

# Plasma Mucin-1 (CA15-3) Levels in Autosomal Dominant Tubulointerstitial Kidney Disease due to *MUC1* Mutations

Petr Vylet<sup>al</sup><sup>a</sup> Kendrah Kidd<sup>a, b</sup> Hannah C. Ainsworth<sup>c</sup> Drahomíra Springer<sup>d</sup>  
Alena Vrbacká<sup>a</sup> Anna Přistoupilová<sup>a</sup> Rebecca P. Hughey<sup>e</sup> Seth L. Alper<sup>f, g</sup>  
Niall Lennon<sup>f</sup> Steven Harrison<sup>f</sup> Maegan Harden<sup>f</sup> Victoria Robins<sup>b</sup>  
Abbigail Taylor<sup>b</sup> Lauren Martin<sup>b</sup> Katrice Howard<sup>b</sup> Ibrahim Bitar<sup>h</sup>  
Carl D. Langefeld<sup>c</sup> Veronika Barešová<sup>a</sup> Hana Hartmannová<sup>a</sup>  
Kateřina Hodaňová<sup>a</sup> Tomáš Zima<sup>d</sup> Martina Živná<sup>a</sup> Stanislav Kmoch<sup>a, b</sup>  
Anthony J. Bleyer<sup>a, b</sup>

<sup>a</sup>Department of Paediatrics and Inherited Metabolic Disorders, Research Unit of Rare Diseases, First Faculty of Medicine, Charles University, Prague, Czech Republic; <sup>b</sup>Section on Nephrology, Wake Forest School of Medicine, Winston-Salem, NC, USA; <sup>c</sup>Division of Public Health Sciences, Department of Biostatistics and Data Science, Wake Forest School of Medicine, Winston-Salem, NC, USA; <sup>d</sup>Institute of Medical Biochemistry and Laboratory Diagnostics, General University Hospital and the First Faculty of Medicine of Charles University, Prague, Czech Republic; <sup>e</sup>Renal-Electrolyte Division, Department of Medicine, University of Pittsburgh, Pittsburgh, PA, USA; <sup>f</sup>The Broad Institute of MIT and Harvard, Cambridge, MA, USA; <sup>g</sup>Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA, USA; <sup>h</sup>Biomedical Center and Institute of Microbiology, Faculty of Medicine in Pilsen of Charles University, Pilsen, Czech Republic

## Keywords

CA15-3 · Mucin-1 · Autosomal dominant tubulointerstitial kidney disease · ADTKD-*MUC1* · rs4072037

## Abstract

**Introduction:** Patients with ADTKD-*MUC1* have one allele producing normal mucin-1 (*MUC1*) and one allele producing mutant *MUC1*, which remains intracellular. We hypothesized that ADTKD-*MUC1* patients, who have only 1 secretory-competent wild-type *MUC1* allele, should exhibit decreased plasma mucin-1 (*MUC1*) levels. To test this hypothesis, we repurposed the serum CA15-3 assay used to measure *MUC1* in

breast cancer to measure plasma *MUC1* levels in ADTKD-*MUC1*. **Methods:** This cross-sectional study analyzed CA15-3 levels in a reference population of 6,850 individuals, in 85 individuals with ADTKD-*MUC1*, and in a control population including 135 individuals with ADTKD-*UMOD* and 114 healthy individuals. **Results:** Plasma CA15-3 levels (mean  $\pm$  standard deviation) were  $8.6 \pm 4.3$  U/mL in individuals with ADTKD-*MUC1* and  $14.6 \pm 5.6$  U/mL in controls ( $p < 0.001$ ). While there was a significant difference in mean CA15-3 levels, there was substantial overlap between the 2 groups. Plasma CA15-3 levels were  $<5$  U/mL in 22% of ADTKD-*MUC1* patients, in 0/249 controls, and in 1% of the reference population. Plasma CA15-3 levels were  $>20$  U/mL in 1/85 ADTKD-

*MUC1* patients, in 18% of control individuals, and in 25% of the reference population. Segregation of plasma CA15-3 levels by the rs4072037 genotype did not significantly improve differentiation between affected and unaffected individuals. CA15-3 levels were minimally affected by gender and estimated glomerular filtration rate. **Discussion/Conclusions:** Plasma CA15-3 levels in ADTKD-*MUC1* patients are approximately 40% lower than levels in healthy individuals, though there is significant overlap between groups. Further investigations need to be performed to see if plasma CA15-3 levels would be useful in diagnosis, prognosis, or assessing response to new therapies in this disorder.

© 2021 The Author(s)  
Published by S. Karger AG, Basel

## Introduction

Mucin-1 (*MUC1*) is a membrane-anchored protein with an extracellular glycosylated domain linked by a disulfide bond to a short cytoplasmic tail. This extracellular domain contains a variable number (20–125 copies) of repetitive segments of 20 amino acids in tandem that have a high threonine and serine content (called a variable number of tandem repeats [VNTR] domain). During *MUC1* biosynthesis, sugar residues are added to the amino acids in the VNTR, resulting in the adherent, mucinous properties of this glycoprotein [1, 2]. *MUC1* is widely expressed in many tissues, including the lung, breast, and the thick ascending limb and distal convoluted tubule of the kidney, providing protection to apical epithelial surfaces [3]. The lungs contribute the majority of *MUC1* detected in plasma or serum of healthy individuals with plasma *MUC1* levels dropping by 36% 1 week after lobectomy [4].

The CA15-3 assay is a widely available commercial assay that measures blood levels of *MUC1* using the DF3 antibody, which detects a nonglycosylated 8-amino acid sequence (DTRPAPGS) within the VNTR [5]. The assay is widely used to monitor breast cancer progression [6], as *MUC1* is expressed by breast cancer cells, and blood levels of *MUC1* increase with rising tumor burden. Development of clinical assays to detect domains of the *MUC1* molecule other than the VNTR has been difficult due to low immunogenicity of other regions of the *MUC1* protein. Any dependence of CA15-3 assay reactivity on *MUC1* VNTR repeat number remains undefined [7].

Autosomal dominant tubulointerstitial kidney disease (ADTKD) due to *MUC1* mutations (ADTKD-*MUC1*) is characterized by a bland urinary sediment and slow, variable progression of CKD leading to end-stage kidney dis-

ease (ESKD) between 30 and 80 years [8–10]. ADTKD-*MUC1* is caused by frameshift mutations that result in a disease-specific *MUC1* frameshift protein (*MUC1fs*) [2, 11, 12]. In ADTKD-*MUC1*, the mutated frameshift protein deposits in cargo vesicles that shuttle immature proteins between the endoplasmic reticulum and the Golgi [13]. This deposition leads to intracellular stress, accelerated apoptosis, and chronic kidney failure. In ADTKD-*MUC1*, all *MUC1*-producing epithelia in the body produce mutant *MUC1*, with deposition of intracellular mutant *MUC1* having been found histologically in a wide variety of tissues [2]. However, for unclear reasons, clinical consequences of this mutant *MUC1* deposition only occur in the kidney.

Due to the high guanosine/cytosine content and repetitive nature of the VNTR region in which *MUC1* frameshift mutations reside [2, 11, 14], routine Sanger sequencing cannot be performed to identify mutations. At present, approved clinical testing for the *MUC1fs* mutation is only performed in the USA at the Broad Institute of MIT and Harvard [15]. Current testing allows for the identification of some, but not all, mutations that cause ADTKD-*MUC1* [11]. Thus, the test is difficult to perform, not available in all locations, and does not identify all causative mutations.

The hypothesis of this investigation was that plasma *MUC1* levels in patients with ADTKD-*MUC1* would be significantly lower than in healthy individuals. As ADTKD-*MUC1* is autosomal dominant, patients with ADTKD-*MUC1* will have one allele producing wild-type *MUC1* that contributes to plasma *MUC1* concentration and another allele producing *MUC1fs* that is retained intracellularly (see online suppl. Fig. 1; for all online suppl. material, see [www.karger.com/doi/10.1159/000515810](http://www.karger.com/doi/10.1159/000515810)). These pathophysiologic changes will occur in all epithelial tissues that produce *MUC1*, resulting in an approximately 50% reduction in normal *MUC1* production and plasma *MUC1* levels. Studying *MUC1* production in ADTKD-*MUC1* could provide insight into this disorder, and plasma *MUC1* levels might be used in the future to determine prognosis or monitor response to new treatments. In addition, measurement of plasma *MUC1* in families with ADTKD could be helpful in directing further genetic testing.

Multiple factors that could affect plasma *MUC1* levels include male gender [16], advancing age [17], and decreased estimated glomerular filtration rate (eGFR) [17, 18]. In addition, the rs4072037-C>T single-nucleotide polymorphism (SNP) located at the 5' end of exon 2 encodes an alternative splice acceptor associated with varia-

tion in serum MUC1 levels (shown in online suppl. Fig. 1) [19]. Allele frequencies of the T and C variants are 53 and 47% in the European (non-Finnish) population of the gnomAD database [20]. The C allele encodes a longer signal peptide of 32 amino acids. The T allele encodes a shorter signal peptide lacking 9 amino acids, which alters cleavage of the *MUC1* signal sequence and the sequence of the mature N-terminus [19, 21]. In population studies, the C allele is associated with longer VNTR lengths (higher numbers of repeat units), and the T allele is associated with shorter VNTR lengths [22]. In a healthy Dutch population and in patients with lung adenocarcinoma, serum MUC1 levels were higher in individuals with the C versus T genotype [16, 23]. A possible explanation for the rs4072037 effect on serum MUC1 levels is the association of the SNP variant with the VNTR length and the possibility that more antibody binds to MUC1 proteins with longer VNTR lengths. The situation is further complicated in ADTKD-*MUC1*, as SNP genotyping will define an individual's rs4072037 genotype, but not that genotype's allelic phase with the mutant *MUC1* allele. ADTKD-*MUC1* patients heterozygous for the rs4072037 SNP (C/T) in whom the rs4072037 C allele is in phase with *wtMUC1* should exhibit higher plasma MUC1 levels than if the T allele is in phase with *wtMUC1* (shown in online suppl. Fig. 1). However, genotyping alone cannot determine which rs4072037 allele is in phase with *wtMUC1*. Family linkage can determine the allelic phase, but this analysis requires disease ascertainment and collection of samples on at least several family members and is often inconclusive. Alternatively, the phase could be determined in single cases using long read single-molecule real-time sequencing (SMRT) targeted to *MUC1* VNTR [24]. To evaluate the effects of these different factors, we compared plasma MUC1 levels in individuals with and without ADTKD-*MUC1*.

## Materials and Methods

This study was approved by the Wake Forest School of Medicine Institutional Review Board (IRB) (IRB00000352) and the IRB of all participating centers. The study was adherent to the Declaration of Helsinki.

### Population

The study population was derived from the Wake Forest Rare Inherited Kidney Disease Registry [25] and includes individuals over age 18 from families with ADTKD who provided both a plasma sample and a sample for genotyping. This is an international cohort, though the majority of the families in the cohort are from the USA. The reference population data were obtained from the

OpenLims laboratory database of the Institute of Medical Biochemistry and Laboratory Diagnostics of General University Hospital and the First Faculty of Medicine of Charles University in Prague and included unrelated individuals >18 years with serum CA15-3 levels, serum creatinine, gender, and age. The eGFR was calculated using the CKD-EPI equation [26].

### CA15-3 Determination

CA15-3 (*MUC1*) concentrations were measured by ARCHITECT CA15-3 (Abbott Laboratories, Chicago, IL, USA) on the Architect i2000 System at the Institute of Medical Biochemistry and Laboratory Diagnostics, General University Hospital, and the First Faculty of Medicine of Charles University. While the reason for testing for individual patients is not known, the assay is primarily used at this center for screening for breast cancer patients. The assay is indicated for use with either plasma or serum. Serum was used for the reference population and plasma for the ADTKD-*MUC1* and control population. The CA15-3 reference range was 0.0–35 U/mL, with a coefficient of variation of <4.9%. Reference population individuals with CA15-3 levels >35 U/mL were excluded from the analysis, as they likely suffered from metastatic cancer.

### Genetic and Bioinformatic Analysis

#### Identification of *MUC1*s Mutations

*MUC1* genotyping was performed as previously described either by mass spectrometry-based probe extension assay at the Broad Institute [2] or by *MUC1* VNTR sequencing using Illumina and single-molecule real-time sequencing methodology at the Charles University [11].

#### rs4072037 Genotyping

All individuals in the study population underwent rs4072037 genotyping by TaqMan<sup>®</sup> SNP Genotyping Assays with TaqMan<sup>®</sup> Genotyping Master Mix (Applied Biosystems). Ten-microliter amplification reactions were set in MicroAmp<sup>®</sup> Fast Optical 96-well reaction plates as follows: 5  $\mu$ L of 2 $\times$  genotyping master mix, 0.5  $\mu$ L of 20 $\times$  genotyping assay working solution, and 4.5  $\mu$ L of 3–10 ng/ $\mu$ L gDNA. Plates were processed in the StepOnePlus<sup>™</sup> Real-Time PCR System (Applied Biosystems, San Francisco, CA, USA) under standard mode thermal cycling conditions: 1 $\times$  95 $^{\circ}$ C/10 min and 40 $\times$  (95 $^{\circ}$ C/15 s and 60 $^{\circ}$ C/1 min). Data were analyzed by using StepOne<sup>™</sup> Software v2.1 (Applied Biosystems, San Francisco, CA, USA) in autocall (default), 2-cluster or single-cluster manual calling analytical mode.

### Statistical Analysis

In the reference population, associations between CA15-3 levels and age and eGFR were analyzed using Spearman's correlation. Spearman's correlation was also used to check for any association between storage time and plasma CA15-3 levels in the study population. Gender differences in CA15-3 levels were analyzed by the *t* test. Study population family structures were accounted for through use of generalized estimating equations (GEE1), assuming exchangeable correlation within a family [27]. The robust (sandwich) estimator of variance was computed to test hypotheses of association of age, gender, and genotype with CA15-3 levels. A similar GEE1 model was computed to test associations between CA15-3 levels and eGFR. We note that means  $\pm$  standard deviation and Spearman's correlation estimates (used as descriptors) for the

**Table 1.** Patient characteristics for reference and study population

	Reference population	Study population		
		ADTKD- <i>MUC1</i>	control population	
			ADTKD- <i>UMOD</i>	genetically unaffected
Individuals, <i>n</i>	6,850	85	135	114
Male, <i>n</i> (%) <sup>a</sup>	1,110 (16)	40 (47)	50 (37)	49 (43)
Age, year <sup>b</sup>	60.5±14.9	43.2±14.1	40.9±13.7	42.0±16.1
CA15-3, U/mL <sup>c</sup>	15.7±6.6	8.6±4.3 <sup>c</sup>	14.2±5.3	15.1±5.9
eGFR, mL/min/1.73 <sup>2d</sup>	80.6±27.8	46.5±29.1	48.0±28.9	96.2±21.3

The study population comprises individuals who have been genetically tested. Continuous values are represented as means ± standard deviation. The reference population has fewer males and is older compared to the study population, reflecting use of CA15-3 testing for breast cancer screening, whereas the study population reflects the Wake Forest ADTKD Registry. As described in the results, age and sex were found to have nominal association with CA15-3 levels. <sup>a</sup>The reference population had a significantly higher proportion of female individuals compared to the study population ( $p < 0.001$ ). Across the study population, sex was comparable among all groups: ADTKD-*MUC1* and ADTKD-*UMOD* ( $p = 0.24$ ), ADTKD-*MUC1* and genetically unaffected ( $p = 0.71$ ), and ADTKD-*UMOD* and genetically unaffected ( $p = 0.41$ ). <sup>b</sup>The reference population had a significantly higher mean age compared to the study data ( $p < 0.001$ ). Within the study data, age was comparable among all groups: ADTKD-*MUC1* and ADTKD-*UMOD* ( $p = 0.21$ ), ADTKD-*MUC1* and genetically unaffected ( $p = 0.73$ ), and ADTKD-*UMOD* and genetically unaffected ( $p = 0.38$ ). <sup>c</sup>The mean CA15-3 level in the ADTKD-*MUC1* group is significantly lower than in the ADTKD-*UMOD* ( $p < 0.001$ ) and genetically unaffected ( $p < 0.001$ ) groups. The mean CA15-3 levels were comparable between ADTKD-*UMOD* and genetically unaffected groups ( $p = 0.30$ ). The mean CA15-3 levels were different between the reference and control population ( $p = 0.007$ ); however, we note that the overall distributions of CA15-3 were comparable between the reference and control groups (shown in Fig. 1 and online suppl. Methods). The mean CA15-3 levels were significantly different between the reference population and ADTKD-*MUC1* individuals ( $p < 0.001$ ). <sup>d</sup>The mean eGFR is significantly higher in the genetically unaffected versus the ADTKD-*MUC1* ( $p < 0.001$ ) and ADTKD-*UMOD* ( $p < 0.001$ ) groups. The mean eGFR is not significantly different between ADTKD-*MUC1* and ADTKD-*UMOD* groups ( $p = 0.69$ ).

study data do not account for family data. Equivalence of CA15-3 distributions between the reference population and the study populations (controls and ADTKD-*MUC1*) was compared using the Kolmogorov-Smirnov's  $D_{\max}$ . To account for family structure in the study data, a resampling approach was applied (shown in online suppl. Methods).

The distribution of CA15-3 levels for each of the 3 datasets (reference, control, and ADTKD-*MUC1*) was estimated using kernel density methods [28], a nonparametric empirical estimate of distribution. For predictive modeling, training ( $n = 267$ ) and validation ( $n = 67$ ) datasets were randomly created based on an 80/20 split of the 334 individuals from the study population (shown in online suppl. Table 1). Families were randomly assigned to either training or validation to ensure independence of the 2 datasets. The training set allows one to develop diagnostic levels to differentiate affected versus unaffected. In the validation dataset, one uses the diagnostic rules developed with the training set to see if they work in another set of data. In the training dataset, for both the upper and lower tails of the empirical distribution function, we estimated the CA15-3 levels showing the maximum separation between ADTKD-*MUC1* and the control populations. These thresholds were used to classify ADTKD-*MUC1* status with high certainty. CA15-3 levels between these thresholds were deemed indeterminate. The thresholds were then applied in the validation dataset.

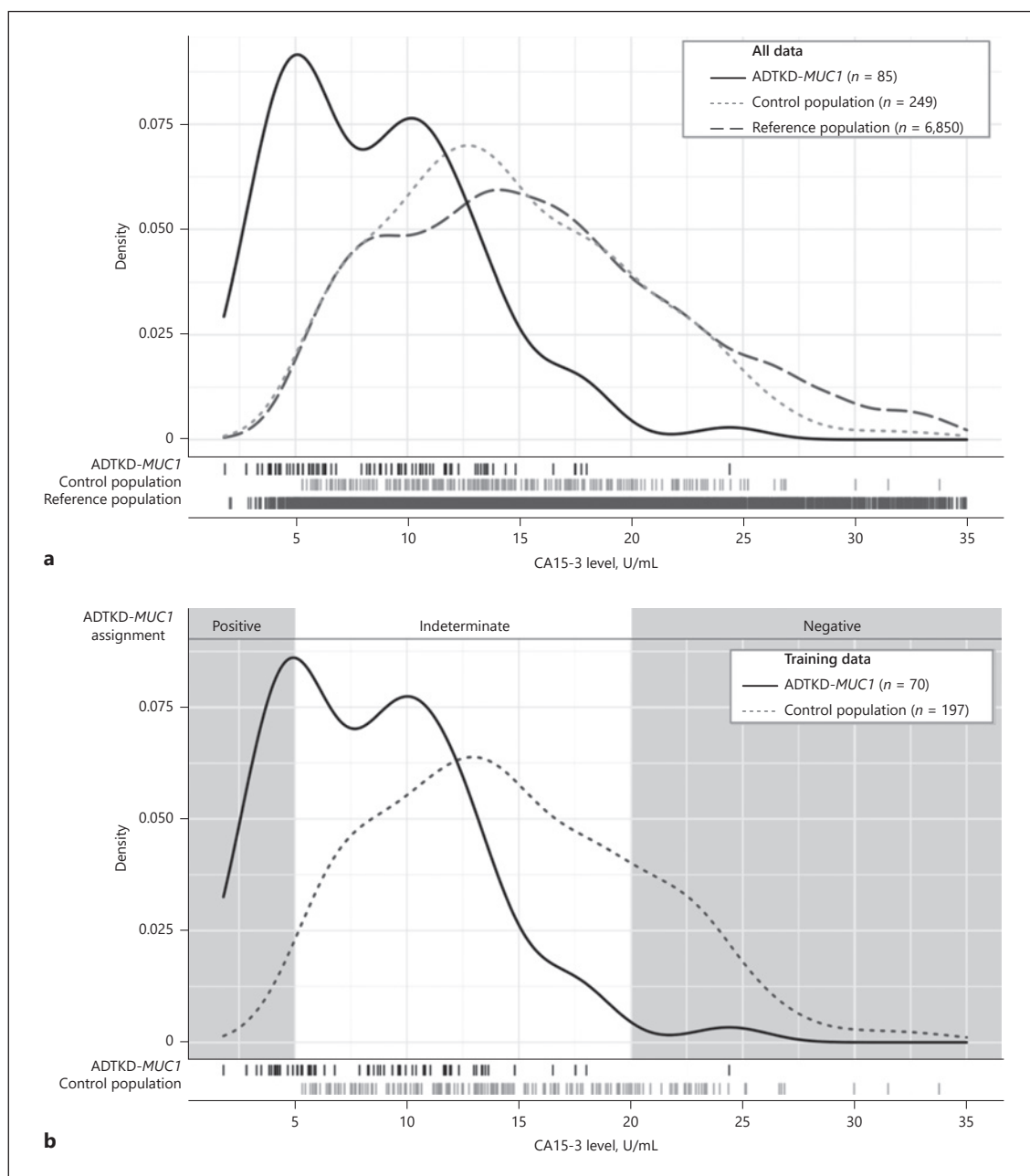
We also computed a logistic regression model in the training dataset to provide a parametric estimate of the probability of ADTKD-*MUC1* (see online suppl. material). Association testing between CA15-3 and clinical traits and the predictive logistic regression models were generated using SAS software version 9.4 (SAS Institute Inc., Cary, NC, USA).

## Results

### CA15-3 Levels in Reference and Study Populations

The reference population included 6,850 individuals (shown in Table 1) who underwent testing from January 02, 2015, to December 31, 2018. The study population included 85 individuals with genetically confirmed ADTKD-*MUC1* and 249 control individuals (135 individuals with ADTKD-*UMOD* and 114 genetically unaffected family members).

The distribution of CA15-3 levels for the reference and study cohorts is shown in Figure 1a. Mean plasma CA15-3 levels in the reference population were slightly higher



**Fig. 1.** Distribution of serum CA15-3 levels in reference and study populations. **a** Distributions of CA15-3 levels were estimated using kernel density methods for 3 populations (ADTKD-*MUC1*, a control population consisting of patients with ADTKD-*UMOD* and of unaffected family members from both ADTKD-*MUC1* and ADTKD-*UMOD* families, and a reference population). The mean CA15-3 levels were comparable in reference and control populations ( $15.7 \pm 6.65$  U/mL vs.  $14.6 \pm 5.6$  U/mL). The mean serum CA15-3 level in the ADTKD-*MUC1* group was significantly lower than in the control population ( $p < 0.001$ ). The ADTKD-*MUC1* distribution curve has 2 peaks. The lower CA15-3 level peak is from patients with the rs4072037 T/T and C/T genotype, and the higher CA15-3 level peak represents the C/C genotype. **b** Distribu-

tions of CA15-3 levels were estimated using kernel density methods for ADTKD-*MUC1* (solid line) and control individuals (dotted line) in the training dataset ( $n = 267$ ). CA15-3 levels (data points) are shown for each group at the x-axis. Using these density estimates, we calculated the CA15-3 levels that provided maximum separation in the lower and upper tails, corresponding to estimated 5 (U/mL) and 20 (U/mL), respectively. These thresholds are shown in gray and were used as classification rules for ADTKD-*MUC1* assignment:  $<5$  = positive,  $>20$  = negative, and values between 5 and 20 were assigned as indeterminate. These thresholds were applied and summarized in the training and independent validation datasets (shown in online suppl. Table 2).

**Table 2.** Predicted ADTKD-*MUC1* status in training, validation, and reference datasets based on CA15-3 thresholds

Predicted ADTKD- <i>MUC1</i> <sup>a</sup>	Training data ( <i>n</i> = 267)		Validation data ( <i>n</i> = 67)		Reference ( <i>n</i> = 6,850)
	ADTKD- <i>MUC1</i>	control population	ADTKD- <i>MUC1</i>	control population	
Positive (CA15-3 ≤ 5)	24% (17)	0 (0)	13% (2)	0 (0)	1% (95)
Negative (CA15-3 ≥ 20)	0.01% (1)	20% (39)	0 (0)	12% (6)	25% (1,726)
Indeterminate (CA15 5-20)	74% (52)	80% (158)	87% (13)	88% (46)	74% (5,029)

<sup>a</sup> Blood CA15-3 thresholds were based on the points of maximum separation between ADTKD-*MUC1* individuals and the control population in the training dataset.

in men versus women ( $16.2 \pm 6.9$  vs.  $15.6 \pm 6.6$  U/mL,  $p = 0.01$ ). Although CA15-3 levels correlated significantly ( $p < 0.001$ ) with age ( $r = 0.128$ ) and eGFR ( $r = -0.11$ ), each factor accounted for only approximately 1% of the variation among the 6,850 CA15-3 reference values (shown in online suppl. Fig. 2, 3). In addition, we found no correlation between plasma CA15-3 levels and storage time prior to sample analysis (see online suppl. Fig. 1).

In the study population, the mean CA15-3 level was  $8.7 \pm 4.3$  U/mL in the ADTKD-*MUC1* group and  $14.6 \pm 5.6$  U/mL in controls (including the healthy and ADTKD-*UMOD* group) ( $p < 0.001$ ). In the ADTKD-*MUC1* group, CA15-3 levels in men and women were similar ( $9.0 \pm 4.8$  U/mL vs.  $8.3 \pm 3.9$  U/mL,  $p = 0.05$ ). In controls, CA15-3 levels in men and women were also similar ( $15.6 \pm 6.0$  U/mL vs.  $14.0 \pm 5.3$  U/mL,  $p = 0.41$ ). Age was modestly associated with CA15-3 levels in the control group ( $p = 0.02$ ) but not in the ADTKD-*MUC1* group ( $p = 0.63$ ; shown in online suppl. Fig. 5). The eGFR was significantly associated with CA15-3 levels in ADTKD-*MUC1* individuals ( $p = 0.001$ ) but not in the ADTKD-*UMOD* group ( $p = 0.28$ ; shown in online suppl. Fig. 6).

Using a resampling approach, the Kolmogorov-Smirnov's  $D_{\max}$  statistic found the CA15-3 distribution in the reference population to be comparable with the control population. A greater deviation was observed between the reference and ADTKD-*MUC1* individuals (shown in online suppl. Fig. 7).

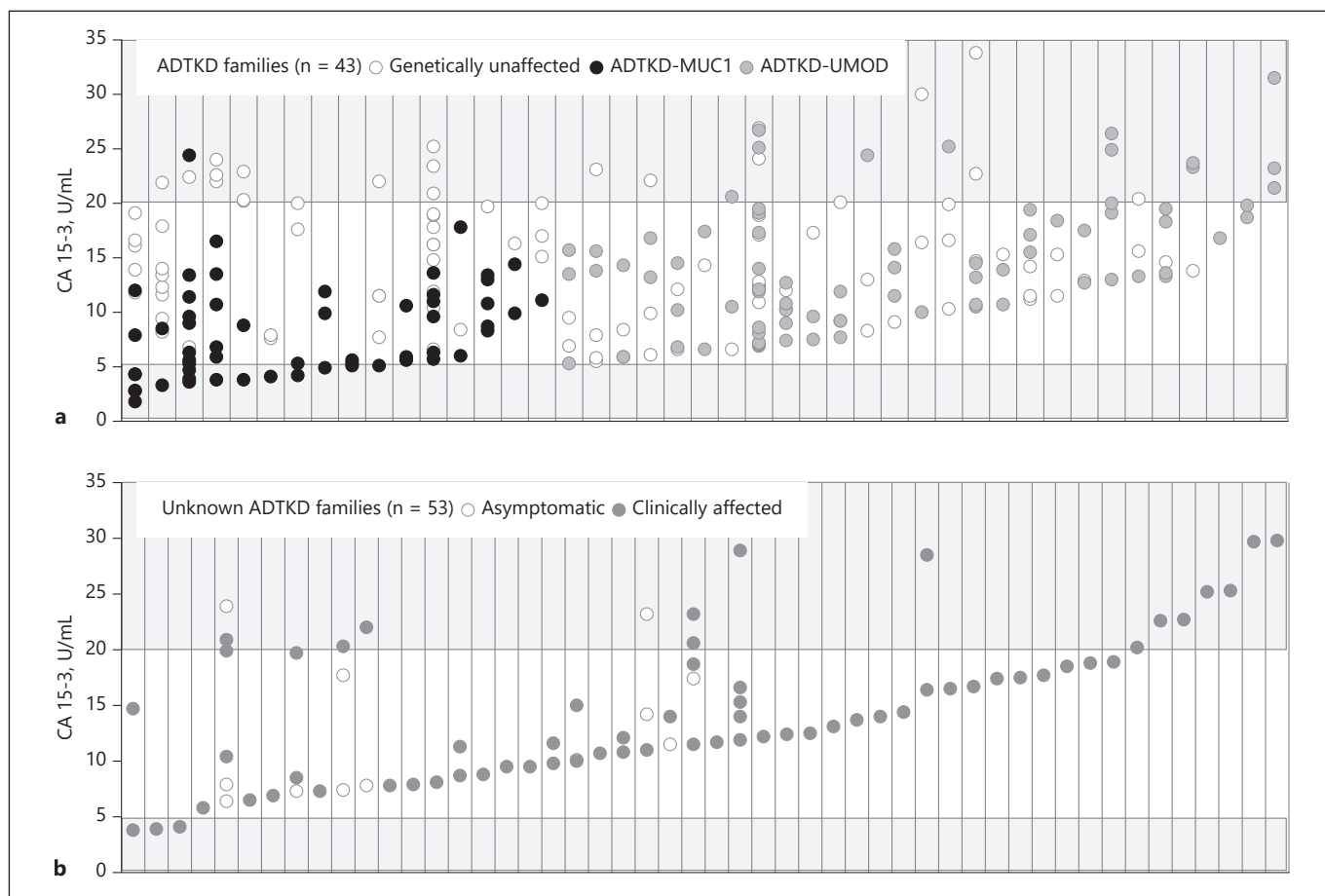
#### Distribution of CA15-3 Levels according to rs4072037 Genotype

To better understand the distribution of CA15-3 levels within the study population, 334 individuals were genotyped for rs4072037. Hardy-Weinberg equilibrium was present in both control ( $p = 0.22$ ) and ADTKD-*UMOD* ( $p = 0.33$ ) populations. Notably, variant genotype frequencies differed between ADTKD-*MUC1* individuals

and controls ( $p = 0.04$ ). The observed CC, CT, and TT allele frequencies of 42, 50, and 8% differed from the expected frequencies of 28, 50, and 22%, with the T allele frequency significantly underrepresented. The rs4072037 genotype was significantly associated with control population plasma CA15-3 levels (shown in online suppl. Fig. 8; online suppl. Table 2). The mean CA15-3 level was  $18.9 \pm 5.0$  IU/mL in the CC genotype versus  $14.9 \pm 4.9$  in the CT genotype ( $p < 0.001$ ) and  $9.9 \pm 4.1$  U/mL in the TT genotype ( $p < 0.001$ ). This pattern was replicated among ADTKD-*MUC1* individuals, with a mean CA15-3 level of  $11.5 \pm 3.9$  in the CC group versus  $6.7 \pm 3.3$  in the CT group ( $p = 0.002$ ) and  $5.8 \pm 2.2$  in the TT group ( $p = 0.02$ ). We could not determine if ADTKD-*MUC1* individuals heterozygous for rs4072037 (CT genotype) had the C or T allele in phase with the *wtMUC1* allele (see online suppl. Fig. 1). CA15-3 levels in the CT group were much closer to those of the TT group (shown in online suppl. Fig. 8), implying that most of the CT group had the C allele in phase with the mutant *MUC1* allele.

#### ADTKD-*MUC1* Prediction Using CA15-3 Thresholds

We used the empirical distribution function to estimate the points of maximum separation between the ADTKD-*MUC1* ( $n = 70$ ) and control ( $n = 197$ ) populations of the training dataset ( $n = 267$ ). CA15-3 levels of 4.3 and 20.3 U/mL were identified for the lower and upper tails of the data (shown in Fig. 1b). Based on these findings, we assigned thresholds of 5 and 20 U/mL for classifying positive and negative ADTKD-*MUC1* status, respectively (shown in Table 2). Although most data fell outside these values (meaning the values were indeterminate and a presumptive diagnosis could not be made), we note that individuals with levels within these thresholds were correctly classified at high rates of certainty (shown in Table 2). The validation dataset ( $n = 67$ ) included no misclassifications, and only 1% of individuals ( $n = 95$ ) were mis-



**Fig. 2.** Plasma CA15-3 levels in ADTKD families. The CA15-3 thresholds identified as predictive of ADTKD-*MUC1* (<5 U/mL) and not ADTKD-*MUC1* (>20 U/mL) are shaded. **a** Families included were identified as having 3 or more individuals present in the study population ( $n = 43$ ); genetically unaffected individuals ( $n = 96$ ) with

white circles, ADTKD-*MUC1* ( $n = 57$ ) with black circles, and ADTKD-*UMOD* ( $n = 78$ ) with gray circles. **b** Families included are negative for mutations in *UMOD* and *MUC1* ( $n = 53$ ); asymptomatic family members are represented by white circles, and clinically affected family members are represented by dark gray circles. ADTKD, autosomal dominant tubulointerstitial kidney disease.

classified in the reference dataset. Although plasma CA15-3 levels were associated with eGFR only in ADTKD-*MUC1* patients (shown in online suppl. Fig. 6), diagnostic low plasma CA15-3 levels were found in individuals at all eGFR values (shown in online suppl. Table 3).

To further assess the potential value of this testing, we looked at the distribution of plasma CA15-3 levels in families with a diagnosis of ADTKD. Figure 2a shows the distribution of plasma CA15-3 levels for families with ADTKD-*MUC1* and ADTKD-*UMOD*. In 8/16 ADTKD-*MUC1* families, there was an affected family member with a plasma CA15-3 level <5 U/mL, with only 1 misclassification due to a CA15-3 level >20. For ADTKD-*UMOD* families, 7/27 families had a plasma CA15-3 level >20, with no

misclassifications. We also looked at ADTKD families with negative *UMOD* and *MUC1* testing (shown in Fig. 2b), who might have a mutation in a new gene or who might have a mutation in *MUC1* that we cannot detect. Of 53 families, 3 had affected family members with CA15-3 levels <5 U/mL and 13 (26%) had levels >20 U/mL, suggesting that another gene may be responsible for ADTKD. Among 10 families with samples from >1 affected individual, a tentative diagnosis was made in 5 families.

We also performed prediction modeling including the rs4072037 SNP variant. However, use of this SNP did not improve our ability to differentiate between patients with ADTKD-*MUC1* and controls (see online suppl. Methods).

## Discussion/Conclusion

The hypothesis of this investigation was that plasma CA15-3 levels in individuals with ADTKD-*MUC1* are lower than plasma CA15-3 levels of unaffected individuals. Analysis of the study population revealed that the mean CA15-3 level for ADTKD-*MUC1* individuals was  $8.6 \pm 4.3$  U/mL, whereas that for control individuals was  $14.6 \pm 5.6$  U/mL ( $p < 0.001$ ). We found that plasma CA15-3 levels were little affected by eGFR, age, or gender in either the control or ADTKD-*MUC1* group. The lack of significant change in plasma CA15-3 levels as a function of age in patients with ADTKD-*MUC1* suggests that the epithelial cells (predominantly from the lung) that produce MUC1 detected in blood can maintain normal MUC1 synthetic ability over time despite deposition of the mutant MUC1 protein in these cells. Plasma CA15-3 levels did not change significantly in a manner related to level of kidney function, implying that the kidney is not a major source of plasma CA15-3 and does not contribute significantly to its degradation.

Significant variation in CA15-3 levels due to the rs4072037 genotype and other unaccounted factors resulted in substantial overlap of CA15-3 levels in affected and unaffected individuals. However, we found plasma CA15-3 levels  $<5$  U/mL in 22% of patients with ADTKD-*MUC1*, in 0 individuals in the control population, and in 1% of individuals in the reference population. In contrast, plasma CA15-3 levels  $>20$  U/mL were present in 1/85 ADTKD-*MUC1* patients, 18% of the control population, and 25% of the reference population. We then genotyped the *MUC1* rs4072037 variant to see if we could improve the differentiation between ADTKD-*MUC1* patients and controls. Variant genotyping did not substantially improve differentiation between affected and unaffected individuals, and the small numbers of individuals with each genotype prevented us from making definitive conclusions.

Genetic testing for ADTKD-*MUC1* remains difficult. First, the genetic test for ADTKD-*MUC1* cannot be performed by routine genetic testing in commercial laboratories, and mutations are not identified by whole genome or whole exome sequencing [2]. Clinically approved *MUC1* genetic testing is available at only 1 location in the USA [15] and available in several centers in Europe. Fortunately, the clinical test is free of charge and available to all individuals with ADTKD (contact [ableyer@wakehealth.edu](mailto:ableyer@wakehealth.edu) for testing). However, while free testing is available, international mailing restrictions limit the ability to test affected individuals. Second, the clinically approved genetic assay (which specifically identifies a cyto-

sine duplication in a heptanucleotide cytosine tract in the *MUC1* VNTR as a cause of ADTKD-*MUC1*) and other advanced research techniques [11, 12] do not identify all potential mutations that cause ADTKD-*MUC1*. Thus, there are families suspected of having ADTKD-*MUC1* on clinical, genetic (linkage), and histochemical bases in whom a mutation cannot be identified.

The results of this study suggest that CA15-3 levels might be somewhat useful if mutational analysis cannot be obtained. CA15-3 levels may be useful in families with ADTKD in whom genetic testing for *UMOD*, *MUC1*, and other mutations is negative. If there is an affected family member with a plasma CA15-3 level  $<5$  U/mL, one might first consider advanced genetic research techniques to identify other *MUC1* mutations, whereas if an affected family member has a plasma CA15-3 level that is  $>20$  U/mL, one might consider whole exome sequencing to identify new causes of ADTKD. For example, in 53 families with undiagnosed ADTKD, we found a plasma CA15-3 level  $<5$  U/mL in 3 families and  $>20$  U/mL in 13 families (shown in Fig. 2b). We have previously reported on histochemical testing of the urine with antibodies to mutant mucin-1 [11]. Urinary testing may provide a better diagnostic yield, but it is dependent on the presence of cells in the urine and the use of an antibody to mutant mucin-1 that is not commercially available. Unlike the CA15-3 level that is immediately available to the clinician, the urine test is a research test that is currently performed at only 1 academic center.

An important finding of this study was the markedly decreased prevalence of the rs4072037 T/T genotype in ADTKD-*MUC1*. Only 8% of ADTKD-*MUC1* individuals had the T/T genotype versus 24% of the control population (23% ADTKD-*UMOD* and 25% genetically unaffected). This finding suggests that the T allele, which creates the truncated signal peptide and is associated with a shorter VNTR, may protect against or attenuate development of ADTKD-*MUC1*, allowing the disease to remain undetected. A shorter mutated VNTR may also lead to less severe pathophysiologic changes by other mechanisms, as with the rs4293393 minor variant linked to missense *UMOD* mutations, for which decreased allele frequency in affected individuals likely reflects decreased mutant uromodulin expression, yielding a milder form of ADTKD-*UMOD* associated with later age of ESKD onset (Kidd K, KI Reports 2020, in press). Alternative explanations include a dominant negative effect of the T allele on the C allele or that survival to ESKD of patients with the T/T genotype could actually be shorter. We are currently exploring these hypotheses.



A limitation of this investigation is that we only used the analytic method employed at 1 center (the First Faculty of Medicine of Charles University, Prague, Czech Republic). As plasma CA15-3 assays, unlike serum creatinine values, are not standardized internationally, these results may not be generalizable. In addition, most laboratories report only the upper limit of normal, as plasma CA15-3 levels are generally used for breast cancer screening, a context in which only high levels are of diagnostic importance. There were also differences between samples in the reference population and the study groups (affected and control groups). The reference samples were recently obtained sera from patients in the Czech Republic. For the affected and control group, plasma CA15-3 determination was performed on participants mostly from the USA and on plasma that had been frozen at  $-80^{\circ}\text{C}$  for up to 4 years. For this reason, statistical comparisons were performed primarily within each study group. The lack of studies on the effect of long-term storage of plasma samples at  $-80^{\circ}\text{C}$  is a weakness of our study, though there was no correlation between sample storage time and plasma CA15-3 levels. While the distribution curve for the study and reference cohorts was similar (shown in Fig. 1a), simultaneous measurement of a large number of plasma and serum samples would be needed to provide further validation. Mean plasma CA15-3 levels in the reference population were  $16.2 \pm 6.9$  and  $15.6 \pm 6.6$  U/mL for men and women versus  $15.6 \pm 6.0$  and  $14.0 \pm 5.3$  U/mL for men and women in the control population.

In conclusion, plasma CA15-3 levels are decreased by approximately 40% in patients with ADTKD-*MUC1*. However, considerable overlap between values in affected and unaffected individuals limits utility of the plasma CA15-3 level as a diagnostic test for ADTKD-*MUC1*.

## Acknowledgement

We thank the patients who participated in this study.

## References

- 1 Fowler JC, Teixeira AS, Vinall LE, Swallow DM. Hypervariability of the membrane-associated mucin and cancer marker MUC1. *Hum Genet.* 2003;113(6):473–9.
- 2 Kirby A, Gnirke A, Jaffe DB, Barešová V, Pochet N, Blumenstiel B, et al. Mutations causing medullary cystic kidney disease type 1 lie in a large VNTR in MUC1 missed by massively parallel sequencing. *Nat Genet.* 2013; 45(3):299–303.
- 3 Patton S, Gendler SJ, Spicer AP. The epithelial mucin, MUC1, of milk, mammary gland and other tissues. *Biochim Biophys Acta.* 1995;1241(3):407–23.
- 4 Sakuma T, Takahashi K, Ohya N, Usuda K, Handa M. Serum KL-6, a novel mucin-like glycoprotein, as an indicator of interstitial pneumonitis following lobectomy. *Surg Today.* 1999;29(2):121–8.
- 5 Klee GG, Schreiber WE. MUC1 gene-derived glycoprotein assays for monitoring breast cancer (CA 15-3, CA 27.29, BR): are they measuring the same antigen? *Arch Pathol Lab Med.* 2004;128(10):1131–5.
- 6 Li X, Dai D, Chen B, Tang H, Xie X, Wei W. Clinicopathological and prognostic significance of cancer antigen 15-3 and carcinoembryonic antigen in breast cancer: a meta-analysis including 12,993 patients. *Dis Markers.* 2018;2018:9863092.

## Statement of Ethics

This study was approved by the Institutional Research Boards of Wake Forest School of Medicine, Winston-Salem, NC, USA, and the First Faculty of Medicine, Charles University, Prague, Czech Republic. All individuals provided informed consent before inclusion to the study. This study was performed in compliance with the Helsinki Declaration.

## Conflict of Interest Statement

The authors had no conflicts of interest to declare.

## Funding Sources

S.K. and colleagues were supported by the Ministry of Health of the Czech Republic (grant NV17-29786A), the Ministry of Education of the Czech Republic (grant LTAUSA19068), and by institutional programs of Charles University in Prague (UNCE/MED/007 and PROGRES-Q26/LF1). The National Center for Medical Genomics (LM2018132) kindly provided sequencing and genotyping. A.J.B. and colleagues were funded by NIH-NIDDK R21 DK106584, CKD Biomarkers Consortium Pilot and Feasibility Studies Program funded by NIH-NIDDK (U01 DK103225), the Slim Health Foundation, the Black-Brogan Foundation, and Soli Deo Gloria.

## Author Contributions

P.V., D.S., A.V., A.P., V.B., H.H., K.H., M.Z., and T.Z. performed genotyping and supervised, performed, and/or interpreted CA15-3 assays. K.K. supervised collection of samples at Wake Forest and analyzed data. H.A. and C.L. performed statistical analysis. R.H., I.B., and S.A. contributed substantially to the design of the study. N.L., S.H., and M.H. performed genotyping. V.R., A.T., L.M., and K.H. contacted patients, obtained consent, and arranged sample collection. S.K. and A.B. came up with the initial concept for the manuscript, oversaw all aspects of genotyping, measurement, and analysis, and developed the first drafts of the manuscripts. All authors contributed to manuscript revisions.

- 7 Duffy MJ. CA 15-3 and related mucins as circulating markers in breast cancer. *Ann Clin Biochem.* 1999;36(Pt 5):579–86.
- 8 Olinger E, Hofmann P, Kidd K, Dufour I, Belge H, Schaeffer C, et al. Clinical and genetic spectra of autosomal dominant tubulointerstitial kidney disease due to mutations in UMOD and MUC1. *Kidney Int.* 2020 Sep; 98(3):717–31.
- 9 Devuyt O, Olinger E, Weber S, Eckardt KU, Kmoch S, Rampoldi L, et al. Autosomal dominant tubulointerstitial kidney disease. *Nat Rev Dis Primers.* 2019;5(1):60.
- 10 Bleyer AJ, Kmoch S, Antignac C, Robins V, Kidd K, Kelsoe JR, et al. Variable clinical presentation of an MUC1 mutation causing medullary cystic kidney disease type 1. *Clin J Am Soc Nephrol.* 2014;9(3):527–35.
- 11 Zivna M, Kidd K, Pristoupilova A, Baresova V, Defelice M, Blumenstiel B, et al. Noninvasive immunohistochemical diagnosis and novel MUC1 mutations causing autosomal dominant tubulointerstitial kidney disease. *J Am Soc Nephrol.* 2018.
- 12 Yamamoto S, Kaimori JY, Yoshimura T, Namba T, Imai A, Kobayashi K, et al. Analysis of an ADTKD family with a novel frameshift mutation in MUC1 reveals characteristic features of mutant MUC1 protein. *Nephrol Dial Transplant.* 2017.
- 13 Dvela-Levitt M, Kost-Alimova M, Emani M, Kohnert E, Thompson R, Sidhom EH, et al. Small molecule targets TMED9 and promotes lysosomal degradation to reverse proteinopathy. *Cell.* 2019;178(3):521–e23.
- 14 Bleyer AJ, Kmoch S, Greka A. Diagnostic utility of exome sequencing for kidney disease. *N Engl J Med.* 2019;380(21):2080.
- 15 Blumenstiel B, Defelice M, Birsoy O, Bleyer AJ, Kmoch S, Carter TA, et al. Development and validation of a mass spectrometry-based assay for the molecular diagnosis of mucin-1 kidney disease. *J Mol Diagn.* 2016;18(4):566–71.
- 16 Janssen R, Kruit A, Grutters JC, Ruven HJ, Gerritsen WB, van den Bosch JM. The mucin-1 568 adenosine to guanine polymorphism influences serum Krebs von den Lungen-6 levels. *Am J Respir Cell Mol Biol.* 2006; 34(4):496–9.
- 17 Li X, Xu Y, Zhang L. Serum CA153 as biomarker for cancer and noncancer diseases. *Prog Mol Biol Transl Sci.* 2019;162:265–76.
- 18 Rani BS, Suchitra MM, Srinivasa Rao PVLN, Kumar VS. Serum tumor markers in advanced stages of chronic kidney diseases. *Saudi J Kidney Dis Transpl.* 2019 Jul–Aug;30(4): 898–904.
- 19 Ng W, Loh AX, Teixeira AS, Pereira SP, Swallow DM. Genetic regulation of MUC1 alternative splicing in human tissues. *Br J Cancer.* 2008;99(6):978–85.
- 20 Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature.* 2020; 581(7809):434–43.
- 21 Hanisch FG, Kinlough CL, Staubach S, Hughey RP. MUC1 membrane trafficking: protocols for assessing biosynthetic delivery, endocytosis, recycling, and release through exosomes. *Methods Mol Biol.* 2012;842:123–40.
- 22 Pratt WS, Islam I, Swallow DM. Two additional polymorphisms within the hypervariable MUC1 gene: association of alleles either side of the VNTR region. *Ann Hum Genet.* 1996;60(Pt 1):21–8.
- 23 Horimasu Y, Ishikawa N, Tanaka S, Hirano C, Iwamoto H, Ohshimo S, et al. MUC1 in lung adenocarcinoma: cross-sectional genetic and serological study. *BMC Cancer.* 2017;17(1): 263.
- 24 Wenzel A, Altmueller J, Ekici AB, Popp B, Stueber K, Thiele H, et al. Single molecule real time sequencing in ADTKD-MUC1 allows complete assembly of the VNTR and exact positioning of causative mutations. *Sci Rep.* 2018;8(1):4170.
- 25 Bleyer AJ, Kidd K, Robins V, Martin L, Taylor A, Santi A, et al. Outcomes of patient self-referral for the diagnosis of several rare inherited kidney diseases. *Genet Med.* 2019.
- 26 Stevens LA, Schmid CH, Greene T, Zhang YL, Beck GJ, Froissart M, et al. Comparative performance of the CKD epidemiology collaboration (CKD-EPI) and the modification of diet in renal disease (MDRD) study equations for estimating GFR levels above 60 mL/min/1.73 m<sup>2</sup>. *Am J Kidney Dis.* 2010;56(3): 486–95.
- 27 Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. *Biometrics.* 1988;44(4): 1049–60.
- 28 Berk R. *Statistical learning from a regression perspective.* Switzerland: Springer International Publisher; 2016.