carefully phenotyped cohorts. Large observational CKD studies such as GENetics of Nephropathy: an International Effort (GENIE) [49], Chronic Renal Insufficiency Cohort (CRIC) [50] and German Chronic Kidney Disease Study (GCKD) [51] will allow more insight into the development of ESRD, and hopefully lead to the identification of a more comprehensive risk profile for both CKD and ESRD.

Supporting information

S1 Table. Characteristics of samples analysed in this study.

(PDF)

S2 Table. SNPs uniquely identified from sequencing data.

(PDF)

S3 Table. List of SNPs uniquely extracted either in the UK/ROI GENetics of Nephrology an International Effort cohort or Wellcome Trust Case- Control 3 Renal Transplant Dysfunction Study cohort.

(PDF)

S4 Table. SNPs that could influence probe binding for certain cg sites in *UMOD* and *UMODL1*. Highlighted in bold font are cg sites that had significant p-values.

(PDF)

S1 Fig. SNPs depicted as blue lines across the gene structure with exons noted.

(PDF)

Acknowledgments

We thank all individuals for participating in this study and our funders for financially supporting this research.

Author Contributions

Conceptualization: CB APM AJM.

Data curation: CB JK APM AJM.

Formal analysis: CB AJM.

Funding acquisition: APM AJM.

Investigation: CB JK APM AJM.

Methodology: CB JK AJM.

Project administration: APM AJM.

Resources: CB JK APM AJM.

Supervision: APM AJM.

Validation: CB AJM.

Visualization: CB AJM.

Writing – original draft: CB APM AJM.

Writing – review & editing: CB JK APM AJM.

References

1. Levey AS, Bosch JP, Lewis JB, Greene T, Rogers N and Roth D. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group. *Ann Intern Med.* 1999; 130: 461–70. PMID: [10075613](https://pubmed.ncbi.nlm.nih.gov/10075613/)
2. Middleton RJ, Foley RN, Hegarty J., Cheung C.M., McElduff P., Gibson J.M. et al. The unrecognized prevalence of chronic kidney disease in diabetes. *Nephrol Dial Transplant.* 2006; 21: 88–92. <https://doi.org/10.1093/ndt/gfi163> PMID: [16221715](https://pubmed.ncbi.nlm.nih.gov/16221715/)
3. Fogarty DG and Taal MW. A stepped care approach to the management of chronic kidney disease. In: Taal MW, Chertow GM, Marsden PA, Skorecki A, Yu ASL, Brenner BM. editors. *Brenner and Rector's The Kidney.* Elsevier; 2011. pp. 2205–2239.
4. KDIGO. KDIGO Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. *Kidney Int Suppl.* 2013; 3: 1–150.
5. Herzog CA, Asinger RW, Berger AK, Charytan DM, Diez J, Hart RG, et al. Cardiovascular disease in chronic kidney disease. A clinical update from Kidney Disease: Improving Global Outcomes (KDIGO). *Kidney Int.* 2011; 80: 572–86. <https://doi.org/10.1038/ki.2011.223> PMID: [21750584](https://pubmed.ncbi.nlm.nih.gov/21750584/)
6. Ekart R, Ferjuc A, Furman B, Gerjevic S, Bevc S and Hojs R. Chronic kidney disease progression to end stage renal disease: A single center experience of the role of the underlying kidney disease. *Therapeutic Apher and Dial.* 2013; 17: 363–367.
7. Shaw C, Pitcher D, Pruthi R, and Fogarty D. UK renal registry 16th annual report: chapter 2 UK RRT prevalence in 2012: national and centre-specific analyses. *Nephron Clin Pract.* 2013; 125: 29–53. <https://doi.org/10.1159/000360021> PMID: [24662166](https://pubmed.ncbi.nlm.nih.gov/24662166/)
8. 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 Gen.* 2009; 41: 712–717.
9. Boger CA, Gorski M, Li M, Hoffman MM, Huang C, Yang Q, et al. Association of eGFR-related loci identified by GWAS with incident CKD and ESRD. *PLoS Genet.* 2011; 7: e1002292. <https://doi.org/10.1371/journal.pgen.1002292> PMID: [21980298](https://pubmed.ncbi.nlm.nih.gov/21980298/)

10. Pattaro C, Kottgen A, Teumer A, Garnaas M, Boger CA, Fuchsberger C, et al. Genome-wide association and functional follow-up reveals new loci for kidney function. *PLoS Genet.* 2012; 8: e1002584. <https://doi.org/10.1371/journal.pgen.1002584> PMID: 22479191
11. Gorski M, Tin A, Garnaas M, McMahon GM, Chu AY, Tayo BO, et al. Genome-wide association study of kidney function decline in individuals of European descent. *Kidney Int.* 2015; 87: 1017–29. <https://doi.org/10.1038/ki.2014.361> PMID: 25493955
12. Liu CT, Garnaas MK, Tin A, Kottgen A, Franceschini N, Peralta CA, et al. Genetic Association for Renal Traits among Participants of African Ancestry Reveals New Loci for Renal Function. *PLoS Genet.* 2011; 7: e1002264. <https://doi.org/10.1371/journal.pgen.1002264> PMID: 21931561
13. 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: e1001177. <https://doi.org/10.1371/journal.pgen.1001177> PMID: 21082022
14. Collins AJ, Vassalotti JA, Wang C, Li S, Gilbertson DT, Liu J, et al. Who should be targeted for CKD screening? Impact of diabetes, hypertension, and cardiovascular disease. *Am J Kidney Dis.* 2009; 53: S71–77. <https://doi.org/10.1053/j.ajkd.2008.07.057> PMID: 19231764
15. Levey AS, Astor BC, Stevens LA and Coresh J. Chronic kidney disease, diabetes, and hypertension: what's in a name? *Kidney Int.* 2010; 78: 19–22. <https://doi.org/10.1038/ki.2010.115> PMID: 20428101
16. McKnight AJ, McKay GJ and Maxwell AP. Genetic and Epigenetic Risk Factors for Diabetic Kidney Disease. *Adv. Chronic Kidney Dis.* 2014; 21: 287–296. <https://doi.org/10.1053/j.ackd.2014.03.010> PMID: 24780457
17. Smyth LJ, McKay GJ, Maxwell AP and McKnight AJ. DNA hypermethylation and DNA hypomethylation is present at different loci in chronic kidney disease. *Epigen.* 2014; 9: 366–376.
18. Wing MR, Devaney JM, Joffe MM, Xie D, Feldman HI, Dominic EA, et al. DNA methylation profile associated with rapid decline in kidney function: findings from the CRIC Study. *Nephrol Dial Transplant.* 2014; 29: 864–872. <https://doi.org/10.1093/ndt/gft537> PMID: 24516231
19. 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: 882–892. <https://doi.org/10.1136/jmg.39.12.882> PMID: 12471200
20. Rampoldi L, Caridi G, Santon D, Boaretto F, Bernascone I, Lamorte G, et al. Allelism of *MCKD1*, *FJHN* and *GCKD* caused by impairment of uromodulin export dynamics. *Hum Mol Genet.* 2003; 12: 3369–3384. <https://doi.org/10.1093/hmg/ddg353> PMID: 14570709
21. Williams SE, Reed AA, Galvanovskis J, Antignac C, Goodship T, Karet FE, et al. Uromodulin mutations causing familial juvenile hyperuricaemic nephropathy lead to protein maturation defects and retention in the endoplasmic reticulum. *Hum Mol Genet.* 2009; 18: 2963–2974. <https://doi.org/10.1093/hmg/ddp235> PMID: 19465746
22. Eckardt KU, Alper SL, Antignac C, Bleyer AJ, Chauveau D, Dahan K, et al. Autosomal dominant tubulointerstitial kidney disease: diagnosis, classification, and management—A KDIGO consensus report. *Kidney Int.* 2015; 88(4): 676–683. <https://doi.org/10.1038/ki.2015.28> PMID: 25738250
23. Bleyer AJ, Hart PS, Knoch S. Autosomal Dominant Tubulointerstitial Kidney Disease, *UMOD*-Related. In: Pagon RA, Adam MP, Ardinger HH, Wallace SE, Amemiya A, Bean LJH, et al. editors. *GeneReviews*® [Internet]. Seattle (WA): University of Washington, Seattle; 1993–2017.
24. Kottgen A, Yang Q, Shimmin LC, Tin A, Schaeffer C, Coresh J, et al. Association of estimated glomerular filtration rate and urinary uromodulin concentrations with rare variants identified by *UMOD* gene region sequencing. *PLoS One.* 2012; 7: e38311. <https://doi.org/10.1371/journal.pone.0038311> PMID: 22693617
25. 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: 1655–1660. <https://doi.org/10.1038/nm.3384> PMID: 24185693
26. Reznichenko A, Boger CA, Sneider H, van den Born J, de Borst MH, Damman J, et al. *UMOD* as a susceptibility gene for end-stage renal disease. *BMC Med Genet.* 2012; 13.
27. Chambers JC, Zhang W, Lord GM, van der Harst P, Lawlor DA, Sehmi JS, et al. Genetic loci influencing kidney function and chronic kidney disease. *Nat Genet.* 2010; 42: 373–375. <https://doi.org/10.1038/ng.566> PMID: 20383145
28. 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: 1869–1882. <https://doi.org/10.1681/ASN.2013070781> PMID: 24578125
29. Alhualia TS, Lindholm E, Groop L, and Melander O. Uromodulin gene variant is associated with type 2 diabetic nephropathy. *J Hypertens.* 2011; 29: 1731–1734. <https://doi.org/10.1097/HJH.0b013e328328349de25> PMID: 21738052

30. 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: e1001039. <https://doi.org/10.1371/journal.pgen.1001039> PMID: 20686651
31. Kottgen A, Hwang SJ, Larson MG, Van Eyk JE, Fu Q, Benjamin EJ, et al. Uromodulin levels associate with a common UMOD variant and risk for incident CKD. *J Am Soc Nephrol.* 2010; 21: 337–344. <https://doi.org/10.1681/ASN.2009070725> PMID: 19959715
32. Renigunta A, Renigunta V, Saritas T, Decher N, Mutig K and Waldegger S. Tamm-Horsfall glycoprotein interacts with renal outer medullary potassium channel ROMK2 and regulates its function. *J Biol Chem.* 2011; 286: 224–35.
33. Mo L, Zhu XH, Huang HY, Shapiro E, Hasty DL and Wu XR. Ablation of the Tamm-Horsfall protein gene increases susceptibility of mice to b ladder colonization by type 1-fimbriated *Escherichia coli*. *Am J Physiol Renal Physiol.* 2004; 286: F795–802. <https://doi.org/10.1152/ajprenal.00357.2003> PMID: 14665435
34. Liu Y, Mo L, Goldfarb DS, Evan AP, Liang F, Khan SR, et al. Progressive renal papillary calcification and ureteral stone formation in mice deficient for Tamm-Horsfall protein. *Am J Physiol Renal Physiol.* 2010; 299: F469–478. <https://doi.org/10.1152/ajprenal.00243.2010> PMID: 20591941
35. Scolari F, Izzi C and Ghiggeri GM. Uromodulin: from monogenic to multifactorial diseases. *Nephrol Dial Transplant.* 2014; 30: 1–7.
36. El-Ackhar TM, Wu XR, Rauchman M, McCracken R, Keifer S and Dagher PC. Tamm-Horsfall protein protects the kidney from ischemic injury by decreasing inflammation and altering TLR4 expression. *Am J Physiol Renal Physiol.* 2008; 295: F534–544. <https://doi.org/10.1152/ajprenal.00083.2008> PMID: 18495803
37. Rockey DC, Bell PD. and Hill JA. Fibrosis—A Common Pathway to Organ Injury and Failure. *N Engl J Med.* 2015; 372: 1138–1149. <https://doi.org/10.1056/NEJMra1300575> PMID: 25785971
38. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a toolset for whole-genome association and population based linkage analysis. *Am J Hum Genet.* 2007; 81: 559–575. <https://doi.org/10.1086/519795> PMID: 17701901
39. Barrett JC, Fry B, Maller J and Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 2005; 21: 263–265. <https://doi.org/10.1093/bioinformatics/bth457> PMID: 15297300
40. Johnson AD, Hansaker RE, Pulit SL, Nizzari MM, O'Donnell CJ and de Bakker PI. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics.* 2008; 24: 2938–2939. <https://doi.org/10.1093/bioinformatics/btn564> PMID: 18974171
41. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet.* 2013; 45: 1238–1243. <https://doi.org/10.1038/ng.2756> PMID: 24013639
42. Lonsdale J, Thomas J and Salvatore M. The genotype-tissue expression (GTEx) project. *Nat Genet.* 2013; 45: 580–58. <https://doi.org/10.1038/ng.2653> PMID: 23715323
43. Ward LD and Kellis M HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucl Acids Res.* 2012; 40: D930–D934 <https://doi.org/10.1093/nar/gkr917> PMID: 22064851
44. Kumar P, Henikoff S and Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc.* 2009; 4: 1073–81. <https://doi.org/10.1038/nprot.2009.86> PMID: 19561590
45. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods.* 2010; 7: 248–249. <https://doi.org/10.1038/nmeth0410-248> PMID: 20354512
46. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet.* 2012; 13: 484–492. <https://doi.org/10.1038/nrg3230> PMID: 22641018
47. Shibuya K, Nagamine K, Okui M, Ohsawa Y, Asakawa S, Minoshima S et al., Initial characterization of an uromodulin-like 1 gene on human chromosome 21q22.3, *Biochem Biophys Res Commun.* 2004; 319(4): 1181–1189. <https://doi.org/10.1016/j.bbrc.2004.05.094> PMID: 15194491
48. Yamada Y, Nishida T, Ichihara S, Kato K, Fujimaki T, Oguri M et al., Identification of chromosome 3q28 and ALPK1 as susceptibility loci for chronic kidney disease in Japanese individuals by a genome-wide association study. *J Med Genet.* 2013; 50(6): 410–418. <https://doi.org/10.1136/jmedgenet-2013-101518> PMID: 23539754
49. Sandholm N, Salem RM, McKnight AJ, Brennan EP, Forsblom C, Isakova T, et al. Susceptibility Loci Associated with Kidney Disease in Type 1 Diabetes. *PLoS Genet.* 2012; 8: e1002921. <https://doi.org/10.1371/journal.pgen.1002921> PMID: 23028342

50. Feldman HI, Appel LJ, Chertow GM, Cifelli D, Cizman B, Daugirdas J, et al. The chronic renal insufficiency cohort (CRIC) study: design and methods. *JASN*, 2003; 14: S148–153. PMID: [12819321](#)
51. 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: 1454–1460. <https://doi.org/10.1093/ndt/gfr456> PMID: [21862458](#)