Large-Scale Proteomics Improve Prediction of Chronic Kidney Disease in People With Diabetes

Ziliang Ye, Yuanyuan Zhang, Yanjun Zhang, Sisi Yang, Panpan He, Mengyi Liu, Chun Zhou, Xiaoqin Gan, Yu Huang, Hao Xiang, Fan Fan Hou, and Xianhui Qin

OBJECTIVE

To develop and validate a protein risk score for predicting chronic kidney disease (CKD) in patients with diabetes and compare its predictive performance with a validated clinical risk model (CKD Prediction Consortium [CKD-PC]) and CKD polygenic risk score.

RESEARCH DESIGN AND METHODS

This cohort study included 2,094 patients with diabetes who had proteomics and genetic information and no history of CKD at baseline from the UK Biobank Pharma Proteomics Project. Based on nearly 3,000 plasma proteins, a CKD protein risk score including 11 proteins was constructed in the training set (including 1,047 participants; 117 CKD events).

RESULTS

The median follow-up duration was 12.1 years. In the test set (including 1,047 participants; 112 CKD events), the CKD protein risk score was positively associated with incident CKD (per SD increment; hazard ratio 1.78; 95% CI 1.44, 2.20). Compared with the basic model (age + sex + race, C-index, 0.627; 95% CI 0.578, 0.675), the CKD protein risk score (C-index increase 0.122; 95% CI 0.071, 0.177), and the CKD-PC risk factors (C-index increase 0.175; 95% CI 0.126, 0.217) significantly improved the prediction performance of incident CKD, but the CKD polygenic risk score (C-index increase 0.007; 95% CI −0.016, 0.025) had no significant improvement. Adding the CKD protein risk score into the CKD-PC risk factors had the largest C-index of 0.825 (C-index from 0.802 to 0.825; difference 0.023; 95% CI 0.006, 0.044), and significantly improved the continuous 10-year net reclassification (0.199; 95% CI 0.059, 0.299) and 10-year integrated discrimination index (0.041; 95% CI 0.007, 0.083).

CONCLUSIONS

Adding the CKD protein risk score to a validated clinical risk model significantly improved the discrimination and reclassification of CKD risk in patients with diabetes.

Division of Nephrology, Nanfang Hospital, Southern Medical University; National Clinical Research Center for Kidney Disease; State Key Laboratory of Organ Failure Research; Guangdong Provincial Institute of Nephrology, Guangdong Provincial Key Laboratory of Renal Failure Research, Guangdong Provincial Clinical Research Center for Kidney Disease, Guangzhou Regenerative Medicine and Health Guangdong Laboratory, Guangzhou, China

Received 10 February 2024 and accepted 2 July 2024

Z.Y. and Yu.Z. contributed equally.

Chronic kidney disease (CKD) is a major and growing global health challenge, affecting ∼10% of adults worldwide (1). The substantial health care costs associated with CKD, along with the looming threat of end-stage kidney disease resulting from CKD, impose a significant financial burden on both individuals and society (2,3). Diabetes is one of the key risk factors for CKD (1,4). Hence, early identification of modifiable risk factors is essential to prevent or delay the development of CKD in patients with diabetes and has important public health implications.

Polygenic risk scores, incorporating sequence variants, show promise in improving screening and prevention efforts (5). However, the clinical applicability of polygenic risk scores in assessing CKD risk in patients with diabetes needs to be validated through prospective studies.

Proteins play a crucial role in regulating biological processes by integrating genetic effects with environmental, age-related, comorbidity-related, behavioral, and drug-related factors (6,7). The potential of proteomic profiling lies in its ability to unveil novel biomarkers that may precede the onset of CKD. As structural components within tissues, organs, and organisms, proteins mediate a wide range of biochemical activities related to metabolism, regulation, signaling, growth, and senescence (6,7). Some studies (8,9) have explored the association between several circulating proteins and the risk of CKD progression in people with diabetic kidney disease. However, there is a lack of research on the predictive value of circulating proteins on the risk of CKD in people with diabetes.

The Chronic Kidney Disease Prediction Consortium (CKD-PC) has developed an equation to predict incident CKD in people with diabetes (10). This equation incorporates a comprehensive set of variables, including age, sex, race/ethnicity, estimated glomerular filtration rate (eGFR), history of cardiovascular diseases, smoking, hypertension, BMI, albuminuria, glycated hemoglobin (HbA_{1c}), and the use of glucose-lowering drugs. We hypothesize that enhancing the validated CKD-PC model with a panel of multiple protein biomarkers and capturing various pathophysiological pathways of CKD may significantly enhance the accuracy of CKD risk prediction—a hypothesis that has not been explored to date.

In this study, we used data from the UK Biobank and its substudy, the UK Biobank Pharma Proteomics Project (UKB-PPP), a large-scale proteomic investigation that measured nearly 3,000 distinct plasma proteins. We aimed to conduct the most extensive proteomic analysis to date to develop and validate a protein risk score for predicting CKD risk in participants with diabetes and compare the predictive capabilities of the CKD protein risk scores, CKD polygenic risk scores, and clinical risk factors (CKD-PC model) of CKD in predicting CKD risk in participants with diabetes.

RESEARCH DESIGN AND METHODS

UK Biobank Sample Population

The UK Biobank, established between 2006 and 2010, stands as a substantial prospective observational study with the primary goal of investigating the effects of diverse exposures on health and diseases. It recruited ~500,000 adult participants aged 37 to 73 from 22 assessment centers across the U.K. The enrollment procedure encompassed participants completing a touch-screen questionnaire, taking physical measurements, and providing biological samples. Detailed information on the study's design and data collection procedures can be found in previous publications (11). Ethical approval was obtained from the North West Multi-centre Research Ethics Committee (Haydock, U.K.), and all participants provided signed informed consent.

The UKB-PPP involved the collaboration of 13 biopharmaceutical companies contributing funding for blood-based proteomic data generation (12). An algorithm to define prevalent diabetes was developed by Eastwood et al. (13), which used information including self-reported medical conditions (Field IDs in the UK Biobank: 4041, 6177, 6153, 2976, 22986, 20002, and 20009) and medications (Field IDs in the UK Biobank: 20003) to define prevalent diabetes and its type (type 1 diabetes or type 2 diabetes). Of the 3,362 participants with diabetes (defined as prevalent diabetes or HbA_{1c} $\geq$6.5%) in the UKB-PPP, those who lacked information on kidney disease status or had CKD at or before baseline (eGFR $<$60 mL/min/1.73 m^2, urine albumin-to-creatinine ratio [UACR] $\geq$30 mg/g, or a history of CKD), or had missing genetic data, $\geq$10 third-degree relatives, mismatched self-reported and genetic sex, and missing covariates in CKD-PC models, were further excluded. The included participants were randomly divided into a training cohort (n = 1,047) and a test cohort (n = 1,047) (Supplementary Fig. 1).

Clinical Risk Factors

Clinical risk factors incorporated into the CKD-PC model (10) included age, sex, race, eGFR, history of cardiovascular disease, smoking status, hypertension, BMI, UACR, HbA_{1c}, and the use of glucose-lowering drugs.

The collection and processing of baseline blood and urine samples was previously reported and validated (14). Serum creatinine level was measured by enzymatic analysis on a Beckman Coulter AU5800, and eGFR was calculated using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation based on serum creatinine (15). Urine albumin and urine creatinine were measured on a Beckman Coulter AU5400 by immunoturbidimetric analysis and enzymatic analysis, respectively. UACR was calculated from urinary albumin and creatinine measurements. The HbA_{1c} assay was performed using a high-performance liquid chromatography method (Bio-Rad Variant II Turbo analyzer, Bio-Rad Laboratories). Detailed information on covariates, such as age, sex, smoking, and the use of glucose-lowering drugs and angiotensin converting enzyme (ACE) inhibitors, was obtained through standardized questionnaires. BMI was calculated as weight divided by height in meters squared. Baseline history of cardiovascular disease was defined as self-reported or physician-diagnosed stroke, coronary heart disease, and heart failure. History of hypertension was defined as a systolic/diastolic blood pressure $\geq$140/90 mmHg or self-reported or physician-diagnosed hypertension. Further details about these measurements can be found in the UK Biobank online protocol (www.ukbiobank.ac.uk).

Proteomics Measurements

The details of the Olink proteomics assay, data processing, and quality control procedures have been previously described (12). In brief, blood plasma underwent proteomic profiling using the antibody-based Olink Explore 3072 proximity extension assay, measuring 2,941 protein analytes and capturing 2,923 unique proteins. The UKB laboratory team executed the randomization and plating of all samples before delivery. The processing of these samples occurred across three NovaSeq 6000 Sequencing Systems. Olink's facilities implemented rigorous quality control measures. Sample controls are used to determine precision within and between plates, and plate controls are used to standardize protein levels within a plate (12). After normalization of protein concentrations, inverse-rank normalized protein expression values were derived for each protein in each participant. These normalized protein expression values, measured on a $\log_2$ scale, represent Olink's relative protein quantification unit.

Protein Risk Score Development

The development of the protein risk score involved the use of a training set that accounted for 50% of the total participants (1,047 participants; 117 events) for training. The remaining 50% of participants (1,047 participants; 112 events) were reserved for testing the protein risk score (a test set).

To enhance the robustness of the prediction models, proteins with >20% missing were excluded (n = 12). The remaining 2,911 proteins with missing measurements were mean-imputed in the training set, and the mean values of the training set were applied to impute the missing measurement of protein in the test set. Of the remaining 2,911 proteins, we first assessed the association between a single protein and the risk of incident CKD in the training set, adjusting for age, sex, race, eGFR, UACR, HbA_{1c}, diabetes duration and types, and use of ACE inhibitor. A Cox proportional hazard model was constructed with those proteins that were statistically significant after Bonferroni correction, age, sex, and race as covariates in the training set. Model selection used a least absolute shrinkage and selection operator penalty with 10-fold cross-validation to determine the penalization strength and the number of variables to be included in the model. The coefficients for the remaining proteins were extracted from the model when the penalization strength was set to the point where the smallest partial likelihood deviance was achieved. Subsequently, the CKD protein risk score was calculated for those participants in the test set as the weighted sum of the remaining proteins with their corresponding coefficients. The sign of corresponding weight (negative for protective factors and positive for harmful factors) accounts for the effect directionality of the protein in the CKD protein risk score.

Genetic Scores of CKD

Comprehensive details about genotyping, imputation, and quality control procedures in the UK Biobank study have been previously documented (16). A polygenetic risk score for CKD was constructed based on a previous study (5), incorporating 39 single nucleotide polymorphisms (SNPs). The CKD polygenetic risk score was calculated using a weighted method:

$$\text{Polygenic risk score} = -\sum_{j=1}^{M} \beta_j \times \text{SNP}_{M,}$$

where each SNP was recoded as 0, 1, or 2 based on the number of risk alleles. The list of SNPs used and their effects is provided in Supplementary Table 1. A higher polygenetic risk score indicates a greater genetic predisposition to CKD.

Study Outcome

The study outcome was incident CKD, defined using the *International Classification of Diseases* 9th Revision codes 585 and 5859, ICD-10 codes I12.0, I13.1, I13.2, N18.0, N18.3, N18.4, N18.5, N18.8, and N18.9, and the Office of Population Censuses and Surveys Classification of Interventions and Procedures, version 4 (OPCS-4), code M01 (17,18). The follow-up period for each participant was calculated from the date of the first assessment until the first date of incident CKD diagnosis, date of death, date of loss to follow-up, or the end of follow-up, whichever came first.

In the sensitivity analysis, incident CKD was also defined using the "first occurrence fields" mapped to the three-character ICD-10 code N18 in the UK Biobank. The "first occurrence fields" delineate each health outcome using the three-character codes within ICD-10's diagnostic chapters across primary care, hospital inpatient data, and death data (https://biobank.ndph.ox.ac.uk/ukb/label.cgi?id=1712).

Statistical Analysis

A Cox proportional hazard model was used to estimate hazard ratios (HRs) and 95% CIs for the risk of incident CKD associated with CKD protein risk score (per SD increment). Using R package *pmsamp size*, given the C-index as 0.802 (10), the mean follow-up of 11.3 years in the current study, and 16 predictor parameters in the new prediction model, the minimum sample size was 522 (19). Model discrimination was evaluated using Harrell C-indices, with bootstrapping applied for estimating the CIs. Statistical testing for nested models relied on changes in model deviance based on log partial likelihood. The 10-year continuous net reclassification improvement (NRI) and integrated discrimination improvement (IDI) were used to address potential finer increments in reclassification (20,21). Bootstrapping was used for estimating CIs in these assessments. The Database for Annotation, Visualization and Integrated Discovery (DAVID) (https://david.ncifcrf.gov/) under the default parameters was used to perform functional enrichment analysis.

All statistical analyses were conducted using R software, with two-sided hypothesis tests, and P values <0.05 were considered statistically significant unless specified otherwise.

RESULTS

Of the 2,094 participants included in the study, 1,305 (62.3%) were men. In both the training and test sets, ~94% of the participants had type 2 diabetes, and 6% had type 1 diabetes. Compared with participants with type 1 diabetes, those with type 2 diabetes were older, and had lower eGFR, lower HbA_{1c}, and shorter diabetes duration (Supplementary Table 2).

Over a median follow-up of 12.1 years, 229 individuals experienced incident CKD events, including 117 in the training cohort and 112 in the test cohort. Participants in the training cohort and test cohort demonstrated similar characteristics (Table 1). Compared with participants without incident CKD events, those with incident CKD events were older and had lower eGFR, higher UACR, and a higher prevalence of cardiovascular disease (Supplementary Table 3).

Of the 2,911 proteins, 12 proteins remained statistically significant ($P < 0.05$) after Bonferroni correction (Supplementary Table 4 and Supplementary File 2). Supplementary Fig. 2 shows the Spearman rank correlation between the 12 proteins. When the penalization strength of least absolute shrinkage and selection operator regression was set to achieve the smallest partial likelihood deviance, 11 proteins were selected (Supplementary Fig. 3 and Supplementary Table 5). Therefore, the CKD protein risk score was computed as the weighted sum of those 11 proteins. Gene set enrichment analysis using DAVID (22) identified cytokine–cytokine receptor interaction and innate immune response as pathways enriched among the 11 proteins (Supplementary Fig. 4). All 11 proteins were replicated within the test set except for 1 with marginal significance (Supplementary Table 6). In the test set, after adjusting for the CKD-PC risk factors and the CKD polygenic risk score, the CKD protein risk score was significantly positively associated with the risk of incident

Table 1—Baseline characteristics of study population

	Total participants (*N* = 2,094)	Training set (*n* = 1,047)	Test set (*n* = 1,047)	*P* value
Age, years	59.4 (7.3)	59.5 (7.4)	59.4 (7.3)	0.848
Sex, *n* (%)	1,305 (62.3)	657 (62.8)	648 (61.9)	0.718
Black, *n* (%)	94 (4.5)	54 (5.2)	40 (3.8)	0.170
BMI, kg/m^2	31.0 (5.7)	31.0 (5.7)	31.0 (5.6)	0.923
Ever smoker, *n* (%)	1,006 (48.0)	489 (46.7)	517 (49.4)	0.238
eGFR, mL/min/1.73 m^2	92.2 (13.0)	92.6 (13.0)	91.7 (13.1)	0.110
UACR, mg/g	10.0 (6.2)	10.1 (6.3)	9.8 (6.2)	0.253
HbA_{1c}, mmol/mol	53.1 (13.7)	53.0 (13.2)	53.2 (14.3)	0.718
Diabetes duration, years	6.0 (2.8, 10.5)	5.5 (2.5, 10.8)	6.0 (3.0, 10.4)	0.570
Type 1 diabetes, *n* (%)	126 (6.0)	62 (5.9)	64 (6.1)	0.927
ACE inhibitor, *n* (%)	744 (35.5)	361 (34.5)	383 (36.6)	0.338
Glucose-lowering drug, *n* (%)				0.956
Insulin medication	353 (16.9)	179 (17.1)	174 (16.6)	
Oral medication	912 (43.6)	454 (43.4)	458 (43.7)	
No medication	829 (39.6)	414 (39.5)	415 (39.6)	
Prevalent disease, *n* (%)				
Cardiovascular disease	410 (19.6)	205 (19.6)	205 (19.6)	1.000
Hypertension	1,696 (81.0)	835 (79.8)	861 (82.2)	0.164
Incident CKD, *n* (%)	229 (10.9)	117 (11.2)	112 (10.7)	0.779

Data are presented as mean (SD) or median (interquartile range), unless indicated otherwise as *n* (%).

CKD (per SD increment; HR 1.78; 95% CI 1.44, 2.20) (Table 2).

In the evaluation of CKD risk prediction performance, compared with the basic model (age + sex + race, C-index 0.627; 95% CI 0.578, 0.675), the CKD protein risk score (C-index 0.748; 95% CI 0.704, 0.793; C-index increase 0.122; 95% CI 0.071, 0.177) and the CKD-PC risk factors (C-index 0.802; 95% CI 0.762, 0.841; C-index increase 0.175; 95% CI 0.126, 0.217) significantly improved the prediction performance of incident CKD, but the CKD polygenic risk score (C-index 0.634; 95% CI 0.586, 0.682; C-index increase 0.007; 95% CI −0.016, 0.025) did not show significant improvement performance (Table 3).

When the CKD polygenic risk score was added to the CKD-PC risk factors (C-index increase 0.000; 95% CI −0.005, 0.005) and the CKD protein risk score (C-index increase 0.005; 95% CI −0.005, 0.014), there was no significant increase in discrimination of CKD risk (Table 3). However, adding the CKD protein risk score into the model with the CKD-PC risk factors (C-index from 0.802 to 0.825; C-index increase 0.023; 95% CI 0.006, 0.044) or the CKD polygenic risk score (C-index from 0.634 to 0.754; C-index increase 0.120; 95% CI 0.072, 0.176) resulted in a significant increase in the C-index of CKD risk. The combination of the CKD-PC risk factors and the CKD protein risk score demonstrated the highest predictive performances for CKD risk with a C-statistic of 0.825 (95% CI, 0.789, 0.861) (Table 3).

Adding the CKD protein risk score to the basic model (age + sex + race) significantly improved the continuous 10-year NRI (80 events during the 10-year follow-up; 0.361; 95% CI 0.266, 0.483) and 10-year IDI (0.083; 95% CI 0.048, 0.130) for CKD risk. Incorporating the CKD protein risk score into the model with the CKD-PC risk factors also led to a significant improvement in the continuous 10-year NRI (0.199; 95% CI 0.059, 0.299) and 10-year IDI (0.041; 95% CI 0.007, 0.083) for CKD risk (Table 3). When the CKD risk threshold was set at 5.5% (half of the CKD incident rate in the study), adding the CKD protein risk score to the CKD-PC risk factors correctly reclassified 1 event and 25 nonevents.

When the CKD protein risk score was computed using proteins with the top five weights, adding the CKD protein risk score into the CKD-PC risk factors had a C-index of 0.817 (C-index from 0.802 to 0.817; difference 0.016; 95% CI 0.002, 0.032) and significantly improved the continuous 10-year NRI (0.189; 95% CI 0.063, 0.338) and 10-year IDI (0.030; 95% CI 0.006, 0.070) (Supplementary Table 7). Similar results were observed when incident CKD was defined by the "first occurrence fields" mapped to the three-character ICD-10 code N18 in the UK Biobank (Supplementary Table 8) or when the study population was restricted to those with type 2 diabetes (Supplementary Table 9).

CONCLUSIONS

In this study, we developed and validated a protein risk score using circulating proteins within a sizable cohort of patients with diabetes. When integrated into the previously validated CKD-PC model, the CKD protein risk score showed significant improvement in predicting CKD events in patients with diabetes.

Some studies (8,9) have explored the relationship of several circulating proteins with the risk of CKD progression in participants with diabetic kidney disease. There are still gaps in understanding the relationship between nontargeted proteins and the risk of CKD in patients with diabetes and whether these circulating

Table 2—Association between protein risk score of CKD and the risk of incident CKD in the test cohort

Exposure	Covariates	*N*/events	Adjusted HR per SD (95% CI)
Protein risk score		1,047/112	
Model 1	Age, sex, race		2.23 (1.87, 2.66)
Model 2	Age, sex, race, polygenic risk score of CKD		2.21 (1.85, 2.63)
Model 3	CKD-PC risk factors[a]		1.79 (1.45, 2.21)
Model 4	CKD-PC risk factors,[a] polygenic risk score of CKD		1.78 (1.44, 2.20)

[a]CKD-PC risk factors represent those variables included in the CKD-PC model, including age, sex, race, eGFR, history of cardiovascular disease, never smoking, hypertension, BMI, UACR, HbA_{1c}, the use of glucose-lowering drugs, and the interaction of HbA_{1c} and the use of glucose-lowering drugs.

proteins can enhance the prediction of CKD risk. Our study addresses this gap by investigating a larger number of nontargeted circulating proteins in a prospective cohort, revealing that the CKD protein risk score significantly improves the prediction of incident CKD in patients with diabetes. After adjusting for variables in the CKD prediction model of CKD-PC for patients with diabetes (10) and the CKD polygenic risk score, the CKD protein risk score was significantly positively associated with the risk of incident CKD. Of note, in this study, the prediction performance of the CKD polygenic risk score was significantly weaker than that of the CKD protein risk score in participants with diabetes. Previous studies have consistently shown that some genetic variants are significantly associated with CKD in participants without diabetes but not in those with diabetes (23). In addition, the genetic variants associated with CKD, either as individual SNPs or polygenetic risk score, could not improve the prediction of CKD given by clinical data in participants with and without diabetes (24).Therefore, due to the limitations of the polygenetic risk score in predicting diseases with multiple modifiable risk factors, enthusiasm for using the polygenetic risk score should be tempered (24,25).

The mean C-statistics of the CKD-PC prediction model for patients with diabetes was 0.801 (interquartile range 0.750–0.819) in the previously reported 5-year specific risk of incident CKD (10), and 0.802 in our current study, suggesting that the performance of the CKD-PC prediction model in patients with diabetes is robust. After adding the CKD protein risk score to the CKD-PC model for participants with diabetes,

Table 3—Performance of risk prediction models for incident CKD[a]

Model	C-index (95% CI)	C-index increase (95% CI)	10-year IDI improvement (95% CI)	10-year continuous NRI (95% CI)
Age + sex + race	0.627 (0.578, 0.675)			
Reference model: age + sex + race				
Age + sex + race + CKD polygenic risk score	0.634 (0.586, 0.682)	0.007 (−0.016, 0.025)	0.002 (0.000, 0.009)	−0.009 (−0.119, 0.118)
Age + sex + race + CKD protein risk score	0.748 (0.704, 0.793)	0.122 (0.071, 0.177)	0.083 (0.048, 0.130)	0.361 (0.266, 0.483)
CKD-PC risk factors	0.802 (0.762, 0.841)	0.175 (0.126, 0.217)	0.133 (0.086, 0.227)	0.485 (0.352, 0.587)
Reference model: age + sex + race + CKD polygenic risk score				
Age + sex + race + CKD polygenic risk score + protein risk score	0.754 (0.710, 0.798)	0.120 (0.072, 0.176)	0.081 (0.043, 0.133)	0.362 (0.246, 0.465)
Reference model: age + sex + race + CKD protein risk score				
Age + sex + race + CKD polygenic risk score + protein risk score	0.754 (0.710, 0.798)	0.005 (−0.005, 0.014)	0.000 (−0.004, 0.012)	−0.013 (−0.251, 0.157)
Reference model: CKD-PC risk factors				
Age + sex + race + CKD polygenic risk score	0.634 (0.586, 0.682)	−0.168 (−0.211, −0.121)	−0.131 (−0.225, −0.084)	−0.454 (−0.577, −0.305)
Age + sex + race + CKD protein risk score	0.748 (0.704, 0.793)	−0.053 (−0.094, −0.000)	−0.050 (−0.131, −0.008)	−0.170 (−0.353, −0.002)
CKD-PC risk factors + CKD polygenic risk score	0.802 (0.763, 0.842)	0.000 (−0.005, 0.005)	−0.000 (−0.003, 0.010)	−0.027 (−0.134, 0.150)
CKD-PC risk factors + CKD protein risk score	0.825 (0.789, 0.861)	0.023 (0.006, 0.044)	0.041 (0.007, 0.083)	0.199 (0.059, 0.299)

[a]CKD-PC risk factors represent those variables included in the CKD-PC model, including age, sex, race, eGFR, history of cardiovascular disease, never smoking, hypertension, BMI, UACR, HbA_{1c}, the use of glucose-lowering drugs, and the interaction of HbA_{1c} and the use of glucose-lowering drugs.

the discrimination and reclassification of CKD risk in patients with diabetes was significantly improved, suggesting that the CKD protein risk score has important clinical application value in predicting CKD risk in people with diabetes. Possible explanations include that proteins are more closely linked to the pathogenesis of disease than traditional risk factors and that multiprotein models accurately capture the biological effects of beneficial or harmful exposures on disease risk (26). Therefore, in patients with diabetes, the CKD protein risk score can supplement traditional risk factors and genetic information for CKD, help clinicians to better identify individuals at high risk for CKD, and promote early monitoring and prevention.

Of the 11 proteins included in the CKD protein risk score, the coefficients of growth/differentiation factor 15 (GDF15), IGF-binding protein 4 (IGFBP4), neutrophil gelatinase-associated lipocalin (NGAL), RNase K6 (RNASE6), and C-type lectin domain family 4 member D (CLEC4D) ranked in the top five. GDF15 belongs to the transforming growth factor-β cytokine superfamily and has been reported to be associated with a higher risk of renal events in patients with type 2 diabetes (27). IGFBP4 binds to both IGF-I and -II. Previous studies have found an inverse relationship between IGFBP4 and eGFR (28), and IGFBP4 was significantly higher in patients with diabetic kidney disease (29). NGAL, an acute reactive protein secreted by neutrophils and expressed in the kidney's medullary tubules (30), has been identified as a biomarker for predicting acute kidney injury (31) and is highly expressed in response to tubular injury (32). In addition, RNASE6 is predominantly found in monocytes and neutrophils, which can be triggered by bacterial infections and exhibit antibacterial activity (33). It is speculated that RNASE6 aggravates glomerular injury in diabetic nephropathy through the renal mononuclear phagocytosis system (34). As a member of the C-type lectin/C-type lectin-like domain superfamily, CLEC4D functions as a pattern recognition receptor within the innate immune framework, identifying both damage-associated and pathogen-associated molecular patterns from bacterial origins (35,36), and it is also integral in mediating cellular adhesion, intercellular communication, and the modulation of inflammatory and immune responses (37). Moreover, CLEC4D expression has been positively correlated with neutrophilic presence and implicated in the promotion of neutrophil extracellular trap formation pathways (38), a process linked to the pathogenesis of proteinuria (39).

Despite these insights, our study has some limitations. First, the observed variation in protein analyte levels across different measurement technologies (40) suggests that our findings should be validated across various panels in future studies.

Second, the UK Biobank is composed primarily of individuals with European ancestry, warranting future investigations to assess the translation of the protein risk score across diverse populations and ethnicities.

Although we have described the important clinical value of proteomics in predicting CKD risk in people with diabetes, these proteins are not easily measured in routine clinical chemistry and are relatively expensive to detect, making it relatively difficult to translate these results into clinical practice. However, with the rapid advancement of protein detection technology, the cost of detection is expected to decline rapidly, and the accessibility will rapidly improve.

Finally, using a test set for CKD prediction from the same population as the one to derive the protein risk score may lead to an overestimation of predictive performance. Further validations outside this cohort would be good to be able to use these as risk protein markers.

Conclusion

We developed and validated a protein risk score for CKD risk from large-scale proteomics as a robust and independent predictor of incident CKD in people with diabetes, which significantly improved the discrimination and reclassification of CKD risk in patients with diabetes, either in combination with age and sex or when added to a validated clinical risk model for CKD. Blood detection of biomarkers is objective, quantifiable, and convenient, whereas the collection of clinical factors is relatively cumbersome, requiring reports or medical records from participants and multiple tests such as physical examinations, blood tests, and urine tests. With the rapid development of protein detection technology, our findings highlight the important clinical value of proteomic analysis, especially combining proteomic analysis with clinical information to screen at-risk populations and better predict CKD risk, thereby promoting early monitoring and prevention in people with diabetes.

Acknowledgments. The authors especially thank all of the UK Biobank participants and all of the people involved in building the UK Biobank study. This research was conducted using the UK Biobank Resource under application number 73201.

Funding. The study was supported by National Key Research and Development Program (2022YFC2009600, 2022YFC2009605 to X.H.Q.), National Natural Science Foundation of China (81973133 to X.H.Q.), Key Technologies R&D Program of Guangdong Province (2023B1111030004 to F.F.H.), National Natural Science Foundation of China (Key Program) (82030022, 82330020 to F.F.H.), Program of Introducing Talents of Discipline to Universities, 111 Plan (D18005 to F.F.H.), and Guangdong Provincial Clinical Research Center for Kidney Disease (2020B1111170013 to F.F.H.).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Duality of Interest. No potential conflicts of interest relevant to this article were reported.

Author Contributions. Z.Y., Yu.Z., F.F.H., and X.Q. designed and conducted the research. Z.Y., Yu.Z., and X.Q. wrote the manuscript. Z.Y. and Ya.Z. performed the data management and statistical analyses. All authors read and approved the final manuscript. X.Q. is the guarantor of this work and, as such, has full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Handling Editors. The journal editors responsible for overseeing the review of the manuscript were John B. Buse and Csaba P. Kovesdy.

References

1. GBD Chronic Kidney Disease Collaboration. Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2020;395:709–733

2. Couser WG, Remuzzi G, Mendis S, Tonelli M. The contribution of chronic kidney disease to the global burden of major noncommunicable diseases. Kidney Int 2011;80:1258–1270

3. Go AS, Chertow GM, Fan D, McCulloch CE, Hsu C-y Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. N Engl J Med 2004;351:1296–1305

4. Deng Y, Li N, Wu Y, et al. Global, regional, and national burden of diabetes-related chronic kidney disease from 1990 to 2019. Front Endocrinol (Lausanne) 2021;12:672350

5. Tremblay J, Haloui M, Attaoua R, et al. Polygenic risk scores predict diabetes complications and their response to intensive blood pressure and glucose control. Diabetologia 2021;64:2012–2025

6. Ganz P, Heidecker B, Hveem K, et al. Development and validation of a protein-based risk score for cardiovascular outcomes among patients with stable coronary heart disease. JAMA 2016;315:2532–2541

7. Williams SA, Kivimaki M, Langenberg C, et al. Plasma protein patterns as comprehensive indicators of health. Nat Med 2019;25::1851–1857
8. Liu C, Debnath N, Mosoyan G, et al. Systematic review and meta-analysis of plasma and urine biomarkers for CKD outcomes. J Am Soc Nephrol 2022;33:1657–1672
9. Schlosser P, Grams ME, Rhee EP. Proteomics: progress and promise of high-throughput proteomics in chronic kidney disease. Mol Cell Proteomics 2023;22:100550
10. Nelson RG, Grams ME, Ballew SH, et al.; CKD Prognosis Consortium. Development of risk prediction equations for incident chronic kidney disease. JAMA 2019;322:2104–2114
11. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 2015;12:e1001779
12. Sun BB, Chiou J, Traylor M, et al.; Regeneron Genetics Center. Plasma proteomic associations with genetics and health in the UK Biobank. Nature 2023;622:329–338
13. Eastwood SV, Mathur R, Atkinson M, et al. Algorithms for the capture and adjudication of prevalent and incident diabetes in UK Biobank. PLoS One 2016;11:e0162388
14. Elliott P, Peakman TC, UK Biobank. The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. Int J Epidemiol 2008;37: 234–244
15. Levey AS, Stevens LA, Schmid CH, et al.; CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration). A new equation to estimate glomerular filtration rate. Ann Intern Med 2009;150:604–612
16. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 2018;562:203–209
17. He P, Ye Z, Liu M, et al. Association of handgrip strength and/or walking pace with incident chronic kidney disease: a UK Biobank observational study. J Cachexia Sarcopenia Muscle 2023;14:805–814
18. Zhang X, Liu Y-M, Lei F, et al. Association between questionnaire-based and accelerometer-based physical activity and the incidence of chronic kidney disease using data from UK Biobank: a prospective cohort study. eClinicalMedicine 2023; 66:102323
19. Riley RD, Snell KI, Ensor J, et al. Minimum sample size for developing a multivariable prediction model: part II - binary and time-to-event outcomes. Stat Med 2019;38:1276–1296
20. Pencina MJ, D'Agostino RB, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Stat Med 2011;30:11–21
21. Pencina MJ, D'Agostino RB, D'Agostino RB, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med 2008;27: 157–172
22. Jiao X, Sherman BT, Huang DW, et al. DAVID-WS: a stateful web service to facilitate gene/ protein list analysis. Bioinformatics 2012;28:1805–1806
23. Uglebjerg N, Ahmadizar F, Aly DM, et al. Four missense genetic variants in CUBN are associated with higher levels of eGFR in non-diabetes but not in diabetes mellitus or its subtypes: a genetic association study in Europeans. Front Endocrinol (Lausanne) 2023; 14:1081741
24. Cañadas-Garre M, Kunzmann AT, Anderson K, et al. Albuminuria-related genetic biomarkers: replication and predictive evaluation in individuals with and without diabetes from the UK Biobank. Int J Mol Sci 2023;24:11209
25. Sud A, Horton RH, Hingorani AD, et al. Realistic expectations are key to realising the benefits of polygenic scores. BMJ 2023;380: e073149
26. Deo R, Dubin RF, Ren Y, et al. Proteomic cardiovascular risk assessment in chronic kidney disease. Eur Heart J 2023;44:2095–2110
27. Delrue C, Speeckaert R, Delanghe JR, Speeckaert MM. Growth differentiation factor 15 (GDF-15) in kidney diseases. Adv Clin Chem 2023;114:1–46
28. Wu T, Xie C, Han J, et al. Insulin-like growth factor binding protein-4 as a marker of chronic lupus nephritis. PLoS One 2016;11:e0151491
29. Wang S, Chi K, Wu D, Hong Q. Insulin-like growth factor binding proteins in kidney disease. Front Pharmacol 2021;12:807119
30. Murtha MJ, Eichler T, Bender K, et al. Insulin receptor signaling regulates renal collecting duct and intercalated cell antibacterial defenses. J Clin Invest 2018;128:5634–5646
31. Marakala V. Neutrophil gelatinase-associated lipocalin (NGAL) in kidney injury - a systematic review. Clin Chim Acta 2022;536:135–141
32. Soni SS, Cruz D, Bobek I, et al. NGAL: a biomarker of acute kidney injury and other systemic conditions. Int Urol Nephrol 2010;42:141–150
33. Pulido D, Arranz-Trullén J, Prats-Ejarque G, et al. Insights into the antimicrobial mechanism of action of human RNase6: structural determinants for bacterial cell agglutination and membrane permeation. Int J Mol Sci 2016;17:552
34. Li T, Wang Y, Zhu C, Yang Y, Long C, Chen Q. Identification of ribonuclease 6 as an immuno-inflammatory key gene associated with the glomerular injury in diabetic nephropathy. Sci Rep 2022;12:19709
35. Zhu L-L, Zhao X-Q, Jiang C, et al. C-type lectin receptors Dectin-3 and Dectin-2 form a heterodimeric pattern-recognition receptor for host defense against fungal infection. Immunity 2013;39:324–334
36. Miyake Y, Toyonaga K, Mori D, et al. C-type lectin MCL is an FcRγ-coupled receptor that mediates the adjuvanticity of mycobacterial cord factor. Immunity 2013 May 23;38:1050–1062
37. Graham LM, Gupta V, Schafer G, et al. The C-type lectin receptor CLECSF8 (CLEC4D) is expressed by myeloid cells and triggers cellular activation through Syk kinase. J Biol Chem 2012; 287:25964–25974
38. Zhao T, Zeng J, Zhang R, et al. Serum Olink proteomics-based identification of protein biomarkers associated with the immune response in ischemic stroke. J Proteome Res 2024;23:1118–1128
39. Zhang Z, Liu M, Zhang Y, et al. Interaction of neutrophil counts and folic acid treatment on new-onset proteinuria in hypertensive patients. Br J Nutr 2021 126:1040–1047
40. Pietzner M, Wheeler E, Carrasco-Zanini J, et al. Synergistic insights into human health from aptamer- and antibody-based proteomic profiling. Nat Commun 2021;12:6822