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.

Senescence Biomarkers Independently Predict Disease-Related Mortality in Chronic Kidney Disease

Thomas McLarnon
Ulster University, Derry-Londonderry Campus
Donya Ghazinader
Ulster University, Derry-Londonderry Campus
Steven Watterson
Ulster University, Derry-Londonderry Campus
Laura Lennox
Western Health and Social Care Trust
Jack Duncan
Western Health and Social Care Trust
Frank McCarroll
Western Health and Social Care Trust
Taranjit Singh Rai
t.rai@ulster.ac.uk

Ulster University, Derry-Londonderry Campus

Article

Keywords: Biomarkers, Machine Learning, Senescence, Mortality, SASP, Survival, Longevity

Posted Date: September 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10539671/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

Chronic kidney disease (CKD) staging systems predict kidney failure progression but fail to provide clinicians with patient level estimates of mortality risk, unlike similar tools that exist in cardiovascular disease, COPD and liver disease. Cellular senescence and its associated SASP have been implicated in CKD progression and comorbidity burden, yet whether these circulating SASP components independently predict long-term mortality has yet to be proven. Using plasma proteomics from a CKD diagnosed UKBiobank cohort, we identify EDA2R, NT-pro-BNP, CTSZ and REN as senescence associated biomarkers that strongly associate with 9-year disease-related mortality and estimated glomerular filtration rate (eGFR). A support vector machine (SVM) trained on nine proteomic features achieved an AUC of 0.832 during training and 0.711 in a held-out test set, outperforming both a clinical covariate model and the Charlson Comorbidity Index (CCI), the gold standard risk-adjustment tool for predicting mortality. Cox proportional hazard models confirmed that EDA2R, NT-pro-BNP, CTSZ and REN independently predict long-term disease related mortality, post-adjustment for age, sex, cardiovascular disease, diabetes and eGFR. Proteomic aging clocks (PACs), which utilized EDA2R, NT-pro-BNP and REN among other proteins, proved that individuals experiencing 9-year disease-related mortality were often more biologically aged, with proteomic accelerated aging being statistically significant between survivors and non-survivors, regardless of disease severity. Lastly, we found that EDA2R, NT-pro-BNP, CTSZ and REN expression was inversely correlated with renal function, while also confirming their direction of expression was consistent across our discovery cohort in the UKBiobank and in our own independently recruited CKD cohort. These findings promote senescence-associated plasma proteins as non-invasive, independent predictors of mortality in CKD, and nominate EDA2R and CTSZ as previously uncharacterised candidates for senomorphic interventions.

Main

Chronic kidney disease (CKD) is a progressive, irreversible condition affecting over 14% of the global population and is directly associated with 1.48 million deaths annually[1]. CKD frequently co-occurs with comorbidities such as cardiovascular diseases (CVD) and diabetes mellitus which are estimated to be responsible for 2.6 million (CVD-CKD)[2] and 0.56 million (Diabetic kidney disease, DKD) deaths annually[3]. CKD patients are characterised by either kidney damage lasting for at least 3-months or a reduction in estimated glomerular filtration rate (eGFR) less than 60mL/min/1.73m^2, or both.

Current strategies for the treatment and management of CKD involves the stratification of patients into groups based on stage of disease and renal function, into five stages of disease severity (G1–G5) according to eGFR and albuminuria category, as defined by the KDIGO 2024 Clinical Practice Guidelines[4,5]. While this clinical stratification framework is effective at guiding nephrological referral, patient monitoring, and the implementation of renoprotective therapies, it is primarily designed to predict kidney failure progression. The current CKD staging does not provide patient-level estimates of disease-related mortality risk, which are already applied across multiple chronic disease conditions such as

Cardiovascular disease (CVD) (SCORE2)[6], Chronic obstructive pulmonary disease (COPD) (BODE index)[7] and Liver disease (MELD Score)[8].

Outside the current clinical staging, cellular senescence has been identified as a key physiological driver of CKD progression[9,10], comorbidity burden[11] and mortality risk[12–14]. Cellular senescence is defined as a state of cell cycle arrest, wherein cells cease division and play an active role during embryonic development[15], tumour suppression[16] and wound healing[17]. Senescence also has an inflammatory phenotype, the senescence associated secretory phenotype or SASP, resulting in the secretion of pro-inflammatory signatures that may worsen patient health via disease progression, organ dysfunction and activation of inflammatory cascades[18,19]. SASP exhibited by senescent cells has already been investigated across numerous disease areas including CVD[20], Diabetes mellitus[21], COPD[22], COVID-19[23,24] and CKD[9,10].

In the kidneys, multiple studies have shown that senescent cells can actively evade apoptosis, and exhibit pronounced SASP secretions, contributing towards inflammation and fibrosis, an effect exacerbated in patients with severe stages of renal disease[25–27]. Senescence is also implicated in the progression of renal aetiologies into CKD, across IgA nephropathies[28], Diabetic nephropathies[29] and acute kidney injury[30,31], with canonical markers of senescence such as $p16^{INK4A}$ identified across the distal tubules, collecting ducts, podocytes, parietal epithelium of the glomerulus, vascular smooth muscle cells and interstitial cells of deteriorating kidney transplant biopsies[26].

Despite the large amount of mechanistic evidence that implicates SASP components as main drivers of worsening disease severity and mortality, there has been no study to date that investigates these markers in CKD related mortality. Circulating SASP components are fully detectable in peripheral blood and provide significant utility as prognostic biomarkers, offering a non-invasive alternative to invasive renal biopsies[13,32]. Certain SASP components have already been utilized in proteomic clocks to predict the biological age of patients, and have been tied to frailty, worsened symptomologies, all-cause mortality, accelerated aging and increased incidence of CKD[33,34].

Recent advances in machine learning and technological development provide the opportunity to apply these methods while incorporating senescence-associated biomarkers to stratify patients into high-risk cohorts, allowing for more guided clinical decision-making regarding patient treatment and disease management. This could involve using senescence-associated biomarkers as potential drug targets. By targeting signatures that correlate the strongest with severity and mortality, this would result in the dampening of SASP, acting as a senomorphic therapy that could improve longevity as well as promote healthy aging[35,36].

No study has systematically identified SASP as an independent predictor of disease-related mortality in a CKD diagnosed cohort. Here we identify EDA2R, NT-pro-BNP, CTSZ and REN as the most significantly elevated senescent biomarkers in CKD patients who experienced disease-related mortality within 9-years

of sample collection. Machine learning frameworks when fitted with 9 proteomic biomarkers achieved an AUC score of 0.832 during model training and 0.711 during model testing, confirming generalisability and real-world performance. These proteins together were able to outcompete clinically centred frameworks such as a clinical model utilizing eGFR, sex, age, diabetes and CVD presence, as well as the full ICD10 Charlson comorbidity index (Myocardial infarction, congestive heart failure, peripheral vascular disease, cerebrovascular disease, dementia, chronic pulmonary disease, rheumatic disease, peptic ulcer disease, mild liver disease, diabetes without chronic complication, diabetes with chronic complication, hemiplegia or paraplegia, renal disease, malignancy (including leukaemia and lymphoma), moderate or severe liver disease, metastatic solid tumor and AIDS/HIV). Cox-proportional hazard models confirmed that these proteins independently predict disease-related mortality after adjusting for clinical covariates, while proteomic clocks utilizing these proteins demonstrate that accelerated aging rather than chronological age holds statistically significant influence in determining what patients will experience mortality within the 9-year time point. Senescence-associated protein biomarker expression was also found to be inversely correlating with eGFR, with directional consistency in both discovery and validation cohorts. These findings collectively identify senescence-associated plasma proteins as clinically accessible, independent predictors of disease-related mortality in CKD patients, providing an alternative personalised medicine-based framework for patient stratification.

Results

Senescence biomarkers and pathways are upregulated in CKD patients yet to experience disease-related mortality

Proteomic profiles of CKD patients were quantified from their plasma and patients were separated into two groups, experienced mortality within 9 years vs no mortality within 9 years. Quantified proteins were then tested in differential expression analysis to identify biomarkers. Interestingly, senescence associated proteins ADM, PGF, EDA2R and NT-pro-BNP were all identified as the most statistically significant biomarkers indicative of mortality (Fig. 1A/1B). Proteins in the volcano-plot were coloured according to direction of expression and senescence association, with red used for standard upregulation, blue for downregulation, green if the proteins corresponding gene transcript was differentially expressed in cellular senescence models of proximal tubule cells or fibroblasts from human kidneys[37] and grey if they were not significant (Fig. 1A). Protein expression for patients visualised in the heatmap was overlaid and separated by clinical factors to identify strong confounders, with the use of both sex and average age of cohort, which ruled out chronological age as influential, suggesting it may be biological age (Fig. 1B). Next, the full OLINK proteomic landscape of these patients was utilized in a dimensionality reduction analysis which focused on the principal components most correlative with mortality status (PC1, PC6). Patient datapoints on the PCA were coloured according to mortality status, with red labelling if the patient experiences mortality withing 9-years of sample collection, and blue colouring if they did not (Fig. 1C). This demonstrated that the proteomic biomarkers shared unique biological variance and were able to separate CKD patients based on their 9-year mortality status

(Fig. 1C). The proteomic landscape was then narrowed to only include potential biomarkers with a false discovery rate below 0.05, using these collective biomarkers for pathway analysis, which revealed key inflammatory and SASP pathways such as cytokine – cytokine receptor interaction, NF-kappa B signalling, viral protein interaction with cytokine / cytokine receptor and chemokine signalling pathways (Fig. 1D). Lastly, the expressional distribution of the top four biomarkers was visualized using violin-boxplots which demonstrated the large statistical significance of ADM, PGF, EDA2R and NT-pro-BNP as heavily upregulated proteomic biomarkers of disease-related mortality (Fig. 1E i-iv).

Baseline senescence biomarkers can predict 9-year disease-related mortality in CKD patients

Multiple machine learning frameworks were tested, having access to all 372 proteins and then using forward feature selection to build the best classifier, which resulted in a linear-SVM utilising expressional values of EDA2R, NT-pro-BNP, CTSZ, CLEC4C, HMOX2, NTRK3, REN, CCL3 and CLEC6A (Fig. 2A). Models were colour coded according to algorithm and kernel type, with linear-SVM coloured yellow, radial basis function-SVM colour coded blue, logistic regression colour coded green and random forest colour coded pink (Fig. 2A). The optimal model and feature subset was then used to build the final model, employing repeated 10-fold cross-validation with 10 repeats for robustness and yielded a training AUC of 0.832, sensitivity of 0.752 and specificity of 0.785 (Fig. 2B). Model explainability and behaviour was then investigated, with the decision boundary of the final model being projected and visualised on a reduced dimensional space confirming the presence of biological signal and confirming no overfitting was present (Fig. 2C). Regions shaded light blue on the decision boundary plot represent $f(x) < 0$ which classifies all points on that side as non-mortality and regions shaded light red on the decision boundary plot represent $f(x) > 0$ which classifies all points on that side as mortality (Fig. 2C). Feature influence was tested with SHAP, which demonstrated clear, linear relationships between a proteins biological expression and the effect on an individuals predicted outcome by removing individual proteins and testing their effects on predictions pre- and post-removal (Fig. 2D). SHAP data points were colour coded according to proteomic expression (feature value), proteins with higher expression were coloured red and proteins with lower expression were coloured blue. Lastly, the model was evaluated on the held-out test subset, achieving an external AUC of 0.711 (Fig. 2E), making 46 true negative predictions, 43 true positive predictions, 24 false positive predictions and 27 false negative predictions (Fig. 2F). Shaded blue regions on the ROC curve represent the 95% confidence intervals, generated via stratified bootstrap resampling (n = 5000), with sensitivity calculated at fixed specificity values (Fig. 2E).

Senescence proteomics improve CKD mortality prediction beyond established clinical models

Current methods for stratifying CKD patients revolve around the use of clinical metrics like age, sex, CVD presence, diabetes mellitus presence and eGFR, or the use of weighted comorbidity formulas like the

Charlson comorbidity index (CCI) which were tested against the fitted SVM with senescence biomarkers. The proteomics SVM outperformed both the CCI and clinical benchmark models, achieving an AUC score of 0.711 coloured red on the ROC-curve, compared to the clinical models 0.613 coloured green on the ROC-curve and the CCI models 0.637 coloured blue on the ROC-curve, demonstrating the utility senescence biomarkers have on future risk stratification (Fig. 3A). Performance of the models on the held-out cohort were tested to see the effectiveness of the senescence biomarkers on the individual level which showed the senescence model outperforms on the entire cohort, but also in most subgroups excluding diabetes where the CCI marginally outperformed it (Fig. 3B). Subgroup comparisons were coloured and shaped according to the specific models, with red circles denoting the proteomics SVM, green squares denoting the clinical baseline model and blue triangles denoting the Charlson comorbidity index (Fig. 3B).

Senescence biomarker upregulation predicts poor survivability beyond clinical measures

The established proteomic panel from the final model was then used to perform Cox-proportional hazard survival analysis to determine the respective hazard ratios (HR) of each protein. Senescence associated proteins EDA2R (HR 2.05), NT-pro-BNP (HR 1.32), REN (HR 1.34) and CTSZ (HR 2.27) were identified as the most statistically significant proteins that are associated with 9-year disease-related mortality (Fig. 4A). Remarkably, when these models were then adjusted to account for variance relating to age, sex, CVD presence, diabetes mellitus presence and eGFR (CKD-EPI), these proteins all remained statistically significant: EDA2R (adjusted HR 2.02), NT-pro-BNP (adjusted HR 1.2), REN (adjusted HR 1.19) and CTSZ (adjusted HR 1.7), demonstrating their predictive capabilities independently from clinical factors (Fig. 4A). It is also worth of note that the top performing biomarkers were all senescence-associated, whereas the bottom performers are not senescence-associated. The point estimates for each protein present on the forest plot are colour coded according to whether a proteins corresponding gene transcript was differentially expressed in cellular senescence models of proximal tubule cells or fibroblasts from human kidneys, with green confirming they are and grey confirming they are not[37]. The shape of the point estimates change according to whether the model was adjusted or unadjusted, with adjusted models appearing as diamond shaped and unadjusted models appearing as univariate in shape (Fig. 4A).These same proteins were then used in Kaplan-Meier (KM) survival analysis to test the effect of expression and survivability. The KM survival curves demonstrates that an increase in protein expression of EDA2R, NT-pro-BNP, REN and CTSZ all correlates with poorer survivability within the 9-year timepoint (Fig. 4B i-iv). The KM survival curves are colour coded based on proteomic biomarker expression, with red colouring confirming above mean expression of the protein and blue colouring confirming below mean expression of the protein (Fig. 4B i-iv).

Proteomic Age Acceleration Is Elevated in Non-Survivors regardless of CKD stage

With confirmation that these protein biomarkers can predict disease-related mortality, we then tested whether accelerated aging was significantly different between survivor and non-survivor patients. Using the coefficients from a previously published study33, we compared the biological ages, or "ProtAges", of survivor (Fig. 5A) and non-survivor (Fig. 5B) patients, and found that patients who experienced disease-related mortality within the 9-year time point had higher biological ages than patients who survived past this time point. Survivor and non-survivor groups had the same average chronological age, but the survivor cohort showed a mean ProtAge deviation (ProtAge − chronological age) of + 9.8 years (Fig. 5A), compared to + 19 years in the non-survivor cohort (Fig. 5B), confirming that biological age, rather than chronological age, is the key association with mortality33. To assess this relationship further, we calculated proteomic age acceleration (PAA) for each patient as their age-adjusted residual deviation from the cohort-wide proteomic age trend. Higher PAA was associated with poorer survivability (Fig. 5C), with survivors showing lower PAA than non-survivors (Fig. 5D). Finally, we examined the distribution of PAA between survivor and non-survivor patients across all stages of CKD severity, and found that non-survivor patients had significantly higher PAA than survivor patients at every disease stage (Fig. 5E), ruling out disease severity as a confounding factor.

Validated Senescence biomarkers directly correlate with renal function

With the identification of senescence associated biomarkers for disease-related mortality, the next step was to evaluate any potential relationships between the proteins themselves and metrics for disease severity such as eGFR. Pearson correlation tests were employed to test the expression of proteins against eGFR, which confirmed EDA2R (Fig. 6A), NT-pro-BNP (Fig. 6B), REN (Fig. 6C) and CTSZ (Fig. 6D) all having inversely proportional relationships with eGFR. Strikingly, these relationships were all preserved within UKBiobank and across a smaller validation cohort, with EDA2R being the most significant and conserved ($r= -0.745$,p-value = 9×10^{-29}) (Fig. 6A). Datapoints and line of best fits were colour coded according to cohort, with blue colouring indicating patients and line of best fit within the UKBiobank and red colouring indicating patients and line of best fit within the Acute Kidney Injury (AKI) outcome study (Fig. 6A – 6D). We next assessed whether these findings replicated in an independent cohort with the same diagnosis of CKD and mortality classification. Differential expression analysis confirmed that ADM (Fig. 6Ei), EDA2R (Fig. 6Eii), NT-pro-BNP (Fig. 6Eiii) and PGF (Fig. 6Eiv) retained the same direction of dysregulation observed in the UK Biobank cohort, demonstrating that the proteomic signal is not specific to a single dataset.

Discussion

In this study, we investigated the differential expression of senescence-associated plasma proteins in CKD patients who experienced disease-related mortality within 9-years post-blood sample collection. To our knowledge, this is the first study that has demonstrated the utility of circulating senescence-associated plasma proteins as potential biomarkers (EDA2R, NT-pro-BNP, CTSZ, CLEC4C, HMOX2,

NTRK3, REN, CCL3 and CLEC6A) of long-term disease-related mortality in CKD. Among the identified senescence-associated proteins, both EDA2R and CTSZ presented the strongest independent associations with disease-related mortality, warranting specific focus. EDA2R (Tumour necrosis factor receptor superfamily member 27), the strongest associate of disease-related mortality and eGFR (Fig. 4A, 4Bii, 6A. 6Eii), is a receptor for the EDA isoform A2, mediating the activation of NF-kappa B and JNK inflammatory pathways[38]. Upregulation of EDA2R has been identified as an indicator of accelerated aging, cellular senescence, frailty and systemic inflammation[39–41]. Interestingly, EDA2R has been used in proteomic based clocks that predict all-cause mortality and associated risk of age-associated diseases[34]. However, EDA2R has not previously been investigated as an independent predictor of disease-related mortality within a CKD-specific cohort, nor has its relationship with renal function been quantified.

CTSZ (Cathepsin Z), the second strongest associate of disease-related mortality and eGFR (4A, 4Bi, 6D), exhibits both carboxy-monopeptidase and carboxy-dipeptidase activity with roles in extracellular matrix remodelling and macrophage-mediated inflammation[38]. CTSZ has been documented to play key roles in the NLRP3 inflammasome, by specifically augmenting NLRP3 inflammasome-derived IL-1β expression, amplifying local and systemic inflammation[42]. Direct links between CTSZ and disease-related mortality have been identified, with high CTSZ expression as part of a senescence-associated signature correlating with poor survivability in glioblastoma[43]. To our knowledge, circulating CTSZ in CKD contexts has yet to be investigated, however, inhibition of specific lysosomal cathepsin members has been shown to reduce renal fibrosis in murine models of CKD[44]. Collectively, these findings suggest that CTSZ-mediated inflammatory signalling may represent a novel and previously uncharacterised link to disease-related mortality in CKD.

Our SVM using inflammatory and senescence-associated proteins as features performs well on both training (AUC: 0.832, Fig. 2B) and testing (AUC: 0.711, Fig. 2E) cohorts. The observed AUC reduction of 0.121 is consistent with real-world biological variance rather than overfitting, reflecting the complexity of predicting 9-year disease-related mortality from a single baseline proteomic timepoint. Repeated 10-fold cross-validation during training (Fig. 2B) and subsequent visualisation of the model's projected decision boundary (Fig. 2C) collectively confirmed that the model captured genuine biological signal. SHAP analysis of the models predictions confirmed directional consistency with protein expression, with an increase in expression of EDA2R, NT-pro-BNP, CTSZ and REN all resulting in increased predictions of disease-related mortality (Fig. 2D).On the held-out test cohort, the model correctly identified 43 true positive and 46 true negative predictions, achieving a sensitivity of 61.4% and specificity of 65.7% (Fig. 2F). Despite these modest absolute metrics, the model's discriminative performance (AUC: 0.711) consistently exceeded that of both the clinical (AUC: 0.613) and Charlson Comorbidity Index benchmark models (AUC: 0.637) (Fig. 3A), while also outperforming both comparators across most patient subgroups (Fig. 3B).

Cox-proportional hazard models when fitted with these proteins demonstrates their association with disease-related mortality with EDA2R (HR: 2.05), NT-pro-BNP (HR: 1.32), CTSZ (HR: 2.27) and REN (HR: 1.34) being the most statistically significant with p-values < 3×10^{-8} (Fig. 4A). When these proteins were adjusted to account for clinical covariates such as age, sex, CVD presence, diabetes mellitus presence and eGFR, they all remained statistically significant with p-values < 0.001. Remarkably, EDA2R specifically retained near-identical hazard ratios following adjustment for clinical covariates (Unadjusted HR: 2.05, Adjusted HR: 2.02), representing a reduction of less than 2%, which demonstrates its strong association with disease-related mortality, independent of key clinical variables. From the entire nine-protein length panel, the top six best performers were associated with senescence, with the less significant remaining three proteins not being associated with senescence. Kaplan-Meier analysis demonstrated the relationship between protein expression and survival outcomes, as CTSZ (Fig. 4Bi), EDA2R (Fig. 4Bii), REN (Fig. 4Biii) and NT-pro-BNP (Fig. 4Biv) all exhibited p-values < 0.0001, with higher expression of these proteins correlating with poorer survival outcomes.

Proteomic age, derived from an externally validated proteomic aging clock[33], was found to deviate from chronological age more in non-survivors than in survivors (mean deviation + 19 vs + 9.8 years), despite both groups being matched for chronological age (63.9 vs 63.9 years). This indicates that the observed difference reflects genuine biological ageing rather than confounding by chronological age itself. When this relationship was further quantified using proteomic age acceleration (PAA, the age-adjusted residual deviation in proteomic age), the pattern held consistently across every stage of CKD severity (KDIGO classification), with non-survivors exhibiting significantly higher PAA than survivors regardless of eGFR-defined stage, suggesting that PAA captures mortality risk independently of renal function decline. The clock itself employs several of the senescence-associated biomarkers identified in our own analysis, most notably EDA2R and NT-pro-BNP, both of which were also among the strongest predictors of disease-related mortality within this cohort. The identification and use of these proteins in both our own analysis and an independently published proteomic clock adds credibility to their biological relevance and potential as biomarkers and drug-target candidates for this senescence-driven ageing phenotype in CKD.

Correlation analysis between protein expression and eGFR using Pearson correlation demonstrated that EDA2R, NT-pro-BNP, CTSZ and REN all exhibited inversely proportional relationships with renal function in the UKBiobank and an independent validation cohort. Both NT-pro-BNP (Fig. 6B) and REN (Fig. 6C) had r coefficients ranging from − 0.30 to − 0.52 with eGFR, confirming their inversely proportional relationship, which supports their dual roles as both cardiac and renal biomarkers due to the co-occurrence of CKD with CVD[45–48]. Interestingly, EDA2R (Fig. 6A) and CTSZ (Fig. 6D) had stronger inversely proportional relationships with eGFR as CTSZ had r coefficients of − 0.42 and − 0.55 across the UKBiobank and the external validation cohort respectively. The strongest relationship observed was between EDA2R and eGFR, with r coefficients of − 0.67 and − 0.74 across the UKBiobank and the external validation cohort respectively. While the relationship between circulating EDA2R and eGFR had not been previously investigated, EDA2R expression has been shown to be upregulated in glomerular and tubular cells of

both diabetic mice and human diabetic nephropathy biopsies, where it promotes podocyte apoptosis through enhanced reactive oxygen species (ROS) generation[49]. These findings together could suggest that EDA2R-mediated inflammation may contribute to progressive renal functional decline in a wider CKD setting. Differential expression analysis of a biologically similar cohort from an independent study revealed that the most statistically significant differentially expressed proteins identified in the UKBiobank shared the same direction of expression in a separate cohort, proving proteomic concordance and validating the findings for ADM (Fig. 6Ei), EDA2R (Fig. 6Eii), NT-pro-BNP (Fig. 6Eiii) and PGF (Fig. 6Eiv).

The ability of these senescence-associated proteins to retain statistically significant hazard ratios following the adjustment for age, sex, CVD, Diabetes and eGFR demonstrates their ability to provide independent prognostic information beyond traditional clinical markers. The independence of these senescence-associated proteins from currently employed clinical markers provides a unique opportunity for a personalised medicine framework to stratify patients according to risk of mortality. Circulating plasma proteins can be non-invasively quantified from routine blood sampling, offering an accessible and scalable approach to employ this mortality risk stratification that compliments the existing staging system. Collectively, these findings support the framework for integrating senescence-associated biomarkers into clinical care pathways for the purpose of risk stratification, allowing for high-risk patients to be identified early on and necessary changes to their care be made.

The significance of EDA2R, NT-pro-BNP, CTSZ and REN at predicting 9-year disease-related mortality, coupled with their inversely proportional relationship with eGFR suggests the biomarkers themselves may hold distinct mechanistic functions. These findings present the opportunity for these biomarkers to be used as druggable targets for a senomorphic therapy, whereby their inhibition could reduce SASP burden while simultaneously promoting longevity, healthy aging and renoprotective mechanisms.

This concept is partially realised in current clinical care through the inhibition of REN: the direct renin inhibitor aliskiren has been evaluated in patients with type 2 diabetes and nephropathy in the AVOID trial (ClinicalTrials.gov: NCT00097955). The study showed that when given in addition with maximal-dose losartan, the result was a marked reduction in urinary albumin-to-creatinine ratio by 20% relative to the placebo, with twice as many patients achieving more than a 50% reduction in albuminuria when also compared to placebo[50]. Unlike REN, where direct inhibition yields a clear renoprotective benefit, the NT-pro-BNP precursors proBNP and BNP require more complex therapeutic strategies. Genetic ablation of BNP in animal models produces progressive hypertension, cardiac hypertrophy and nephropathy, which demonstrates the protective roles of BNP[51]. Despite this, NT-pro-BNP, which stems from the precursor proBNP is cleared independently of BNP's primary degradation pathway, which offers a distinct opportunity to indirectly reduce its levels by treating the cardiac strain driving its production, which is seen commonly in NT-pro-BNP guided clinical trials. For both EDA2R and CTSZ, there have been no studies or trials to date that have attempted to reduce their levels of expression and document the clinical endpoints, which further illustrates the need for targeted senomorphic therapies involving these biomarkers.

The present study despite its findings does acknowledge several limitations. The UKBiobank cohort comprising 468 participants is small when employed for machine learning and lacks a direct external validation cohort for model predictions despite rigorous methods to mitigate this effect. Heterogeneity of the cohort itself is limited as included study participants are solely from the United kingdom and Republic of Ireland, which could influence the observed biological effects. Additionally, proteomic measurements were taken at a single baseline timepoint, making it impossible to determine whether the senescence-associated proteins drive mortality or reflect the burden of the existing CKD. Finally, the model using the protein biomarkers achieved a modest performance on unseen data, with a sensitivity and specificity of 61.4% and 65.7% respectively, limiting its current clinical deployability.

Overall, this study demonstrates that circulating senescence-associated plasma proteins are independent predictors of disease-related mortality in CKD, most notably with EDA2R and CTSZ identified as novel and previously uncharacterised contributors to long-term mortality risk and renal function correlates. Additional studies in larger, more heterogenous CKD cohorts are needed to validate these associations and to determine whether senescence biomarker profiling can be translated into clinical risk stratification frameworks to improve patient outcomes.

Methods

UKBiobank Study Population

The UK Biobank recruited 500,000 participants within the United Kingdom between 2006–2010 for baseline assessment, with approximately 54,000 participants having proteomic profiling carried out using the OLINK Explore 3072 platform. Cohort generation using data from the UKBiobank was carried out under the CardioScope project (UKBiobank project DI: 104404). Proteomics were filtered to identify overlapping proteins between the UKBiobank and validation cohort, identifying 372 shared proteins to be analysed. From this cohort, we identified participants with prevalent chronic kidney disease (CKD; ICD-10 code N18) diagnosed prior to baseline blood sample collection who subsequently died from CKD-related causes during follow-up, yielding 468 participants for analysis. CKD-related mortality was defined as death with an underlying or contributory cause of death coded as:

Direct renal causes

N17 (acute kidney injury), N18 (chronic kidney disease), N19 (unspecified renal failure), N03 (chronic nephritic syndrome), N05 (unspecified nephritic syndrome), I12 (hypertensive chronic kidney disease), I13 (hypertensive heart and chronic kidney disease), T86.1 (kidney transplant complications), Z94.0 (kidney transplant status), Y84.1 (dialysis complications), Z99.2 (dependence on renal dialysis).

Cardiovascular complications of CKD

I11 (hypertensive heart disease), I21 (acute myocardial infarction), I25 (chronic ischemic heart disease), I27 (pulmonary heart disease), I42 (cardiomyopathy), I50 (heart failure), I64 (stroke).

Diabetes mellitus with renal complications

E11.2 (type 2 diabetes with renal complications), E11.7 (type 2 diabetes with multiple complications), E14.2 (unspecified diabetes with renal complications), E14.7 (unspecified diabetes with multiple complications).

To determine the optimal timepoint for the CKD-related mortality classifier, power analysis was performed across 15 time points (2–17 years post-baseline). At each time point, participants were stratified into those who died before versus after said threshold. Statistical power was calculated using two-sample t-tests with unequal sample sizes (pwr.t2n.test in R) for each protein individually, using the observed Cohen's d effect size, sample sizes at that time point, and an α = 0.05. Mean power was calculated by averaging across all proteins, and separately for proteins showing univariate significance ($p < 0.05$) versus non-significant proteins ($p \geq 0.05$). The 9-year time point was selected as it provided perfect class balancing (n = 234 v n = 234) while maintaining mean statistical power > 80% across most proteins (Supplementary Fig. 1).

Validation cohort recruitment and quantification

Our validation cohort comprised 155 CKD patients (mean age 59 years; 63% male; CKD stages 2–5), with 86 CKD patients having recorded mortality information, recruited from Altnagelvin and Letterkenny University hospitals between 2018–2020, as part of the Acute Kidney Injury (AKI) outcome study: A prospective cohort study to investigate the value of biomarkers in predicting adverse outcomes following acute kidney injury, as previously described[10,52]. Plasma proteomics was performed using the OLINK Proseek Multiplex proximity extension assay across six panels (cardiovascular II & III, immune response, inflammation, neuro-exploratory, neurology), quantifying 476 unique proteins after quality control. Senescence associated protein identification was performed using RNA-sequencing data from human derived proximal tubule and fibroblast cell lines from the kidney[37]. Senescence association was determined if an OLINK proteins corresponding gene transcript was differentially expressed in either the proximal tubule or fibroblast senescent model. Validation cohorts were used to test correlations between senescence protein biomarkers against baseline renal function (eGFR), and to ensure their direction of expression between survivor and non-survivor patients matched our discovery (UK Biobank) cohort.

Demographic analysis of cohort

Demographic and clinical profiling was carried out at baseline comparing cases (mortality event within 1–9 years) and controls (survival beyond 9 years). Differences in continuous variables (age, eGFR, creatinine) were assessed via t-tests, and differences in categorical variables (sex, CVD, diabetes) were assessed via chi-square tests. Statistical significance was defined as $p < 0.05$ and was assessed in R using $tab \leq o \neq$ package (v0.13.2). The CKD-related mortality variable was tested across multiple time points, with the 9-year time point selected based upon case ~ control ratio balancing and statistical power (Supplementary Fig. 1).

Differential expression analysis

Proteomic differential expression analysis was carried out in R using Welch's two-tailed *t*-test with the $stats$ package (v4.5.1). Due to the expected biological variation between participants with 9-year mortality compared to > 9-year mortality, we employed Welch's two-tailed *t*-test to account for unequal variance of proteomic recordings between these cohorts. Volcano plots were employed to visualise the overall differences in protein expression between cohorts, heatmaps were employed to visualise the individual patient proteomic differences between cohorts, PCA was utilised to visualize the entire proteomic variation between cohorts, and violin boxplots were used to visualise individual proteomic differences between cohorts for the most statistically significant proteins. Pathway analysis was carried out in R using the $pathf \in dR$ package (v2.7.0) which used the proteins corresponding gene symbol, log2FC and FDR (Benjamini-Hochberg correction).

Machine Learning framework for Mortality prediction

Supervised Machine Learning for classification was carried out in R, primarily with the use of $caret$ (v7.0.1), $\prec rec$ (v0.14.5) and $pROC$ (v1.19.0.1). Mortality-9-year status was used as the binary classifier, with OLINK proteins as potential features. Data was first partitioned into train and test cohorts using a 70:30 split. Four independent models were created to ensure performance and feature identification was consistent and utilized linear support vector machines (linear SVM), radial basis function support vector machines (non-linear SVM), logistic regression and random forest. Feature selection for each model was carried out using greedy-forward feature selection, wherein every protein was tested alone and the protein with the highest area under curve from a receiver operator characteristic curve (AUC-ROC) was selected as the first feature. This feature was then tested with every other feature, with the feature that increased AUC-ROC the most being retained. This was repeated several times until n = 20 to prevent overfitting. Feature selected final models were evaluated against each other with a linear SVM with 12 features being identified as the most optimal model. All models underwent 10-fold cross validation to prevent overfitting and used the same training / testing splits. The optimal linear SVM model that was identified was then rebuilt this time undergoing both 10-fold cross validation with 3 external repeats and hyperparameter optimisation via random grid search. Performances of all 10 folds across the 3 repeats were visualized to ensure metrics like AUC-ROC, sensitivity and specificity all passed random chance.

Model behaviour and Explainability

The SVM decision boundary was visualised in principal component (PC) space to demonstrate model behaviour and explainability. Training data underwent PCA, and a coordinate grid of 150 × 150 points (22,500 total) was constructed spanning the range of PC1 and PC2. This grid was inverse-transformed to the original dimensional space by setting PC1 and PC2 to their grid values while holding PC3-PC17 at zero. The trained SVM model predicted class labels for all grid points, and predictions were visualized in PC1-PC2 coordinates to reveal the decision boundary (f(x) = 0).

Evaluation of both feature importance and individual-level prediction influence for the SVM was performed using SHAP (Shapley Additive Explanations). The trained SVM model was first wrapped using a model-agnostic predictor interface, which allows interactions with the model to extract interpretations of the predictions without altering the model itself.

SHAP values were then computed for each individual observation by quantifying the contribution of each input feature to the SVM's prediction for that specific case. SHAP values were calculated by employing:

$\phi i = \sum(S \subseteq N \setminus \{i\}) [|S|!(|N|-|S|-1)!]/|N|! [f(S \cup \{i\}) - f(S)]$

- ϕi = SHAP value of feature i,
- N = set of all features,
- $S \subseteq N \setminus \{i\}$ = subset of all other features,
- f(S) = model prediction using only features in subset S, and
- $f(S \cup \{i\})$ = prediction with feature i added to the subset.

This approach provided an individual-level explanation of model behaviour, highlighting how specific features influenced the SVM's predicted future mortality status. For consistent comparison across features, the corresponding input feature values were individually scaled to have mean zero and unit variance. This normalization removed differences in feature range, allowing SHAP values to be visualized with a colour gradient representing standardized feature magnitude, thereby clarifying the directional influence of each feature on model predictions.

Survival analysis

To test the association between our optimally identified biomarkers and CKD related mortality, Cox proportional hazards models were constructed using $survival$ (v3.8.3) and $surv \min er$ (v0.5.1). Univariate models were created using only the expressional values of each biomarker (Biomarker ~ Mortality), then adjusted models were created which accounted for confounding clinical variables (Biomarker ~ Age + Sex + CVD + Diabetes + CKD-EPI eGFR). Kaplan-Meier curves were then generated to visualise the relationship between biomarker expression and survival outcomes. Participants were assigned to either High or Low expressional groupings and then visualized to see the direct association between biomarker expression and survival outcomes. Expression groupings were created wherein High expression > median expression and Low expression = < median expression. Participants who survived beyond the 9-year follow-up period were treated as censored at the end of follow-up, which are indicated by tick marks on the Kaplan-Meier curves.

Proteomic Aging Clocks

Estimation of CKD patient biological age was carried out using a previously published proteomic aging clock (PAC), which was built using UK Biobank data with penalised Cox regression and Gompertz survival models, converting a mortality-associated risk score into an estimate of age[33]. We applied the same protein coefficients and conversion formula from the published manuscript directly to our own

patient proteomics to generate proteomic age (ProtAge) estimates. Proteomic age deviation was calculated for each patient as ProtAge minus chronological age. Proteomic age acceleration (PAA) was additionally calculated as the residual from a linear regression of ProtAge on chronological age within our cohort, representing each patient's proteomic age relative to the age-adjusted cohort trend.

Correlations and directionality of mortality biomarkers

Identified biomarker concentration underwent correlation analysis to test individual relationships with baseline renal function (eGFR CKD-EPI) using Pearson's correlation test. This was carried out in two separate CKD cohorts, the UKBiobank cohort (n = 468) and a cross-border validation cohort (n = 155). The same mortality encoding present in the UKBiobank cohort was then applied to the cross-border validation cohort (n = 86), which was used to validate the directionality of expression for EDA2R, NT-pro-BNP, REN and CTSZ.

Mortality model comparison with alternative models

The optimal machine learning model was then compared to more clinically oriented models, such as a logistic regression model which used clinical features (Mortality ~ Age + Sex + eGFR + CVD status + Diabetes status) and another logistic regression model that utilised the Charlson Comorbidity Index (CCI) (Mortality ~ Age + CCI). Training and testing participants were consistent across all machine learning models to ensure fair comparisons. The full CCI was used according to ICD10 codes[53,54], retaining the original Charlson weights, with Age included as a covariate in the logistic regression, instead of an age-adjusted CCI framework. Only comorbidities with a recorded diagnosis date preceding blood sample collection were included.

Declarations

Author Contributions

All the authors warrant that this article is the authors' original work, has not received prior publication and is not under consideration for publication elsewhere. Thomas McLarnon wrote the manuscript under the supervision of Taranjit Singh Rai with scientific and clinical input from Frank McCarroll and Steven Watterson. Thomas McLarnon and Taranjit Singh Rai conceived the idea and designed the research. Taranjit Singh Rai secured funding. Thomas McLarnon performed the analysis with input from Donya Ghazinader under the supervision of Taranjit Singh Rai and Steven Watterson. Frank McCarroll, Laura Lennox and Jack Duncan acted as clinical leads.

Data Availability

UK Biobank data are available to approved researchers via application to the UK Biobank Access Management System (https://www.ukbiobank.ac.uk). This research was conducted under UK Biobank application number [104404]. Proteomic data were accessed as part of the UK Biobank Pharma Proteomics Project. Data from the Acute Kidney Injury (AKI) outcome study validation cohort are not

publicly available but may be available from the corresponding author upon reasonable request, subject to institutional data sharing agreements and ethical approval. The processed bulk RNA-seq count matrix data in the article are available at https://zenodo.org/records/10532131.

Code Availability

The underlying code for this study is not publicly available but may be made available to qualified researchers on reasonable request from the corresponding author.

Acknowledgements

We would like to sincerely thank all the study participants for their invaluable contribution. We would also like to express our sincere gratitude to both the Western Health Social Care Trust, Northern Ireland and Letterkenny University hospitals (Northwest of Ireland) for facilitating access to patients and clinics. This research was mostly funded by grants to T.S.R from Poscidon and Northern Ireland Kidney Research Fund (NIKRF). We thank Drs Sean McCallion and Ying Kuan for the collection and creation of the in-house validation dataset. We also thank Professor Anthony J Bjourson for securing original grant that started the kidney research program at Ulster University under the EU Regional Development Fund EU Sustainable Competitiveness Programme for NI & the NI Public Health Agency.

Ethic Declarations

UK Biobank received ethical approval from the Northwest Multi-Centre Research Ethics Committee (REC reference: 21/NW/0157), and all participants provided written informed consent at recruitment. This study was conducted under UK Biobank application number [104404]. The Acute Kidney Injury (AKI) outcome study was approved by Greater Manchester West Research Ethics Committee and then NHS GRAMPIAN, North of Scotland Research Ethics Service; REC ref 18/NW/0348 and 18/NS/0067. and all participants provided written informed consent prior to enrolment. All research was conducted in accordance with the Declaration of Helsinki.

Clinical trial number: not applicable

Competing Interests

The authors declare no competing interests.

References

1. Mark, P. B. *et al.* Global, regional, and national burden of chronic kidney disease in adults, 1990–2023, and its attributable risk factors: a systematic analysis for the Global Burden of Disease Study 2023. *The Lancet* 406, 2461–2482 (2025).
2. 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. *The Lancet* 395, 709–733 (2020).

3. Zhang, L. *et al.* Epidemiological research on diabetic nephropathy at global, regional, and national levels from 1990 to 2021: an analysis derived from the global burden of disease 2021 study. *Front. Endocrinol. (Lausanne)*. 16, (2025).
4. Stevens, P. E. *et al.* KDIGO 2024 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. *Kidney Int.* 105, S117–S314 (2024).
5. Chen, T. K., Hoenig, M. P., Nitsch, D. & Grams, M. E. Advances in the management of chronic kidney disease. *BMJ* e074216 (2023) doi:10.1136/bmj-2022-074216.
6. Hageman, S. *et al.* SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. *Eur. Heart J.* 42, 2439–2454 (2021).
7. Celli, B. R. *et al.* The Body-Mass Index, Airflow Obstruction, Dyspnea, and Exercise Capacity Index in Chronic Obstructive Pulmonary Disease. *New England Journal of Medicine* 350, 1005–1012 (2004).
8. Kamath, P. S. *et al.* A Model to Predict Survival in Patients With End–Stage Liver Disease. *Hepatology* 33, 464–470 (2001).
9. Tan, H., Xu, J. & Liu, Y. Ageing, cellular senescence and chronic kidney disease: experimental evidence. *Curr. Opin. Nephrol. Hypertens.* 31, 235–243 (2022).
10. McLarnon, T. *et al.* Sendotypes predict worsening renal function in chronic kidney disease patients. *Clin. Transl. Med.* 15, (2025).
11. Schroth, J., Thiemermann, C. & Henson, S. M. Senescence and the Aging Immune System as Major Drivers of Chronic Kidney Disease. *Front. Cell Dev. Biol.* 8, 564461 (2020).
12. McHugh, D. & Gil, J. Senescence and aging: Causes, consequences, and therapeutic avenues. *Journal of Cell Biology* 217, 65–77 (2018).
13. St Sauver, J. L. *et al.* Biomarkers of cellular senescence and risk of death in humans. *Aging Cell* 22, e14006 (2023).
14. Cawthon, R. M., Smith, K. R., O'Brien, E., Sivatchenko, A. & Kerber, R. A. Association between telomere length in blood and mortality in people aged 60 years or older. *The Lancet* 361, 393–395 (2003).
15. Muñoz-Espín, D. *et al.* Programmed Cell Senescence during Mammalian Embryonic Development. *Cell* 155, 1104–1118 (2013).
16. Serrano, M., Lin, A. W., McCurrach, M. E., Beach, D. & Lowe, S. W. Oncogenic ras Provokes Premature Cell Senescence Associated with Accumulation of p53 and p16INK4a. *Cell* 88, 593–602 (1997).
17. Jun, J.-I. & Lau, L. F. The matricellular protein CCN1 induces fibroblast senescence and restricts fibrosis in cutaneous wound healing. *Nat. Cell Biol.* 12, 676–685 (2010).
18. Hayflick, L. & Moorhead, P. S. The serial cultivation of human diploid cell strains. *Exp. Cell Res.* 25, 585–621 (1961).
19. Coppé, J.-P. *et al.* Senescence-Associated Secretory Phenotypes Reveal Cell-Nonautonomous Functions of Oncogenic RAS and the p53 Tumor Suppressor. *PLoS Biol.* 6, e301 (2008).

20. Hu, C., Zhang, X., Teng, T., Ma, Z.-G. & Tang, Q.-Z. Cellular Senescence in Cardiovascular Diseases: A Systematic Review. *Aging Dis.* 13, 103–128 (2022).
21. Narasimhan, A., Flores, R. R., Robbins, P. D. & Niedernhofer, L. J. Role of Cellular Senescence in Type II Diabetes. *Endocrinology* 162, (2021).
22. Barnes, P. J. Senescence in COPD and Its Comorbidities. *Annu. Rev. Physiol.* 79, 517–539 (2017).
23. Lee, S. *et al.* Virus-induced senescence is a driver and therapeutic target in COVID-19. *Nature* 599, 283–289 (2021).
24. Schmitt, C. A. *et al.* COVID-19 and cellular senescence. *Nat. Rev. Immunol.* 23, 251–263 (2023).
25. Docherty, M. H., Baird, D. P., Hughes, J. & Ferenbach, D. A. Cellular Senescence and Senotherapies in the Kidney: Current Evidence and Future Directions. *Front. Pharmacol.* 11, (2020).
26. Melk, A., Schmidt, B. M. W., Vongwiwatana, A., Rayner, D. C. & Halloran, P. F. Increased Expression of Senescence-Associated Cell Cycle Inhibitor p16 in Deteriorating Renal Transplants and Diseased Native Kidney. *American Journal of Transplantation* 5, 1375–1382 (2005).
27. Sis, B. *et al.* Accelerated expression of senescence associated cell cycle inhibitor p16INK4A in kidneys with glomerular disease. *Kidney Int.* 71, 218–226 (2007).
28. Liu, J. *et al.* Accelerated senescence of renal tubular epithelial cells is associated with disease progression of patients with immunoglobulin A (IgA) nephropathy. *Translational Research* 159, 454–463 (2012).
29. Chen, K. *et al.* Optineurin-mediated mitophagy protects renal tubular epithelial cells against accelerated senescence in diabetic nephropathy. *Cell Death Dis.* 9, 105 (2018).
30. Chen, J. *et al.* Cellular senescence of renal tubular epithelial cells in acute kidney injury. *Cell Death Discov.* 10, 62 (2024).
31. Li, C., Shen, Y., Huang, L., Liu, C. & Wang, J. Senolytic therapy ameliorates renal fibrosis postacute kidney injury by alleviating renal senescence. *The FASEB Journal* 35, (2021).
32. Carrasco-Zanini, J. *et al.* Proteomic signatures improve risk prediction for common and rare diseases. *Nat. Med.* 30, 2489–2498 (2024).
33. Kuo, C. *et al.* Proteomic aging clock (< scp>PAC</scp>) predicts age-related outcomes in middle-aged and older adults. *Aging Cell* 23, (2024).
34. Argentieri, M. A. *et al.* Proteomic aging clock predicts mortality and risk of common age-related diseases in diverse populations. *Nat. Med.* 30, 2450–2460 (2024).
35. Zhang, L., Pitcher, L. E., Prahalad, V., Niedernhofer, L. J. & Robbins, P. D. Targeting cellular senescence with senotherapeutics: senolytics and senomorphics. *FEBS J.* 290, 1362–1383 (2023).
36. Chen, S. *et al.* Multi-omics profiling reveals systemic rejuvenation of the aged kidney through senolytic therapy. *NPJ Regen. Med.* 11, 31 (2026).
37. Shao, X. *et al.* Generation of a conditional cellular senescence model using proximal tubule cells and fibroblasts from human kidneys. *Cell Death Discov.* 10, 364 (2024).

38. Bateman, A. *et al.* UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* 51, D523–D531 (2023).
39. Farrington, G. *et al.* The Roles of EDA2R in Ageing and Disease. *Aging Cell* 24, (2025).
40. Barbera, M. C. *et al.* Increased ectodysplasin-A2-receptor EDA2R is a ubiquitous hallmark of aging and mediates parainflammatory responses. *Nat. Commun.* 16, 1898 (2025).
41. Perez, K. *et al.* Single nuclei profiling identifies cell specific markers of skeletal muscle aging, frailty, and senescence. *Aging* https://doi.org/10.18632/aging.204435 (2022) doi:10.18632/aging.204435.
42. Campden, R. I. *et al.* Extracellular cathepsin Z signals through the α5 integrin and augments NLRP3 inflammasome activation. *J. Biol. Chem.* 298, 101459 (2022).
43. Salam, R. *et al.* Cellular senescence in malignant cells promotes tumor progression in mouse and patient Glioblastoma. *Nat. Commun.* 14, 441 (2023).
44. Fox, C. *et al.* Inhibition of lysosomal protease cathepsin D reduces renal fibrosis in murine chronic kidney disease. *Sci. Rep.* 6, 20101 (2016).
45. Alshahrani, S. Renin-angiotensin-aldosterone pathway modulators in chronic kidney disease: A comparative review. *Front. Pharmacol.* 14, 1101068 (2023).
46. Poznyak, A. V *et al.* Renin-Angiotensin System in Pathogenesis of Atherosclerosis and Treatment of CVD. *Int. J. Mol. Sci.* 22, (2021).
47. Parikh, N. I. *et al.* Plasma renin and risk of cardiovascular disease and mortality: the Framingham Heart Study. *Eur. Heart J.* 28, 2644–2652 (2007).
48. Sasaki, T. *et al.* N-Terminal Pro–B-Type Natriuretic Peptide and Incident CKD. *Kidney Int. Rep.* 6, 976–985 (2021).
49. Lan, X. *et al.* EDA2R mediates podocyte injury in high glucose milieu. *Biochimie* 174, 74–83 (2020).
50. Parving, H.-H., Persson, F., Lewis, J. B., Lewis, E. J. & Hollenberg, N. K. Aliskiren Combined with Losartan in Type 2 Diabetes and Nephropathy. *New England Journal of Medicine* 358, 2433–2446 (2008).
51. Holditch, S. J. *et al.* B-Type Natriuretic Peptide Deletion Leads to Progressive Hypertension, Associated Organ Damage, and Reduced Survival. *Hypertension* 66, 199–210 (2015).
52. McCallion, S. *et al.* Senescence Biomarkers CKAP4 and PTX3 Stratify Severe Kidney Disease Patients. *Cells* 13, 1613 (2024).
53. Charlson, M. E., Pompei, P., Ales, K. L. & MacKenzie, C. R. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *J. Chronic Dis.* 40, 373–383 (1987).
54. Quan, H. *et al.* Coding Algorithms for Defining Comorbidities in ICD-9-CM and ICD-10 Administrative Data. *Med. Care* 43, 1130–1139 (2005).

Figures

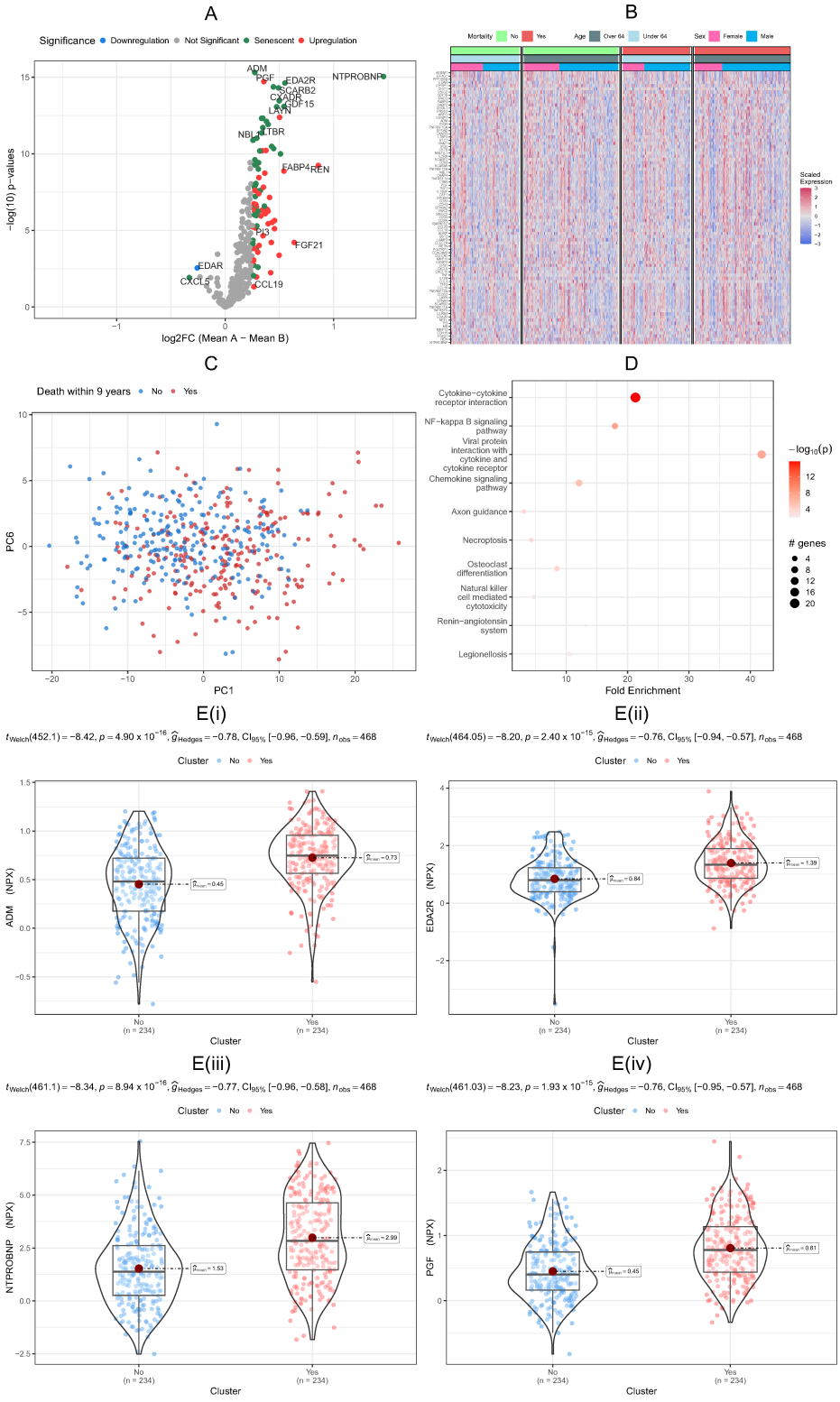


Figure 1

Senescence biomarkers and pathways are upregulated in CKD patients who subsequently experience disease-related mortality

(A) Volcano plot of differential protein expression between CKD patients who experienced disease-related mortality within 9 years and those who did not (Welch's two-tailed t-test). Points are colour-coded

by significance and senescence status: blue, significantly downregulated; grey, not significant; green, senescence-associated (protein's corresponding gene transcript was differentially expressed in senescent proximal tubule or fibroblast cell models); red, significantly upregulated. The x-axis shows log2 fold-change (Mean mortality group − Mean survivor group); the y-axis shows −log10(p-value).

(B) Heatmap of scaled protein expression (z-score) across all CKD patients (columns), annotated by 9-year mortality status (green, No; red, Yes), chronological age category (grey, Over 64, mean age of cohort; light blue, Under 64, mean age of cohort), and sex (pink, Female; blue, Male). Rows represent individual proteins. Clinical annotation tracks confirm that clustering by protein expression is not driven by age or sex, ruling them out as potential confounding factors.

(C) Principal component analysis (PCA) of the full proteomic landscape (372 proteins), showing the two components most associated with mortality status (PC1 and PC6). Each point represents a patient, coloured by 9-year mortality status (blue, No; red, Yes), demonstrating partial separation of patients by proteomic profile.

(D) Pathway enrichment analysis (pathfindR) of proteins with FDR < 0.05, showing the top enriched pathways. Fold enrichment is plotted on the x-axis; dot size represents the number of genes contributing to each pathway; colour represents −log10(p-value). The most enriched pathways include cytokine–cytokine receptor interaction, NF-κB signalling, viral protein interaction with cytokine/cytokine receptor, and chemokine signalling, consistent with a senescence-associated secretory phenotype (SASP).

(E) Violin-boxplots showing the expression distribution (NPX) of the four most statistically significant senescence-associated biomarkers between patients who did (Yes, n = 234) and did not (No, n = 234) experience 9-year disease-related mortality: **(i)** ADM $p = 4.90 \times 10^{-16}$, **(ii)** EDA2R $p = 2.40 \times 10^{-15}$, **(iii)** NT-pro-BNP $p = 8.94 \times 10^{-16}$ **(iv)** PGF $p = 1.93 \times 10^{-15}$. Dark red points indicate group means; boxplots show median and interquartile range.

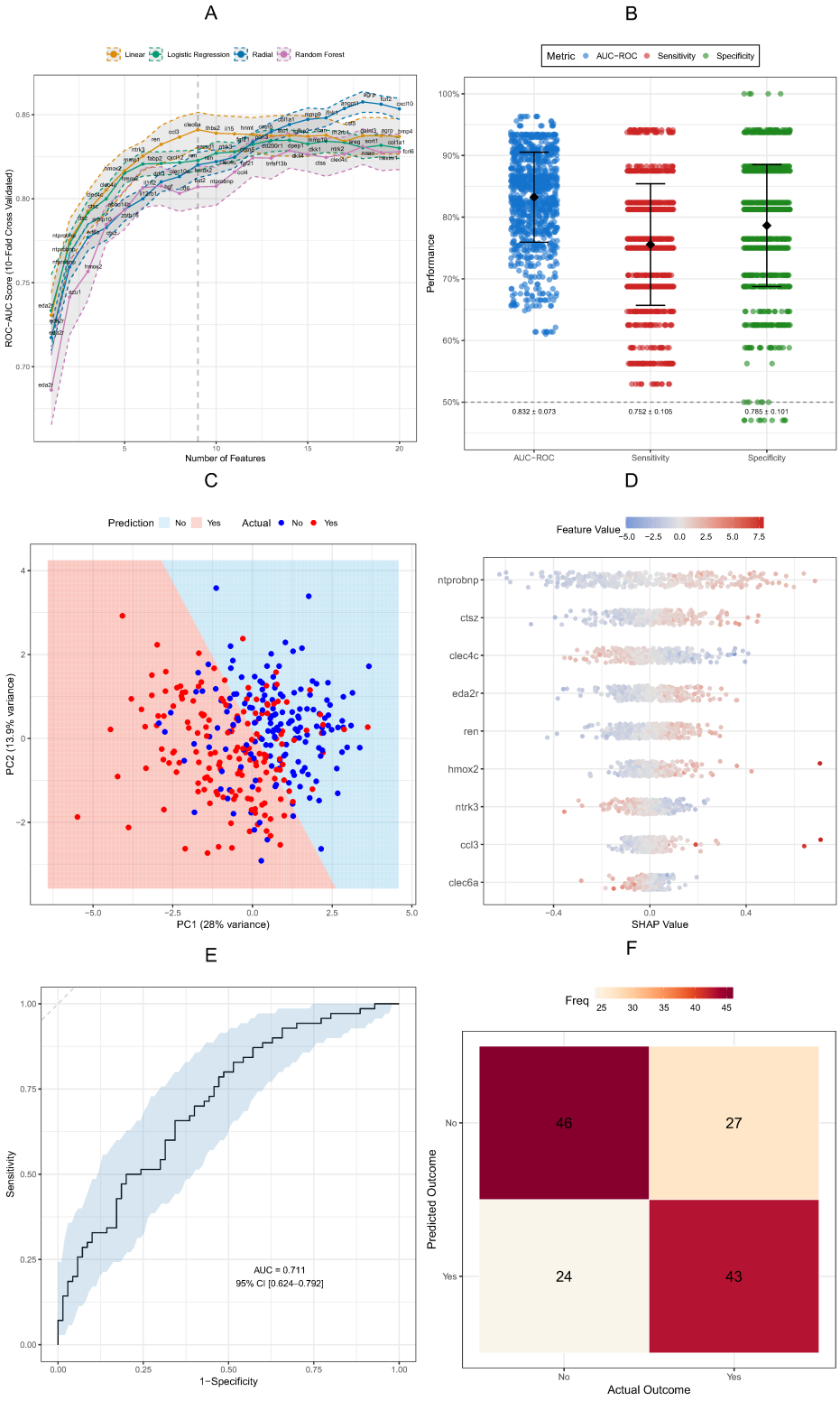


Figure 2

A senescence-associated proteomic signature predicts 9-year disease-related mortality in CKD

(A) Forward feature selection results across four machine learning algorithms, including linear SVM (orange), logistic regression (green), radial basis function SVM (blue), and random forest (pink), showing 10-fold cross-validated ROC-AUC for each number of features included. Individual protein names are

annotated at their point of inclusion. The dashed vertical line indicates the optimal 9-feature linear SVM model (EDA2R, NT-pro-BNP, CTSZ, CLEC4C, HMOX2, NTRK3, REN, CCL3, CLEC6A).

(B) Performance of the final linear SVM model across repeated 10-fold cross-validation (10 external repeats), showing individual fold-level values for AUC-ROC (blue), sensitivity (red), and specificity (green). Black diamonds and error bars represent mean ± SD (AUC-ROC: 0.832 ± 0.073; sensitivity: 0.752 ± 0.105; specificity: 0.785 ± 0.101). Dashed line indicates chance-level performance of 50%.

(C)SVM decision boundary visualised in principal component space (PC1 vs PC2, training data). Shaded regions denote model predictions: light blue, f(x) < 0 (predicted non-mortality); light red, f(x) > 0 (predicted mortality). Points represent individual patients coloured by true/actual 9-year mortality status (blue, No; red, Yes), illustrating separation of classes consistent with genuine biological signal rather than overfitting.

(D)SHAP (Shapley Additive Explanations) summary plot for the 9 features of the final SVM model, ranked by overall importance. Each point represents an individual patient; position on the x-axis indicates the SHAP value (impact on model output) for that feature in that patient; colour represents the patient's standardised feature expression (blue, low; red, high).

(E) Receiver operating characteristic (ROC) curve for the final SVM model evaluated on the held-out test cohort (AUC = 0.711, 95% CI [0.624–0.792]). Shaded band represents the 95% confidence interval generated via stratified bootstrap resampling (n = 5000).

(F) Confusion matrix of model predictions on the held-out test cohort, showing 46 true negatives, 43 true positives, 24 false positives, and 27 false negatives. Colour intensity corresponds to prediction frequency.

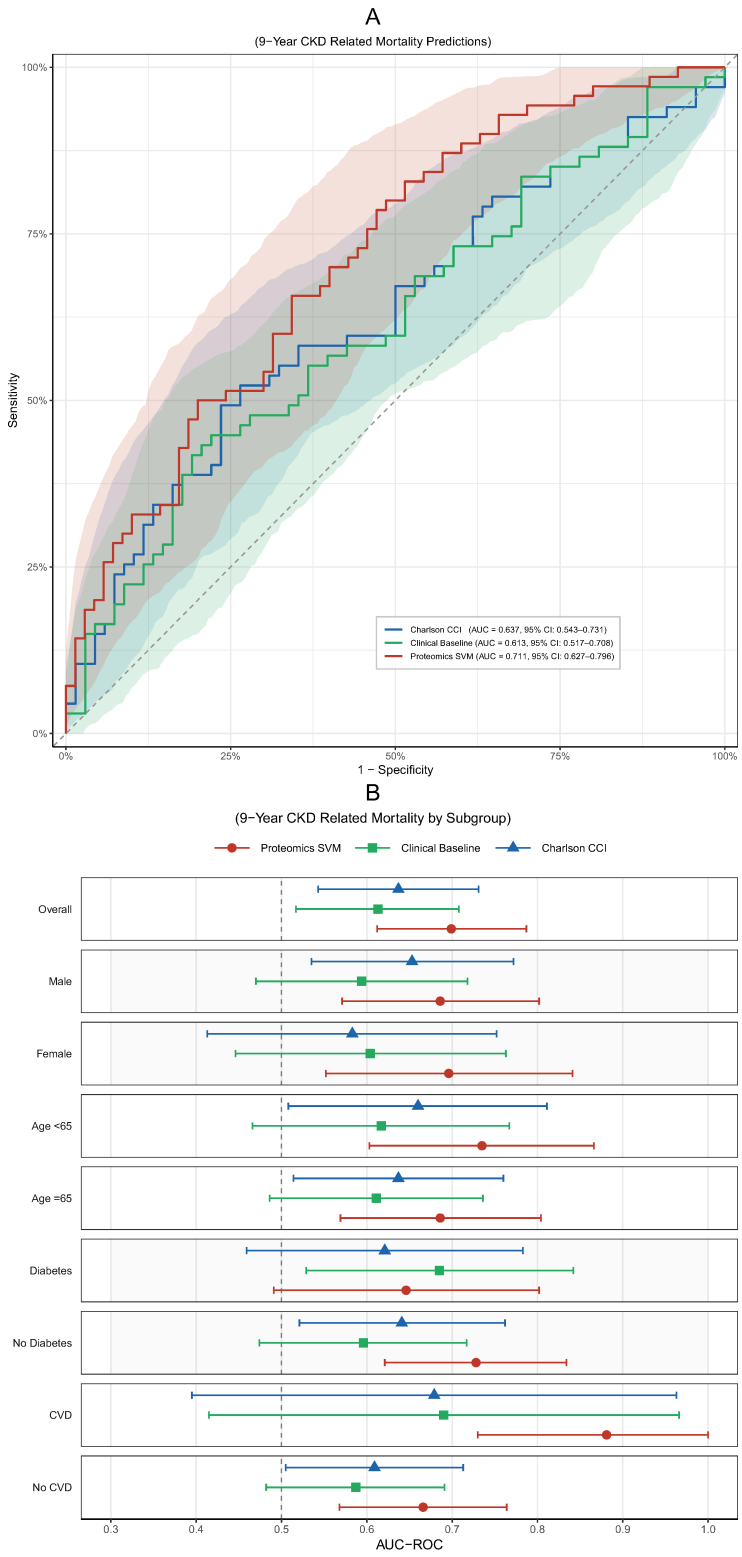


Figure 3

Senescence proteomics improve CKD mortality prediction beyond established clinical models

(A)ROC curves comparing the discriminative performance of the proteomics SVM (red; AUC = 0.711) against a clinical baseline model (age, sex, eGFR, CVD status, diabetes status; (green; AUC = 0.613) and the Charlson Comorbidity Index model (blue; AUC = 0.637) for prediction of 9-year CKD-related mortality.

(B) Subgroup analysis of model performance (AUC-ROC) across the overall held-out cohort and clinically relevant subgroups (sex, age <65 vs ≥65, diabetes status, CVD status). Points are coloured and shaped by model: red circles, proteomics SVM; green squares, clinical baseline; blue triangles, Charlson CCI. The proteomics SVM outperforms both comparator models across the overall cohort and in most subgroups, with the exception of the diabetes subgroup, where the Charlson CCI performs marginally better.

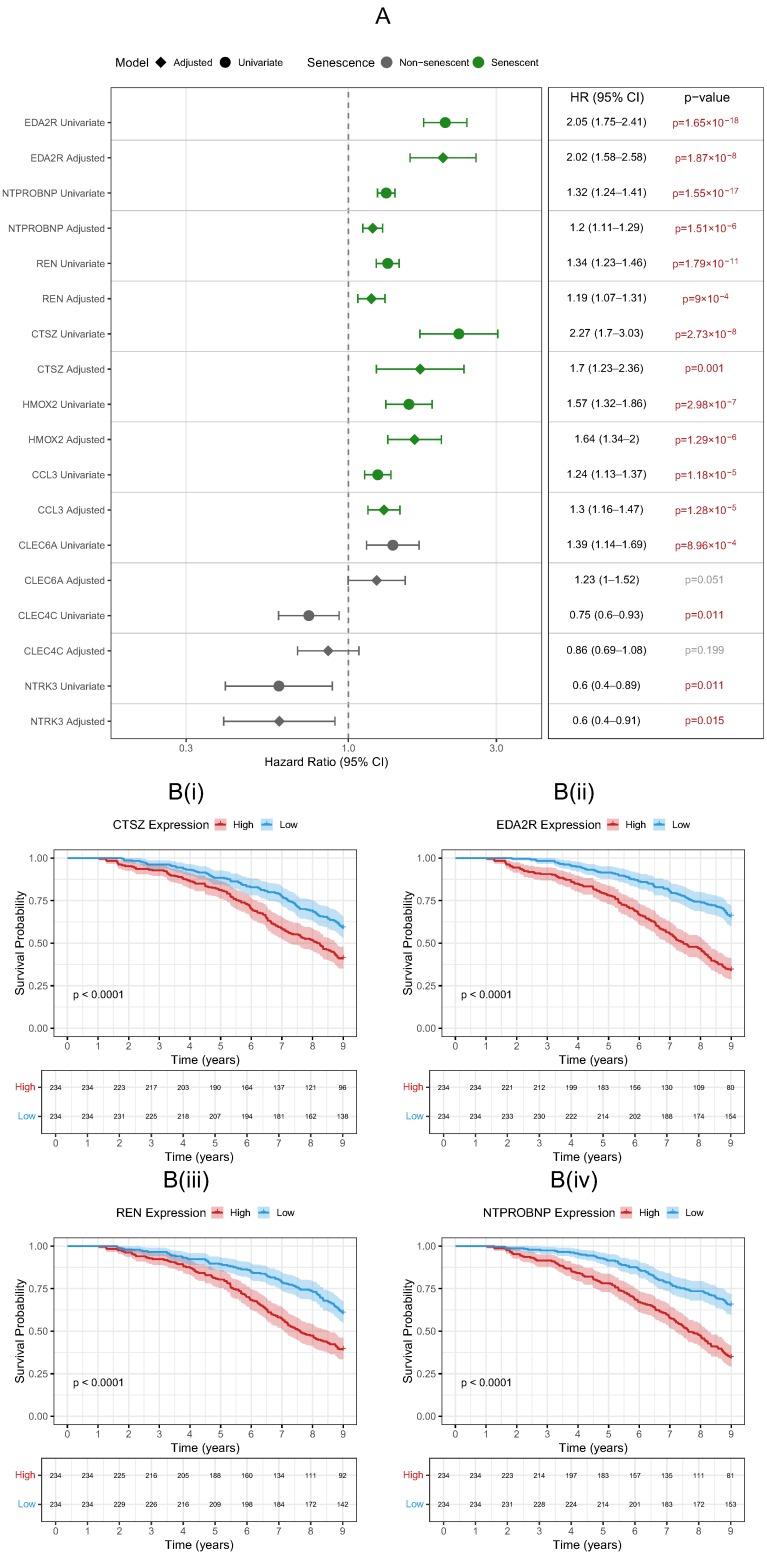


Figure 4

Senescence biomarker upregulation independently predicts poor survivability

(A) Forest plot of hazard ratios (HR, 95% CI) from Cox proportional hazards models for each of the 9 proteins in the final SVM panel, shown as both univariate (circle) and clinically adjusted (diamond; adjusted for age, sex, CVD presence, diabetes presence, and eGFR) models. Point colour indicates senescence association (green, senescence-associated; grey, not senescence-associated). EDA2R (HR 2.05, adjusted HR 2.02), CTSZ (HR 2.27, adjusted HR 1.7), REN (HR 1.34, adjusted HR 1.19), and NT-pro-BNP (HR 1.32, adjusted HR 1.2) were the most significant senescence-associated predictors (all $p < 0.001$, adjusted).

(B) Kaplan-Meier survival curves for 9-year disease-related mortality, stratified by median expression (High, red; Low, blue) of **(i)** CTSZ, **(ii)**EDA2R, **(iii)** REN, and **(iv)** NT-pro-BNP (all log-rank $p < 0.0001$). Tick marks indicate patients censored at the end of the 9-year follow-up. Numbers-at-risk tables are shown below each panel.

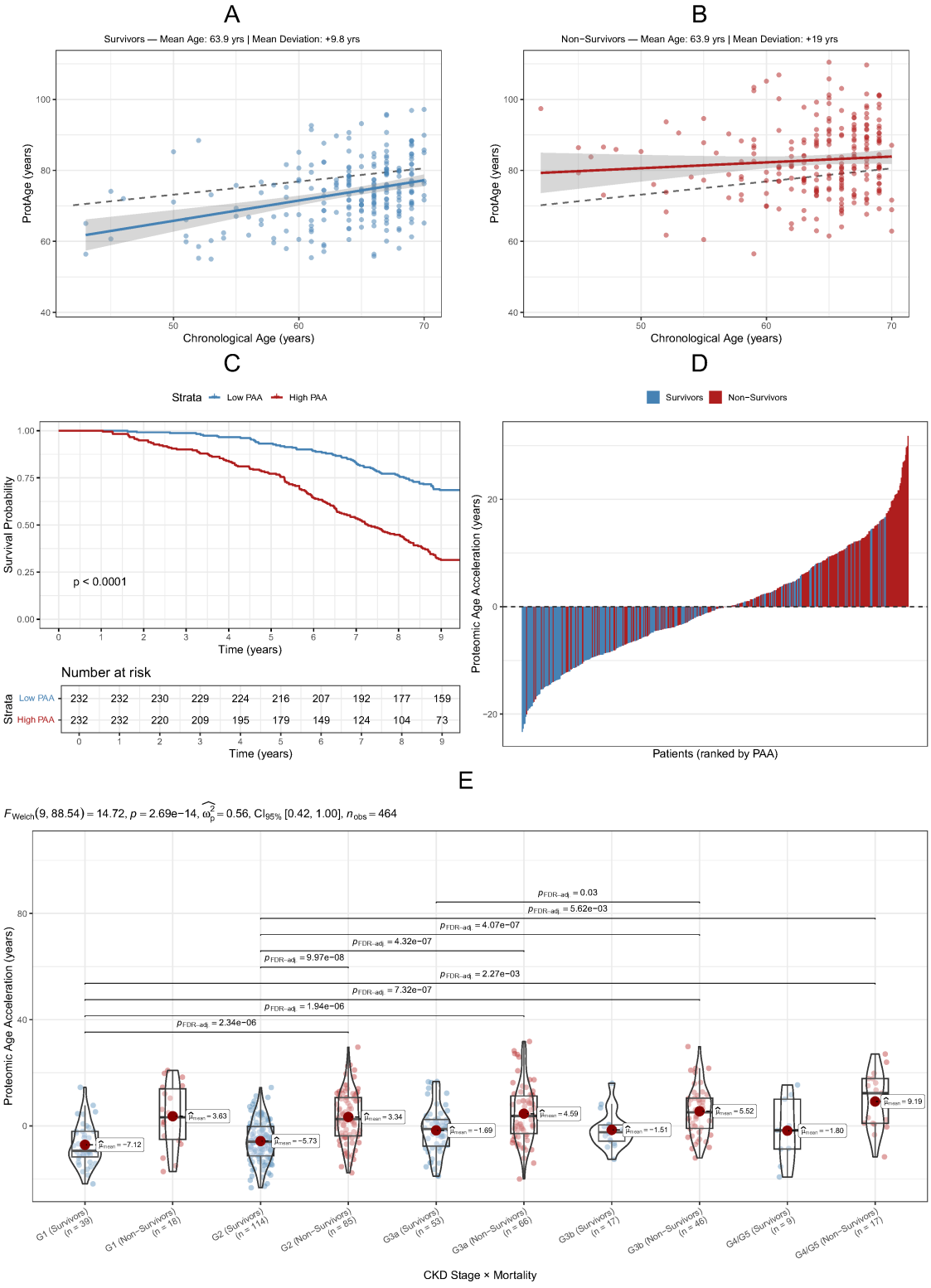


Figure 5

Proteomic age acceleration is elevated in non-survivors, independent of CKD stage

(A) Scatter plot of ProtAge (predicted biological age) against chronological age (years) for survivor patients (n = 232), with linear fit (blue) and 95% confidence band. Mean chronological age and mean

proteomic age deviation (ProtAge − chronological age) for this group are shown in the subtitle (Mean Age: 63.9 yrs; Mean Deviation: +9.8 yrs).

(B) As in (A), for non-survivor patients (n = 232) (red fit line). Mean chronological age was equivalent to the survivor group (63.9 yrs), while mean proteomic age deviation was approximately twice as large (+19 yrs), indicating greater biological age acceleration despite matched chronological age.

(C) Kaplan-Meier survival curves for 9-year disease-related mortality, stratified by median split of proteomic age acceleration (PAA; High PAA, red; Low PAA, blue) (log-rank $p < 0.0001$). Numbers-at-risk are shown below the plot.

(D) Waterfall plot of individual patient PAA values (years), ranked in ascending order, coloured by 9-year mortality status (blue, Survivors; red, Non-Survivors). Non-survivors are disproportionately represented at higher PAA values.

(E) Distribution of PAA across CKD disease stage (G1–G4/G5, by eGFR category) and mortality status, shown as paired boxplots (Survivors vs Non-Survivors within each stage). Pairwise FDR-adjusted p-values are shown for each stage comparison (all $p < 0.05$, ranging from $p = 2.34 \times 10^{-6}$ to $p = 0.03$ [G4/G5]). Overall group difference: $p = 2.69 \times 10^{-14}$. Non-survivors show significantly higher PAA than survivors at every CKD stage, indicating that disease severity does not confound the association between PAA and mortality.

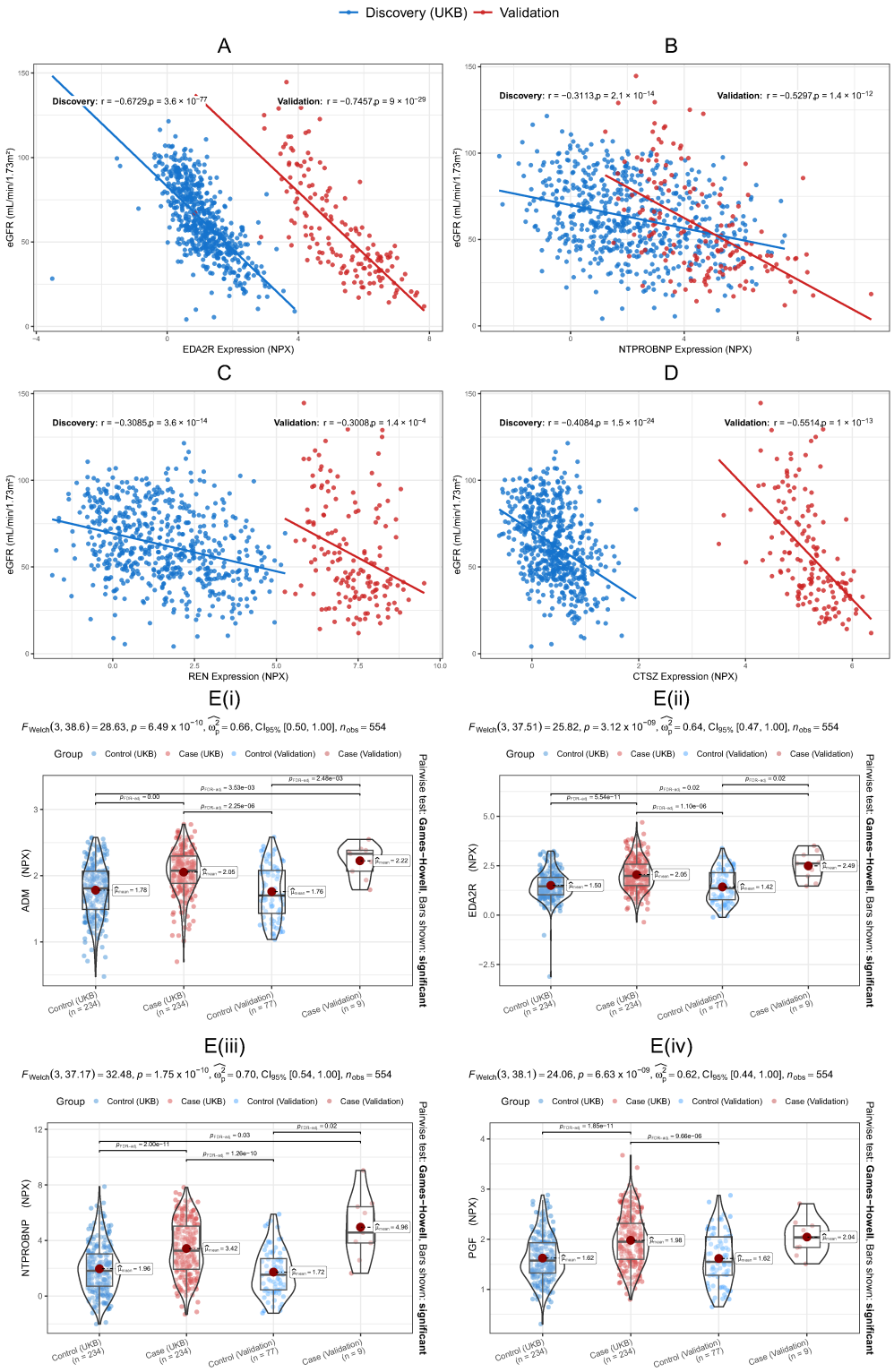


Figure 6

Senescence-associated biomarkers correlate with renal function and are validated in an independent cohort

(A–D) Pearson correlation between baseline biomarker expression (NPX) and estimated glomerular filtration rate (eGFR, CKD-EPI, mL/min/1.73m²) for **(A)** EDA2R, **(B)** NT-pro-BNP, **(C)** REN, and **(D)** CTSZ.

Points and regression lines are coloured by cohort: blue, UK Biobank discovery cohort; red, independent validation (AKI outcome study) cohort. Correlation coefficients and p-values are shown for each cohort. EDA2R showed the strongest and most consistent inverse correlation with eGFR across both cohorts (Discovery: $r = -0.673$, $p = 3.6 \times 10^{-77}$; Validation: $r = -0.746$, $p = 9 \times 10^{-29}$).

(E) Violin-boxplots validating the direction of differential expression for **(i)**ADM, **(ii)** EDA2R, **(iii)** NT-pro-BNP, and **(iv)** PGF across four groups: UK Biobank controls (survived beyond 9 years), UK Biobank cases (9-year mortality), validation cohort controls, and validation cohort cases. Overall group differences assessed via Welch's ANOVA (all $p < 6.5 \times 10^{-9}$). Direction of expression (elevated in cases relative to controls) was consistent across both cohorts for all four proteins, confirming the proteomic signal is not specific to a single dataset.

Supplementary Files

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

- SupplementaryFigure1Finalversion.pdf