Preprints.org

Article

Genetic Association of *UMOD* Gene Variant (16:20353266A>G) and Bioinformatics Analyses of the Uromodulin Protein in CKD Patients of Pakistani Origin

Mehmoona Zahid , Rashid Saif [*] , Amna Younus , Mehnaz Ghulam Hussain , Sadia Jabeen , Hooria Younas

Article

Genetic Association of *UMOD* Gene Variant (16:20353266A>G) and Bioinformatics Analyses of the Uromodulin Protein in CKD Patients of Pakistani Origin

Mehmoona Zahid [1], Rashid Saif [2, 3], Amna Younus [1], Mehnaz Ghulam Hussain [1], Sadia Jabeen [3] and Hooria Younas [1]

[1] Department of Biochemistry, Kinnaird College for Women, Lahore, Pakistan
[2] Department of Biotechnology, Qarshi University, Lahore, Pakistan
[3] Decode Genomics, Punjab University Employees Housing Scheme, Lahore, Pakistan
* Correspondence: rashid.saif37@gmail.com

Abstract: Chronic Kidney Disease (CKD) is one of the worldwide health threats, kidneys may be affected by internal pathologies and environmental factors as well as a number of genes are also involved in the kidney dysfunctioning, so *UMOD* is one of these genes which has significant effect in regulating urea and creatinine level in blood and decreased excretion of uric acid which causes interstitial damages in kidney and may result CKD. Hence, the present study is envisaged to observe the association of this gene locus (r.1264T>C) with CKD in Pakistani population using ARMS-PCR genotyping, for the purpose, a total of 75 samples were investigated, out of which 45 were found homozygous wild-type, 21 heterozygous and 09 homozygous mutants. PLINK data analysis toolset showed that our sampled population does not obeys Hardy-Weinberg Equilibrium by χ^2 (2, $N = 75$) = 18.87, p = 0.031 and Chi-square statistics with p = 1.40×10^{-005} showing a significant genotypic association with the subject phenotype with an alternative allele frequency of 0.15 and 0.48 within cases & controls respectively. Furthermore, an odds-ratio of 0.2 predicts that the risk of wild-type allele (A) is 0.8 times higher in patient as compared to controls and mutant allele is 0.2 times preventive in patients. Additionally, a few bioinformatics tools e.g., ProtParam, PsiPred, TMHMM-2.0, Motif Scan, ScanProsite, NetPhos3.1, GPS PAIL2.0, NetOGlyc4.0, PRmePRed, PDB-RCSB and STRING were applied to predict different characteristics of Uromodulin protein including physicochemical properties, secondary/transmembrane structures, protein conserved domains and posttranslational modifications respectively. This pilot scale study provided a comprehensive insight of the association of aforementioned genetic variant with CKD to identify and evaluate the genetic risk factors along with functional effects of the subject variant which may be used as CKD prevention by adopting DNA screening and genetic counselling recommendations.

Keywords: *UMOD*; CKD; creatinine; urea; eGFR; uromodulin; genotyping; ARMS-PCR

Introduction

Chronic Kidney Disease (CKD) a persistent and permanent loss of kidney functions in humans, that is distinguished by reduction in glomerular filtration up to eGFR<60mL/min/1.73m² (1), which is one of the leading causes of health problems around the globe, affecting 15-20% of adults and elderly population respectively, resulting 1.2 million deaths each year worldwide (2). Particularly in Pakistan, recent studies showed that the overall CKD prevalence is upto10.5% in young and 43.6% in aged people (3,4). CKD is caused by both environmental and clinical factors however, a number of genes are also involved from which *UMOD, SLC47A1, SHROOM3, RNF128, GATM, SLC6A19, RNF186* are more associated with renal failure, *UMOD* is the most promising gene to explore in CKD

because of its enormous effect on glomerular filtration rate (eGFR) and its risk variant (A) has high frequency of 80% in European and other populations (5-7).

Uromodulin (*UMOD)* gene is 23251bp long, encodes Uromodulin (640 amino acids) a glycoprotein produced by kidney in the epithelial cells of the TAL (thick ascending limb of loop of henle) & DCT (distal convoluted tubules) and through proteolytic cleavage by hepsin   released in urine, where it is present in highly aggregated state and forms filamentous gel that acts as barrier permeability (8). Furthermore, four EGF-like domains found in its N-terminal region including EGFI and EGFII which are involved in attachment of Ca^{+2}, D8C a domain rich in cysteine and an elastase resistant zona pellucida domain at C-terminal that causes protein polymerization (9, 10). Uromodulin protein binds a variety of ligands, regulates salt transport, play a role in innate immunity, gives protection against urinary tract infections and kidney stones (7).

UMOD gene ID: 7369 positioned on chromosome 16p12.3 is composed of 11 exons and 11 introns over a genomic region of about 20kb having 125 mutations (95% localized in exon 3 and 4) of which 78(62%) effect cysteine residues, 42(34%) influence residues other than cysteine and 5(4%) are inframe-deletions (7). The *UMOD* variant rs493393A>G was selected here because of its higher Serum creatinine (Scr) ($p=2.5x10^{-38}$) and CKD effect ($p=9.1x10^{-38}$) in general population (11). As far as the possible effect of this mutation is misfolding of Uromodulin which gets trapped in endoplasmic reticulum, as a result cellular membranes release and express it in a less efficient manner and (Scr) level is elevated subsequently contributing in the pathogenesis of autosomal dominant tubulointerstitial kidney disease (ADTKD) eventually leading to CKD and renal failure (12, 13).

The current research is designed to investigate the association of *UMOD* subject variant rs4293393A>G with CKD patients in Pakistani population and characterization of Uromodulin protein using different bioinformatics tools. This variant is located at genomic position 16:20353266A>G in intron 1 according to Chr. accession ID: NC-000016.10 and nucleotide position of r.1264T>C according to mRNA transcript ID: NM_003361.4 is in assembly: GRCh38.p14 (GCF-000001405.40).

Material and Methods:

Sample collection and DNA extraction:

A total of 75 samples (50 cases and 25 controls) were collected. CKD cases of both genders of age 19-65 years with urea level (71-217mg/dL) and SCr level (4.3- 12.3mg/dL) were selected from various hospitals in Lahore-Pakistan between June-August 2022 to find the association of *UMOD* gene variant g.16:20353266A>G with CKD. Blood from the patients were drawn into EDTA-K3 vancutainers and preserved at 4°C. DNA extraction from each sample was performed using GDSBio (https://www.gdsbio.com/en/) genomic DNA extraction kit in accordance with the manufacturer's guidelines. Samples details are given in Supplementary Table S1 (S1).

Primer designing:

Based on *UMOD* gene accession ID: NC_000016.10:20333051-20356301 with complete CDS against transcript ID: NM_003361.4 (r.1264T>C) five PCR primers were designed using OligoCalc software, three of which were ARMS-PCR primers used for amplifying normal and mutant alleles, labeled as reverse normal (RN) and reverse mutant (RM) with a mismatch inserted at the fourth nucleotide position from 3' end to improve accuracy and efficacy, and a forward common primer (FC). Further two more primers were designed with amplicon size of 618bp that served as internal control (IC). Table 1 is providing primers sequences and other attributes.

Table 1. Details of primers for ARMS-PCR.

Primers ARMS / IC	5'-3' Sequence	Tm (ºC)	Length (bases)	GCs %	Product size (bp)
Forward common	TGGGAAAGCAGTGCCAAGGT	60.5	20	55	
Reverse normal	GGGTCAGGTCCAGTGATTTCT	61.2	21	52	282
Reverse mutant	GGTCAGGTCCAGTGATTTCC	60.5	20	55	
Forward IC	TAACCCACAGCCTCCTACAC	60.5	20	55	
Reverse IC	GGTCAGGTCCAGTGATTTCC	60.5	20	55	618

PCR amplification:

A SimpliAmp thermal cycler from Applied Biosystems was used to perform the ARMS-PCR reaction. Where each sample was examined in two independent reactions, one using a common forward primer and a normal ARMS reverse primer and the other using a common forward primer and a mutant-ARMS reverse primer. To assess the fidelity of the PCR, internal control primers were also added to each tube at the same time, this primer-set amplify the targeted region of known size.

A total of 16uL reaction mixture was prepared using the following ingredients: 2uL of genomic DNA that had been extracted at a concentration of 50ng/uL, 10mM of each primer (forward common, N-reverse or M-reverse primers, forward and reverse (IC) primers), 0.05IU of *Taq* polymerase, 2.5mM MgCl₂, 2.5mM dNTPs, 1x *Taq* buffer and molecular grade water. Following PCR protocol was used with a 5-minute initial denaturation at 95°C, 30 cycles of denaturation at 95°C for 30 sec, annealing at 60°C for 30 sec, and extension at 72°C for 45 sec, with the final extension at 72°C for 10 minutes and then stored at 4°C for infinity (Figure 1).
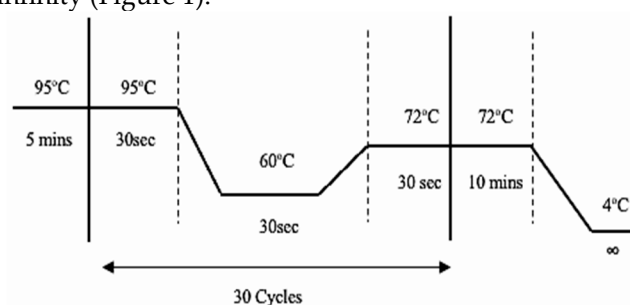

Figure 1. ARMS-PCR protocol.

Statistical analysis:

PLINK data toolset was used to evaluate the Hardy Weinberg Equilibrium (HWE) using equation ($p^2 + 2pq + q^2 = 1$) to check if sampled population is in accordance with HWE or not and Chi-Square association using $\chi2 = \sum \frac{(O-E)^2}{E}$ equation to find the association of subject variant with CKD after screening the rs4293393A>G locus in total of seventy-five samples. Furthermore odds-ratio and allele frequencies in cases and controls were also computed for the overall selected population.

Additionally, Pearson correlation test on SPSS was performed to analyze whether age has any correlation with increasing creatinine or urea level and to examine a correlation between urea to creatinine level.

Bioinformatics Analyses:

Multiple tools/databases are available for information about genes or proteins of any species. These databases include Expert Protein Analysis System (EXPASY), European Bioinformatics Institute (EBI), National Center of Biotechnology Information (NCBI) and Genome Database (GDB). Transcript of *UMOD* was obtained from NCBI (https://www.ncbi.nlm.nih.gov/) and sequence

of the subject protein was generated from Sequence Translation EMBL-EBI tool from Genome Data Viewer (GDV), it was analyzed that the variant is in the intronic region and there is no mutation in protein sequence so, different tools were applied for characterization of native protein.

Physicochemical properties:

PePDraw and Expasy ProtParam tools were used for the prediction of physicochemical properties. In this study ProtParam tool (http://web.expasy.org/protparam/) was applied for the prediction of properties like number of amino acids, molecular weight, atomic composition, theoretical PI, instability and aliphatic index, extension coefficient, grand average of hydropathicity (GRAVY) and estimated half-life of the protein.

Secondary structure prediction:

Beta strands, coils and helices make up secondary structure. This feature of the protein was analyzed via PsiPred tool (http://bioinf.cs.ucl.ac.uk/psipred/).

Conserved domains valuations:

Prediction of protein domains was performed by EBI-InterPro, EBL-SMART, Pfam, Motif server (http://myhits.isb-sib.ch/cgi-bin/motif_scan), CD Server (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi), Interproscan 5 (http://www.ebi.ac.uk/Tools/pfa/iprscan5/) and Motif scan tool. Here the domains of Uromoduline were observed using Motif scan tool (https://www.genome.jp/tools-bin/search_motif_lib).

Transmembrane structure investigation:

Transmembrane structure of protein can be predicted using Protscale (http://web.expasy.org/protscale/), CTU-TMHMM Server v.2 (https://services.healthtech.dtu.dk/service.php?TMHMM-2.0) and Expasy- TMpred. CTU-TMHMM Server v.2 was used to find transmembrane structure of UMOD protein in the current study.

Three-dimensional (3D) structure prediction:

PDB RCSB (https://www.rcsb.org/) was applied for 3D structure of Uromodulin. The Protein Data Bank collection, which contains details about the 3D forms of proteins, nucleic acids and complex assemblies, is the source of the material used in this site. This database aids students and researchers in understanding many facets of biomedicine and agriculture, from protein production to health and illness. The RCSB PDB organizes and records PDB data as a member of the wwPDB.

Post-translational modifications:

ScanProsite (http://prosite.expasy.org/scanprosite/) tool was applied for analyzing post translational changes such as binding with metals, prosthetic groups and organelles were analyzed. For acetylation GPS_PAIL 2.0 (http://bdmpail.biocuckoo.org/prediction.php), methylation **PRmePRed (**http://bioinfo.icgeb.res.in/PRmePRed/#) and glycosylation NetOGlyc-4.0 (https://services.healthtech.dtu.dk/service.php?NetNGlyc-1.0) tools were applied.

Protein-protein interaction and functional analysis:

STRING database was applied for analyzing protein-protein interaction networks of Uromodulin protein, all these associated functional networks of Uromodulin help to perform the proper function.

Results:

Variant genotyping:

In the current study, *UMOD* variant (rs4293393) revealed an association with CKD in the Pakistani population. Genotyping of 75 samples were performed (cases=50, controls=25) (Figure 2). In CKD patients, 36 were homozygous wild-type, 13 heterozygous and 01 was homozygous mutant. Similarly, the controls revealed 09 homozygous wild-type, 08 heterozygous and 08 homozygous mutant. The overall sampled population showed that 45(60%) were homozygous wild-type, 21(28%) heterozygous and 9(12%) homozygous-mutant. Moreover, alternative allele (G) frequency was 0.15 and 0.48 in cases and controls respectively.

Figure 2. Genotyping of 50 CKD cases (A) and 25 controls (B).

Statistical association/correlation and odds-ratio analyses:

Using the PLINK data analysis toolset, HWE and Chi-Square association analysis were performed, a significant $p-value = 0.031$ was observed which manifests that our population is not obeying the HWE as it is below the set threshold confidence interval of 0.05, which is rejecting the null hypothesis. Moreover, alternative allele (G) frequency was 0.48 in controls and 0.15 in cases showing that mutant allele (G) is under-selection in CKD patients.

Further analysis revealed a significant association of *UMOD* variant (rs4293393) with CKD patients by the outcomes ($p = 1.40×10^{-005}$) and OR of 0.2 showing that there is 0.2 times lesser chances of the occurrence of mutant allele (G) in the cases which is protective allele and there is 0.8 times more chances of occurrence of wild-type allele (A) in patients than controls. Hence, Risk of the wild-type allele is 0.8 times higher for developing CKD as compared to mutant allele (Table 2).

Table 2. Analysis of the associations between study variants and the Pakistani population.

# of samples	Chr. position	cDNA variant NM_003361.4	Protein variant NP_003352.2	Genotypic information			Alternative allele frequency		*p*-value (OR)
				Homo-wild (AA/%)	Hetero (AG/%)	Homo-mutant (GG/%)	Cases	Controls	

| 75 | 16:20353266 | r.1264T>C | p.= | 45/60 | 21/28 | 9/12 | 0.15 | 0.48 | 1.40×10^{-005} (0.2) |

Ultimately, using SPSS software Pearson correlation test was applied to check correlation between the variables (age, urea and creatinine). It was analysed that there is moderate positive correlation of urea and creatinine, weak negative correlation in age and creatinine level and very weak positive correlation in age and urea levels (Table 3).

Table 3. Pearson Correlation of age, urea and creatinine.

Variable	R-value
Urea/creatinine	.435
Age/creatinine	-0.117
Age/urea	.099

UMOD protein analysis using different bioinformatics tools:

I.   *Physicochemical findings using Protparam:*

The physicochemical characteristics of the Uromodulin protein were estimated using ProtParam tool. The statistics showed that there were total 640 aa with molecular weight of 69760.86, theoretical isoelectric point (PI) was 5.05, negatively charged residues (Asp + Glu) were 69, while positively charged residues (Arg + Lys) were less that is 46. The molecular formula of protein was $C_{3011}H_{4654}N_{832}O_{952}S_{63}$, with a total number of atoms 9512. The extension coefficient can be observed either considering all pairs of Cys form Cystines in protein or all pairs of Cys are reduced. When all pairs of Cys form Cystines observed value was 101780 but with reduced pairs the value was 98780. Under in-vitro conditions the mammalian proteins half-life was estimated to be 30hrs, while in vivo environment like yeast and *E. coli*, the half-life was predicted as greater than 20 and 10 respectively. For the laboratory experiments the stability of protein is critical so here instability score was also predicted which is 40.53, indicates that the studied protein is stable. Moreover, Aliphatic index was 70.69 and Grand average of hydropathicity (GRAVY) is -0.111 (Table 4).

Table 4. Physicochemical properties of Uromodulin protein.

Physicochemical properties	Parameters
Number of amino acids	640aa
Theoretical PI	5.05
Molecular weight	69760.86
(Asp + Glu)Negative charged residues	69
(Arg + Lys)Positive charged residues	46
Molecular formula	$C_{3011}H_{4654}N_{832}O_{952}S_{63}$
No. of atoms	9512
Extension coefficient	101780 (considering that all pairs of cysteines are formed) 98780 (supposing that all pairs of cysteines are reduced)
Estimated half life	30hrs
Instability index	40.53
Aliphatic index	70.69

GRAVY	-0.111

If we look into the detailed structure of protein, the two amino acids were observed to be higher in number that is 55(8.60%) of Serine and Leucine. Further Glycine and Threonine with 52(8.10%) were second highest. The remaining aa were Cys, Ala, Val, Asp, Arg, Gln and Pro were 48(7.50%), 44(6.9%), 44(6.9%), 39(6.1%), 30 (4.7%), 30(4.7%), 29 (4.5%) respectively. Similarly ratio of carbon, hydrogen, nitrogen, oxygen and sulfur in the protein is 31%, 49%, 9%, 10%, and 1% respectively (Figure 3).
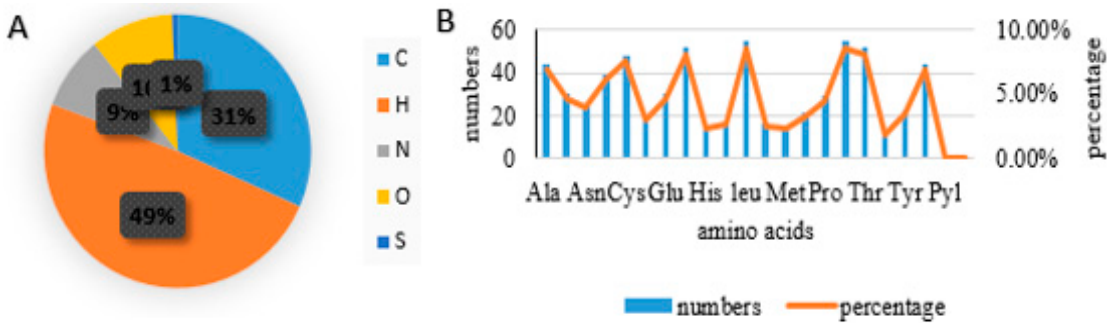

Figure 3. Atomic ratio of protein (A), number of amino acids with percentages (B).

II. *Secondary structure prediction of UMOD using PsiPred:*

This tool predicts the secondary structure of studied protein. In the following (Figure 4), the amino acids region highlighted in pink shade (4) represents the helix, similarly yellow (34) symbolizes the strands and grey shades specifies the coils of protein (Figure 4).
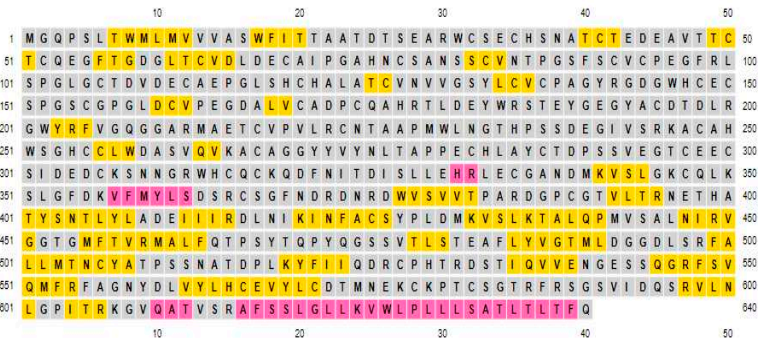

Figure 4. Secondary structure prediction in Uromodulin. .

III. *Transmembrane structure prediction using TMHMM - 2.0:*

According to this tool 3.49475aa are anticipated to be present in outside of membrane and 1.5012 are expected to be present in first 60 aa on an average that could play a part in transmembrane helical structure (Figure 5).
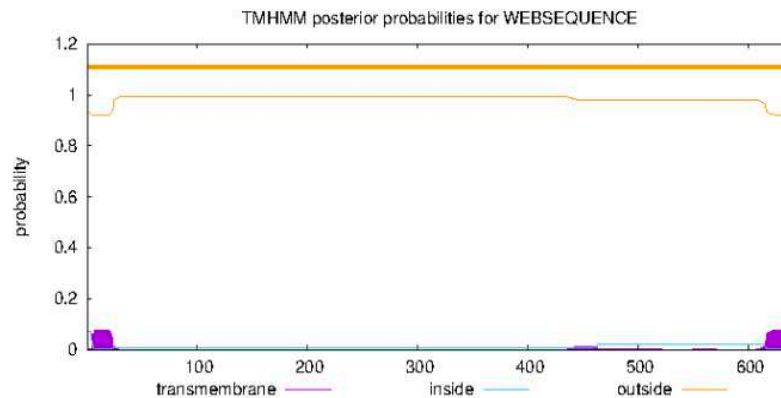
Figure 5. Transmembrane helix in Uromodulin.

IV.   *Protein motifs prediction using Motif finder:*

This tool predicted that there are eight motifs in the protein which are Zona pellucida, EGF-3, EGF_CA, cEGF, EGF, hEGF, FXa_inhibition, and EGF_MSP1_1. Zona pellucida like domain located from 335 aa position to 583 aa, EGF domain is present in three regions in the whole sequence (32 to 63, 73 to 99, 118 to 148 and 299 to 321). Calcium binding EGF domain is present in two regions from 65 aa position to 99 and from 108 to 148. Furthermore, each domain's position and i-Evalue is given in the protein sequence is given in the (Table 5).

Table 5. Motif information of Uromodulin.

Pfam ID:	Position	Description	i-Evalue
Zona_pellucida	335..583	PF00100, Zona pellucida-like domain	9.5e-52
EGF_3	34..63	PF12947, EGF domain	5.7e-08
	73..99		1.4e-06
	118..148		3.2e-10
	299..321		0.88
EGF_CA	65..99	PF07645, Calcium-binding EGF domain	1.5e-12
	108..148		6.6e-10
cEGF	49..68	PF12662, Complement Clr-like EGF-like	0.0001
	89..111		8.1e-06
	132..149		0.0071
EGF	34..59	PF00008, EGF-like domain	0.021
	75..99		0.00011
	115..144		7.4e-06
hEGF	35..56	PF12661, Human growth factor-like EGF	0.87
	77..98		0.0084
	124..141		9.9e-05
FXa_inhibition	76..102	PF14670, Coagulation Factor Xa inhibitory site	0.0033
	120..143		0.25
EGF_MSP1_1	75..98	PF12946, MSP1 EGF domain 1	0.25

119..148	0.0053

V.   *Post-translational modifications using ScanProsite:*

This tool gives detailed insight of post-translational changes like phosphorylation, glycosylation, cell attachment sequences occur in the protein. Furthermore, it gives the count of disulphide bridges present in different domains of protein. 32-174 aa is Cysteine rich region of protein and mostly disulfide bridges are present at this region. In complete protein sequence 48 Cysteine residues and 24 disulfide bonds exists, location of disulphide bonds present within the domains of Uromodulin protein are depicted in (Figure 6).
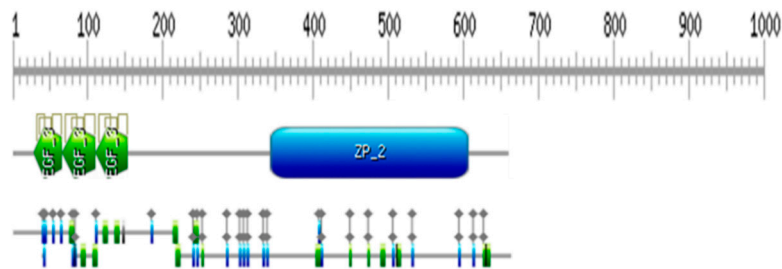

Figure 6. Disulphide bridges in Uromodulin domain.

Results showed the phosphorylation prediction for all the residues of Serine (S), Threonine (T) and Tyrosine (Y) present in the complete sequence. Threonine sites where phosphorylation occurs are 40, 42, 51, 62, 107, 179, 296, 394, 457, 489, 573, 605 and phosphorylation at Serine residues occur at position 301, 237, 244, 291, 327,434 and 591. The kind of glycosylation which might be N-linked, C-linked, S-linked and O-linked, depends on the type of aa atom that bonds the carbohydrate chain. In Uromodulin there is only Asparagine N-linked glycosylation and sites are 38, 76, 80, 232, 275, 322, 396, and 513.

Table 6. N-glycosylation and phosphorylation sites of Uromodulin.

Amino acid position	Glycosylation	Phosphorylation
38	N-linked glycosylation at asparagine	
40	--	Phospho-threonine
42	--	Phospho-threonine
51	--	Phospho-threonine
62	--	Phospho-threonine
76	N-linked glycosylation at asparagine	--
80	N-linked glycosylation at asparagine	--
107	--	Phospho-threonine
179	--	Phospho-threonine
232	N-linked glycosylation at asparagine	
237	--	Phospho-serine
275	N-linked glycosylation at asparagine	--
291	--	Phospho-serine
296	--	Phospho-threonine
301	--	Phospho-serine
322	N-linked glycosylation at asparagine	--
327	--	Phospho-serine

394	--	Phospho-serine
396	N-linked glycosylation at asparagine	--
434	--	Phospho-serine
457	--	Phospho-threonine
489	--	Phospho-threonine
513	N-linked glycosylation at asparagine	--
573	--	Phospho-threonine
591	--	Phospho-serine
605	--	Phospho-threonine

Covalent addition myristate a C14-saturated fatty acid causes acetylation to the N-terminal of a number of proteins via an amide linkage and N-myristoylation site of Uromodulin protein are 73-78, 88-93, 103-108, 116-121, 131-136, 207-212, 210-215, 233-238, 390-395, 474-479, 493-498, 608-613. The Arg-Gly-Asp (RGD) sequences range from 142-144.

VI.   *Phosphorylation prediction using NetPhos 3.1:*

It displayed the Serine, Threonine and Tyrosine phosphorylation sites in the sequence of uromodulin. Pink lines show the threshold limit which is usually 0.5. Higher the threshold value of a residue higher is its possibility to be phosphorylated. Red lines indicate phosphorylation on the serine, green line on threonine and blue line shows phosphorylation at tyrosine residue (Figure 7).
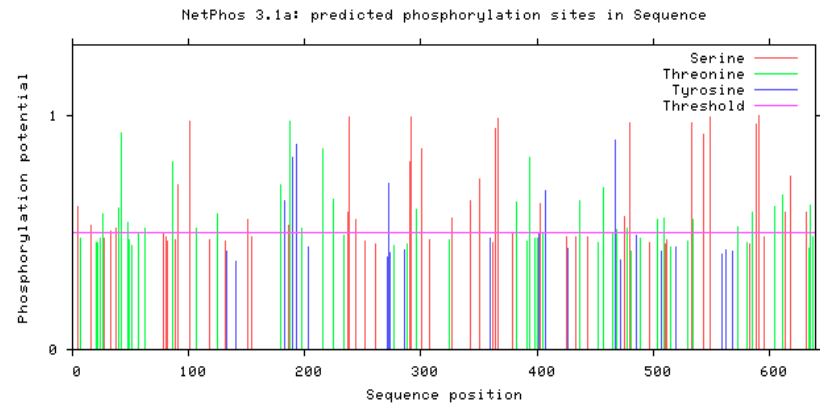

Figure 7. Serine, Threonine and Tyrosine phosphorylation sites.

VII.   *Acetylation prediction using GPS PAIL 2.0:*

This tool predicts acetylation of all the Lysine (K) residues in the subject protein sequence at position 246, 265, 307, 318, 341, 346, 350, 356, 420, 432, 436, 519, 577, 579 and 624 (Figure 8).
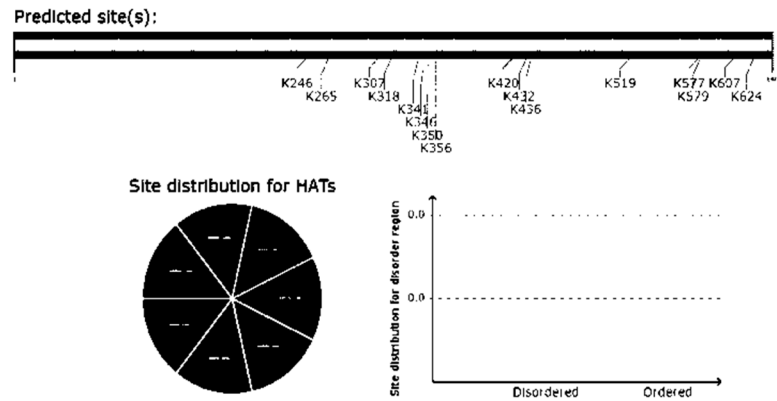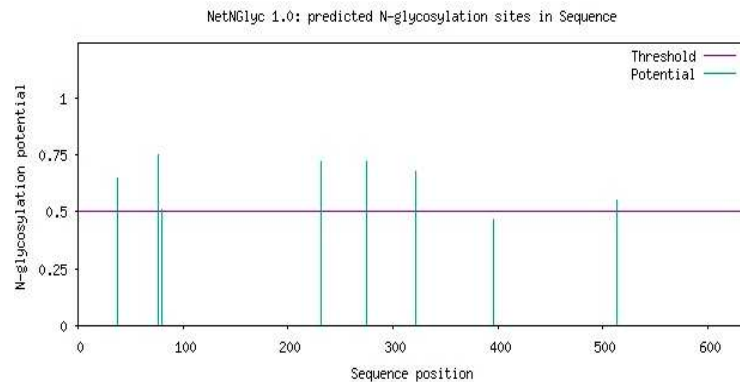
Figure 8. Acetylation of Lysine residue within Uromodulin.

VIII.   Glycosylation prediction of Uromodulin using NetOGlyc-4.0:

This tool predicted glycosylation of carbohydrates in Uromodulin. Purple line gives an indication of threshold value (0.5). Higher threshold value of any residue showed the greater possibility of glycosylation (Figure 9).

Figure 9. Glycosylation sites of Uromodulin protein.

IX.   Methylation prediction of Uromodulin using PRmePRed:

This tool provides the prediction for all K and R residues for methylation in subject protein. Uromodulin residues for methylation are 99, 142, 200, 204, 212, 245, 365, 385, 449, 459, 498,547, 586, 588, 597, 606, and 615.

Table 7. Methylation sites of Uromodulin.

Position	Peptide	Prediction score
99	FSCVCPEGF**R**LSPGLGCTD	0.692402
142	YLCVCPAGY**R**GDGWHCECS	0.785598
200	EGYACDTDL**R**GWYRFVGQG	0.780858
204	CDTDLRGWY**R**FVGQGGARM	0.823455
212	YRFVGQGGA**R**MAETCVPVL	0.788485
245	PSSDEGIVSR**K**CAHWSGH	0.556259
365	KVFMYLSDS**R**CSGFNDRDN	0.543877
385	DWVSVVTPA**R**DGPCGTVLT	0.708181
449	QPMVSALNI**R**VGGTGMFTV	0.763262
459	VGGTGMFTV**R**MALFQTPSY	0.515069
498	TMLDGGDLS**R**FALLMTNCY	0.602271
547	VENGESSQG**R**FSVQMFRFA	0.932742
586	KCKPTCSGT**R**FRSGSVIDQ	0.766291
588	KPTCSGTRF**R**SGSVIDQSR	0.817883
597	RSGSVIDQS**R**VLNLGPITR	0.676464
606	RVLNLGPITR**K**GVQATVSR	0.831847
615	RKGVQATVS**R**AFSSLGLLK	0.729294

X.   Prediction of 3D structure using PDB RCSB:

This tool predicted 3D structure of Uromodulin where side chains showing the carbohydrate chains and inner chains represent polymer structure.
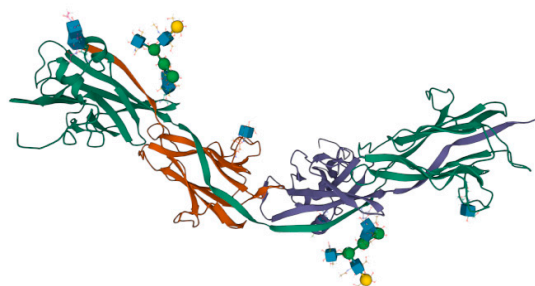

Figure 10. 3D structure of Uromodulin protein.

XI.   Prediction of Uromodulin interaction networks using STRING database:

This tool is used to predict the interaction sites of Uromodulin protein which are B4GALNT2, TNF, SLC12A1, AQP2, ALB, GTPBP1, CD36, NPC1, SLc22A12, and SHROOM3 that help Uromodulin to perform its proper function (Figure 11).
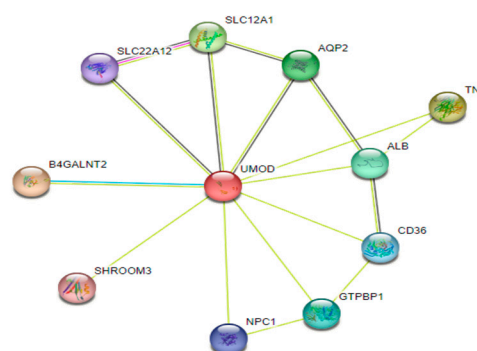

Figure 11. Associated functional networks of Uromodulin.

Around 11 nodes are present connecting Uromodulin protein, nodes represent protein networking. Query proteins and the first shell of interactors are represented by coloured network nodes, whereas the second shell of interactors is represented by white network nodes. Edges reflect protein-protein connections that are intended to be precise and meaningful, i.e., proteins work together to perform a common task; this does not imply that the proteins are physically bound to one another. Anticipated interactions, Black: coexpression, light blue: protein homology, light green: textmining, red: gene fusions, blue: gene co-occurrence and red represents gene neighbourhood.

Discussion:

Average rate of protein synthesis within the kidney is 40% per day (14) and Uromodulin is the most abundant protein synthesized in TAL of kidney (mean rate = 50-150mg/24hrs) (15). All the transcripts of *UMOD* are present in kidney (16) so it determines the structural integrity of the distal nephron and kidney function (17). Polymorphism in *UMOD* gene has a great association with developing CKD and the screened variant rs4293393A>G causes ADTKD which is triggered by the accumulation of Uromodulin in ER in epithelial cells of TAL resulting in fibrosis. Subject variants acting together with other promoter region variants effect transcriptional activity of *UMOD* gene promoter and double the expression of Uromodulin (5, 18). Uromodulin consists of 48 conserved Cysteine residues (24 disulphide bond), thus its release takes time but immature Uromodulin takes even more time to excrete and gets trapped (Figure 13) (19, 20), ADTKD-*UMOD* is symptomized as higher level of Scr, uric acid and urea in blood and decreased uric acid (21).
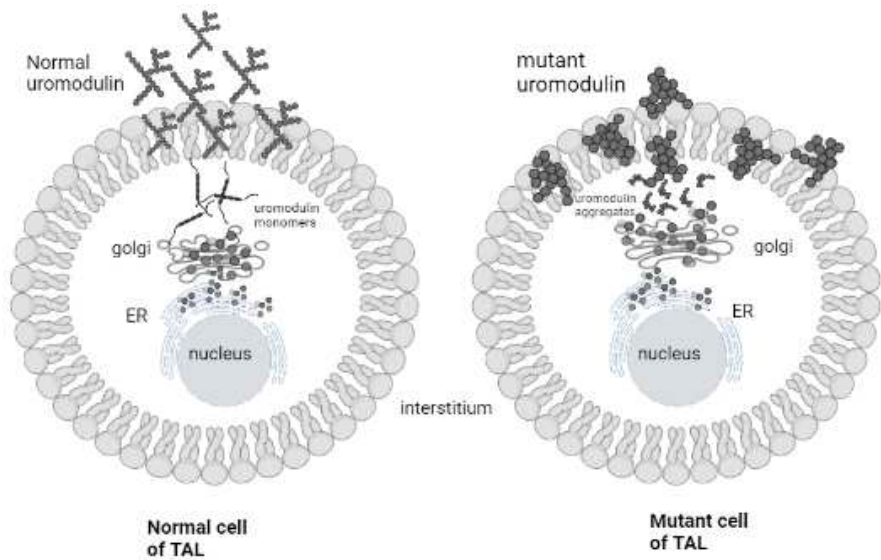
Figure 13. Pathophysiology of ADTKD, Disruption of disulphide bonds causes accumulation of Uromodulin in ER as well as in cell membrane and results interstitial damage in kidnies.

The primary objective of this research endeavor was to determine the association of variant rs4293393 with CKD in Pakistani population and it gives us enough evidence that undoubtedly rs4293393A>G is associated with developing CKD in Pakistan. Our findings revealed that 45 were homozygous wild-type, 21 heterozygous, 09 homozygous-mutant in overall sampled population. The locus 16:20353266 is highly conserved, and this conserved nature is observed by multiple sequence alignment (MSA) in ten different species (Figure 14), which means any change in this locus may cause abnormalities in the body.
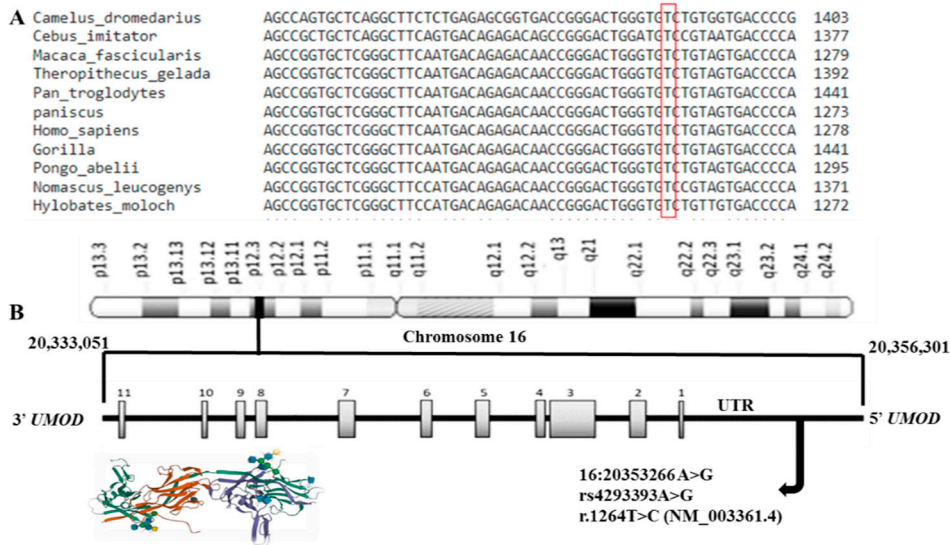

Figure 14. MSA of genotyped locus in ten different species (A), Genetic map of *UMOD*, subject variant location & structure of Uromodulin protein (B).

Previously a number of genome-wide association studies (GWAS) have also confirmed the fact that rs4293393A>G is associated with developing CKD (22-24). Moreover, a number of studies reported that T allele at rs4293393 is the risk allele for CKD, increasing urea and creatinine level so it is more common in patients than controls with almost 80% of the population frequency (7). In accordance with the current study another study in Iraq showed that (A) allele was more common in controls (25). Similarly, one of anothter study where 720 pateints in Belgium showed that (G) is in 19% of the population (18). Similarly in Europeans the same variant has 19% guanosine and 81% adenosine residues (5).

Bioinformatics analysis provide us a complete insight of the structural and functional features of the Uromodulin protein. Physical and chemical parameters was done using ProtParam tool and Psipred tool predicted four helices, 34 beta strands and all the remaining amino acids make coils. TMHMM - 2.0 predicted no transmembrane structure in the subject protein.

Motif finder tool characterized that eight motifs within Uromodulin and for the detailed characterization of post-translational changes ScanProsite tool was used which gave the information that disulphide bridges are present in only EGF like domains of Uromodulin. Moreover, it predicted the glycosylation, phosphorylation cell attachment sites of Uromodulin. GPS PAIL 2.0 predicted acetylation, PRmePRed tool predicted methylation sites and Protein-protein interaction of Uromodulin was predicted by STRING database.

In the current study *UMOD* gene variant (rs4393393A>G) is found associated with CKD in Pakistani population with $p = 1.40 \times 10^{-005}$. In future, screening of the subject variant could be used as a biomarker to evaluate risk factor and genetic councelling may be opted to harness this disease by limiting the segreggation of risk allele in next generations. Furtermore, indiginous population could be studied by whole-genome/exome sequencing and associated variants may be highlighted.

Conclusion:

Aforementioned *UMOD* gene variant is associated with CKD in sampled population of Pakistani origin. Mutant allele frequency is 0.15 and risk of wild-type allele is 0.8 times higher in cases vs controls. Hence, the screening of studied variant rs4293393A>G will be helpful to know genomic architecture of the indigenous population in order to prevent this life-threatening condition.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org.

Conflict of Interest: The authors have no conflict of interest.

Ethical statement: The samples were obtained by the consent of the Akhtar Saeed Trust Hospital, Lahore, Pakistan. The concerned Institutional Review Board reviewed the study protocol which complies with the laid-down rules and regulations of their Ethics Committee.

References

1.   Chen TK, Knicely DH, Grams ME. Chronic kidney disease diagnosis and management: a review. Jama. 2019;322(13):1294-304.

2.   Levin A, Tonelli M, Bonventre J, Coresh J, Donner J-A, Fogo AB, et al. Global kidney health 2017 and beyond: a roadmap for closing gaps in care, research, and policy. The Lancet. 2017;390(10105):1888-917.

3.   Alam A, Amanullah F, Baig-Ansari N, Lotia-Farrukh I, Khan FS. Prevalence and risk factors of kidney disease in urban Karachi: baseline findings from a community cohort study. BMC research notes. 2014;7(1):1-11.

4.   Jessani S, Bux R, Jafar TH. Prevalence, determinants, and management of chronic kidney disease in Karachi, Pakistan-a community based cross-sectional study. BMC nephrology. 2014;15(1):1-9.

5.   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. Nature medicine. 2013;19(12):1655-60.

6.   Devuyst O, Pattaro C. The *UMOD* locus: insights into the pathogenesis and prognosis of kidney disease. Journal of the American Society of Nephrology. 2018;29(3):713-26.

7.   Devuyst O, Olinger E, Rampoldi L. Uromodulin: from physiology to rare and complex kidney disorders. Nature Reviews Nephrology. 2017;13(9):525-44.

8.   Brunati M, Perucca S, Han L, Cattaneo A, Consolato F, Andolfo A, et al. The serine protease hepsin mediates urinary secretion and polymerisation of Zona Pellucida domain protein Uromodulin. Elife. 2015;4:e08887.

9.   Bokhove M, Nishimura K, Brunati M, Han L, De Sanctis D, Rampoldi L, et al. A structured interdomain linker directs self-polymerization of human Uromodulin. Proceedings of the National Academy of Sciences. 2016;113(6):1552-7.

10. Jovine L, Qi H, Williams Z, Litscher E, Wassarman PM. The ZP domain is a conserved module for polymerization of extracellular proteins. Nature cell biology. 2002;4(6):457-61.

11. Sveinbjornsson G, Mikaelsdottir E, Palsson R, Indridason OS, Holm H, Jonasdottir A, et al. Rare mutations associating with serum creatinine and chronic kidney disease. Human molecular genetics. 2014;23(25):6935-43.

12. Bernascone I, Vavassori S, Di Pentima A, Santambrogio S, Lamorte G, Amoroso A, et al. Defective intracellular trafficking of Uromodulin mutant isoforms. Traffic. 2006;7(11):1567-79.

13. Smith GD, Robinson C, Stewart AP, Edwards EL, Karet HI, Norden AG, et al. Characterization of a recurrent in-frame *UMOD* indel mutation causing late-onset autosomal dominant end-stage renal failure. Clinical Journal of the American Society of Nephrology. 2011;6(12):2766-74.

14. Garibotto G, Tessari P, Robaudo C, Zanetti M, Saffioti S, Vettore M, et al. Protein turnover in the kidney and the whole body in humans. Mineral and electrolyte metabolism. 1997;23(3-6):185-8.

15. Lee JW, Chou C-L, Knepper MA. Deep sequencing in microdissected renal tubules identifies nephron segment–specific transcriptomes. Journal of the American Society of Nephrology. 2015;26(11):2669-77.

16. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. Science (New York, NY). 2015;347(6220):1260419-.

17. 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. Nephrology Dialysis Transplantation. 2018;33(2):284-95.

18. Kidd K, Vylet'al P, Schaeffer C, Olinger E, Živná M, Hodaňová K, et al. Genetic and clinical predictors of age of ESKD in individuals with autosomal dominant tubulointerstitial kidney disease due to *UMOD* mutations. Kidney international reports. 2020;5(9):1472-85.

19. Utami SB, Mahati E, Li P, Maharani N, Ikeda N, Bahrudin U, et al. Apoptosis induced by an Uromodulin mutant C112Y and its suppression by topiroxostat. Clinical and experimental nephrology. 2015;19(4):576-84.

20. 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. Human molecular genetics. 2009;18(16):2963-74.

21. Lin Z, Yang J, Liu H, Cai D, An Z, Yu Y, et al. A novel Uromodulin mutation in autosomal dominant tubulointerstitial kidney disease: a pedigree-based study and literature review. Renal Failure. 2018;40(1):146-51.

22. Kumar V, Yadav AK, Kumar V, Bhansali A, Jha V. Uromodulin rs4293393 T> C variation is associated with kidney disease in patients with type 2 diabetes. The Indian Journal of Medical Research. 2017;146(Suppl 2):S15.

23. Wang J, Liu L, He K, Gao B, Wang F, Zhao M, et al. *UMOD* Polymorphisms Associated with Kidney Function, Serum Uromodulin and Risk of Mortality among Patients with Chronic Kidney Disease, Results from the C-STRIDE Study. Genes. 2021;12(11):1687.

24. Devuyst O, Bochud M, Olinger E. *UMOD* and the architecture of kidney disease. Pflügers Archiv-European Journal of Physiology. 2022:1-11.

25. Al-khalidi RAA, Alwan IA, Razaaq HAA. The extent of *UMOD* gene polymorphism and its level in type 2 diabetes patients. Annals of Tropical Medicine and Public Health. 2020;23:231-803.