Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Circulating Proteomic Signatures Mediate the Association Between Allostatic Load and Incident CKD: Implications for Risk Stratification and Personalized Screening

Xu Hu
First Affiliated Hospital of Nanchang University
Yue Wang
First Affiliated Hospital of Nanchang University
Hongxi Wang
Nanchang University
Weiwen Ye
First Affiliated Hospital of Nanchang University
Lizhi Zhou
First Affiliated Hospital of Nanchang University
Xiaoqiang Liu
ndyfy08564@ncu.edu.cn

First Affiliated Hospital of Nanchang University

Article

Keywords: Allostatic load, Chronic kidney disease, Proteomics, Risk prediction, Personalized screening

Posted Date: May 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9437706/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Introduction: Allostatic load (AL), a composite measure of chronic multisystem physiological stress, has been cross-sectionally linked to chronic kidney disease (CKD), but longitudinal evidence and molecular mediators remain unclear. We evaluated the association between AL and incident CKD, identified mediating circulating proteins, and developed a proteomic risk score (ProS) for risk-adapted CKD screening.

Methods: We analyzed 368,806 UK Biobank participants free of CKD at baseline. AL scores were derived from 13 multisystem biomarkers. Cox regression assessed AL–CKD associations. In a proteomic subcohort (n = 39,343), causal mediation analysis identified mediating proteins, and LASSO-Cox regression selected core proteins for ProS construction. Predictive performance was evaluated using time-dependent ROC analysis, calibration curves, and decision curve analysis in an independent validation set. Risk advancement periods and personalized screening ages were subsequently derived.

Results: Over a mean follow-up of 13.1 years, higher AL was associated with increased CKD risk (highest vs. lowest quartile: HR = 2.03, 95% CI: 1.91–2.16). Ten core proteins (e.g., FSTL3, HAVCR1, GDF15) mediated this association, jointly accounting for 67.7% of the AL effect on CKD. The ProS model achieved robust discrimination (10-year AUC = 0.834). ProS-based risk stratification advanced the recommended CKD screening initiation age to 46.0 years in the high-risk group, compared with 70.3 years in the low-risk group.

Conclusions: AL independently predicts incident CKD, operating largely through circulating proteomic alterations. The 10-protein ProS enables effective risk stratification and supports personalized, age-adapted CKD screening strategies.

1. Introduction

Chronic kidney disease (CKD) represents a significant global public health challenge. Moreover, contemporary burden estimates suggest that CKD affects hundreds of millions of people worldwide, with a persistent upward trajectory in prevalence and downstream consequences for kidney failure, cardiovascular disease, mortality, healthcare expenditure, and health-system capacity [1–4]. Because early-stage CKD is often clinically silent, many patients are first recognized only after substantial renal impairment has already occurred. Accordingly, clarifying upstream risk factors, identifying mechanistic mediators, and developing tools for earlier risk stratification remain priorities for precision prevention and timely intervention.

The concept of allostatic load (AL) was introduced by McEwen and Stellar to describe the cumulative biological wear and tear imposed by chronic stress across multiple regulatory systems [5]. Subsequent work has refined AL as a multidimensional construct spanning neuroendocrine, inflammatory, metabolic, and cardiovascular dysregulation, and has linked higher AL to adverse outcomes including cardiovascular disease, diabetes, cognitive decline, and mortality [6–9]. More recently, cross-sectional

evidence from NHANES 2015–2018 suggested that higher AL is associated with prevalent CKD, but ProRSpective evidence and mechanistic data remain limited [10].

Advances in high-throughput proteomics now provide an opportunity to interrogate circulating proteins that may sit along the pathway linking chronic physiological stress to renal dysfunction. Proximity extension assay (PEA) based platforms can quantify thousands of plasma proteins with high sensitivity from small sample volumes, making them particularly well suited to large biobank studies [12, 13]. Meanwhile, genome-wide risk profiling has improved the characterization of inherited susceptibility to reduced kidney function and clinically diagnosed CKD [14]. Together, these resources create a strong foundation for integrative epidemiologic and translational analyses.

Against this background, we used large-scale ProRSpective data from the UK Biobank to pursue four aims: (1) to evaluate the independent association and dose-response relationship between AL and incident CKD; (2) to identify circulating proteins that mediate the AL to CKD pathway; (3) to construct a composite protein risk score (ProRS) based on the selected core proteins and assess its predictive performance alongside a polygenic risk score (PRS); and (4) to explore the clinical utility of these risk scores for risk-adapted CKD screening and individualized screening initiation ages.

2. Methods

2.1 Study Design and Participants

This study was conducted within the UK Biobank, a large ProRSpective cohort that recruited approximately 500,000 adults aged 40–69 years from 22 assessment centers across the United Kingdom between 2006 and 2010 [11]. At baseline, participants completed questionnaires, physical examinations, and biospecimen collection and consented to linkage with longitudinal health records. This study was conducted under UK Biobank Application Number: 532564, and all participants provided written informed consent in accordance with the Declaration of Helsinki. We excluded participants with a history of CKD at baseline, baseline estimated glomerular filtration rate (eGFR) < 60 mL/min/1.73 m^2, or missing variables required to compute AL. After exclusions, 368,806 participants were included in the full analytic cohort (Fig. 1). For proteomic analyses and model development, we identified 39,343 participants with complete plasma proteomic data and randomly allocated them in a 7:3 ratio to a training set (n = 27,541) and validation set (n = 11,802).

Figure 1. Flowchart of study population selection from the UK Biobank cohort.

2.2 Plasma Proteomic Measurements

UK Biobank plasma proteomic profiling was performed using the Olink Explore 3072 platform, which is based on PEA technology and enables simultaneous quantification of nearly 3,000 human plasma proteins [12]. Protein concentrations were reported as normalized protein expression (NPX) values on a log2 scale after batch correction and quality control. In this study, proteins with > 20% missingness were

excluded, remaining missing values were imputed using k-nearest neighbors (KNN), and all retained NPX values were standardized to a mean of 0 and an SD of 1 to facilitate model fitting and comparison of effect sizes [22]. After quality control, 2,911 proteins were included in downstream mediation analyses.

2.3 Exposure Assessment

The allostatic load (AL) score was constructed using a multidomain framework adapted from prior AL literature [5, 7, 8]. Biomarkers were drawn from five domains: metabolic function (HbA1c, triglycerides, LDL cholesterol, HDL cholesterol), cardiovascular function (systolic blood pressure, diastolic blood pressure, pulse rate), inflammatory status (C-reactive protein and insulin-like growth factor-1), liver function (alanine aminotransferase, aspartate aminotransferase, gamma-glutamyl transferase), and kidney-related physiology (serum creatinine). Each biomarker was categorized into age- and sex-specific quartiles, with the lowest quartile treated as high risk for HDL cholesterol and IGF-1 and the highest quartile treated as high risk for all other markers. Domain-specific scores were calculated as the average of constituent marker scores, and the final AL score was obtained by averaging across domains and rescaling to 0-100, with higher values indicating greater cumulative physiological dysregulation.

2.4 Outcome Assessment

The primary endpoint was incident CKD. Participants with baseline CKD or baseline eGFR < 60 mL/min/1.73 m2 were excluded before follow-up. Incident CKD during follow-up was ascertained from linked inpatient hospital records using ICD codes and from self-reported physician diagnoses; the specific coding definitions are summarized in Supporting Table 9. Follow-up time was defined from baseline assessment to the earliest occurrence of incident CKD, death, loss to follow-up, or the last available health-record update (October 31, 2022).

2.5 Covariates

Potential confounders were selected on the basis of prior CKD literature and clinical relevance [3, 4]. These included age, sex, ethnicity, Townsend Deprivation Index, educational attainment, smoking status, alcohol consumption, physical activity, body mass index (BMI), baseline eGFR, and history of cardiovascular disease. Baseline eGFR was estimated using the CKD-EPI equation [15]. Covariates with > 20% missingness were excluded, and remaining missing values were imputed using multiple imputation by chained equations (MICE) [21]. For baseline descriptions, continuous variables are presented as mean +/- SD and categorical variables as counts (percentages).

2.6 Polygenic Risk Score Construction

To minimize sample overlap and overfitting, we used external summary statistics that excluded UK Biobank participants. The CKD PRS was derived from the CKDGen Consortium meta-analysis reported by Wuttke et al. [14]. Independent genome-wide significant loci ($P < 5 \times 10^{-8}$) were selected after linkage disequilibrium clumping ($R2 < 0.001$; 10,000-kb window), and participant-level PRS values were calculated by weighting effect alleles according to the published summary statistics. Variants with minor allele frequency < 0.01 or genotype missingness > 5% were excluded from the target dataset.

2.7 Statistical Analysis

2.7.1 Association Analysis Between AL and CKD Risk

We used Cox proportional hazards regression to evaluate the association between AL and incident CKD [16]. Model 1 adjusted for age, sex, and ethnicity; Model 2 additionally adjusted for Townsend Deprivation Index, education, BMI, smoking, and alcohol intake; and Model 3 further adjusted for baseline eGFR, cardiovascular disease history, and CKD PRS. Associations were summarized as hazard ratios (HRs) and 95% confidence intervals (CIs).

To examine dose-response patterns, AL was analyzed both in quartiles and as a standardized continuous variable. Trend tests were performed by entering quartile groups as ordinal terms. Restricted cubic spline models with four knots were used to explore nonlinearity. Kaplan-Meier methods were used to visualize cumulative CKD incidence across AL groups, and between-group differences were assessed with log-rank tests [27].

2.7.2 Subgroup and Sensitivity Analyses

We conducted subgroup analyses stratified by age, sex, ethnicity, smoking status, alcohol use, and educational attainment. Effect modification was evaluated by including interaction terms between AL and the stratification variable in the fully adjusted model. Sensitivity analyses included complete-case analysis, exclusion of CKD events occurring within 2 years of baseline, and competing-risk analysis using the Fine-Gray subdistribution hazards model to account for death as a competing event [17].

2.7.3 Proteomic Mediation Analysis

Mediation analyses were performed in the proteomic training set. For each of the 2,911 candidate proteins, we applied a causal mediation framework to estimate total, direct, and indirect effects of AL on incident CKD [25]. Linear regression was used for the exposure-to-mediator path and Cox regression for the mediator-to-outcome path, adjusting for age, sex, Townsend Deprivation Index, BMI, and baseline eGFR. Indirect effects and proportions mediated were derived using bootstrap resampling (500 iterations). False discovery rate (FDR) correction was applied across all proteins, and FDR < 0.05 with a positive mediation proportion was considered statistically significant [20].

2.7.4 Functional Enrichment Analysis

Gene Ontology and Kyoto Encyclopedia of Genes and Genomes pathway enrichment analyses were conducted for significant mediator proteins using clusterProfiler, with hypergeometric testing and FDR correction [26]. Biological Process, Cellular Component, and Molecular Function categories were all evaluated.

2.7.5 Core Protein Selection and Composite Score Construction

Proteins with significant mediation effects were entered into a least absolute shrinkage and selection operator (LASSO)-penalized Cox model for feature reduction [23, 24]. Tuning was performed by 10-fold cross-validation, and the most parsimonious model within one standard error of the minimum deviance (lambda.1se) was selected. Proteins with non-zero coefficients were retained as core proteins. The composite protein risk score (ProRS) was defined as the weighted sum of standardized NPX values using the corresponding LASSO-Cox coefficients. ProRS was standardized in the training set and then transformed in the validation set using the training-set mean and SD.

2.7.6 Plasma Protein Dynamic Trajectory Modeling and Hierarchical Clustering

To characterize temporal patterns before diagnosis, we reverse-aligned participants who subsequently developed CKD by the interval between baseline blood sampling and clinical diagnosis, setting the diagnosis date as year 0. Population-level trajectories of standardized protein concentrations were reconstructed over the 15 years preceding diagnosis. Quantitative trajectory features, including baseline level, end level, extrema, and overall magnitude of change, were extracted and used for unsupervised clustering with hierarchical clustering and K-means methods.

2.7.7 Prediction Model Construction and Validation

Three Cox-based prediction models were constructed in the training set: a PRS model, a ProRS model, and a combined model incorporating both PRS and ProRS. Discrimination was evaluated using time-dependent receiver operating characteristic curves at 5, 10, and 15 years [18]. Calibration was assessed with calibration plots, and clinical utility was examined by decision curve analysis [19]. All performance metrics were evaluated in the independent validation set.

2.7.8 Risk Stratification and Individualized Screening Age Determination

Participants were categorized into low-, intermediate-, and high-risk groups according to tertiles of each score in the training set. Kaplan-Meier methods were used to compare cumulative CKD incidence across risk groups [27]. Risk advancement periods (RAPs) were then calculated to quantify how many years earlier a higher-risk group reached the same cumulative incidence observed in the low-risk group [28]. To derive risk-adapted screening ages, we estimated age-specific 10-year cumulative CKD risks and identified the age at which each risk group crossed a prespecified 10-year risk threshold of 10%.

2.7.9 Statistical Software

All analyses were performed using R version 4.4.0 on the RAP online platform. Major packages included survival, rms, cmprsk, mediation, glmnet, timeROC, clusterProfiler, and ggplot2.

3. Results

3.1 Baseline Characteristics

This study included 368,806 participants free of CKD at baseline. Over a mean follow-up of 13.1 years, 15,446 incident CKD cases (4.19%) were identified (Table 1). Compared with non-cases, participants who developed CKD were older, more likely to be male, more likely to have obesity, and had a higher Townsend Deprivation Index. Educational attainment, smoking status, and alcohol consumption also differed significantly between groups (all $P < 0.001$).

Table 1
Baseline Characteristics of Study Participants According to Incident CKD Status

Characteristics	**Overall (N = 368,806)**	**Non-CKD (n = 353,306)**	**Incident CKD (n = 15,446)**	***P* value**
Age, years, mean (SD)	56.42 (8.08)	56.18 (8.07)	61.95 (6.08)	< 0.001
Sex, n (%)				0.001
Female	198,427 (53.8)	191,016 (54.1)	7,411 (48.0)	
Male	170,379 (46.2)	162,344 (45.9)	8,035 (52.0)	
Ethnicity, n (%)				0.227
White	349,528 (94.8)	334,889 (94.8)	14,639 (94.8)	
Non-White	19,278 (5.2)	18,471 (5.2)	807 (5.2)	
TDI, median [IQR]	−1.34 (3.05)	−1.36 (3.04)	−0.95 (3.25)	< 0.001
Education level, n (%)				< 0.001
College or university degree	119,316 (32.4)	116,250 (32.9)	3,066 (19.8)	
Other levels	183,780 (49.8)	176,533 (50.0)	7,247 (46.9)	
Unknown	65,710 (17.8)	60,577 (17.1)	5,133 (33.2)	
BMI category, n (%)				< 0.001
Underweight (< 18.5 kg/m²)	1,913 (0.5)	1,881 (0.5)	32 (0.2)	
Normal (18.5–24.9 kg/m²)	121,038 (32.8)	118,194 (33.4)	2,844 (18.4)	
Overweight (25.0–29.9 kg/m²)	157,320 (42.7)	150,865 (42.7)	6,455 (41.8)	
Obese (≥ 30.0 kg/m²)	88,535 (24.0)	82,420 (23.3)	6,115 (39.6)	
Smoking status, n (%)				< 0.001
Never	201,783 (54.7)	194,809 (55.1)	6,974 (45.2)	
Previous	127,971 (34.7)	121,360 (34.3)	6,611 (42.8)	
Current	39,052 (10.6)	37,191 (10.5)	1,861 (12.1)	
Alcohol consumption, n (%)				< 0.001
Never	16,018 (4.3)	14,998 (4.2)	1,020 (6.6)	
Previous	13,006 (3.5)	12,091 (3.4)	915 (5.9)	

Abbreviations: CKD, chronic kidney disease; SD, standard deviation; IQR, interquartile range; TDI, Townsend Deprivation Index; BMI, body mass index; eGFR, estimated glomerular filtration rate; CVD, cardiovascular disease; PRS, polygenic risk score.

Characteristics	Overall (N = 368,806)	Non-CKD (n = 353,306)	Incident CKD (n = 15,446)	*P* value
Current	339,782 (92.1)	326,271 (92.3)	13,511 (87.5)	
eGFR, mL/min/1.73 m², mean (SD)	91.56 (11.92)	92.07 (11.64)	79.95 (12.21)	< 0.001
History of CVD, n (%)				< 0.001
No	349,460 (94.8)	336,691 (95.3)	12,769 (82.7)	
Yes	19,346 (5.2)	16,669 (4.7)	2,677 (17.3)	
PRS for CKD, mean (SD)	24.75 (17.06)	24.28 (16.87)	35.39 (17.85)	< 0.001
Genetic risk category, n (%)				< 0.001
Low risk	92,202 (25.0)	90,780 (25.7)	1,422 (9.2)	
Intermediate risk	92,202 (25.0)	89,695 (25.4)	2,507 (16.2)	
High risk	92,201 (25.0)	88,089 (24.9)	4,112 (26.6)	

Abbreviations: CKD, chronic kidney disease; SD, standard deviation; IQR, interquartile range; TDI, Townsend Deprivation Index; BMI, body mass index; eGFR, estimated glomerular filtration rate; CVD, cardiovascular disease; PRS, polygenic risk score.

[Insert Table 1 **here]**

For proteomic analyses and prediction modeling, 39,343 participants with complete proteomic data were randomly split into a training set (n = 27,541) and a validation set (n = 11,802). As shown in Supporting Table 1, the two subsets were well balanced with respect to age, sex, and most baseline covariates.

3.2 Association Between AL and CKD

AL was independently associated with incident CKD. In the fully adjusted model, compared with the lowest AL quartile (Q1), the HRs for Q2, Q3, and Q4 were 1.33 (95% CI: 1.25–1.42), 1.55 (95% CI: 1.46–1.65), and 2.03 (95% CI: 1.91–2.16), respectively, with P for trend < 0.001 (Table 2). Per 1-SD increase in AL, CKD risk increased by 27% (HR = 1.27, 95% CI: 1.25–1.30; P < 0.001). Restricted cubic spline analyses suggested a monotonic increase in risk with higher AL values, with a steeper gradient at the upper end of the distribution (Supporting Fig. 1). Kaplan-Meier curves also showed progressively wider separation in cumulative CKD incidence across AL quartiles during follow-up (Supporting Fig. 2).

Table 2
Multivariable Cox Proportional Hazards Regression Analysis for the Association Between Allostatic Load and Incident CKD

Exposure	Model 1			Model 2			Model 3		
	HR	**95% CI**	***P***	**HR**	**95% CI**	***P***	**HR**	**95% CI**	***P***
AL, continuous (per SD increase)	1.81	1.79–1.84	< 0.001	1.67	1.64–1.69	< 0.001	1.27	1.25–1.30	< 0.001
AL quartile									
Q1	Reference			Reference			Reference		
Q2	1.64	1.53–1.75	< 0.001	1.51	1.42–1.61	< 0.001	1.33	1.25–1.42	< 0.001
Q3	2.75	2.59–2.92	< 0.001	2.39	2.25–2.54	< 0.001	1.55	1.46–1.65	< 0.001
Q4	5.37	5.07–5.68	< 0.001	4.19	3.95–4.44	< 0.001	2.03	1.91–2.16	< 0.001
P for trend	< 0.001			< 0.001			< 0.001		

Model 1: Adjusted for age, sex, and ethnicity.

Model 2: Adjusted for Model 1 covariates plus TDI, education level, BMI category, smoking status, and alcohol consumption.

Model 3: Adjusted for Model 2 covariates plus eGFR, history of CVD, and PRS for CKD.

Abbreviations: AL, allostatic load; CKD, chronic kidney disease; HR, hazard ratio; CI, confidence interval; SD, standard deviation; TDI, Townsend Deprivation Index; BMI, body mass index; eGFR, estimated glomerular filtration rate; CVD, cardiovascular disease; PRS, polygenic risk score.

3.3 Subgroup and Sensitivity Analyses

The positive association between AL and incident CKD was largely consistent across subgroups (Supporting Fig. 3). Interaction testing indicated significant effect modification by age and sex. Results were materially unchanged after complete-case analysis, after excluding CKD events within the first 2 years of follow-up, and after accounting for death as a competing risk in Fine-Gray models (Supporting Tables 2–4).

3.4 Mediation Analysis of Circulating Proteins and Trajectory Analysis

Among the 2,911 proteins screened, 747 showed significant mediation of the AL-CKD association (mediation proportion > 0 and FDR < 0.05; Fig. 2a and Supporting Table 5).

Figure 2. Mediation effects of individual plasma proteins and functional enrichment analysis in the allostatic load–CKD pathway.(a)Volcano plot of mediation effects for 2,911 plasma proteins. The x-axis shows the proportion mediated; the y-axis shows -log10(FDR). Red, significant positive mediators; blue, significant negative mediators; gray, non-significant. Dashed line, FDR = 0.05. (b) Top enriched Gene Ontology (BP, green; CC, orange; MF, blue) and KEGG (red) terms among significant mediator proteins. The x-axis shows -log10(P value); circle size indicates gene count.

Functional enrichment analyses indicated that these mediator proteins were primarily involved in extracellular matrix organization, inflammatory regulation, cytokine signaling, receptor-ligand activity, and growth factor binding (Fig. 2b), supporting the possibility that chronic physiological stress contributes to renal decline through broad inflammatory and microenvironmental remodeling pathways.

LASSO-Cox feature reduction identified 10 core proteins: EDA2R, FSTL3, GDF15, HAVCR1, IGFBP4, RNASE1, SCARB2, SHISA5, WFDC2, and YAP1 (Fig. 3a-b). In the fully adjusted model, ProRS was strongly associated with incident CKD (HR per 1-SD increment = 2.01, 95% CI: 1.91–2.11; P < 0.001; Supporting Table 8). The 10-protein composite mediated 67.7% of the total AL effect on CKD (95% CI: 56.2%-89.0%; P < 0.001; Supporting Fig. 4).

Trajectory analyses suggested two broad patterns before diagnosis. One cluster, including FSTL3, EDA2R, SCARB2, RNASE1, YAP1, and HAVCR1, showed relatively stable but persistently abnormal levels; another cluster, including WFDC2, IGFBP4, GDF15, and SHISA5, showed sharper late-stage increases approaching diagnosis (Fig. 3c-d and Supporting Tables 6–7).

- **Figure 3**. LASSO-Cox selection of core mediator proteins and pre-diagnostic temporal trajectories. (a)Cross-validation curve for the LASSO-Cox model. Partial likelihood deviance versus − log(λ); error bars, standard errors from 10-fold cross-validation. Left dashed line, λ_min; right dashed line, λ_1SE. Numbers above indicate non-zero coefficients retained. (b) Coefficient paths of candidate proteins as a function of log(λ). Dashed line marks the selected λ_1SE. (c) Hierarchical clustering heatmap of standardized protein expression (Z-scores) for the 10 core proteins over years before CKD diagnosis. Rows, proteins; columns, time points. Three clusters are color-coded: Cluster 1 (red), Cluster 2 (green), Cluster 3 (blue). (d) Temporal trends of protein Z-scores before diagnosis. Upper panel, overall trend across all core proteins; lower panels, cluster-specific trends with shaded 95% CI.

3.5 CKD Prediction Model Based on AL-Mediated Proteins

We developed three prediction models: PRS only, ProRS only, and the combined PRS + ProRS model.

The protein-based model showed robust discrimination across follow-up horizons, with AUCs of 0.843, 0.834, and 0.834 at 5, 10, and 15 years, respectively (Fig. 4a). In contrast, the PRS model showed limited discrimination, with AUCs close to 0.50. The combined model performed similarly to the ProRS model.

Calibration curves showed good agreement between predicted and observed risks for both the ProRS and combined models (Fig. 4b). Decision curve analysis at 10 years indicated meaningful clinical net benefit across a wide range of threshold probabilities for the ProRS and combined models, and both outperformed PRS alone (Fig. 4c).

Figure 4. Predictive performance of PRS, ProRS, and combined models for incident CKD at 5-, 10-, and 15-year horizons.(a) Time-dependent ROC curves for PRS-only (gray), Protein-only (orange), and Combined (dark orange) models. AUC values are shown in the legend; dashed diagonal, AUC = 0.5. (b) Calibration plots of predicted versus observed CKD risk. Points represent deciles of predicted risk; dashed diagonal indicates perfect calibration. (c) Decision curve analysis showing net benefit across threshold probabilities. Black line, "treat all"; gray line, "treat none."

3.6 CKD Risk Stratification and Individualized Screening Strategy

Kaplan-Meier curves demonstrated significant separation of low-, intermediate-, and high-risk groups for all three scoring systems, with stronger discrimination for ProRS and the combined score than for PRS alone (Fig. 5a-c). Relative to the low-risk group, the high-PRS group reached equivalent CKD risk 2.4 years earlier, whereas the high-ProRS and high-combined-score groups reached equivalent risk 42.6 and 40.3 years earlier, respectively (Table 3).

Using age-specific 10-year risk curves, the estimated screening initiation ages differed modestly across PRS groups but substantially across ProRS and combined-score groups. In the ProRS model, recommended screening ages were 70.3, 62.7, and 46.0 years for low-, intermediate-, and high-risk groups, respectively. The combined model yielded a similar pattern, with screening beginning at 46.0 years in the high-risk group and 70.3 years in the low-risk group (Fig. 5d-f and Table 3).

Figure 5. CKD risk stratification and age-specific 10-year risk by PRS, ProRS, and combined scores.(a-c) Kaplan-Meier cumulative incidence curves of CKD stratified into low, intermediate, and high-risk groups by (a) PRS, (b) ProRS, and (c) ComRS. Log-rank P values are shown. (d-f) Age-specific 10-year cumulative CKD risk by attained age, stratified by (d) PRS, (e) ProRS, and (f) ComRS. Dashed horizontal line indicates the 10% risk threshold for recommended screening initiation.

Table 3
Risk-Advancement Period and Risk-Adapted Starting Age for CKD Screening According to Risk Score Strata

Risk Score Category	Cases (n)	Controls (n)	Incidence (%)	RAP (95% CI), Years	Screening Starting Age (Years)
Polygenic Risk Score (PRS)					
Low risk	511	12,552	3.91	Reference	62.7
Intermediate risk	590	12,472	4.52	1.5 (0.5–2.6)	58.6
High risk	650	12,411	4.98	2.4 (1.4–3.5)	58.5
Proteomic Risk Score (ProRS)					
Low risk	90	12,972	0.69	Reference	70.3
Intermediate risk	307	12,755	2.35	16.8 (11.2–22.4)	62.7
High risk	1,354	11,708	10.37	42.6 (33.5–51.6)	46
Combined Risk Score (ComRS)					
Low risk	96	12,966	0.73	Reference	70.3
Intermediate risk	302	12,760	2.31	14.9 (9.6–20.1)	62.4
High risk	1,353	11,709	10.36	40.3 (31.8–48.7)	46

Abbreviations:CI, confidence interval; CKD, chronic kidney disease; ComRS, combined risk score; PRS, polygenic risk score; ProRS, proteomic risk score; RAP, risk-advancement period

4. Discussion

In this large ProRSpective cohort, higher allostatic load was associated with substantially greater risk of incident CKD, extending earlier cross-sectional observations with longitudinal evidence from a much larger population base [10, 11]. The result is biologically plausible because AL aggregates the cumulative burden of metabolic dysfunction, elevated blood pressure, inflammatory activation, and other chronic stress-related physiological disturbances that are themselves central contributors to kidney injury and progression [3–9]. The persistence of the association across multiple sensitivity analyses further supports the robustness of the finding.

The observed effect modification by age and sex is also noteworthy and broadly consistent with prior AL research, including the recent CKD analysis based on NHANES[10]. From a clinical standpoint, these interactions suggest that AL may be particularly informative for identifying subgroups in whom earlier surveillance of renal function could be warranted.

Proteomic mediation analyses identified 747 significant mediators and distilled these signals into a 10-protein core panel. Among these proteins, FSTL3 showed the largest individual mediation proportion. FSTL3 is an activin-binding glycoprotein that rises with worsening renal function and has also been shown experimentally to attenuate mesangial cell proliferation and fibronectin expression under high-glucose conditions [34, 35]. These observations make FSTL3 a plausible marker of stress-related fibrotic remodeling in the kidney.

HAVCR1 (KIM-1) is one of the most established markers of tubular injury. It is minimally expressed in healthy kidneys, is markedly induced in injured proximal tubular epithelial cells, and can be detected in blood or urine after injury [29]. More recent cohort evidence further suggests that circulating KIM-1 is associated with tubulointerstitial lesions and progression to kidney failure in CKD [30]. In the present study, the strong mediating role of HAVCR1 is consistent with a pathway in which chronic multisystem stress contributes to persistent subclinical tubular injury before overt CKD becomes clinically apparent.

GDF15 also emerged as a major mediator. As a stress-responsive cytokine, GDF15 rises in response to inflammation, oxidative stress, mitochondrial dysfunction, and tissue injury. Prior studies have linked higher circulating GDF15 to CKD progression, incident CKD, diabetic kidney disease, and adverse cardiovascular outcomes [31–33]. In the Framingham Heart Study, GDF15 also mediated associations between diet quality and downstream cardiometabolic outcomes, underscoring its role as a broader integrative signal of chronic physiological stress [39].

Several of the remaining proteins in the 10-protein panel are also biologically credible. WFDC2/HE4 has been implicated in renal fibrosis and is elevated in patients with CKD and biopsy-evaluable fibrotic burden [36, 37]. YAP1 has recently been linked to mitochondrial quality control in diabetic kidney disease [38]. EDA2R, meanwhile, has emerged as a cross-tissue aging and parainflammatory signal, which may align with the chronic stress and inflammaging dimensions captured by AL [40]. For proteins such as IGFBP4, RNASE1, SCARB2, and SHISA5, the current findings are hypothesis-generating and warrant further mechanistic work.

Collectively, the 10 selected proteins mediated more than two-thirds of the AL-CKD association, suggesting that chronic stress-related renal injury is not conveyed through a single pathway but through a coordinated network of inflammatory activation, extracellular matrix remodeling, tubular injury, and altered cellular homeostasis.

From a prediction standpoint, the ProRS model showed strong and stable discrimination over 5, 10, and 15 years. The PRS-only model, by contrast, showed limited clinical discrimination despite being derived from a large and well-powered kidney-function GWAS [14]. This pattern is consistent with the idea that

proteins sit closer to current pathophysiological state than inherited susceptibility alone and may therefore capture more clinically actionable short-to-medium-term risk information.

The combined model performed similarly to the protein-only model, suggesting that the incremental predictive contribution of the current PRS was modest once the circulating protein panel was known. This does not negate the biological relevance of genetic susceptibility, but it does indicate that the selected proteins may already summarize much of the downstream risk information that matters for near-term CKD prediction in this setting.

The screening-age analyses provide a clinically intuitive translation of these findings. A high-risk participant defined by ProRS or the combined score reached the same CKD risk profile decades earlier than a low-risk participant, supporting the idea that biomarker-based stratification could be used to tailor when screening begins [4]. In practical terms, a risk-adapted strategy could intensify early surveillance in individuals with high proteomic risk while deferring screening in those with consistently low risk.

This study has several strengths. It leverages the scale and phenotypic depth of the UK Biobank, integrates epidemiologic, proteomic, and genetic information, applies internal validation for risk modeling, and shows consistent findings across multiple sensitivity analyses [11].

Several limitations should also be acknowledged. First, UK Biobank participants are predominantly of European ancestry, which may limit generalizability to other populations [11]. Second, AL and proteomic markers were measured at baseline only, precluding direct assessment of within-person temporal change. Third, mediation analysis supports but does not prove causality, and experimental or genetic triangulation would strengthen mechanistic inference. Finally, although the selected proteins are biologically plausible and analytically measurable, translation into routine clinical testing will require external validation, assay harmonization, and cost-effectiveness evaluation.

5. Conclusions

In summary, chronic multisystem physiological stress, quantified here as allostatic load, was significantly associated with incident CKD. Circulating proteomic signatures mediated a substantial share of this relationship, and a 10-protein risk score showed strong performance for CKD risk stratification and individualized screening-age estimation. These findings support the potential value of risk-adapted CKD screening strategies while highlighting the need for external validation and implementation-focused research.

Abbreviations

AL
Allostatic load
ALT
Alanine aminotransferase

AST
Aspartate aminotransferase
AUC
Area under the curve
BMI
Body mass index
BP
Biological process
CC
Cellular component
CI
Confidence interval
CKD
Chronic kidney disease
CKD-EPI
Chronic Kidney Disease Epidemiology Collaboration
ComRS
Combined risk score
CRP
C-reactive protein
CVD
Cardiovascular disease
EDA2R
Ectodysplasin A2 receptor
eGFR
Estimated glomerular filtration rate
FDR
False discovery rate
FSTL3
Follistatin-like 3
GDF15
Growth differentiation factor 15
GGT
Gamma-glutamyl transferase
GO
Gene Ontology
GWAS
Genome-wide association study
HbA1c
Glycated hemoglobin A1c

HAVCR1
Hepatitis A virus cellular receptor 1 (KIM-1)
HDL
High-density lipoprotein
HR
Hazard ratio
ICD
International Classification of Diseases
IGF-1
Insulin-like growth factor 1
IGFBP4
Insulin-like growth factor binding protein 4
IQR
Interquartile range
KDIGO
Kidney Disease: Improving Global Outcomes
KEGG
Kyoto Encyclopedia of Genes and Genomes
KIM-1
Kidney injury molecule 1
KNN
K-nearest neighbors
LASSO
Least absolute shrinkage and selection operator
LDL
Low-density lipoprotein
MF
Molecular function
MICE
Multiple imputation by chained equations
NHANES
National Health and Nutrition Examination Survey
NPX
Normalized protein expression
PEA
Proximity extension assay
ProRS
Protein risk score (composite)
PRS
Polygenic risk score

RAP
Risk advancement period
RCS
Restricted cubic spline
RNASE1
Ribonuclease A family member 1
ROC
Receiver operating characteristic
SCARB2
Scavenger receptor class B member 2
SD
Standard deviation
SHISA5
Shisa family member 5
TDI
Townsend Deprivation Index
WFDC2
WAP four-disulfide core domain protein 2 (HE4)
YAP1
Yes-associated protein 1

Declarations

Acknowledgments

The data used in this study were obtained from the UK Biobank(Application Number: 532564). The UK Biobank data were accessedand analyzed in accordance with the terms of the UK Biobank's Accessand Use Agreement.

Author contributions

XH , WY and YW participated in study design and drafting the manuscript. HX and LZ contributed to the collection and analysis of data. XL and HW contributed to the discussion and revision of the manuscript. All authors read and approved the submitted version

Ethics Approval Statement

The UK Biobank study obtained ethical approval from the North West Multi-Centre Research Ethics Committee (REC reference: 21/NW/0157)

Funding

National Natural Science Foundation of China (No. 82400848) and the Medical Interdisciplinary Innovation Fund of Nanchang University (No. NCUJCCX-2024-03)

Data availability

Data used in this project are available from the UK Biobank by submitting a data request proposal (www.ukbiobank.ac.uk) under Application Number: 532564.

Conflict of interest

The authors declare that they have no conflict of interest

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT in order to assist with language refinement and rephrasing to improve readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

1. Bikbov, B. et al. Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* **395** (10225), 709–733 (2020).
2. Jager, K. J. et al. A single number for advocacy and communication-worldwide more than 850 million individuals have kidney diseases. *Kidney Int.* **96** (5), 1048–1050 (2019).
3. Webster, A. C., Nagler, E. V., Morton, R. L. & Masson, P. Chronic kidney disease. *Lancet* **389** (10075), 1238–1252 (2017).
4. Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO 2024 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. *Kidney Int.* 10.1016/j.kint.2023.10.018 (2024).
5. McEwen, B. S. & Stellar, E. Stress and the individual: mechanisms leading to disease. *Arch. Intern. Med.* **153** (18), 2093–2101 (1993).
6. McEwen, B. S. Protective and damaging effects of stress mediators. *N Engl. J. Med.* **338** (3), 171–179 (1998).
7. McEwen, B. S. Allostasis and allostatic load: implications for neuropsychopharmacology. *Neuropsychopharmacology* **22** (2), 108–124 (2000).
8. Juster, R. P., McEwen, B. S. & Lupien, S. J. Allostatic load biomarkers of chronic stress and impact on health and cognition. *Neurosci. Biobehav Rev.* **35** (1), 2–16 (2010).
9. Guidi, J., Lucente, M., Sonino, N. & Fava, G. A. Allostatic load and its impact on health: a systematic review. *Psychother. Psychosom.* **90** (1), 11–27 (2021).
10. Li, Q., Zhao, K., Wu, L. & Zhang, H. Association between allostatic load and chronic kidney disease: findings from the NHANES 2015–2018 study. *Int. Urol. Nephrol.* 10.1007/s11255-026-05014-6 (2026).

11. Sudlow, C. 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.* **12** (3), e1001779 (2015).
12. Assarsson, E. et al. Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLoS One*. **9** (4), e95192 (2014).
13. Sun, B. B. et al. Genomic atlas of the human plasma proteome. *Nature* **558** (7708), 73–79 (2018).
14. Wuttke, M. et al. A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat. Genet.* **51** (6), 957–972 (2019).
15. Levey, A. S. et al. A new equation to estimate glomerular filtration rate. *Ann. Intern. Med.* **150** (9), 604–612 (2009).
16. Cox, D. R. Regression models and life-tables. *J. R Stat. Soc. Ser. B Stat. Methodol.* **34** (2), 187–220 (1972).
17. Fine, J. P. & Gray, R. J. A proportional hazards model for the subdistribution of a competing risk. *J. Am. Stat. Assoc.* **94** (446), 496–509 (1999).
18. Blanche, P., Dartigues, J. F. & Jacqmin-Gadda, H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat. Med.* **32** (30), 5381–5397 (2013).
19. Vickers, A. J. & Elkin, E. B. Decision curve analysis: a novel method for evaluating prediction models. *Med. Decis. Mak.* **26** (6), 565–574 (2006).
20. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R Stat. Soc. Ser. B Stat. Methodol.* **57** (1), 289–300 (1995).
21. van Buuren, S. & Groothuis-Oudshoorn, K. mice: multivariate imputation by chained equations in R. *J. Stat. Softw.* **45** (3), 1–67 (2011).
22. Troyanskaya, O. et al. Missing value estimation methods for DNA microarrays. *Bioinformatics* **17** (6), 520–525 (2001).
23. Tibshirani, R. The Lasso method for variable selection in the Cox model. *Stat. Med.* **16** (4), 385–395 (1997).
24. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33** (1), 1–22 (2010).
25. Tingley, D., Yamamoto, T., Hirose, K., Keele, L. & Imai, K. mediation: R package for causal mediation analysis. *J. Stat. Softw.* **59** (5), 1–38 (2014).
26. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16** (5), 284–287 (2012).
27. Kaplan, E. L. & Meier, P. Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* **53** (282), 457–481 (1958).
28. Brenner, H., Gefeller, O. & Greenland, S. Risk and rate advancement periods as measures of exposure impact on the occurrence of chronic diseases. *Am. J. Epidemiol.* **137** (7), 760–770 (1993).

29. Han, W. K., Bailly, V., Abichandani, R., Thadhani, R. & Bonventre, J. V. Kidney injury molecule-1 (KIM-1): a novel biomarker for human renal proximal tubule injury. *Kidney Int.* **62** (1), 237–244 (2002).
30. Chronic Kidney Disease Biomarkers Consortium and the CRIC Study Investigators. Plasma Kidney Injury Molecule 1 in CKD: Findings From the Boston Kidney Biopsy Cohort and CRIC Studies. *Am. J. Kidney Dis.* **79** (2), 231–243e1 (2022).
31. Nair, V. et al. Growth differentiation factor-15 and risk of CKD progression. *J. Am. Soc. Nephrol.* **28** (7), 2233–2240 (2017).
32. Bao, X. et al. Growth differentiation factor-15 and incident chronic kidney disease: a population-based cohort study. *BMC Nephrol.* **22** (1), 351 (2021).
33. Carlsson, A. C. et al. Growth differentiation factor 15 (GDF-15) is a potential biomarker of both diabetic kidney disease and future cardiovascular events in cohorts of individuals with type 2 diabetes: a proteomics approach. *Ups J. Med. Sci.* **125** (1), 37–43 (2020).
34. Ebert, T. et al. FSTL3 is increased in renal dysfunction. *Nephrol. Dial Transpl.* **32** (10), 1637–1644 (2017).
35. Wang, X., Shi, L., Han, Z. & Liu, B. Follistatin-like 3 suppresses cell proliferation and fibronectin expression via p38MAPK pathway in rat mesangial cells cultured under high glucose. *Int. J. Clin. Exp. Med.* **8** (9), 15214–15221 (2015).
36. LeBleu, V. S. et al. Identification of human epididymis protein-4 as a fibroblast-derived mediator of fibrosis. *Nat. Med.* **19** (2), 227–231 (2013).
37. Wan, J. et al. Elevated serum concentrations of HE4 as a novel biomarker of disease severity and renal fibrosis in kidney disease. *Oncotarget* **7** (42), 67748–67759 (2016).
38. Ye, S. et al. YAP1 preserves tubular mitochondrial quality control to mitigate diabetic kidney disease. *Redox Biol.* **78**, 103435 (2024).
39. Kim, Y., Vasan, R. S. & Ma, J. Proteins as mediators of the association between diet quality and incident cardiovascular disease and all-cause mortality: the Framingham Heart Study. *J. Am. Heart Assoc.* **10** (17), e021245 (2021).
40. Barbera, M. C. et al. Increased ectodysplasin-A2-receptor EDA2R is a ubiquitous hallmark of aging and mediates parainflammatory responses. *Nat. Commun.* 10.1038/s41467-025-56918-3 (2025).

Figures

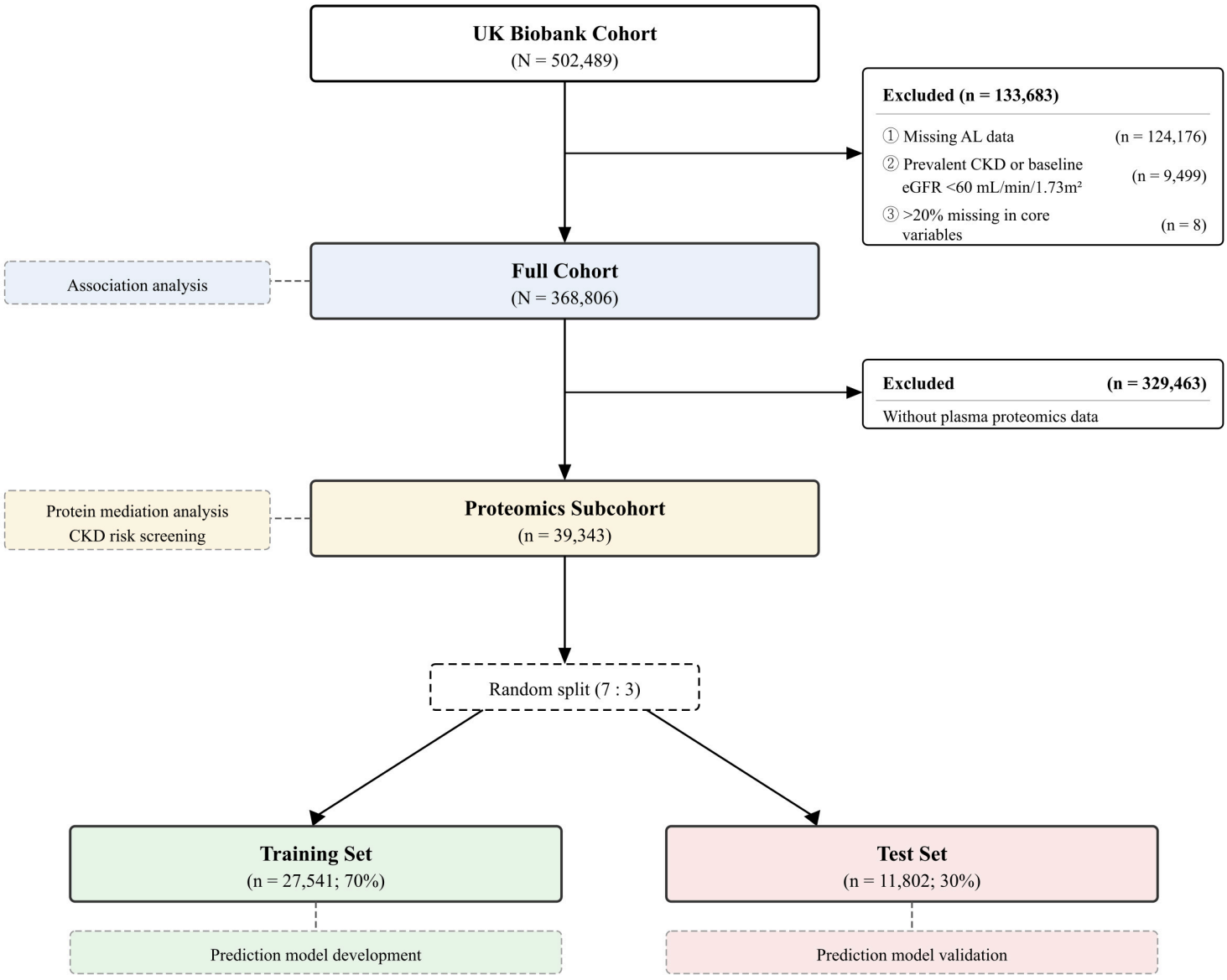


Figure 1

Flowchart of study population selection from the UK Biobank cohort.

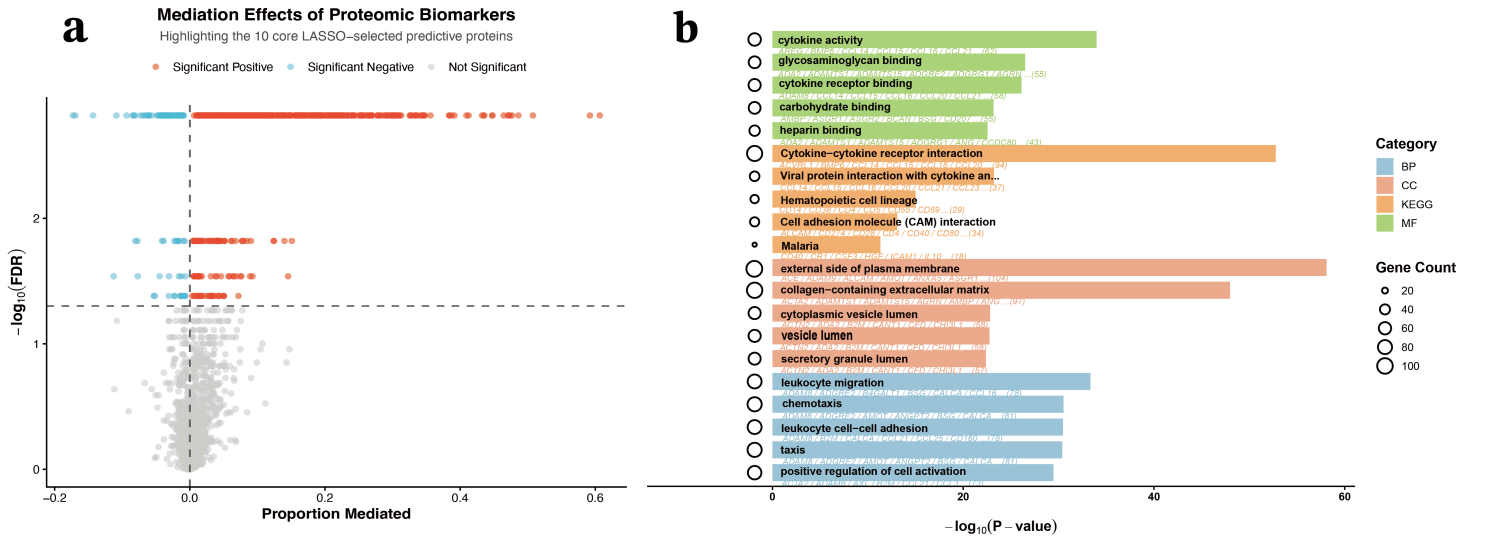


Figure 2

Mediation effects of individual plasma proteins and functional enrichment analysis in the allostatic load–CKD pathway.(a)Volcano plot of mediation effects for 2,911 plasma proteins. The x-axis shows the proportion mediated; the y-axis shows -log10(FDR). Red, significant positive mediators; blue, significant negative mediators; gray, non-significant. Dashed line, FDR = 0.05. (b) Top enriched Gene Ontology (BP, green; CC, orange; MF, blue) and KEGG (red) terms among significant mediator proteins. The x-axis shows -log10(P value); circle size indicates gene count.

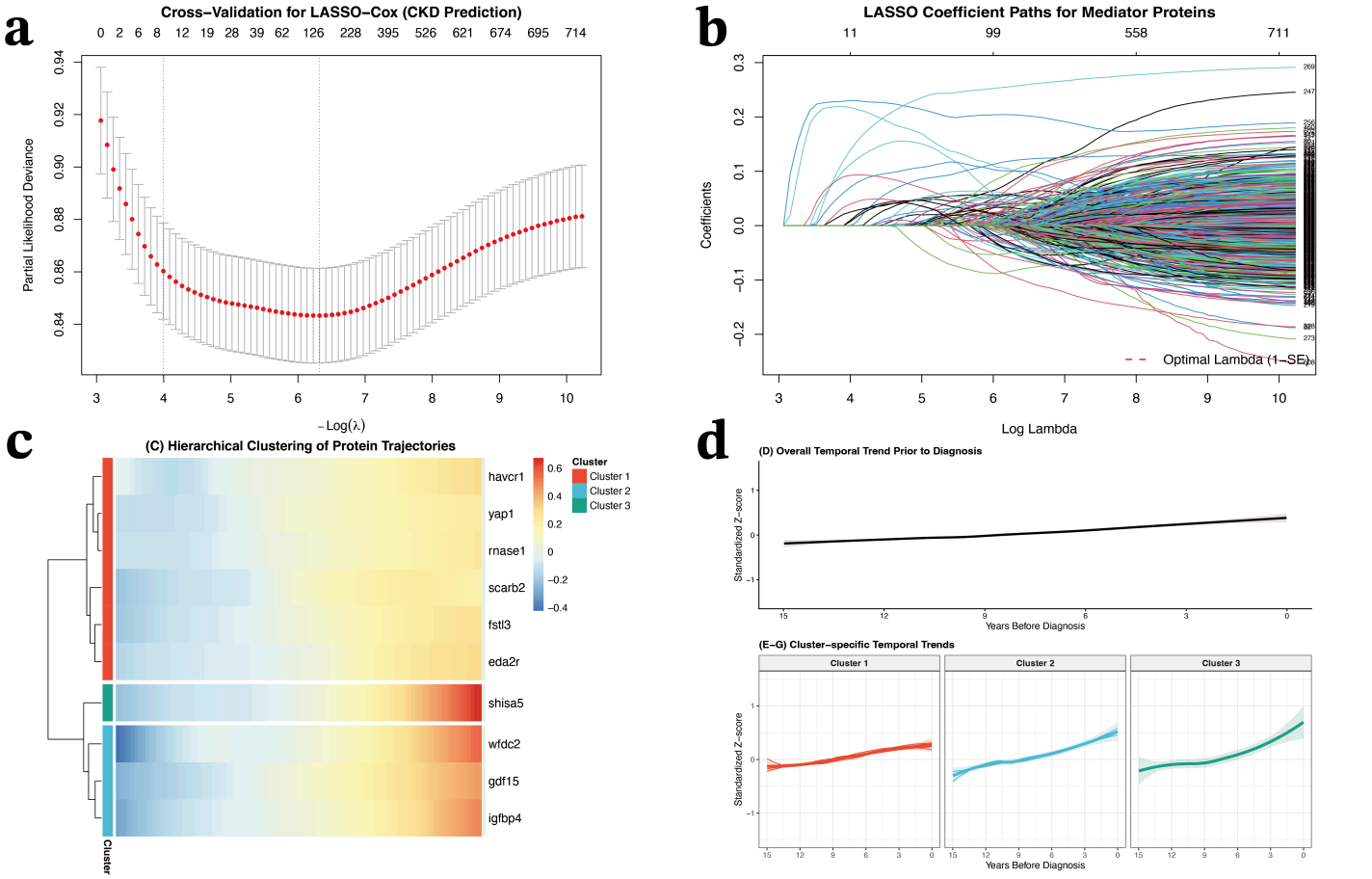


Figure 3

LASSO-Cox selection of core mediator proteins and pre-diagnostic temporal trajectories.(a)Cross-validation curve for the LASSO-Cox model. Partial likelihood deviance versus −log(λ); error bars, standard errors from 10-fold cross-validation. Left dashed line, λ_min; right dashed line, λ_1SE. Numbers above indicate non-zero coefficients retained. (b) Coefficient paths of candidate proteins as a function of log(λ). Dashed line marks the selected λ_1SE. (c) Hierarchical clustering heatmap of standardized protein expression (Z-scores) for the 10 core proteins over years before CKD diagnosis. Rows, proteins; columns, time points. Three clusters are color-coded: Cluster 1 (red), Cluster 2 (green), Cluster 3 (blue). (d) Temporal trends of protein Z-scores before diagnosis. Upper panel, overall trend across all core proteins; lower panels, cluster-specific trends with shaded 95% CI.

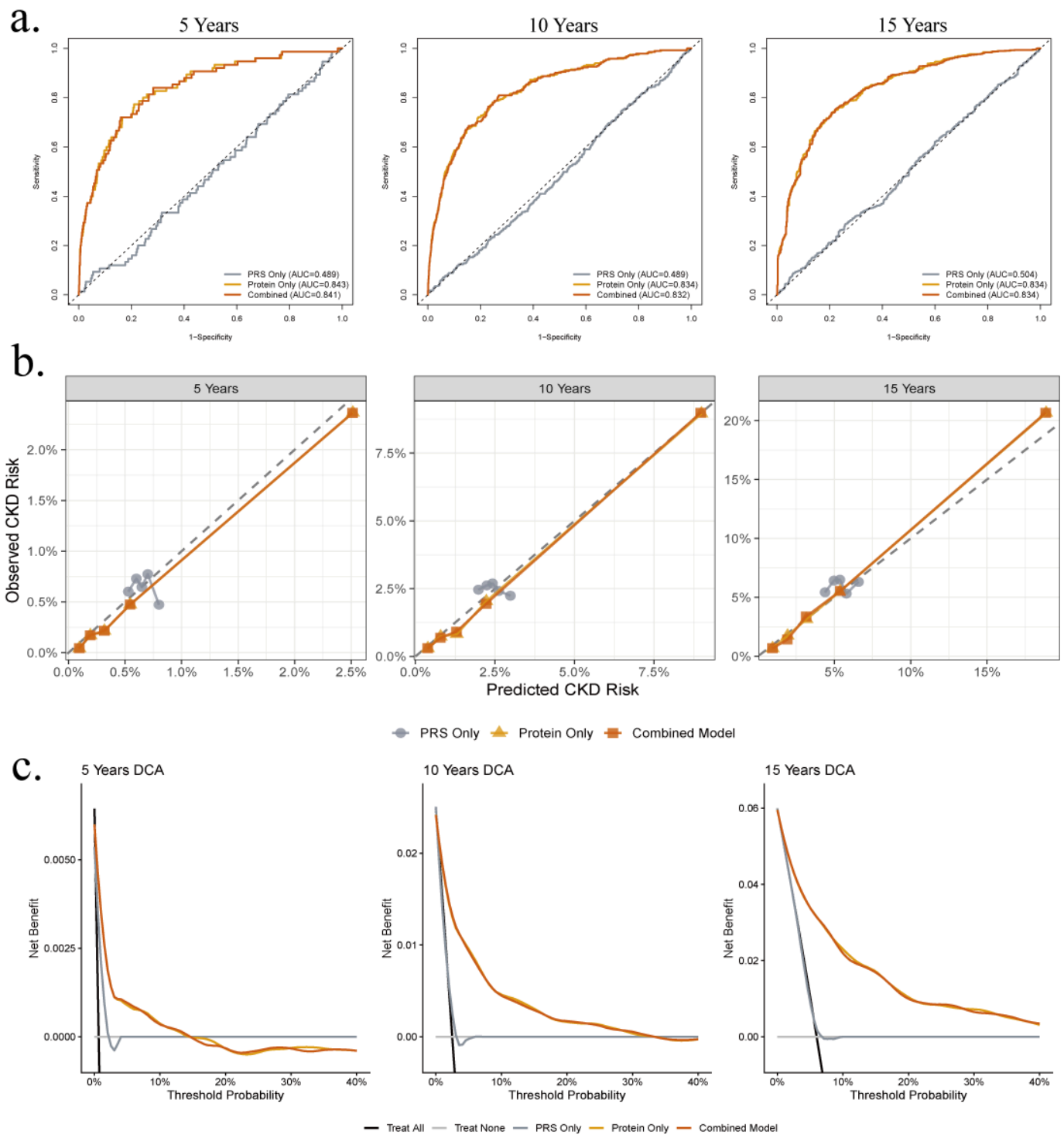


Figure 4

LASSO-Cox selection of core mediator proteins and pre-diagnostic temporal trajectories.(a)Cross-validation curve for the LASSO-Cox model. Partial likelihood deviance versus −log(λ); error bars, standard errors from 10-fold cross-validation. Left dashed line, λ_min; right dashed line, λ_1SE. Numbers above indicate non-zero coefficients retained. (b) Coefficient paths of candidate proteins as a function of log(λ). Dashed line marks the selected λ_1SE. (c) Hierarchical clustering heatmap of standardized

protein expression (Z-scores) for the 10 core proteins over years before CKD diagnosis. Rows, proteins; columns, time points. Three clusters are color-coded: Cluster 1 (red), Cluster 2 (green), Cluster 3 (blue). (d) Temporal trends of protein Z-scores before diagnosis. Upper panel, overall trend across all core proteins; lower panels, cluster-specific trends with shaded 95% CI.

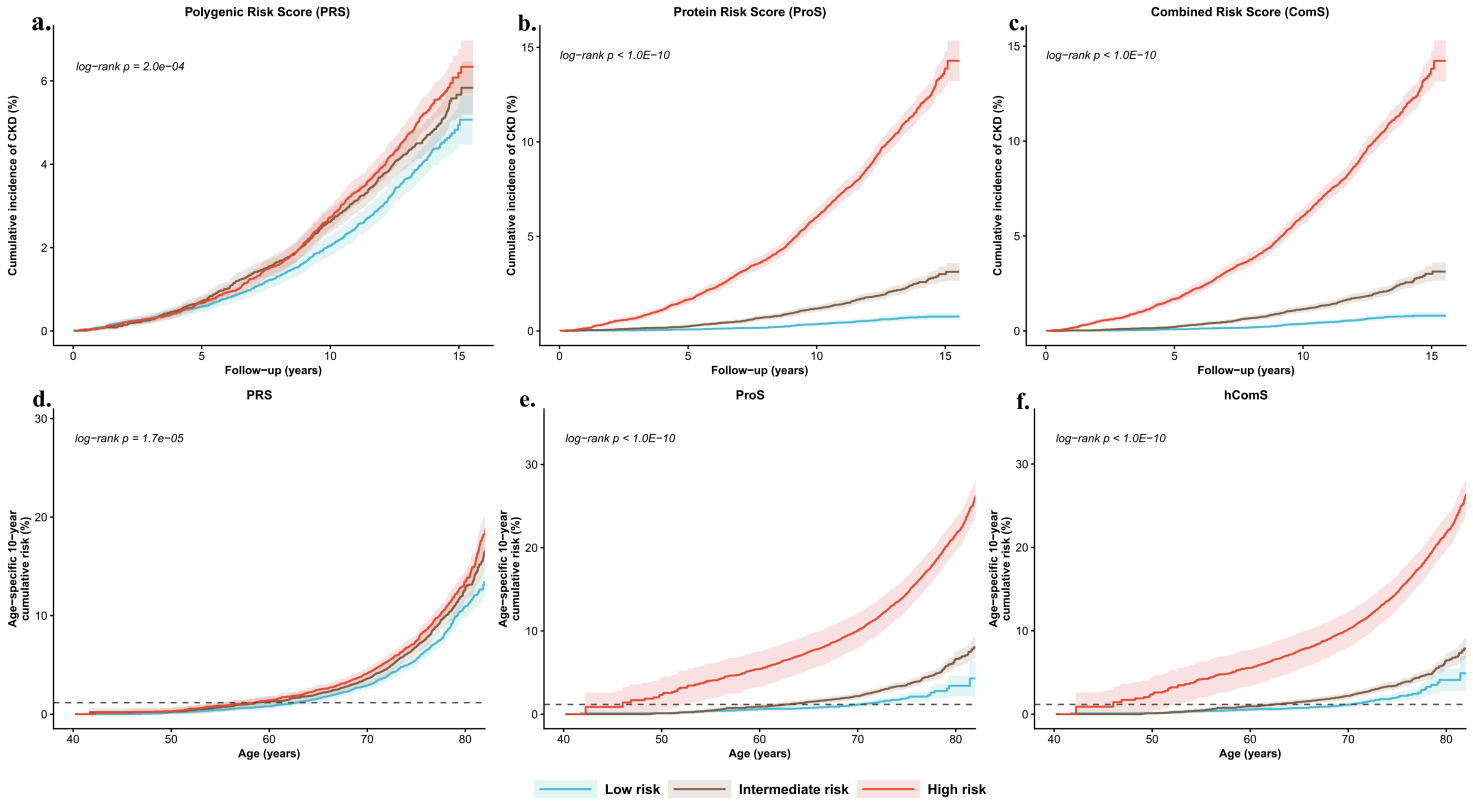


Figure 5

CKD risk stratification and age-specific 10-year risk by PRS, ProRS, and combined scores.(a-c) Kaplan-Meier cumulative incidence curves of CKD stratified into low, intermediate, and high-risk groups by (a) PRS, (b) ProRS, and (c) ComRS. Log-rank P values are shown. (d-f) Age-specific 10-year cumulative CKD risk by attained age, stratified by (d) PRS, (e) ProRS, and (f) ComRS. Dashed horizontal line indicates the 10% risk threshold for recommended screening initiation.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupportingFigure1.png
- SupportingFigure2.png
- SupportingFigure3.png
- SupportingFigure4.png
- SupportingTable1.docx

- Suooprtingtable2.xlsx
- Supportingtable3.xlsx
- Supportingtable4.xlsx
- Supportingtable5.xlsx
- SupportingTable6.docx
- SupportingTable7.docx
- SupportingTable8.docx
- SupportingTable9.xlsx