**Nov. 11, 2022**

**Circulating proteins enriching TNF receptor signaling pathways and apoptosis processes predict risk of fast progression to end stage diabetic kidney disease; Results of global proteomics analysis.**

Katsuhito Ihara1,2,3, Eiichiro Satake1,2, Jonathan Wilson, Parker Wilson, Hiroki Kobayashi1,2, Zaipul I. Md Dom1,2, Bozena Krolewski1,2, Joseph Ricca1, Marcus Pezzolesi, Marlon Pragnell, Benjamin Humphreys, Kevin Duffin, and Andrzej S. Krolewski1,2\*

1 Research Division, Joslin Diabetes Center, Boston, MA, USA; 2 Department of Medicine, Harvard Medical School, Boston, MA, USA; 3 Department of Nephrology, Tokyo Medical and Dental University, Tokyo, Japan;

Corresponding author: Dr. Andrzej S. Krolewski, Section on Genetics and Epidemiology, Research Division, Joslin Diabetes Center, One Joslin Place, Boston, MA, 02215:

E-mail:[Andrzej.Krolewski@joslin.harvard.edu](mailto:Andrzej.Krolewski@joslin.harvard.edu)

**Running headline: Predictors of fast progression to ESKD in diabetes**

**Word count;** Abstract: 340 words (limited by 250 words)

Main text: XXXX words (limited by 4,000 words without reference, tables and figures)

Reference: 102 (no limit)

Figures and tables: 5 Figures and 3 Tables (no limit)

**ABSTRACT (283, too long and will be changed to include snRNA sequencing data and images of immunostaining of kidney biopsies)**

Two case-control studies, one in T1D (discovery) (n=195) and another in T2D (replication) (n=206), were conducted using global proteomics analysis to identify circulating proteins and biological pathways associated with disease process(es) underlying fast progression to end stage diabetic kidney disease (ESKD). At baseline, all individuals had normal kidney function and albuminuria. During 7-15 years of follow-up, 149 individuals progressed to ESKD. Baseline plasma concentrations of 455 proteins were measured using proximity extension assay applied in OLINK proteomics platform. In total, elevated concentrations of 46 circulating proteins were associated with development of ESKD (p < 10-5) in T1D and T2D studies, independent from relevant clinical characteristics. In pathway analysis these proteins enriched four clusters of biological pathways: cluster #1 comprised TNF receptor signaling and apoptosis processes (enrichment score = 8.9), #2 comprised receptors and plasma membranes proteins (enrichment score 4.2), and #3 and #4 were redundant with cluster #1 (enrichment score 2.2 and 1.9 accordingly). Cluster #1, #3 and #4 were enriched with the same 20 proteins that comprised multiple TNF receptors/apoptotic proteins.

Using these TNF receptor/apoptosis proteins and logistic Lasso regression analysis 5-6 proteins were identified, allowing us to build five alternative and equally efficient ESKD risk scores to predict fast progression to ESKD. These scores had the best predictive performance. When the other 26 proteins were included in the same analyses, none significantly added to the predictive value of the ESKD risk model.

In conclusion, although many circulating proteins were individually associated with risk of fast progression to ESKD, only a combination of several proteins that enriched TNF receptor signaling and apoptosis pathways were sufficient to predict fast progression to ESKD.

**INTRODUCTION**

Diabetes is responsible for more than 40% of end-stage kidney disease (ESKD) cases in the U.S. (). The majority of individuals with diabetes who develop ESKD have fast kidney function decline (). Such decline is a unidirectional process commencing while individuals have normal kidney function and, in the majority, are progressing steadily (linearly) to ESKD (). While an individual’s annual rate of kidney function decline is usually constant, this rate of decline, estimated as glomerular filtration rate (GFR) slope, varies widely among individuals from -50 to -5.0 ml/min/year (). Kidney function decline from normal GFR at this annual rate results in progression to ESKD within 2 to 15 years of follow-up. Therefore, it is important to have tools to identify these individuals early when they have still normal kidney function. Treatment with reno-protective drugs at such an early stage may prevent or significantly delay progression to ESKD.

Over the last decade, ongoing intensive research has sought to find biomarkers to identify individuals at high risk of fast kidney function decline and fast progression to ESKD. Many studies used a so-called targeted approach and examined specific proteins as candidate biomarkers (). Previously, using such an approach and antibody-based ELISAs, we discovered three new circulating proteins, TNFR1, TNFR2 and KIM, that were shown to be good predictors of fast kidney function decline (). We confirmed our findings in multiple studies (). However, measurements of these proteins do not provide optimal prediction of risk of ESKD. Recently, the search for new biomarkers of diabetic kidney disease was accelerated by the development and application of high throughput proteomics platforms. These platforms enable the simultaneous quantification of hundreds to thousands of circulating proteins that can be examined as candidate biomarkers to predict risk of ESKD.

Using an aptamer-based SOMAscan proteomics platform, we performed a study among individuals with diabetes, in which we identified 43 circulating proteins associated with risk of ESKD (). Several other studies were conducted on individuals in the general population in which the SOMAscan platform was used. In those studies, multiple circulating proteins were found to be associated with annual changes in estimated kidney function during longitudinal observation. Many of the proteins were the same as in our findings. Unfortunately, the analyses were limited to examination of individual proteins in all these studies and we did not attempt to further explore the disease processes underlying fast kidney function decline.

The present global proteomic study aimed not only to find circulating proteins associated with fast progression to ESKD in diabetes, but also to carry out pathway analyses to identify biological pathways/disease processes that underlie fast kidney function decline. We used the OLINK Proseek Multiplex proteomics platform, which measures proteins by real-time qPCR using proximity extension assay (PEA) technology (16). This platform has better precision and is more specific than SOMAscan and also comprises different proteins, but represents many of the same pathways.

In our study, 456 proteins present on 5 OLINK panels were measured in plasma obtained at baseline examination from 405 individuals with Type 1 diabetes (T1D) and with Type 2 diabetes (T2D). These proteins were examined for association with fast progression to ESKD during 15-years of follow-up. In the earlier report, we used a targeted approach to analyze data for 35 TNF related proteins that were measured in individuals with T1D to examine their involvement in early kidney function decline (). In the present report, all 456 proteins were examined. We applied an unbiased/untargeted analysis so we could discover all strongly associated proteins with fast progression to ESKD and, at the same time, we could examine various biological pathways and disease processes that underlie fast progression to ESKD.

**METHODS**

**Study design**

Participants for this research were selected from among individuals who were enrolled into the Joslin Kidney Study, a prospective study that aims to investigate the determinants and describe the natural history of kidney function decline in T1D and T2D. To make this research more cost-effective, we conducted two nested case-control studies. This approach significantly reduced the number of individuals for whom proteomics measurements were performed.

**Joslin Kidney Study (JKS)**

The Joslin Clinic is a large referral center for treatment of individuals with diabetes. These individuals are mainly residents of Eastern Massachusetts and are referred to the Clinic because of diagnosis of T1D or T2D. The majority are referred early in the course of diabetes and remain under the care of the clinic for long time. The JKS is a longitudinal observation study that aims to investigate the determinants of, and to describe the natural history of kidney function decline in, T1D and T2D in the Joslin Clinic population. The Joslin Diabetes Center Committee on Human Studies approved the informed consent, recruitment and examination procedures for the JKS. Previous results from this study and protocols used were published ().

Residents of New England were eligible for enrollment into the JKS. Individuals who were on chronic dialysis, had renal transplant, or had a history of HIV or hepatitis C infection were excluded. The JKS aimed to enroll all eligible participants with albuminuria and a similar number of individuals taken randomly from a much larger pool of eligible participants with normoalbuminuria. Albuminuria categories were determined according to the median values in urinary albuminuria to creatinine ratio (ACR) determined from 2 or more consecutive urine samples obtained during the 2-year period preceding rolling enrollment (baseline). Three categories of ACR were considered: Macro-Albuminuria (ACR ≥300 mg/g), Micro-Albuminuria (30≤ ACR <300 mg/g), and Normo-Albuminuria (ACR <30 mg/g) (). Briefly, 1,884 participants with T1D were recruited from among 3,500 adults aged 20-64 years who attended the Joslin Clinic between 1991 and 2009. Similarly, 1,474 participants with T2D were recruited from among 9,000 individuals aged 35-64 years who attended the Joslin Clinic between 2003 and 2009.

All enrolled individuals were followed until 2015, unless they developed earlier ESKD, died due to unrelated-ESKD deaths or were lost to follow-up. All enrolled individuals had baseline examination and biannual examinations afterward with blood and urine specimens taken for laboratory determinations and storage in -85°C. Individuals with less frequent clinic visits and those who stopped coming to the clinic were examined at their homes. The goal of follow-up was to ascertain eGFR slope and date of onset of ESKD.

Measurements of serum creatinine performed at routine clinic visits or during special examinations were used to determine kidney function at baseline and its changes during follow-up. Protocols to calibrate serum creatinine measurements over time were described previously (). The Chronic Kidney Disease Epidemiology Collaboration formula was used to estimate eGFR (). The primary outcome was onset of ESKD within 15 years of follow-up. To identify ESKD and deaths in the individuals participating in the JKS, we queried the United States Renal Data System (USRDS) and the National Death Index (NDI) covering all events up to the end of 2015. The USRDS maintains a roster of US patients receiving renal replacement therapy that includes dates of dialysis and transplantation. The NDI is a comprehensive roster of deaths in the United States that includes the date and cause of death.

**Selection of individuals for current studies:**

The current research consists of two case-control studies nested in the JKS, one included individuals with T1D and Macro-Albuminuria (T1D study) and the other included individuals with T2D and Albuminuria (Macro & Micro) (T2D study). The selection of individuals for the T1D study was described previously (). Briefly, out of 526 individuals with Macro-albuminuria enrolled into the JKS, we selected individuals (n=103) who had baseline eGFR >45 ml/min and progressed to ESKD during 15 years of follow-up as cases. As controls for these cases, we randomly selected 94 individuals who had Macro-albuminuria and eGFR >45 ml/min at baseline but did not progress to ESKD during follow-up. In T2D individuals enrolled into the JKS, there were 743 individuals with albuminuria (143 with Macro & 600 with Micro) (). From those individuals, we selected those (n=43) who had had baseline eGFR >45 ml/min and progressed to ESKD during 15 years of follow-up as cases. As controls, we randomly selected 163 individuals with baseline Albuminuria and eGFR >45 ml/min but did not progress to ESKD during follow-up. All relevant clinical and research data as well as baseline plasma specimens archived in -85 C were available for this research.

**Proteomics platforms**

The OLINK Proseek Multiplex panels® (Uppsala, Sweden) were used to measure proteins in archived plasma samples using real-time qPCR through PEA technology (). In total, the OLINK (Discovery) platform measures 1,061 proteins, and these proteins are organized into 13 panels. In this research, we measured 455 proteins contained on 5 panels: Cardiovascular II, Cardiovascular III, Development, Neurology, and Oncology II. The measurements were performed at the OLINK laboratory (Uppsala Sweden). The measurements were presented as relative values on a log2-scale. Quality control (QC) was performed in 2 steps: 1) each sample plate was evaluated on the standard deviation of the internal controls, and 2) the quality of each sample was assessed by evaluating the deviation from median value of the internal controls. The proportions of samples passing QC were 92-100% and 96-100% in T1D and T2D cohorts, respectively. Average intra-assay % coefficients of variation in T1D and T2D cohorts were 4-21% and 11-18%, respectively. The results for 25 TNF related proteins in T1D were previously reported ().

**Clinical Covariates**

**FOR ASK to prepare.**

**Pathway analyses**

**For Eiichro to write**

**mRNA sequencing of white blood cells**

**For Eiichiro to write, you may ask Markus for help**

**Single-nucleus RNA and ATAC sequencing**

Single cell RNA (snRNA-seq) and assay for transposase accessible chromatin (snATAC-seq) sequencing quantifies gene expression and chromatin accessibility in individual cells. These technologies can interrogate transcriptional and chromatin accessibility states and signaling pathways in multiple cell types in health and disease. We downloaded a previously-published single cell atlas of diabetic kidney disease (DKD) to compare cell-specific differentially expressed genes (DEGs) and differentially accessible chromatin regions (DAR) with our circulating biomarkers. The snRNA-seq and snATAC-seq libraries were generated from kidney cortex samples obtained from six control adults and seven with DKD. These individuals ranged in age from 50 to 78 years (median 57 years) and included seven men and six women. Mean eGFR of DKD donors (66 +/− 25 ml/min/1.73 m2) and control samples (74 +/− 15 ml/min/1.73 m2) was not statistically different. Two donors with DKD had mild to moderate proteinuria and moderate interstitial fibrosis and glomerulosclerosis identified on histology. Cell-specific DEG and DAR for *VCAM1*+ proximal tubule cells (PT\_VCAM1) compared to control proximal tubule (PT) were downloaded from supplemental data (DEG - Supplemental Data 7, DAR - Supplemental Data 4) and intersected with our list of 46 circulating biomarkers, or a list of hallmark apoptosis genes obtained from the msigdbr R package (v7.5.1). For the correlation between our list of 46 circulating biomarkers and apoptosis genes, we downloaded the aggregated snRNA-seq seurat object from a publicly-available website on cellxgene (<https://cellxgene.cziscience.com/collections/b3e2c6e3-9b05-4da9-8f42-da38a664b45b>) . The snRNA-seq RNA assay was log-normalized using the NormalizeData function in Seurat (v4.1.0) and hallmark apoptosis transcript counts were summed for the PT\_VCAM1 cell annotation. We computed a Pearson correlation between total apoptosis transcript counts (“Apoptosis Index”) and biomarker transcript counts using the rcorr function in the Hmisc R package (v4.7.0) and visualized results using ggplot2.

**Methods used by Jonathan Wilson:**

**Statistical analysis**

Baseline characteristics were presented as median and interquartile range or number and percent, as applicable. To correct for multiple testing, we applied the Bonferroni adjustment. Bonferroni-adjusted p values were calculated from 455 proteins in T1D and from the selected proteins in T2D, respectively. Bonferroni-adjusted p <0.01 in both studies was considered statistically significant for the investigations of proteins in T1D and T2D. To consider the inter-variation between batches, we applied batch-specific quartiles to proteins in each batch, and odds ratio (OR) was estimated as one quartile change in baseline concentration of proteins. Univariate and multivariable logistic regression models were used to estimate the effect of proteins on onset of ESKD within 15 years. Multivariable models were adjusted for baseline eGFR, urinary ACR, and HbA1c, and for cohort indicator (T1D or T2D) in analyses through the combined cohorts of T1D and T2D. Heterogeneity of the effects of proteins on onset of ESKD within 15 years between the cohorts were assessed by random effects model using *I²*,after adjusting for eGFR, ACR, and HbA1c in each cohort.

We examined how much degrees the proteins could improve prediction of onset of ESKD during 15 years of follow-up in the combined cohorts. As a preliminary assessment of variable importance, we applied random forests for 500 trees to classify the importance of proteins by batch-specific quartiles using ‘randomForest’ R package. Random forests are a combination of tree predictors. As with the regression analyses, the contributions of proteins to models were evaluated as one-quartile increase. Secondary, we applied the least absolute shrinkage and selection operator (LASSO) logistic regression for protein selection by batch-specific quartiles using ‘glmnet’ R package. To minimize the complexity of the model, this method penalized the sum of the absolute values of the regression coefficients leading to some coefficients shrinking to zero and thus simultaneously performed variable selection (). Lambda, a penalty factor for penalized maximum likelihood estimation, was chosen by 10-fold cross-validation at its minimum level. After selecting the proteins in the model through the LASSO logistic regression, we developed the risk score from the selected proteins using a logistic regression model of onset of ESKD within 15 years as an outcome variable. The risk scores were calculated as sum of levels of the contributing proteins with batch-specific quartiles weighted by their β coefficients, according to the formula:

Risk score = βA × A + βB × B + βC × C + …

When there were other proteins apart from the objects for LASSO logistic regression, we performed variable selections (forward selection, backward elimination, or stepwise selection) in the model with the risk score, in order to find other proteins contributing to prediction of onset of ESKD within 15 years. Model performance in inclusion of selected variables was evaluated in the following ways. Receiver operating characteristic (ROC) curves were constructed for all possible cutoffs for the models. The area under the ROC curve (AUC), or C-statistic, was estimated as a measure of the probability that a randomly selected individual who experienced an event had a higher predicted risk than an event-free person. Two alternative measures of risk discrimination were calculated. One is the category-free net reclassification index (NRI) (), which examines whether the predicted probabilities of individuals with and without events move in the right directions (upward and downward, respectively) from the old to the new model. The other is the relative integrated discrimination improvement (rIDI) (), which represents the increase in discrimination slopes provided by the new model relative to the old model. The analyses testing the improvement in predictive performance were computed using logistic regression models with methods developed by Kennedy and Pencina ().

To examine possible determinants of variation of the ESKD risk score, Spearman rank correlation was performed between clinical characteristics and ESKD risk score computed in individuals considered as controls in the T2D study is shown in Figure XA. Distribution of risk score in the combination of eGFR and ACR categories in the T2D controls is shown in Figure XB. Regarding the risk score as a categorical variable, we estimated the ORs for onset of ESKD within 15 years according to the ESKD risk score by its quartiles without adjustment and with adjustment for relevant covariates pattern 1 (eGFR, ACR, and HbA1c) and 2 (eGFR, ACR, HbA1c, Age, and Systolic blood pressure) among the all individuals in the combined cohorts. Cumulative incidence of ESKD according to the ESKD risk score by its quartiles was examined using time-to-event analysis. Two-sided p <0.05 were considered as statistical significance through the analyses. All analyses were performed by SAS 9.4 (SAS Institute Inc, Cary, NC, USA), Stata 15 (StataCorp LLC, TX, USA), and R version 4.0.3 (R Core Team 2020).

**RESULTS**

**Characteristics of study groups.**

This investigation comprises two nested case-control studies. Participants for these studies were selected from among individuals who were enrolled into the Joslin Kidney Study and were followed for up to 15 years to determine rate of kidney function decline assessed by estimated glomerular filtration rate (eGFR) slope and to ascertain the onset of ESKD. For the T1D Macro-albuminuria study, we selected 103 cases with ESKD and 93 randomly selected individuals who did not develop ESKD during follow-up (controls.) For the T2D Albuminuria (included Macro- and Micro-) study, we selected 46 cases with ESKD and 163 randomly selected non-cases without ESKD. Ninety-five percent of participants in the studies were Caucasian.

**Table 1** presents characteristics at baseline and during follow-up for cases and controls in each study. Individuals with T1D were younger and had longer duration of diabetes than those with T2D. At baseline, cases had higher HbA1c and urinary ACR and lower eGFR in comparison with controls in both studies. The latter values, however, were in the normal range in most cases. During follow-up, median eGFR loss was approximately 10 ml/min/1.73m2/year in cases in both studies. This indicates that cases with ESKD in T1D and T2D had similar fast kidney function decline. In both studies, there were only few deaths unrelated to ESKD.

**Table 1**

**Search for proteins associated with fast progression to ESKD**

Baseline plasma specimens from both studies were subjected to proteomics analysis using Olink Proseek platform. Concentration of 455 proteins on 5 Olink panels were measured. In the T1D Discovery study, baseline plasma concentrations of 50 proteins were associated with risk of fast progression to ESKD as indicated by significant OR (-log10p >3.6, nominal p value after Bonferroni correction). Distribution of ORs according to p values in this study are shown as a volcano plot in **Fig. 1.** In the T2D study, 43 proteins were confirmed (indicated as red dots in **Fig. 1**) and they are considered in further analyses. These proteins were supplemented by 3 proteins (TNF-R4, TNF-R6 and TNF-R21) that were found in our previous study as associated with ESKD, though they did not reach statistical significance in the current study due to more stringent criteria. In total, 46 circulating proteins were analyzed in the current studies, and they were referred to as ESKD associated proteins. The results for the 409 examined protein not included in the further analysis are provided in **Table S1**.

**Fig. 1 Volcano plot**

The names of the 46 ESKD associated proteins and Odds Ratio (OR) for each of them in each study are shown in **Table 2**. The proteins were grouped according to functional categories of proteins and ranked according to OR in the combined cohorts of T1D and T2D study. Although ORs for many of these proteins seem to be higher in T2D than in the T1D study, heterogeneity testing showed that only adjusted ORs for EPHB4 (*I2* = 84%, p = 0.013) showed evidence of heterogeneity of results across the studies (see legend to **Table 2**). Therefore, the results for T1D and T2D were combined. After adjustment for important clinical covariates such as baseline HbA1c, ACR and eGFR, ORs for almost all proteins, except for TNF-R6, TNF-R21 and AMBP, remained highly statistically significant.

**Table 2**

**Enrichment of biological pathways**

In total, there were 46 ESKD associated proteins. To explore the possible role of these proteins in the disease process that underlies fast progression to ESKD, we examined which biological pathways were enriched with these proteins. We performed formal pathway analysis using the Database for Annotation, Visualization and Integrated Discovery (DAVID) Bioinformatics Resources 6.8 (<https://david.ncifcrf.gov/>). **Fig. 2A** shows the results of the analysis. Four clusters of pathways or terms showed highly statistically significant enrichment. The 1st cluster had the highest score of enrichment (= 8.9) and comprised TNF receptor signaling pathways and apoptotic processes. The 2nd cluster comprised receptors and membrane proteins as biological terms with score of enrichment (= 4.2.) The 3rd and 4th cluster comprised TNF receptor signaling and apoptotic processes and had moderate scores of enrichment (2.2 and 1.9.) No other cluster was identified in the pathway analysis. It is important to emphasize that the same clusters and with similar enrichment scores were found when the DAVID analysis was performed using 1,012 proteins on 11 OLINK panels that were available at the time of implementing this study and 455 examined proteins on 5 OLINK panels (data not shown). This finding indicates that our selection of the 5 OLINK panels out of 11 available was not biased toward selecting TNF receptors.

Inspection of proteins that contributed to enrichment in cluster 1, 3 and 4 revealed 20 proteins. **Fig. 2B** shows thatamong these proteins, 12 were common for TNF receptor signaling and apoptosis processes, 5 were specific for TNF receptor signaling and 3 specific for apoptosis. The remaining 26, except for 5, were enriched categories of receptors or membranes proteins. It is important to point out that most proteins in the TNF receptor signaling and apoptosis processes were included in cluster 2 as well.

**Figure 2A & 2B**

**Source of ESKD associated proteins in circulation**

**Expression of genes encoding circulating ESKD associated proteins in peripheral blood cells:** To examine the possibility that the peripheral blood cell can be a source of elevated concentration in circulation of ESKD associated proteins, a genomic study was conducted in which mRNA was extracted from blood collected to PAX gene and sequenced. PAX gene tubes with blood were collected at baseline examination from 120 individuals included in the T1D (n=56) and T2D (n=76) studies. Clinical characteristics of cases (decliners, n=65) and controls (non-decliners, n=57) are shown in **Table S2** and the comparison of the expression of 46 genes encoding the ESKD associated proteins are shown in **Fig. 3**.

**Fig. 3**

**Expression of 46 genes encoding circulating ESKD associated proteins in proximal tubule injury:**

**Notes for Parker:**

1. **In addition to the comparison of the expression of the 46 genes in damaged vs. healthy proximal tubule (MAIN FINDINGS) you need to show**
2. **Expression of the 46 genes in infiltrating cells from kidney from healthy individuals vs. kidney with advanced diabetic kidney disease (notice that we are showing above your section Figure 3 which demonstrates that 46 genes had the same expression in circulating WBC from decliners and non-decliners). You may need a additional pane to show negative(?) Findings.**
3. **You need to do the same for fibroblasts from kidney from healthy individuals vs. fibroblasts from kidney of patients with advanced diabetic kidney disease.**

**Single Cell Sequencing in Diabetic Kidney Disease**

Diabetic kidney disease (DKD) is associated with proximal tubular injury and progression to CKD. Single cell sequencing of kidney cortex obtained from donors with DKD identified a subset of injured proximal tubule cells that has a distinct transcriptional and chromatin accessibility profile. These injured proximal tubule cells often express VCAM1, but also variably express other injury-associated markers like CD24 and CD133. These injured proximal tubule cells have been called a variety of things in the literature, including failed repair proximal tubule, maladaptive, and degenerative. They are detectable in control kidney samples in a “scattered” distribution and increase in proportion in aging and CKD. Proximal tubule injury and the transition to the PT\_VCAM1 cell state is associated with increased expression and enrichment of pro-inflammatory pathways, including TNF signaling. We compared a list of previously-published differentially expressed genes (DEG) and differentially accessible chromatin regions (DAR) that distinguish between control proximal tubule and injured proximal tubule (PT\_VCAM1). PT\_VCAM1 showed differential expression of 13 of 46 circulating biomarkers compared to control proximal tubule by snRNA-seq (Figure 4A). We correlated expression of apoptosis pathway genes with circulating biomarkers and discovered that increased apoptosis pathway expression shows a significant positive correlation with 17 of 46 circulating biomarkers in the PT\_VCAM1 injured cell state (Figure 4B). PT\_VCAM1 also showed differential accessibility of 32 ATAC peaks by snATAC-seq. These peaks were associated with 16 circulating biomarker genes and were located in promoters, enhancers, gene bodies, and intergenic regions (Figure 4C). We hypothesize that proximal tubule injury in DKD leads to an increased proportion of injured proximal tubule cells that show elevated expression of VCAM1, CD24, CD133 and other injury-associated markers. The PT\_VCAM1 cell state has increased expression of TNF-family genes that escape into the circulation and are detected in peripheral blood. The PT\_VCAM1 cell state may retain the ability to repair itself, but a subset of cells likely progresses to apoptosis (Figure 5D). The proximal tubule is the most abundant cell type in the kidney cortex and TNF signaling and apoptotic genes are not significantly expressed in infiltrating immune cells, which makes the proximal tubule the most likely source of these circulating biomarkers.

**Fig. 4 - Single Cell Sequencing in Diabetic Kidney Disease.** **A)** Differentially expression of 46 circulating biomarker genes for PT\_VCAM1 vs. control proximal tubule are displayed. Only genes that meet the Benjamini-Hochberg adjusted p-value threshold (padj < 0.05) are displayed. **B)** Pearson correlation between an aggregate measure of hallmark apoptosis genes (“Apoptosis Index”) and circulating biomarker genes was computed for the PT\_VCAM1 cell state. The “\*” symbol indicates biomarker genes that are significantly correlated with apoptosis index using an unadjusted pval < 0.05. Biomarker genes that are filled gray were not detected in the snRNA-seq dataset in the PT\_VCAM1 cell state. **C)** Differentially accessibility of 46 circulating biomarker gene peaks for PT\_VCAM1 vs. control proximal tubule are displayed. Only peaks that meet the Benjamini-Hochberg adjusted p-value threshold (padj < 0.05) are displayed. Peaks were annotated with their nearest genomic feature using ChIPSeeker. **D)** Model of proximal tubule injury in DKD. Healthy control proximal tubule expresses markers like *SLC34A1* that mediate normal homeostatic functions. When proximal tubules are injured, they begin to express markers like VCAM1, CD24, and CD133. These injury-associated markers are also associated with expression of TNF-family ligands that are released into circulation. A subset of injured proximal tubule cells ultimately progress to apoptosis.

**Co-localization of immunostaining of ESKD associated proteins with apoptotic cells in kidney biopsies:**

**Fig. 5 Jon will prepare this Figure.**

**Development of ESKD risk score**

In total, 46 circulating proteins were found to be associated with risk of fast progression to ESKD. Many of these proteins, however, represented dysregulation/enrichment of the same pathways. Such proteins may provide redundant information regarding prediction of fast progression to ESKD. Among the 46 ESKD-associated proteins, we performed Random Forest in order to rank the variable importance of these proteins (see **Fig. 7**.) KIM-1 was ranked as the most important protein in terms of predicting fast progression to ESKD, followed by TNF-R19, LAYN, and IL-1RT1. Additionally, to reduce the number of proteins with redundant information and select the most informative proteins LASSO logistic regression analysis was used. The analysis was performed using two different sets of proteins.

Since the enrichment of TNF receptor signaling and apoptosis pathways were such dominant features of the pathway analysis, the informative proteins were selected from among 20 proteins representing these pathways in the first Lasso analysis. This analysis retained 6 proteins, as the most informative exemplars of these pathways: KIM-1, TNF-R27, IL-1RT1, TNF-R11A, TNF-R6B, and TNF-R19 (see **Table 3**). Based on their β coefficients, the TNF receptor/apoptosis ESKD risk score was developed. Since the TNF receptor/apoptosis proteins were highly inter-correlated (data not shown), we searched for alternative combinations/models of these proteins to build ESKD risk scores of similar prognostic performance as that one in Model #1. Performing LASSO analysis by removing one protein at a time from the list of 20 proteins, 4 alternative models were found with similar efficiencies to predict fast progression to ESKD. The overall combination of 5 or 6 out of 8 informative TNF receptor/apoptosis proteins was sufficient to build ESKD risk scores that were equally efficient to predict fast progression to ESKD because none of the 26 proteins were selected in the model through the variable selections for predicting ESKD when the model includes these 5 or 6 proteins. The magnitude of OR per quartile change of the risk score varied between 2.71 and 2.89 without clinical variables and increased accordingly to 2.92 to 3.49 when combined with clinical variables such as HbA1c, ACR and eGFR (data are not shown.) Similarly, the c-statistics of these models were very similar (data not shown.) We need to emphasize that KIM-1 and TNF-R27 are the two proteins that were not replaceable by any of the TNF/apoptosis proteins. To consider building the parsimonious model, we employed the model with 5 out of 6 proteins that the Lasso logistic regression firstly retained. The names of the proteins used to build the score and the indices of its prognostic performance are shown in **Table 3** in the column headed as Model#1. As shown in **Table 3** in columns labeled Models #3 with “Known Biomarkers,” ESKD risk scores without these proteins were less efficient in predicting fast progression to ESKD.

In addition to 20 TNF receptors/apoptosis proteins, 26 other circulating proteins were strongly associated with fast progression to ESKD. To test the importance/usefulness of these proteins to predict fast progression to ESKD, we performed Lasso regression analysis using all 46 ESKD associated proteins as candidate predictors. Five proteins selected for model #1 and five proteins representing other pathways were selected by the Lasso analysis (see **Table 3**). It is important to notice that in this global model, other proteins did not replace the TNF receptor/apoptosis proteins but their effects were only diminished to make room for other proteins (columns Model #1 vs. Global Model #2 in **Table 3**). Importantly, ESKD risk score created out of these 10 proteins did not provide better discrimination of risk of fast progression to ESKD than Model #1 (difference in c statistics between Model #1 vs. Global Model #2 in **Table 3**, -0.003 (95%C.I. -0.020, 0.015) (p = 0.7535). The prognostic performance of this model increased after adding clinical variables HbA1c, ACR and eGFR in similar ways as in Model #3. When we applied the same equation of risk score with an outcome of annual eGFR loss equal to or greater than 5 mL/min/1.73m2 in the T1D Micro-Albuminuria study (**Table S3**), OR of the risk score was 2.30 and its c-statistic was 0.730.

**Table 3**

**Variation of ESKD risk score and prediction of progression to ESKD**

To examine possible determinants of variation of the ESKD risk score based on the 5 TNF receptor/apoptosis proteins, a correlation analysis was run between clinical characteristics and ESKD risk score computed in 163 individuals considered as controls in the T2D study. **Fig. 8A** shows the results. The values of the risk score were not correlated with BMI, only weakly positive with HbA1c, age, and systolic blood pressure, and moderately positive with ACR and negative with eGFR. It is important to recognize that both of the latter are an intermittent outcomes and may be considered as predictors but not as causal exposures. To further tease out the association between risk score and eGFR and ACR, we examined the variation of values in risk scores in the same T2D controls, according to the combination of the categories of eGFR (60-90, 90-120, 120≤) and ACR (30<, 30-300, 300≤) (see in **Fig. 8B**). As eGFR declined and ACR increased, the risk score presented the dose-response relationship with the combination of these categories.

**Fig. 8A and 8B**

To assess the effect of the ESKD risk score on the development of ESKD during 15 years of follow-up, the odds ratios (OR) according to quartiles of the score were estimated with and without adjustment for clinical covariates. The results are shown in **Fig. 8C.** Risk of fast progression to ESKD increased linearly with increased quartiles of risk score. This linear increase was independent from the clinical variables.

**Fig. 8C**

To consider the censorship of follow-ups in the study population, we examined the cumulative incidence of ESKD according to quartile of score using time-to-event analysis (see **Fig. 8D**). Cumulative incidence of ESKD increased explicitly with increased quartiles of risk score.

**Fig. 8D**

**Comparison of OLINK findings with SOMAscan findings.**

In our previous studies using SOMAscan platform, we found 40 circulating proteins that were strong risk predictors of progression to ESKD ( ). In those studies, the risk predictors were identified first in the cohort studies of individuals with impaired kidney function and validated in cohort studies in individuals with normal kidney function. The current study used the OLINK platform and focused on healthier individuals with normal kidney function. Forty-six circulating proteins were found to be strong risk predictors of progression to ESKD. We used two different ways to compare the results from the two sets of studies.

First, we compared the results for individual proteins. Out of 455 proteins examined using OLINK platform and 560 proteins examined using SOMAscan, 138 proteins were present on both platforms and 33 were associated with risk of progression to ESKD at least on one of the platforms. **Table S4** shows the comparison of ORs for progression to ESKD after adjustment for clinical characteristics for each of the 33 proteins obtained in the two sets of studies. Out of 33 proteins, 22 showed statistically significant association with risk of ESKD in both sets of studies. Six proteins, including four TNF receptors, showed strong statistically significant association with risk of ESKD, but only for measurements done on OLINK platform and no association at all when SOMAscan was used. An additional five proteins were significant on SOMAscan but not significant on OLINK. It is important to emphasize that OR for 4 proteins were positive on OLINK but were not statistically significant. However, the discrepant results for NBL1 were striking; very high OR and extremely significant p value for measurements done by SOMAscan and no association when measured by OLINK.

Second, we compared results of enrichment of biological pathways analysis for 46 ESKD associated proteins measured on OLINK as shown in **Fig. 3A** with similar analysis performed for 40 ESKD associated proteins measured on SOMAscan (see **Fig. S1**). In the latter analysis, the results were similar to the first although with different ranking. The top cluster for the 40 SOMAscan proteins comprised terms receptors and membranes. The second cluster comprised TNF receptor signaling pathways and the third cluster comprised apoptosis terms. Overall, the enrichment scores and statistical significances for the pathways and terms in the three clusters in SOMAscan proteins were weaker than in the results of the analysis for the 46 proteins measured on OLINK.

**Discussion not ready**

**REFERENCES**

1. Andrzej S Krolewski. Progressive renal decline: the new paradigm of diabetic nephropathy in type 1 diabetes. Diabetes Care. 2015; 38 (6): 954-962.

2. Andrzej S Krolewski, Jan Skupien, Peter Rossing, James H Warram. Fast renal decline to end-stage renal disease: an unrecognized feature of nephropathy in diabetes. Kidney Int. 2017; 91 (6): 1300-1311.

3. Andrzej S Krolewski, Monika A Niewczas, Jan Skupien, Tomhito Gohda, Adam Smiles, Jon H Eckfeldt, Alessandro Doria, James H Warram. Early progressive renal decline precedes the onset of microalbuminuria and its progression to macroalbuminuria. Diabetes Care. 2014; 37 (1): 226-234.

4. Dorte Vistisen, Gregers Stig Andersen, Adam Hulman, Frederik Persson, Peter Rossing, Marit Eika Jørgensen. Progressive Decline in Estimated Glomerular Filtration Rate in Patients With Diabetes After Moderate Loss in Kidney Function-Even Without Albuminuria. Diabetes Care. 2019; 42 (10): 1886-1894.

5. Jan Skupien, James H Warram, Adam M Smiles, Robert C Stanton, Andrzej S Krolewski. Patterns of Estimated Glomerular Filtration Rate Decline Leading to End-Stage Renal Disease in Type 1 Diabetes. Diabetes Care. 2016; 39 (12): 2262-2269.

6. Diabetes Control and Complications Trial Research Group; D M Nathan, S Genuth, J Lachin, P Cleary, O Crofford, M Davis, L Rand, C Siebert. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. N Engl J Med. 1993; 329 (14): 977-986.

7. DCCT/EDIC Research Group; Ian H de Boer, Wanjie Sun, Patricia A Cleary, John M Lachin, Mark E Molitch, Michael W Steffes, Bernard Zinman. Intensive diabetes therapy and glomerular filtration rate in type 1 diabetes. N Engl J Med. 2011: 365 (25): 2366-2376.

8. Jan Skupien, James H Warram, Adam Smiles, Andrzej Galecki, Robert C Stanton, Andrzej S Krolewski. Improved glycemic control and risk of ESRD in patients with type 1 diabetes and proteinuria. J Am Soc Nephrol. 2014; 25 (12): 2916-2925.

9. Bruce A Perkins, Ionut Bebu, Ian H de Boer, Mark Molitch, William Tamborlane, Gayle Lorenzi, William Herman, Neil H White, Rodica Pop-Busui, Andrew D Paterson, Trevor Orchard, Catherine Cowie, John M Lachin, Diabetes Control and Complications Trial (DCCT)/Epidemiology of Diabetes Interventions and Complications (EDIC) Research Group. Risk Factors for Kidney Disease in Type 1 Diabetes. Diabetes Care. 2019; 42 (5): 883-890.

10. Hermann Haller, Sadayoshi Ito, Joseph L Izzo Jr, Andrzej Januszewicz, Shigehiro Katayama, Jan Menne, Albert Mimran, Ton J Rabelink, Eberhard Ritz, Luis M Ruilope, Lars C Rump, Giancarlo Viberti, ROADMAP Trial Investigators. Olmesartan for the delay or prevention of microalbuminuria in type 2 diabetes. N Engl J Med. 2011; 364 (10): 907-917.

11. B M Brenner, M E Cooper, D de Zeeuw, W F Keane, W E Mitch, H H Parving, G Remuzzi, S M Snapinn, Z Zhang, S Shahinfar, RENAAL Study Investigators. Effects of losartan on renal and cardiovascular outcomes in patients with type 2 diabetes and nephropathy. N Engl J Med. 2001; 345 (12): 861-869.

12. E J Lewis, L G Hunsicker, W R Clarke, T Berl, M A Pohl, J B Lewis, E Ritz, R C Atkins, R Rohde, I Raz, Collaborative Study Group. Renoprotective effect of the angiotensin-receptor antagonist irbesartan in patients with nephropathy due to type 2 diabetes. N Engl J Med. 2001; 345 (12): 851-860.

13. Piero Ruggenenti, Anna Fassi, Anelja Parvanova Ilieva, Simona Bruno, Ilian Petrov Iliev, Varusca Brusegan, Nadia Rubis, Giulia Gherardi, Federica Arnoldi, Maria Ganeva, Bogdan Ene-Iordache, Flavio Gaspari, Annalisa Perna, Antonio Bossi, Roberto Trevisan, Alessandro R Dodesini, Giuseppe Remuzzi, Bergamo Nephrologic Diabetes Complications Trial (BENEDICT) Investigators. Preventing microalbuminuria in type 2 diabetes. N Engl J Med. 2004; 351 (19): 1941-1951.

14. Michael Mauer, Bernard Zinman, Robert Gardiner, Samy Suissa, Alan Sinaiko, Trudy Strand, Keith Drummond, Sandra Donnelly, Paul Goodyer, Marie Claire Gubler, Ronald Klein. Renal and retinal effects of enalapril and losartan in type 1 diabetes. N Engl J Med. 2009; 361 (1): 40-51.

15. Monika A Niewczas, Meda E Pavkov, Jan Skupien, Adam Smiles, Zaipul I Md Dom, Jonathan M Wilson, Jihwan Park, Viji Nair, Andrew Schlafly, Pierre-Jean Saulnier, Eiichiro Satake, Christopher A Simeone, Hetal Shah, Chengxiang Qiu, Helen C Looker, Paolo Fiorina, Carl F Ware, Jennifer K Sun, Alessandro Doria, Matthias Kretzler, Katalin Susztak, Kevin L Duffin, Robert G Nelson, Andrzej S Krolewski. A signature of circulating inflammatory proteins and development of end-stage renal disease in diabetes. Nat Med. 2019; 25 (5): 805-813.

16. Zaipul I. Md Dom, Eiichiro Satake, Jan Skupien, Bozena Krolewski, Kristina O’Neil, Jill A. Willency, Simon T. Dillon, Jonathan M. Wilson, Hiroki Kobayashi, Katsuhito Ihara, Towia A. Libermann, Marlon Pragnell, Kevin L. Duffin, Andrzej S. Krolewski. Circulating Proteins Protect Against Renal Decline and Progression to End-Stage Renal Disease in Diabetes: Results of Global Proteomics Profiling. Sci Transl Med. 2021; XX (XXX): XXX.

17. Eiichiro Satake, Pierre-Jean Saulnier, Hiroki Kobayashi, Manoj K. Gupta, Helen C. Looker, Jonathan M. Wilson, Zaipul I. Md Dom, Katsuhito Ihara, Kristina O’Neil, Bozena Krolewski, Caterina Pipino, Meda E. Pavkov, Viji Nair, Markus Bitzer, Monika A. Niewczas, Matthias Kretzler, Michael Mauer, Alessandro Doria, Behzad Najafian, Rohit N. Kulkarni, Kevin L. Duffin, Marcus G. Pezzolesi, C. Ronald Kahn, Robert G. Nelson, Andrzej S. Krolewski. Profiles of circulating miRNAs and Axon Guidance Pathway proteins associated with progression to end stage kidney disease in diabetes. J Am Soc Nephrol. 2021; XX (XX): XXX.

18. Katsuhito Ihara, Jan Skupien, Bozena Krolewski, Zaipul I Md Dom, Kristina O'Neil, Eiichiro Satake, Hiroki Kobayashi, Narges M Rashidi, Monika A Niewczas, Andrzej S Krolewski. A profile of multiple circulating tumor necrosis factor receptors associated with early progressive kidney decline in Type 1 Diabetes is similar to profiles in autoimmune disorders. Kidney Int. 2020; S0085-2538 (20) 30834-6.

19. Assarsson E, Lundberg M, Holmquist G, et al. Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. PLoS One. 2014; 9: e95192.

20. P Fioretto, M Mauer, E Brocco, M Velussi, F Frigato, B Muollo, M Sambataro, C Abaterusso, B Baggio, G Crepaldi, R Nosadini. Patterns of renal injury in NIDDM patients with microalbuminuria. Diabetologia. 1996; 39 (12): 1569-1576.

21. K V Lemley, I Abdullah, B D Myers, T W Meyer, K Blouch, W E Smith, P H Bennett, R G Nelson. Evolution of incipient nephropathy in type 2 diabetes mellitus. Kidney Int. 2000; 58 (3): 1228-1237.

22. Jan Skupien, James H Warram, Adam M Smiles, Monika A Niewczas, Tomohito Gohda, Marcus G Pezzolesi, Diego Cantarovich, Robert Stanton, Andrzej S Krolewski. The early decline in renal function in patients with type 1 diabetes and proteinuria predicts the risk of end-stage renal disease. Kidney Int. 2012; 82 (5): 589-597.

23. Andrew S Levey, Lesley A Stevens, Christopher H Schmid, Yaping Lucy Zhang, Alejandro F Castro 3rd, Harold I Feldman, John W Kusek, Paul Eggers, Frederick Van Lente, Tom Greene, Josef Coresh, CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration). A new equation to estimate glomerular filtration rate. Ann Intern Med. 2009; 150 (9): 604-612.

24. Parker C Wilson, Benjamin D Humphreys. Single-cell genomics and gene editing: implications for nephrology. Nat Rev Nephrol. 2019; 15 (2): 63-64.

25. Parker C Wilson, Haojia Wu, Yuhei Kirita, Kohei Uchimura, Nicolas Ledru, Helmut G Rennke, Paul A Welling, Sushrut S Waikar, Benjamin D Humphreys. The single-cell transcriptomic landscape of early human diabetic nephropathy. Proc Natl Acad Sci U S A. 2019; 116 (39): 19619-19625.

26. Tim Stuart, Andrew Butler, Paul Hoffman, Christoph Hafemeister, Efthymia Papalexi, William M Mauck 3rd, Yuhan Hao, Marlon Stoeckius, Peter Smibert, Rahul Satija. Comprehensive Integration of Single-Cell Data. Cell. 2019; 177 (7): 1888-1902.e21.

27. Christopher S McGinnis, Lyndsay M Murrow, Zev J Gartner. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. Cell Syst. 2019; 8 (4): 329-337.e4.

28. Etienne Becht, Leland McInnes, John Healy, Charles-Antoine Dutertre, Immanuel W H Kwok, Lai Guan Ng, Florent Ginhoux, Evan W Newell. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol. 2018; 37: 38-44.

29. Robert Tibshirani. Regression shrinkage and selection via the lasso: a retrospective. R Stat Soc B. 2011; 73 (3): 273-282.

30. Michael J Pencina, Ralph B D'Agostino Sr, Ralph B D'Agostino Jr, Ramachandran S Vasan. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med. 2008; 27 (2): 157-172.

31. Kennedy KF, Pencina MJ. A SAS macro to compute added predictive ability of newmarkers predicting a dichotomous outcome [article online], 2010. NC State University. Available from <http://analytics.ncsu.edu/sesug/2010/SDA07.Kennedy.pdf>.

Accessed 4 January 2021.

32. Leo Breiman. Random forests. J Mach Learn 2001; 45: 5-32.

33. Bjoern H Menze, B Michael Kelm, Ralf Masuch, Uwe Himmelreich, Peter Bachert, Wolfgang Petrich & Fred A Hamprecht. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. BMC Bioinformatics. 2009; 10: 213. <https://doi.org/10.1186/1471-2105-10-213>.

34. Maria D Sanchez-Niño, Alberto Benito-Martin, Sara Gonçalves, Ana B Sanz, Alvaro C Ucero, Maria C Izquierdo, Adrian M Ramos, Sergio Berzal, Rafael Selgas, Marta Ruiz-Ortega, Jesus Egido, Alberto Ortiz. TNF superfamily: a growing saga of kidney injury modulators. Mediators Inflamm. 2010; 2010: 182958.

35. J Dutta, Y Fan, N Gupta, G Fan, C Gélinas. Current insights into the regulation of programmed cell death by NF-kappaB. Oncogene. 2006; 25 (51): 6800-6816.

36. Extrinsic apoptosis Pathway, [www.creative-diagnostics.com/extrinsic-apoptosis-pathway.htm](http://www.creative-diagnostics.com/extrinsic-apoptosis-pathway.htm)

37. Corina Lorz, Alberto Benito-Martín, Anissa Boucherot, Alvaro C Ucero, Maria Pia Rastaldi, Anna Henger, Silvia Armelloni, Beatriz Santamaría, Celine C Berthier, Matthias Kretzler, Jesus Egido, Alberto Ortiz . The death ligand TRAIL in diabetic nephropathy. J Am Soc Nephrol. 2008; 19 (5): 904-914.

38. Diana C Spierings, Elisabeth G de Vries, Edo Vellenga, Fiona A van den Heuvel, Jan J Koornstra, Jelle Wesseling, Harry Hollema, Steven de Jong. Tissue distribution of the death ligand TRAIL and its receptors. J Histochem Cytochem. 2004; 52 (6): 821-831.

39. Dinender Kumar, Susan Robertson, Kevin D Burns. Evidence of apoptosis in human diabetic kidney. Mol Cell Biochem. 2004; 259 (1-2): 67-70.

40. M Lorenzi, D F Montisano, S Toledo, A Barrieux. High glucose induces DNA damage in cultured human endothelial cells. J Clin Invest. 1986; 77 (1): 322-325.

41. G Wolf, F N Ziyadeh. The role of angiotensin II in diabetic nephropathy: emphasis on nonhemodynamic mechanisms. Am J Kidney Dis. 1997; 29 (1): 153-163.

42. J J Morrissey, S Klahr. Effect of AT2 receptor blockade on the pathogenesis of renal fibrosis. Am J Physiol. 1999; 276 (1): F39-45.

43. A Ortiz, C Lorz, M P Catalán, T M Danoff, Y Yamasaki, J Egido, E G Neilson. Expression of apoptosis regulatory proteins in tubular epithelium stressed in culture or following acute renal failure. Kidney Int. 2000; 57 (3): 969-981.

44. Zhiwei Zhang, Benjamin D Humphreys, Joseph V Bonventre. Shedding of the urinary biomarker kidney injury molecule-1 (KIM-1) is regulated by MAP kinases and juxtamembrane region. J Am Soc Nephrol. 2007; 18 (10): 2704-2714.

45. Natalia Nowak, Jan Skupien, Monika A Niewczas, Masayuki Yamanouchi, Melissa Major, Stephanie Croall, Adam Smiles, James H Warram, Joseph V Bonventre, Andrzej S Krolewski. Increased plasma kidney injury molecule-1 suggests early progressive renal decline in non-proteinuric patients with type 1 diabetes. Kidney Int. 2016; 89 (2): 459-467.

46. Rohit N Kulkarni, C Ronald Kahn. Ephs and Ephrins Keep Pancreatic β Cells Connected. Cell. 2007; 129 (2): 241-243.

47. Armando Caseiro, António Barros, Rita Ferreira, Ana Padrão, Miguel Aroso, Cláudio Quintaneiro, Amélia Pereira, Rosário Marinheiro, Rui Vitorino, Francisco Amado. Pursuing type 1 diabetes mellitus and related complications through urinary proteomics. Transl Res. 2014; 163 (3): 188-199.

48. Xue Bao, Yan Borné, Iram Faqir Muhammad, Jan Nilsson, Lars Lind, Olle Melander, Kaijun Niu, Marju Orho-Melander, Gunnar Engström. Growth differentiation factor 15 is positively associated with incidence of diabetes mellitus: the Malmo Diet and Cancer-Cardiovascular Cohort. Diabetologia 2019; 62 (1): 78–86.

49. Merel E Hellemons, Magdalena Mazagova, Ron T Gansevoort, Robert H Henning, Dick de Zeeuw, Stephan J L Bakker, Hiddo J Lambers-Heerspink, Leo E Deelman. Growth-differentiation factor 15 predicts worsening of albuminuria in patients with type 2 diabetes. Diabetes Care. 2012; 35 (11): 2340–2346.

50. Merel E Hellemons, Magdalena Mazagova, Ron T Gansevoort, Robert H Henning, Dick de Zeeuw, Stephan J L Bakker, Hiddo J Lambers-Heerspink, Leo E Deelman. Growth-differentiation factor 15 predicts worsening of albuminuria in patients with type 2 diabetes. Diabetes Care. 2012; 35 (11): 2340–2346.

51. Maria Lajer, Anders Jorsal, Lise Tarnow, Hans-Henrik Parving, Peter Rossing. Plasma growth differentiation factor-15 independently predicts all-cause and cardiovascular mortality as well as deterioration of kidney function in type 1 diabetic patients with nephropathy. Diabetes Care. 2010; 33 (7): 1567-1572.

52. Axel C Carlsson, Christoph Nowak, Lars Lind, Carl Johan Östgren, Fredrik H Nyström, Johan Sundström, Juan Jesus Carrero, Ulf Riserus, Erik Ingelsson, Tove Fall, Johan Ärnlöv. 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. 2020; 125 (1): 37-43.

53. Nehal M Elsherbiny, Mohammed M H Al-Gayyar. The role of IL-18 in type 1 diabetic nephropathy: The problem and future treatment. Cytokine. 2016; 81: 15-22.

54. Francesco Volta, M Julia Scerbo, Anett Seelig, Robert Wagner, Nils O'Brien, Felicia Gerst, Andreas Fritsche, Hans-Ulrich Häring, Anja Zeigerer, Susanne Ullrich, Jantje M Gerdes. Glucose homeostasis is regulated by pancreatic β-cell cilia via endosomal EphA-processing. Nat Commun. 2019; 10 (1): 5686.

55. R Bergholdt, A E Karlsen, J Johannesen, P M Hansen, C A Dinarello, J Nerup, F Pociot. Characterization of polymorphisms of an interleukin 1 receptor type 1 gene (IL1RI) promotor region (P2) and their relation to insulin-dependent diabetes mellitus (IDDM). The Danish Study Group of Diabetes in Childhood. Cytokine. 1995; 7 (7): 727-733.

56. K A Metcalfe, G A Hitman, F Pociot, R Bergholdt, E Tuomilehto-Wolf, J Tuomilehto, M Viswanathan, A Ramachandran, J Nerup. An association between type 1 diabetes and the interleukin-1 receptor type 1 gene. The DiMe Study Group. Childhood Diabetes in Finland. Hum Immunol. 1996; 51 (1): 41-48.

57. Jingchun Qu, Hui-Qi Qu, Jonathan P Bradfield, Joseph T Glessner, Xiao Chang, Lifeng Tian, Michael E March, Jeffrey D Roizen, Patrick M Sleiman, Hakon Hakonarson. Association of DLL1 with type 1 diabetes in patients characterized by low polygenic risk score. Metabolism. 2020 4; 114: 154418.

58. Donna M D'Souza, Sarah Zhou, Irena A Rebalka, Blair MacDonald, Jasmin Moradi, Matthew P Krause, Dhuha Al-Sajee, Zubin Punthakee, Mark A Tarnopolsky, Thomas J Hawke. Decreased Satellite Cell Number and Function in Humans and Mice With Type 1 Diabetes Is the Result of Altered Notch Signaling. Diabetes. 2016; 65 (10): 3053-3061.

59. Ilse M Rood, Michael L Merchant, Daniel W Wilkey, Terry Zhang, Vlad Zabrouskov, Johan van der Vlag, Henry B Dijkman, Brigith K Willemsen, Jack F Wetzels, Jon B Klein, Jeroen K Deegens. Increased expression of lysosome membrane protein 2 in glomeruli of patients with idiopathic membranous nephropathy. Proteomics. 2015; 15 (21): 3722-3730.

60. J Jonklaas, Srk Murthy, D Liu, J Klubo-Gwiezdzinska, J Krishnan, K D Burman, L Boyle, N Carrol, E Felger, Y P Loh. Novel biomarker SYT12 may contribute to predicting papillary thyroid cancer outcomes. Future Sci OA. 2017; 4 (1): FSO249.

61. J Reiser, W Kriz, M Kretzler, P Mundel. The glomerular slit diaphragm is a modified adherens junction. J Am Soc Nephrol. 2000; 11 (1): 1-8.

62. Marijana Dajak, Svetlana Ignjatović, Biljana Stojimirović, Snezana Gajić, Nada Majkić-Singh. Evaluation of renal damage by urinary beta-trace protein in patients with chronic kidney disease. Clin Lab. 2011; 57 (1-2): 29-36.

63. Manousos Makridakis, Georgia Kontostathi, Eleni Petra, Rafael Stroggilos, Vasiliki Lygirou, Szymon Filip, Flore Duranton, Harald Mischak, Angel Argiles, Jerome Zoidakis, Antonia Vlahou. Multiplexed MRM-based protein quantification of putative prognostic biomarkers for chronic kidney disease progression in plasma. Sci Rep. 2020; 10 (1): 4815.

64. Rik Derynck, Ying E Zhang. Smad-dependent and Smad-independent pathways in TGF-beta family signaling. Nature. 2003; 425 (6958): 577-584.

65. S Busch, A Acar, Y Magnusson, P Gregersson, L Rydén, G Landberg. TGF-beta receptor type-2 expression in cancer-associated fibroblasts regulates breast cancer cell growth and survival and is a prognostic marker in pre-menopausal breast cancer. Oncogene. 2015; 34 (1): 27-38.

66. Alan R Schenkel, Eric M Dufour, Tina W Chew, Emily Sorg, William A Muller. The murine CD99-related molecule CD99-like 2 (CD99L2) is an adhesion molecule involved in the inflammatory response. Cell Commun Adhes. 2007; 14 (5): 227-237.

67. Oliver J Harrison, Julia Brasch, Gorka Lasso, Phinikoula S Katsamba, Goran Ahlsen, Barry Honig, Lawrence Shapiro. Structural basis of adhesive binding by desmocollins and desmogleins. Proc Natl Acad Sci U S A. 2016; 113 (26): 7160-7165.

68. Yoshimi Takai, Jun Miyoshi, Wataru Ikeda, Hisakazu Ogita. Nectins and nectin-like molecules: roles in contact inhibition of cell movement and proliferation. Nat Rev Mol Cell Biol. 2008; 9 (8): 603-615.

69. Lorenz Vogt, Nicole Schmitz, Michael O Kurrer, Monika Bauer, Heather I Hinton, Silvia Behnke, Dominique Gatto, Peter Sebbel, Roger R Beerli, Ivo Sonderegger, Manfred Kopf, Philippe Saudan, Martin F Bachmann. VSIG4, a B7 family-related protein, is a negative regulator of T cell activation. J Clin Invest. 2006; 116 (10): 2817-2826.

70. Xiao-Dong Geng, Wei-Wei Wang, Zhe Feng, Ran Liu, Xiao-Long Cheng, Wan-Jun Shen, Zhe-Yi Dong, Guang-Yan Cai, Xiang-Mei Chen, Quan Hong, Di Wu. Identification of key genes and pathways in diabetic nephropathy by bioinformatics analysis. J Diabetes Investig. 2019; 10 (4): 972-984.

71. Seung-Mi Kim, Se Won Oh, Sang Hyun Park, Dae Young Hur, Seung-Woo Hong, Sang Youb Han. Epstein-Barr virus-encoded latent membrane protein 1 induces epithelial to mesenchymal transition by inducing V-set Ig domain containing 4 (VSIG4) expression via NF-kB in renal tubular epithelial HK-2 cells. Biochem Biophys Res Commun. 2017; 492 (3): 316-322.

72. Eun-Yeung Gong, Hyung Ah Jo, Sang Hyun Park, Dae Ryong Cha, Dae Young Hur, Sang Youb Han. VSIG4 Induces Epithelial-Mesenchymal Transition of Renal Tubular Cells under High-Glucose Conditions. Life (Basel). 2020; 10 (12): 354.

73. Rohit Khurana, Lieve Moons, Shahida Shafi, Aernout Luttun, Désiré Collen, John F Martin, Peter Carmeliet, Ian C Zachary. Placental growth factor promotes atherosclerotic intimal thickening and macrophage accumulation. Circulation. 2005; 111 (21): 2828-2836.

74. Oskar Zakiyanov, Marta Kalousová, Tomáš Zima, Vladimír Tesař. Placental growth factor in patients with decreased renal function. Ren Fail. 2011; 33 (3): 291-297.

75. Lise Tarnow, Anne Sofie Astrup, Hans-Henrik Parving. Elevated placental growth factor (PlGF) predicts cardiovascular morbidity and mortality in type 1 diabetic patients with diabetic nephropathy. Scand J Clin Lab Invest Suppl. 2005; 240: 73-79.

76. Axel C Carlsson, Johan Sundström, Juan Jesus Carrero, Stefan Gustafsson, Markus Stenemo, Anders Larsson, Lars Lind, Johan Ärnlöv. Use of a proximity extension assay proteomics chip to discover new biomarkers associated with albuminuria. Eur J Prev Cardiol. 2017; 24 (4): 340-348.

77. Piyada Na Nakorn, Supitcha Pannengpetch, Patcharee Isarankura-Na-Ayudhya, Chadinee Thippakorn, Ratana Lawung, Nuankanya Sathirapongsasuti, Chagriya Kitiyakara, Piyamitr Sritara, Prin Vathesatogkit, Chartchalerm Isarankura-Na-Ayudhya. Roles of kininogen-1, basement membrane specific heparan sulfate proteoglycan core protein, and roundabout homolog 4 as potential urinary protein biomarkers in diabetic nephropathy. EXCLI J. 2020; 19: 872-891.

78. Valerie S LeBleu, Yingqi Teng, Joyce T O'Connell, David Charytan, Gerhard A Müller, Claudia A Müller, Hikaru Sugimoto, Raghu Kalluri. Identification of human epididymis protein-4 as a fibroblast-derived mediator of fibrosis. Nat Med. 2013; 19 (2): 227-231.

79. Katsuhito Ihara, Jan Skupien, Hiroki Kobayashi, Zaipul I Md Dom, Jonathan M Wilson, Kristina O'Neil, Hannah S Badger, Lenden M Bowsman, Eiichiro Satake, Matthew D Breyer, Kevin L Duffin, Andrzej S Krolewski. Profibrotic Circulating Proteins and Risk of Early Progressive Renal Decline in Patients With Type 2 Diabetes With and Without Albuminuria. Diabetes Care. 2020; 43 (11): 2760-2767.

80. Hirata Ki, T Ishida, K Penta, M Rezaee, E Yang, J Wohlgemuth, T Quertermous. Cloning of an immunoglobulin family adhesion molecule selectively expressed by endothelial cells. J Biol Chem. 2001; 276 (19): 16223-16231.

81. Ina M Kacso, Gabriel Kacso. Endothelial cell-selective adhesion molecule in diabetic nephropathy. Eur J Clin Invest. 2012; 42 (11): 1227-1234.

82. Ting-yi Du, Hui-ming Luo, Hai-chun Qin, Fang Wang, Qing Wang, Yang Xiang, Yun Zhang. Circulating serum trefoil factor 3 (TFF3) is dramatically increased in chronic kidney disease. PLoS One. 2013; 8 (11): e80271.

83. Hong Yang, Dezhe Qin, Shuanshuan Xu, Chen He, Jing Sun, Jinlian Hua, Sha Peng. Folic acid promotes proliferation and differentiation of porcine pancreatic stem cells into insulin-secreting cells through canonical Wnt and ERK signaling pathway. J Steroid Biochem Mol Biol. 2021; 205: 105772.

84. Mohamed A Elhadad, Christian Jonasson, Cornelia Huth, Rory Wilson, Christian Gieger, Pamela Matias, Harald Grallert, Johannes Graumann, Valerie Gailus-Durner, Wolfgang Rathmann, Christine von Toerne, Stefanie M Hauck, Wolfgang Koenig, Moritz F Sinner, Tudor I Oprea, Karsten Suhre, Barbara Thorand, Kristian Hveem, Annette Peters, Melanie Waldenberger. Deciphering the Plasma Proteome of Type 2 Diabetes. Diabetes. 2020; 69 (12): 2766-2778.

85. Takayuki Adachi, Mitsumi Arito, Naoya Suematsu, Atsuko Kamijo-Ikemori, Kazuki Omoteyama, Toshiyuki Sato, Manae S Kurokawa, Kazuki Okamoto, Kenjiro Kimura, Yugo Shibagaki, Tomohiro Kato. Roles of layilin in TNF-α-induced epithelial-mesenchymal transformation of renal tubular epithelial cells. Biochem Biophys Res Commun. 2015; 467 (1): 63-69.

86. Yeawon Kim, Sun-Ji Park, Scott R Manson, Carlos Af Molina, Kendrah Kidd, Heather Thiessen-Philbrook, Rebecca J Perry, Helen Liapis, Stanislav Kmoch, Chirag R Parikh, Anthony J Bleyer, Ying Maggie Chen. Elevated urinary CRELD2 is associated with endoplasmic reticulum stress-mediated kidney disease. JCI Insight. 2017; 2 (23): e92896.

87. Sangeeta Hingorani, Laura S Finn, Emily Pao, Rick Lawler, Gary Schoch, George B McDonald, Behzad Najafian, Brenda Sandmaier, Ted Gooley. Urinary elafin and kidney injury in hematopoietic cell transplant recipients. Clin J Am Soc Nephrol. 2015; 10 (1): 12-20.

88. Hiroyuki Takatsu, Koji Hase, Masumi Ohmae, Sayaka Ohshima, Koji Hashimoto, Naoko Taniura, Akitsugu Yamamoto, Hiroshi Ohno. CD300 antigen like family member G: A novel Ig receptor like protein exclusively expressed on capillary endothelium. Biochem Biophys Res Commun. 2006; 348 (1): 183-191.

89. Qiuyue Wang, Yuqing Liu, Ying Chen, Ke Wang, Wensheng Xie, Dapeng Wei, Lijuan Hu. CD300LG improves the cytotoxic activity of CIK. Cent Eur J Immunol. 2017; 42 (2): 117-122.

90. Weiming Ouyang, Jiangnan Xue, Jingmei Liu, Wei Jia, Zhouli Li, Xin Xie, Xuesong Liu, Jinlong Jian, Qi Li, Yong Zhu, Angang Yang, Boquan Jin. Establishment of an ELISA system for determining soluble LAIR-1 levels in sera of patients with HFRS and kidney transplant. J Immunol Methods. 2004; 292 (1-2): 109-17.

91. Cristina M Tato, Barbara Joyce-Shaikh, Antara Banerjee, Yi Chen, Manjiri Sathe, Sarah E Ewald, Man-Ru Liu, Daniel Gorman, Terrill K McClanahan, Joseph H Phillips, Paul G Heyworth, Daniel J Cua. The myeloid receptor PILRβ mediates the balance of inflammatory responses through regulation of IL-27 production. PLoS One. 2012; 7 (3): e31680.

92. Jing Wang, Ikuo Shiratori, Junji Uehori, Masahito Ikawa, Hisashi Arase. Neutrophil infiltration during inflammation is regulated by PILRα via modulation of integrin activation. Nat Immunol. 2013; 14 (1): 34-40.

93. Qiong Lu, Guangwen Lu, Jianxun Qi, Han Wang, Yifang Xuan, Qihui Wang, Yan Li, Yanfang Zhang, Chunfu Zheng, Zheng Fan, Jinghua Yan, George F Gao. PILRα and PILRβ have a siglec fold and provide the basis of binding to sialic acid. Proc Natl Acad Sci U S A. 2014; 111 (22): 8221-8226.

94. Ying Jie Ma, Estrid Hein, Lea Munthe-Fog, Mikkel-Ole Skjoedt, Rafael Bayarri-Olmos, Luigina Romani, Peter Garred. Soluble Collectin-12 (CL-12) Is a Pattern Recognition Molecule Initiating Complement Activation via the Alternative Pathway. J Immunol. 2015; 195 (7): 3365-3373.

95. Miltiadis Paliouras, Carla Borgono, Eleftherios P Diamandis. Human tissue kallikreins: the cancer biomarker family. Cancer Lett. 2007; 249 (1): 61-79.

96. Wei Du, Wenzhen Yu, Lvzhen Huang, Min Zhao, Xiaoxin Li. Ephrin-a4 is involved in retinal neovascularization by regulating the VEGF signaling pathway. Invest Ophthalmol Vis Sci. 2012; 53 (4): 1990-1998.

97. David L Allen, Allison S Cleary, Kristin J Speaker, Sarah F Lindsay, Jill Uyenishi, Jason M Reed, Molly C Madden, Ryan S Mehan. Myostatin, activin receptor IIb, and follistatin-like-3 gene expression are altered in adipose tissue and skeletal muscle of obese mice. Am J Physiol Endocrinol Metab. 2008; 294 (5): E918-927.

98. Claus Brandt, Maria Pedersen, Anders Rinnov, Anne S Andreasen, Kirsten Møller, Pernille Hojman, Bente K Pedersen, Peter Plomgaard. Obesity and low-grade inflammation increase plasma follistatin-like 3 in humans. Mediators Inflamm. 2014; 2014: 364209.

99. Gabriel Xavier Ramalho, Marcio Garrison Dytz. Diabetes of the Exocrine Pancreas Related to Hereditary Pancreatitis, an Update. Curr Diab Rep. 2020; 20 (6):16.

100. Steven G Coca, Girish N Nadkarni, Yuan Huang, Dennis G Moledina, Veena Rao, Jane Zhang, Bart Ferket, Susan T Crowley, Linda F Fried, Chirag R Parikh. Plasma Biomarkers and Kidney Function Decline in Early and Established Diabetic Kidney Disease. J Am Soc Nephrol. 2017; 28 (9): 2786-2793.

101. Sarah J Schrauben, Haochang Shou, Xiaoming Zhang, Amanda Hyre Anderson, Joseph V Bonventre, Jing Chen, Steven Coca, Susan L Furth, Jason H Greenberg, Orlando M Gutierrez, Joachim H Ix, James P Lash, Chirag R Parikh, Casey M Rebholz, Venkata Sabbisetti, Mark J Sarnak, Michael G Shlipak, Sushrut S Waikar, Paul L Kimmel, Ramachandran S Vasan, Harold I Feldman, Jeffrey R Schelling, CKD Biomarkers Consortium and the Chronic Renal Insufficiency Cohort (CRIC) Study Investigators. Association of Multiple Plasma Biomarker Concentrations with Progression of Prevalent Diabetic Kidney Disease: Findings from the Chronic Renal Insufficiency Cohort (CRIC) Study. J Am Soc Nephrol. 2021; 32 (1): 115-126.

102. Stephen A Williams, Mika Kivimaki, Claudia Langenberg, Aroon D Hingorani, J P Casas, Claude Bouchard, Christian Jonasson, Mark A Sarzynski, Martin J Shipley, Leigh Alexander, Jessica Ash, Tim Bauer, Jessica Chadwick, Gargi Datta, Robert Kirk DeLisle, Yolanda Hagar, Michael Hinterberg, Rachel Ostroff, Sophie Weiss, Peter Ganz, Nicholas J Wareham. Plasma protein patterns as comprehensive indicators of health. Nat Med. 2019; 25 (12): 1851-1857.

103. Merlin C Thomas, Michael Brownlee, Katalin Susztak, Kumar Sharma, Karin A M Jandeleit-Dahm, Sophia Zoungas, Peter Rossing, Per-Henrik Groop, Mark E Cooper. Diabetic kidney disease. Nat Rev Dis Primers. 2015; 1: 15018.

**FIGURE LEGENDS**

**Figure 1.** Biological pathways enriched with the 54 proteins. Only the biological pathways that fulfilled with p for enrichment <0.0001 are presented in this figure. MAPK, mitogen-activated protein kinase.

**Figure 2.** Venn diagrams of 54 proteins in relation to biological pathways presented in Figure 1.

**Figure 3.** Integratedsingle-nucleus RNA sequencing of kidney cortex between control and diabetic samples. (A) Uniform manifold approximation and projection (UMAP) of all types of kidney cells identified in the aggregated dataset. (B)-(D) Comparisons of gene expressions across 14 types of kidney cells. Nuclear dissociation and single nucleus RNA sequencing (snRNA-seq) with 10X Genomics 5’ chemistry were performed between healthy controls (n= 3) and individuals with early diabetic kidney disease (n= 3). Libraries were processed with cellranger and Seurat. For cluster visualization, UMAP dimensional reduction was performed in Seurat. Panels were separated by 1) Tumor necrosis factor (TNF) receptors and immunoregulatory receptors, 2) other receptors and 3) enzymes, ligands, inhibitors and others. A box is put around a comparison that is statistically different between control and diabetic samples. The vast majority are not different between conditions probably because they’re expressed at relatively low levels. Scale represents normalized log-fold-change and was prepared with the Seurat DotPlot function.

DCT1, early distal convoluted tubule. DCT2-CNT, late distal convoluted tubule and connecting tubule. PC, principal cell. ICA, type A intercalated cell. ICB, type B intercalated cell. PT-VCAM1, proximal tubule cell that expresses VCAM1. PEC, parietal epithelial cell. LEUK, leukocyte. PODO, podocyte. ENDO, endothelial cell. LH, loop of Henle. PT, proximal tubule. FIB, fibroblast. MC-VMSC, mesangial and vascular smooth muscle cell.

**Figure 4.** Variable importance for fast progression to end-stage kidney disease (ESKD) among 46 proteins in the combined cohorts of type 1 (T1D) and type 2 diabetes (T2D) cohorts (*n=* 405), using random forest analysis for 500 trees. The higher the value of mean decrease Gini Index, the higher the importance of the variable in the model. Gini Index is technically a by-product in the training of the random forest classifier, and provides a relative ranking of the spectral features. At each node of the random forest, the optimal split is sought using the Gini Index, indicating how often a particular feature was selected for a split, and how large its overall discriminative value was for the classification problem under study. In the plot shown above, KIM-1 is most important protein among 46 for predicting fast progression to ESKD. ACR, albumin to creatinine ratio.

**Figure 5.** Evaluation of predictive ability for fast progression to end-stage kidney disease (ESKD), using multivariable logistic regression models, with selected proteins from 46, in addition to eGFR, urinary ACR, and HbA1c in the combined cohorts of Type 1 diabetes (T1D) and Type 2 diabetes (T2D) (n= 405). (A) Paths of regression coefficient for 9 selected proteins shrinking toward zero using penalized Lasso logistic regression. Maximum likelihood estimates absolute values of regression coefficients (on the vertical axis). Degree of shrinkage depends on values of lambda, and the minimum values of lambda are applied to determine the final results for Lasso. Each curve corresponds to a protein selected as a result of shrinkage for selection, and draws shrinkage during estimation of regression coefficient. Lasso penalizes the sum of the absolute values of regression coefficients, and a predictor with a coefficient of zero was excluded from the model and was not presented in the figure.

ACR, albumin to creatinine ratio. log, natural logarithm. (B) ROC curves for the predictive performance of Initial Model (blue) and New Model (red). Using logistic regression model for onset of end-stage kidney disease (ESKD) within 15 years, Initial Model includes eGFR, urinary ACR, HbA1c and type of cohort (T1D or T2D). New Model includes 9 selected proteins in Figure 4A (KIM-1, TNF-R27, IL-1RT1, TNF-R11A, CRELD2, PI-3, LAYN, CDH3, and TNF-R6B) in addition to the variables in the Initial Model. Youden Index is commonly used overall measure of test accuracy and calculated as following equation: Sensitivity+Specificity-1 (1 is perfect prediction). Positive and significant NRI value suggests that the new model classified patients into the correct risk strata compared to the old model. ROC, receiver-operating characteristics. AUC, area under the curve. PPV, positive predictive value. NPV, negative predictive value.