Proteomic cardiovascular risk assessment in chronic kidney disease

January 7th, 2026
Journal Club

Guillaume Onyeaghala

Division of Nephrology
University of Texas Medical Branch

utmb Health

Outline

- Introduction
  - Role of proteomics in CVD risk assessment
  - What is aptamer-based technology?
- Methods
  - Cohort selection
  - Proteomics analysis
  - Machine learning modeling
  - Study validation
- Results
  - Protein model based on elastic net regression
  - Mendelian randomization Analyses and Ingenuity Pathway Analyses
- Discussion
  - Main results
  - Limitations

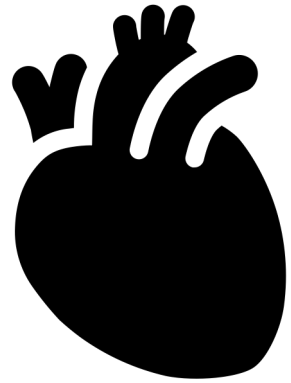
INTRODUCTION

utmb Health

ROLE OF PROTEOMICS IN CARDIOVASCULAR RISK ASSESMENT

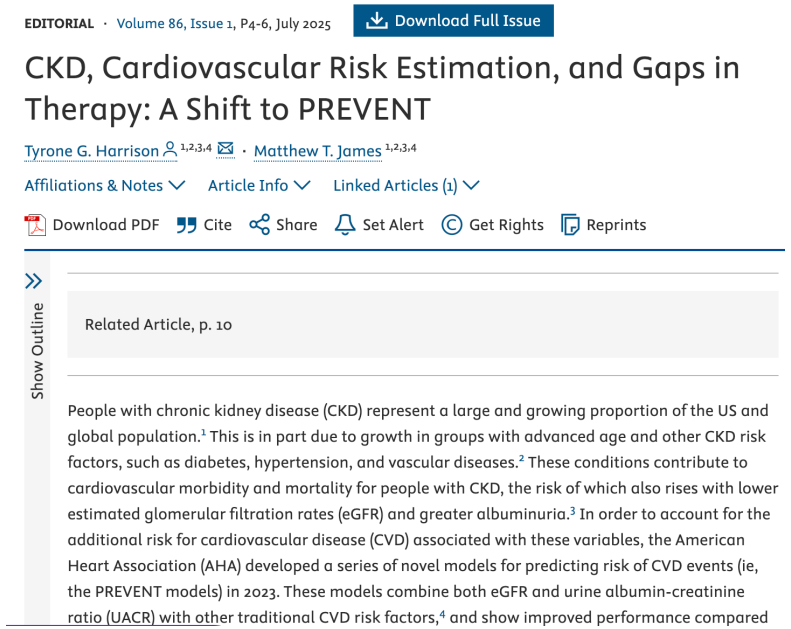
- Chronic kidney disease (CKD) patients experience increased cardiovascular risk across all CKD stages
- Clinicians are asked to risk stratify CKD patients, however cardiovascular risk prediction tools derived in the general population perform poorly in CKD populations.

- Myocardial Infarction
- Stroke
- Heart failure hospitalization
- Cardiovascular death



ROLE OF PROTEOMICS IN CARDIOVASCULAR RISK ASSESMENT

- Existing population-based model was defined by the 2013 American College of Cardiology/American Heart Association (ACC/AHA) Pooled Cohort Equations (PCE) for cardiovascular risk assessment
  - Age, sex, race/ethnicity, systolic blood pressure, use of anti-hypertensive therapies, total cholesterol, high-density lipoprotein (HDL) cholesterol, history of diabetes and current smoking
  - A subsequent model build on the PCE with eGFR < 60 mL/min/1.73 m2 for CKD populations



EDITORIAL · Volume 86, Issue 1, P4-6, July 2025 Download Full Issue

CKD, Cardiovascular Risk Estimation, and Gaps in Therapy: A Shift to PREVENT

Tyrone G. Harrison [1,2,3,4] · Matthew T. James [1,2,3,4]

Affiliations & Notes · Article Info · Linked Articles (1)

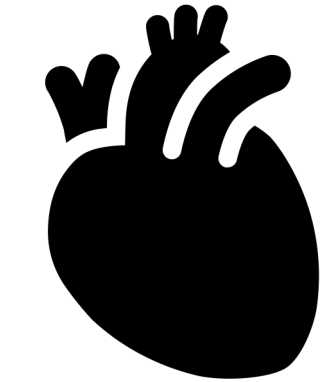
Download PDF · Cite · Share · Set Alert · Get Rights · Reprints

Show Outline

Related Article, p. 10

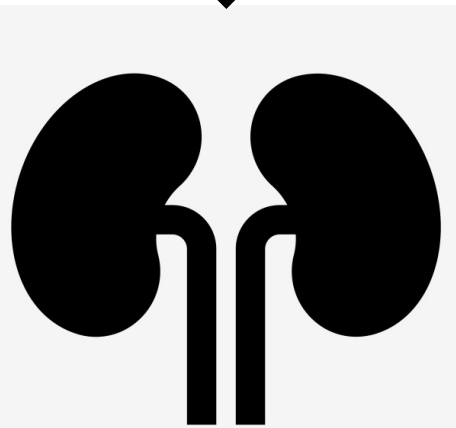
People with chronic kidney disease (CKD) represent a large and growing proportion of the US and global population.[1] This is in part due to growth in groups with advanced age and other CKD risk factors, such as diabetes, hypertension, and vascular diseases.[2] These conditions contribute to cardiovascular morbidity and mortality for people with CKD, the risk of which also rises with lower estimated glomerular filtration rates (eGFR) and greater albuminuria.[3] In order to account for the additional risk for cardiovascular disease (CVD) associated with these variables, the American Heart Association (AHA) developed a series of novel models for predicting risk of CVD events (ie, the PREVENT models) in 2023. These models combine both eGFR and urine albumin-creatinine ratio (UACR) with other traditional CVD risk factors,[4] and show improved performance compared

- Myocardial Infarction
- Stroke
- Heart failure hospitalization
- Cardiovascular death



- PCE model + eGFR risk assessment

- CKD defined as eGFR < 60 mL/min/1.73 m2



utmb Health

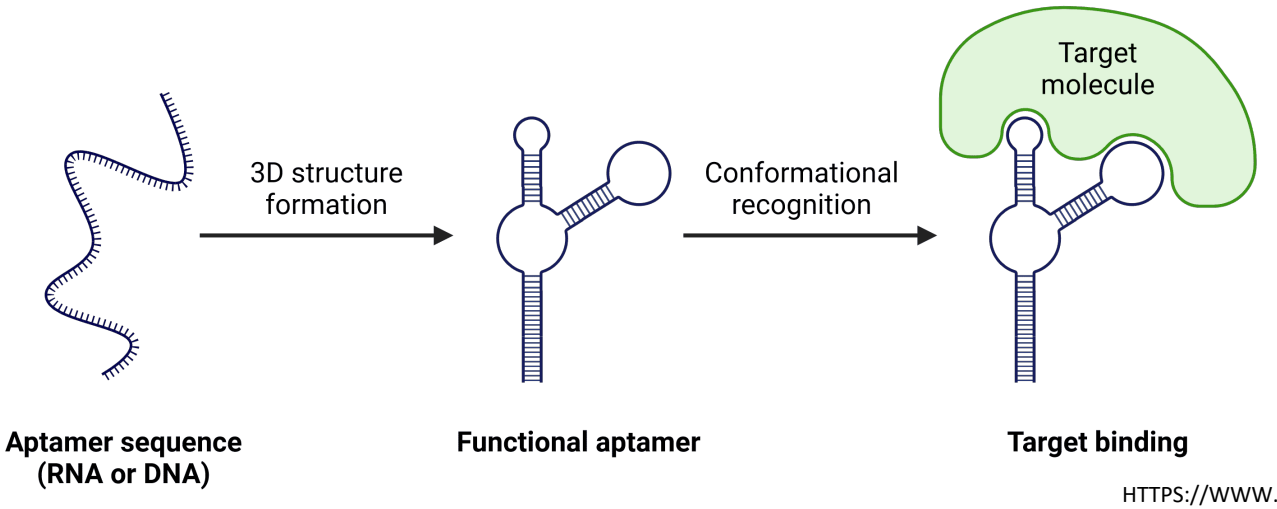
Role of Proteomics in Cardiovascular Risk Assessment

- Proteomics are one of the most upstream biomarkers of risk and disease available
  - Proteins are the targets of 95% of all known drugs
  - Proteomics data can integrate the effects of genes with the environment, comorbidities, lifestyle, and drugs.
  - Better suited to risk modelling as they are potentially modifiable, in contrast to risk factors such as age or sex

WHAT IS APTAMER BASED TECHNOLOGY?

- Aptamers are single-stranded nucleic acid molecules that are selected to bind tightly to a specific target molecule
- Aptamers fold into unique three-dimensional shapes that are highly stable and resistant to degradation, for example, by nucleases
- *Somalogic* aptamers were developed to produce a new class of protein binding reagents with high sensitivity and specificity

- Popular alternatives to this proteomic assay approach have included mass spectrometry (MS) and antibody biding technologies
  - *Somalogic* aptamers offer a multiplex assay which is more cost effective than MS, especially for low abundance proteins
  - Antibodies are not available for the majority of the ~5000 proteins in the circulating proteome
  - OLINK assays are another example of large-scale proteomics assays



utmb Health

HTTPS://WWW.ANTIBODIES.COM/APTAMERS
HTTPS://JOURNALS.SAGEPUB.COM/DOI/10.1016/J.JALA.2009.05.003

Background

Need for a better tool to evaluate CVD risk in the CKD population

Technological advances allowing thousands of circulating proteins to be measured simultaneously & Machine learning techniques to analyze the resulting data

Cohort study of CKD individuals with cryopreserved plasma samples at baseline, and longitudinal follow-up without prior cardiovascular disease.

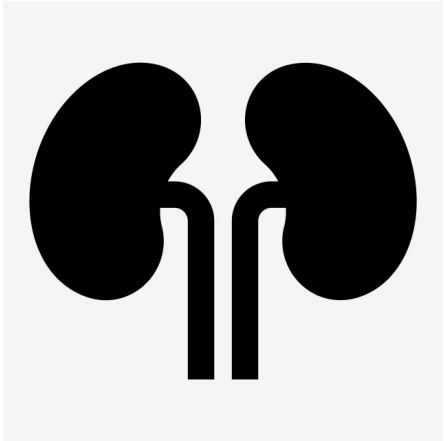
METHODS

utmb Health

Cohort Selection

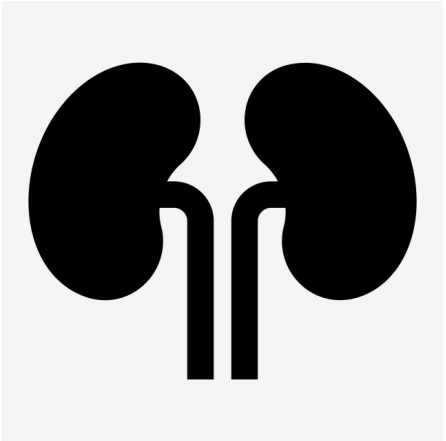
Derivation Cohort

- Chronic Renal Insufficiency Cohort (CRIC). N = 2182
  - Individuals age 21 - 74
  - CKD defined as eGFR < 60 mL/min/1.73 m2
  - CVD free at baseline (Visit 1)
  - Individuals with end-stage renal disease or on dialysis were excluded



External Validation Cohort

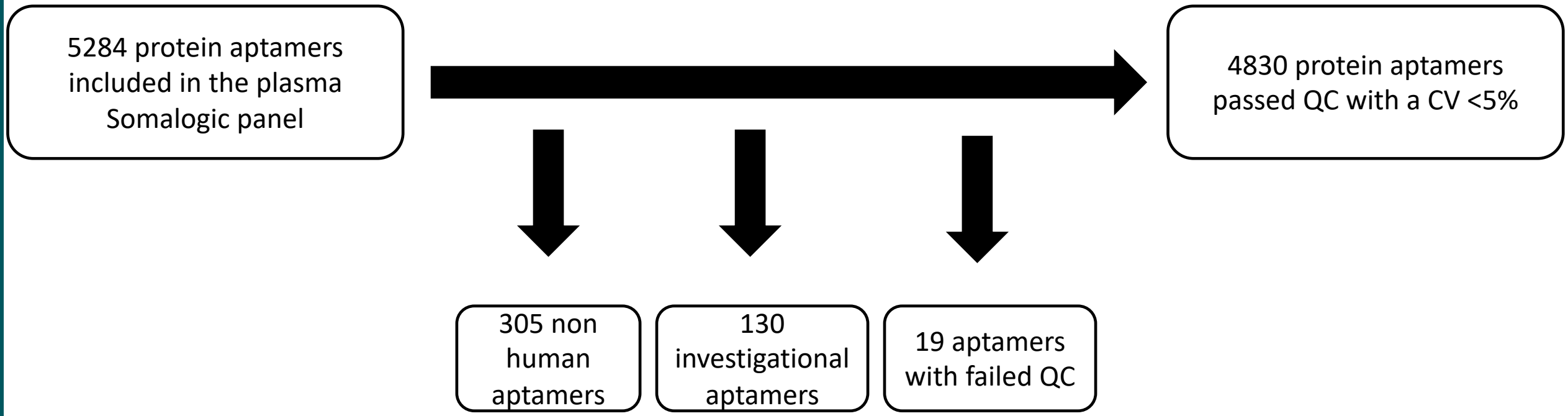
- Atherosclerosis Risk in Communities (ARIC) study. N = 485
  - CKD defined as eGFR < 60 mL/min/1.73 m2
  - CVD free at baseline (Visit 3)
  - Individuals with end-stage renal disease or on dialysis were excluded



utmb Health

CRIC: https://repository.niddk.nih.gov/study/15
ARIC: https://www.nhlbi.nih.gov/science/atherosclerosis-risk-communities-aric-study

Proteomics Analysis in CRIC and ARIC Cohorts



utmb Health

Machine Learning Modeling: Training

Chronic Renal Insufficiency Cohort (CRIC). N = 2182

80%
N = 1746

Elastic Net Regression

- Works well when we have a lot of potentially correlated / colinear covariates (i.e. 4000 + proteins)
- Minimizes prediction errors in the regression model
- Does not randomly pick one out of many features when a group of correlated variables are identified

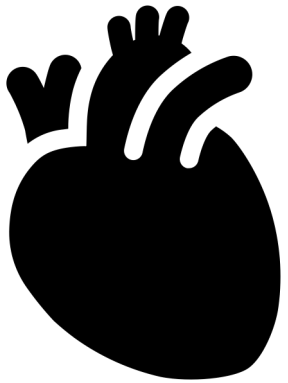
Protein only model

Pooled equation cohort (PCE) model

Protein + PCE model

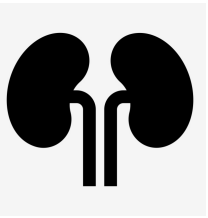
Composite CVD outcome during a 10 year follow up

- Myocardial Infarction
- Stroke
- Heart failure hospitalization
- Cardiovascular death

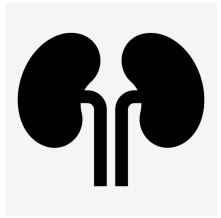


utmb Health

Study Validation

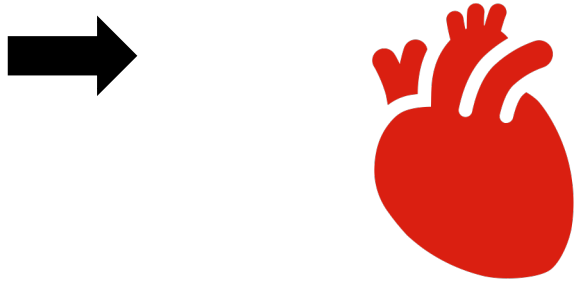


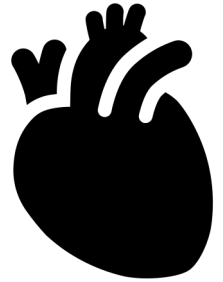
CRIC
(N=436)

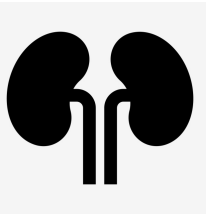


ARIC
(N=485)

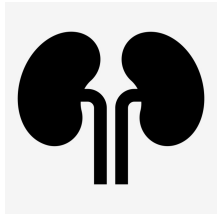
Protein only model vs PCE Model







CRIC
(N=436)

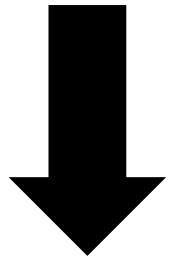


ARIC
(N=485)

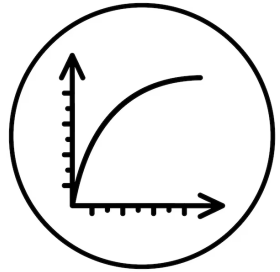
Protein + PCE model vs PCE only model

Observed CVD outcome during a 10 year follow up

Expected CVD outcome during a 10 year follow up



Evaluated receiver operating characteristic (ROC) area under the curve (AUC) annually between 1- and 10-years using time-dependent AUC methodology for the respective predictive models



utmb Health

RESULTS

utmb Health

PROTEIN BASED MODEL PERFORMANCE

- CRIC Training set
  - 373 CVD events
  - 32 proteins were retained in the primary proteomics risk model (16 had not been previously reported in the literature in association with CVD).

PROTEIN BASED MODEL PERFORMANCE

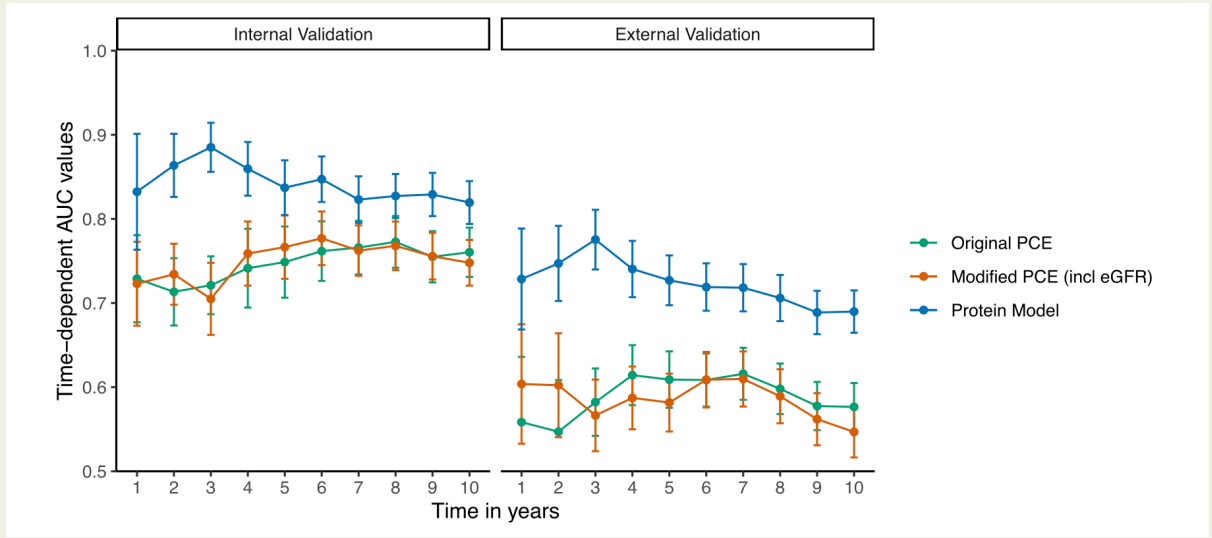


Figure 2 Time-dependent area under the curve values across 10 years of follow-up. The area under the curve values are depicted for each model in the Chronic Renal Insufficiency Cohort testing set (internal validation) and the Atherosclerosis Risk in Communities cohort (external validation). Area under the curve values for the Pooled Cohort Equation were the weighted sum of four different equations representing Black women, White women, Black men, and White men. The modified Pooled Cohort Equation included estimated glomerular filtration rate and the Pooled Cohort Equation risk factors (age, sex, race, systolic blood pressure, use of anti-hypertensive therapies, total cholesterol, high-density lipoprotein cholesterol, history of diabetes, and current smoking), all of which were refit to the Chronic Renal Insufficiency Cohort training set. The protein model was derived in the Chronic Renal Insufficiency Cohort training set using elastic net regression and consisted of 32 proteins.

10X risk difference between Q1 and Q5

4.3X risk difference between Q1 and Q5

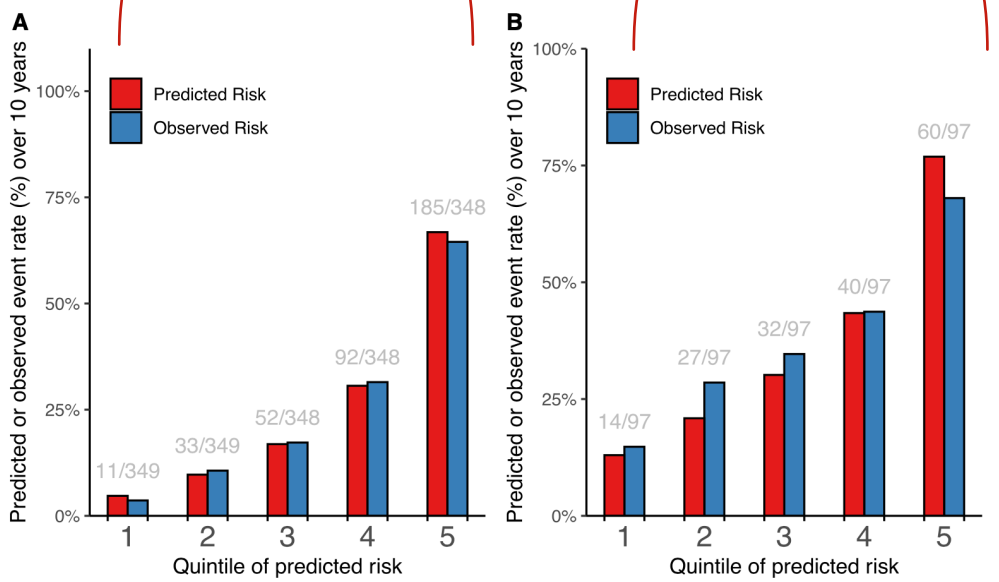


Figure 3 Calibration of proteomic risk model in the Chronic Renal Insufficiency Cohort and Atherosclerosis Risk in Communities cohort. (*A*) The Chronic Renal Insufficiency Cohort training set was comprised of 1742 participants without a self-reported history of coronary heart disease, myocardial infarction, stroke, or heart failure. Over the 10-year follow-up period, there were 373 incident cardiovascular events. (*B*) Similarly, the Atherosclerosis Risk in Communities study population was comprised of 485 participants with chronic kidney disease and without a history of cardiovascular disease at the time of proteomic measurements. Over a 10-year follow-up, there were 173 incident cardiovascular events.

CRIC ARIC CRIC ARIC

utmb Health

PROTEIN BASED MODEL PERFORMANCE

- CRIC Training set
  - 373 CVD events
  - 32 proteins were retained in the primary proteomics risk model (16 had not been previously reported in the literature in association with CVD).

- CRIC Validation set
  - 86 CVD events in the validation set
  - Proteomics risk model was superior to both the original PCE and the modified PCE ($P < 0.05$)
  - CVD risk was 10 times higher in the top quintile vs the lowest quintile of the protein risk score in CRIC

- ARIC validation set
  - 173 CVD events during 10 years of follow up
  - Proteomics risk model was superior to both the original PCE and the modified PCE ($P < 0.05$)
  - CVD risk was 4.3 times higher in the top quintile vs the lowest quintile of the protein risk score in CRIC

***Hybrid models (11 proteins + 15 clinical covariates) performed better than the clinical only models, however the authors focused on the protein only models likely due to the fact that the machine learning algorithms prioritized protein only data when selecting from both protein and clinical covariates

Mendelian randomization and ingenuity pathway analyses for individual biomarkers of CVD risk

Individual cox regression analysis of 4830 proteins with CVD outcome in CRIC

2258 proteins had a q-value <0.05 after univariate analysis

396 proteins had a q-value <0.05 after eGFR adjustment

74 proteins had a q-value <0.05 after adjustment for PCE clinical covariates

36 proteins found in the PheWAS database (phenome-wide association studies)

18 proteins had causal association with one of the endpoints in the CVD outcomes

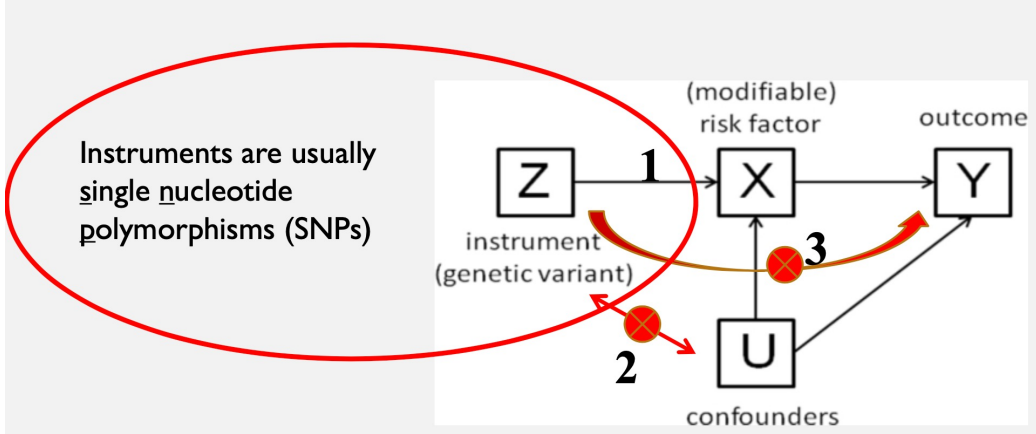


Table 3 Canonical pathways associated with incident CVD

IPA canonical pathway	*P*-value	Ratio
LXR/RXR heterodimer activation	4.04×10^{-8}	0.434 (36/83)
Hepatic fibrosis/hepatic stellate cell activation	7.63×10^{-7}	0.368 (42/114)
Acute phase response signalling	1.77×10^{-4}	0.308 (40/130)
Atherosclerosis signalling	3.13×10^{-4}	0.337 (28/83)
Complement system	4.35×10^{-4}	0.464 (13/28)
FXR/RXR activation	6.32×10^{-4}	0.338 (25/74)
Osteoarthritis pathway	0.00174	0.292 (33/113)
Coagulation system	0.00247	0.414 (12/29)
Extrinsic prothrombin activation pathway	0.00348	0.538 (7/13)
Granulocyte adhesion and diapedesis	0.00362	0.29 (29/100)

Ratio indicates the number of significant proteins that map to the canonical pathway divided by the total number of proteins measured in our study that map to the same pathway. For the IPA analysis, all protein associations with the primary cardiovascular outcome were adjusted for eGFR.
IPA, Ingenuity Pathway Analysis.

utmb Health

DISCUSSION

utmb Health

Main Takeaways

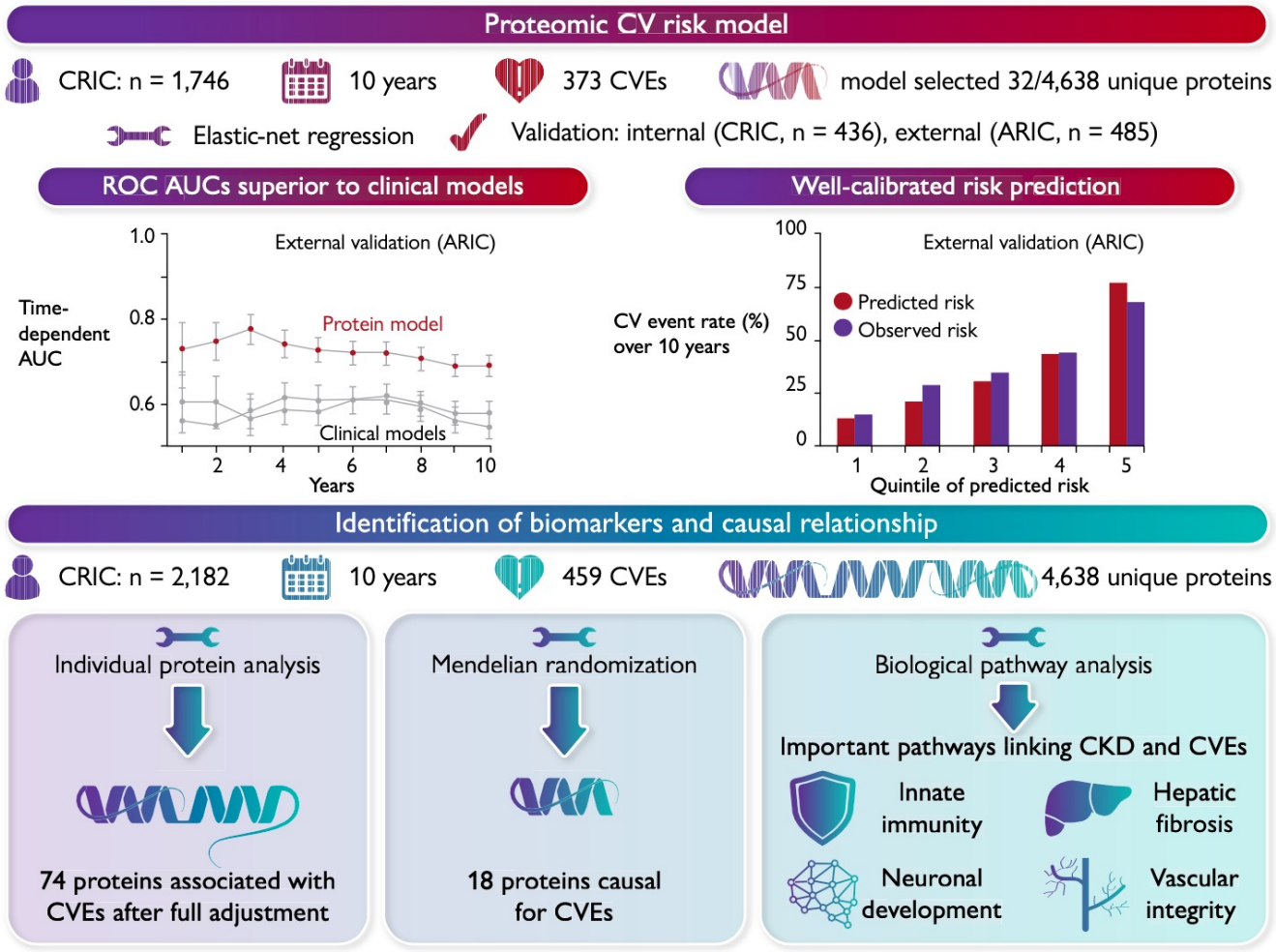
- The currently recommended cardiovascular risk prediction model (2013 PCE) is not adequate for the CKD population
  - It was developed for fatal and non-fatal stroke, fatal coronary disease and non fatal myocardial infarction
  - It was not trained on a population representative of CKD
  - It does not include important covariates such as eGFR
  - There is now a newer 2025 PREVENT model which includes eGFR and urine albumin-creatinine ratio (UACR)

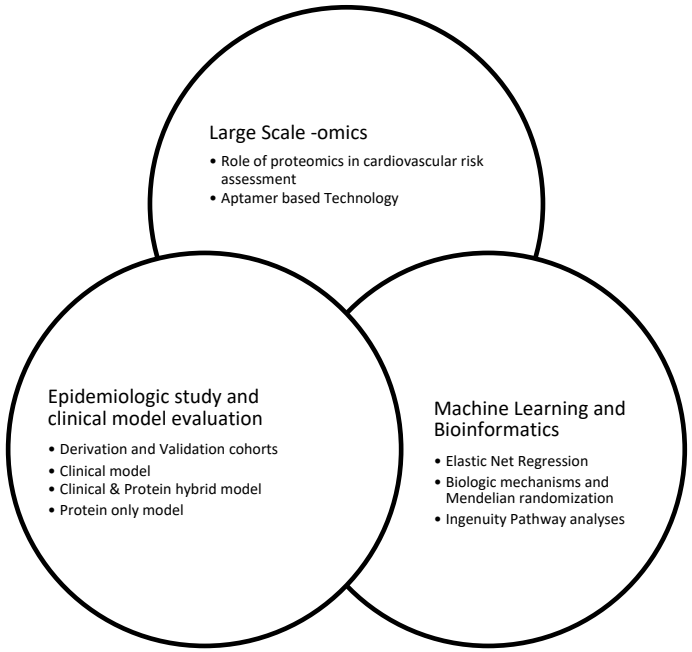
- Researchers derived a 32-protein cardiovascular risk score from a SOMALOGIC proteomics dataset.
  - The proteomics model outperformed traditional clinical models in the internal as well as external validation cohort
  - The proteomics level data can also be integrated with approaches such as mendelian randomization or ingenuity pathway analyses for biomarker discovery analyses.
  - Proteins in the immune regulation, fibrosis and vascular pathways were identified in the mendelian randomization analysis, consistent with the cardiovascular literature

Limitations

- The main CKD etiologies in the study were type-2 diabetes and hypertension
  - The 32-protein cardiovascular risk score model may not be generalizable to other types of CKD etiologies.
- The statistical models did not include important covariates
  - N-terminal pro-B-type natriuretic peptide and troponin as cardiac risk factors, SGLT2 inhibitors, GLP-1 RAs
  - Albuminuria for CKD (not available in the ARIC validation dataset)
- SOMALOGIC technology is not yet widely available for research or considered for clinical care

SUMMARY





utmb Health

THANK YOU!

QUESTIONS?

utmb Health

BONUS SLIDES

utmb Health

WHAT IS APTAMER BASED TECHNOLOGY?

Bind: •Protein of interest binds to aptamer + photocleavable marker

Catch 1: •Any bound protein + aptamer complexes are captured on a bead using the photocleavable marker. Unbound proteins are washed and bound compounds are biotin tagged

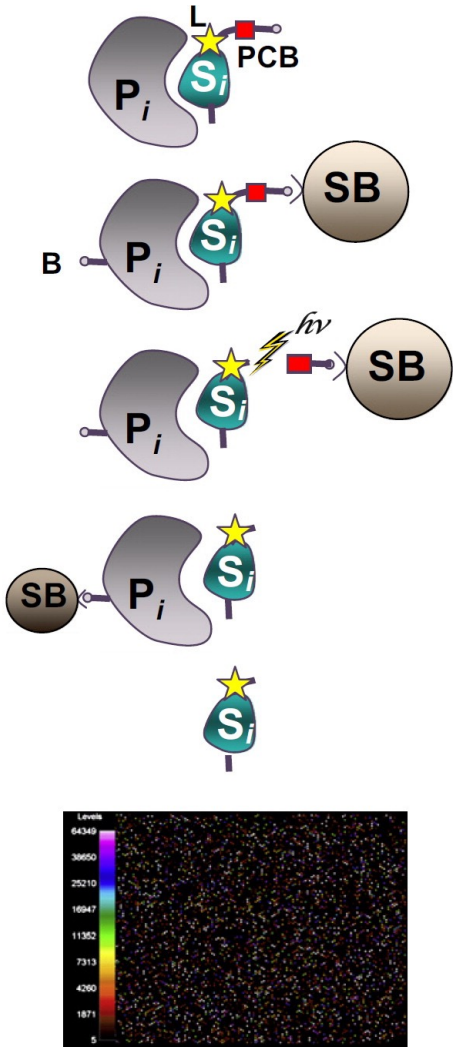
Cleave: •Ultraviolet light cleaves the photocleavable biotin and releases the protein + aptamer in the solution

Catch 2: •Protein + aptamer complexes are captured onto streptavidin-coated magnetic beads

Elute •The aptamers are eluted into the solution

Quantification •Aptamers are hybridized unto a custom DNA chip with probe spot contains DNA with sequence complementary to a specific aptamer, and the fluorescent intensity of each probe spot is proportional to the amount of aptamer recovered, and thus directly proportional to the amount of protein present in the original sample.

1. **Bind**
2. **Catch 1**
3. **Cleave**
4. **Catch 2**
5. **Elute**
6. **Quantify**



utmb Health

HTTPS://JOURNALS.SAGEPUB.COM/DOI/10.1016/J.JALA.2009.05.003