medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Multi-scalar data integration links glomerular angiopoietin-tie signaling pathway activation with progression of diabetic kidney disease

Jiahao Liu[1†], Viji Nair[1], Yi-yang Zhao[2], Dong-yuan Chang[2], Felix Eichinger[1], Emily C. Tanner[1], Damian Fermin[1], Keith A. Bellovich[3], Susan Steigerwalt[4], Zeenat Bhat[5], Jennifer J. Hawkins[1], Lalita Subramanian[1], Sylvia E. Rosas[6], John R. Sedor[7], Miguel A. Vasquez[8], Sushrut S. Waikar[9], Markus Bitzer[1], Subramaniam Pennathur [1], Frank Brosius[10], Min Chen[2], Matthias Kretzler*[1,11], Wenjun Ju*[1,11], for the Kidney Precision Medicine Project and Michigan Translational Core C-PROBE Investigator Group.

[1] Division of Nephrology, Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA.

[2] Renal Division, Department of Medicine, Peking University First Hospital, Beijing, China

[3] Nephrology Sub-section, Ascension St John, Detroit, MI, USA.

[4] Division of cardiovascular medicine, Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA.

[5] Department of Nephrology and Hypertension, Department of Medicine, Wayne State University, Detroit, MI, USA.

[6] Kidney and Hypertension Unit, Joslin Diabetes Center and Harvard Medical School, Boston, MA, USA.

[7] Department of Medicine, Cleveland Clinic, Cleveland, OH, USA.

[8] Department of Internal Medicine, UT Southwestern Medical Center, Dallas, TX, USA.

[9] Department of Medicine, Section of Nephrology, Boston University School of Medicine and

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Boston Medical Center, Brookline, MA, USA.

[10] Division of Nephrology, The University of Arizona College of Medicine Tucson, Tucson, AZ, USA.

[11] Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA.

†Currently at Department of Nephrology, The Second Xiangya Hospital, Central South University, Changsha, Hunan Province, China.

***Corresponding authors:**

MATTHIAS KRETZLER, M.D.
Warner-Lambert/Parke-Davis Professor of Medicine
Nephrology/Internal Medicine and
Computational Medicine and Bioinformatics
University of Michigan
MSRB II, 4544-D
1150 W. Medical Center Dr. Ann Arbor, MI 48109
734-615-5757
fax: 734-763-0982
E-MAIL: kretzler@umich.edu

WENJUN JU, PhD
Research Scientist
Nephrology/Internal Medicine and
Computational Medicine and Bioinformatics
Michigan Diabetes Research and Training Center
University of Michigan
MSRB II, 4544-D
1150 W. Medical Center Dr. Ann Arbor, MI 48109
734-615-0480
Email: wenjunj@med.umich.edu

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Abstract

Diabetes is the leading cause of chronic kidney disease. Prognostic biomarkers reflective of underlying molecular mechanisms are critically needed for effective management of diabetic kidney disease (DKD). In the Clinical Phenotyping and Resource Biobank study, an unbiased, machine learning approach identified a three-marker panel from plasma proteomics which, when added to standard clinical parameters, improved the prediction of outcome of end-stage kidney disease (ESKD) or 40% decline in baseline glomerular filtration rate (GFR) in a discovery DKD group (N=58) and was validated in an independent group (N=68) who also had kidney transcriptomic profiles available. Of the three markers, plasma angiopoietin 2 (ANGPT2) remained significantly associated with composite outcome in 210 Chinese Cohort Study of Chronic Kidney Disease participants with DKD. The glomerular transcriptional Angiopoietin/Tie (ANG-TIE) activation scores, derived from the expression of 154 literature-curated ANG-TIE signaling mediators, positively correlated with plasma ANGPT2 levels and outcome, explained by substantially higher TEK receptor expression in glomeruli and higher ANG-TIE activation scores in endothelial cells in DKD by single cell RNA sequencing. Our work suggests that activation of glomerular ANG-TIE signaling in the kidneys underlies the association of plasma ANGPT2 with disease progression, thereby providing potential targets to prevent DKD progression.

Keywords

Prognostic biomarker, Angiopoietin 2, glomerular endothelial cells, diabetic kidney disease

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Introduction

Diabetes is a primary cause of chronic kidney disease (CKD) (1). Early identification of patients at high risk of disease progression is critical for effective management to prevent or delay CKD (2). A recent study suggests that biomarker testing when combined with more effective treatments present a cost-effective way to improve patient outcomes in the diabetic population (3). However, routinely used clinical measures such as estimated glomerular filtration rate (eGFR) and albuminuria have limited sensitivity and specificity for disease prognosis (4; 5). Advances in kidney research over the past decade have revealed multiple biomarker candidates based on transcriptomic (6; 7), proteomic (4) and metabolomic features (8-13) associated with progression of DKD. Due to ease of sample processing, measurement and quantification methods, non-invasive protein markers showing significant association with disease progression are attractive for clinical applications (14-18). However, statistically significant association between biomarker levels and the outcome of interest is only a first step.

A clear underpinning of molecular mechanisms connecting the biomarker with clinical outcome is particularly important when biomarkers are not directly measured in the disease-affected tissue or organ, but rather in body fluids, such as serum or plasma. The unclear origin of these circulating markers raises concerns about the relevance of these markers to kidney pathophysiology. For example, 17 proteins enriched for TNF receptor superfamily members showed strong association with risk of ESKD in patients with DKD (14), where TNFR 1 and 2, despite robust association with ESKD, were likely not produced by the kidney (19). A better understanding of the molecular mechanisms underlying the association of the biomarker with long-term clinical outcomes is needed to prioritize biomarkers for further validation, support potential clinical application for risk

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

stratification and reveal putative therapeutic targets.

Kidney biopsy transcriptomic profiling can help bridge the gap between circulating factors and kidney function impairment. Histopathological and molecular analysis of human kidney biopsy tissue has been the cornerstone of kidney disease diagnosis and prognosis and has provided insights into disease pathophysiology (20-22). Genome-wide transcriptomic profiling of kidney biopsy samples from humans and animal models have helped define disease-associated pathways and better understand DKD pathogenesis (23-25). Recent development of single cell RNA sequencing (scRNAseq) allows comprehensive transcriptomic profiling of kidney biopsies from healthy and disease at single cell-level resolution, providing valuable insights into molecular pathways previously underappreciated due to cellular heterogeneity (26; 27).

In this study, by integrating plasma proteomics and kidney bulk- and single cell-transcriptomic data, we aimed to identify specific circulating markers and delineate the intrarenal signaling cascade that may mediate the association between the circulating biomarker and DKD progression.

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Methods

The overall workflow of the study is depicted in Figure 1.

Study population

The Clinical Phenotyping and Resource Biobank Core (C-PROBE) is a multiethnic, prospective observational cohort of patients with CKD stages I-IV, collecting clinical information and biospecimens at enrollment and yearly thereafter at six sites in the United States (http://kidneycenter.med.umich.edu/cprobe)(28). The C-PROBE discovery group (A) included patients with CKD and diabetes who had three or more eGFR measurements during the follow-up period and had plasma samples available from time of enrollment. Patients with baseline eGFR < 30 mL/min/1.73m ^{2} were excluded based on advanced disease. eGFR was estimated using the CKD-EPI equation (29). Albuminuria was assessed by Albuwell Hu ELISA Kit (Ethos Biosciences, Newtown Square, PA). eGFR slopes were calculated using a linear mixed-effects models (28). Patients with eGFR slopes less than -3%/year and above 3%/year were matched for major confounders including age, sex, race, and baseline eGFR. The independent C-PROBE validation group (B) included participants with kidney biopsy transcriptomic profiles.

The Chinese Cohort Study of Chronic Kidney Disease (C-STRIDE) is a multicenter prospective cohort of CKD patients (30). Participants with: (i) diagnosis of diabetes mellitus or had fasting blood glucose ≥7 mmol/L; (ii) eGFR<60 and >30 mL/min/1.73m^2 or uACR >30mg/g for at least three months; (iii) had two or more available eGFR measurements during follow-up and (iv) had plasma samples available from the baseline visit were included. eGFR was estimated using the Modification of Diet in Renal Disease equation (31).

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Kidney outcomes and covariates

The outcome was defined as ESKD or more than 40% reduction from baseline eGFR, whichever occurred first. Participants who reached the outcome were defined as progressors. Covariates included age, gender, race, eGFR and uACR documented at study enrollment. Participants missing key covariates were not included in the analyses.

SOMAScan Measurement

SOMAmer-based proteomic assay using the SOMAscan 1.3K Plasma kits (19; 32) per standard experimental and data analysis protocols (33) were performed by the Genome Technology Access Center at the McDonnell Genome Institute at Washington University School of Medicine. Detailed protocol and analysis are included in supplementary methods.

Measurement of ANGPT2 using ELISA

Human Angiopoietin-2 Quantikine ELISA (R&D Systems, DANG20) was used according to the manufacturer's instructions. Plasma samples (1:10) and standards were assayed in duplicate. Three quality controls with high, medium, and low concentrations were included in each plate to control for inter-plate variability. Absorbance at 450 nm was read with the VersaMax Microplate Reader (Molecular Devices) and results were calculated with the SoftMax Pro software. The quality control measures, % recovery mean and % coefficient of variation (CV), were within acceptable ranges and the inter-assay coefficient of variation was below the 4.0% threshold.

Statistical analysis

The descriptive statistics are presented as median (interquartile range) or proportions. Cox

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

proportional hazard model was used to determine the association of biomarker with composite endpoint. Statistically significant biomarkers ($P<0.05$) in a univariate model were used as input for the machine learning approach. Lasso Cox method (34) was used for feature reduction (details in supplementary methods). A multivariate Cox model based on the markers selected by lasso was constructed to evaluate the performance of each marker. Model performance was further evaluated using C statistic (35) and time dependent AUC (area under the curve) for survival data (36). The performance of two competing models were compared using delta C statistic based on a perturbation-resampling method (35), and the Likelihood Ratio (LR) test for its sensitivity in capturing the differences in prediction ability. R package glmnet (37; 38) was used to construct the lasso Cox model. Cox proportional model was constructed using R package survival (Therneau, Grambsch, 2020), C statistic was calculated using R package survC1 (35), time dependent AUC was calculated using R package timeROC (36). All analyses were completed using R version 3.6.2. (R Core Team, 2019).

Gene expression data analysis

Gene expression profiling from micro-dissected kidney biopsies was performed as previously described (39) using Affymetrix 2.1 ST chips (Affymetrix, Santa Clara, CA). Probe sets were annotated to Entrez Gene IDs using custom CDF V.19 generated from the University of Michigan Brain Array group (40). Expression data were normalized across genes using robust multi-array average (RMA) (41). The Cel files can be accessed through the Gene Omnibus, GSE180395 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE180395).

RNAseq data, GSE142025, fastq files were downloaded from GEO (42; 43), reads quality

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

inspected using FastQC, qualitied reads were mapped to human reference genome GRCh38 using STAR 2.5.2b (44). The mapped reads were then examined for distribution across exons, introns, UTR, and intergenic regions using Picard Tools (https://broadinstitute.github.io/picard/). Gene level expression quantification was performed using HTSeq (45), and the resultant count data were normalized in voom (46). PCA and hierarchical clustering were used to identify and remove samples with abnormal expression profiles due to technical issues, and the mapping statistics obtained from STAR.

scRNASeq data from 10 patients with DKD and 18 living kidney transplant donors (LD) were downloaded from Kidney Precision Medicine Project (KPMP) kidney tissue atlas (https://atlas.kpmp.org/repository) and processed according to the KPMP single cell protocol as previously described (18; 47). In the dot plot analysis, the size of the dots represents the percentage of cells expressing *TEK* (*TEK* (+) cells). *TEK* (+) expression was defined based on greater than zero (*TEK* > 0) normalized gene expression, remaining cells were considered *TEK*(-).

TEK/TIE2 gene expression in heathy donors were extracted from Nephroseq database (https://www.nephroseq.org/) using the Lindenmeyer normal kidney tissue panel (48).

Gene set curation, network visualization and functional module detection

The union of ANG-TIE signaling pathway genes curated from three databases, namely the "Angiopoietin receptor Tie2-mediated signaling pathway" from the Pathway Interaction Database (PID) (49), the "Tie-2 signaling pathway" from the REACTOME database (50) and "TIE2-angiopoietin signaling pathway" from the NETPATH database (51) provided a set of 154 genes

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

(Figure S1). Cytoscape (52) was used to visualize the genes and their database sources. Community clustering in HumanBase database (53) was applied to the ANG-TIE signatures within the kidney functional network to identify cohesive gene modules.

Creation of gene-set based pathway activation score

Log_2 transformed gene expression at bulk transcriptome level was used to compute Z-scores, and the average of Z-scores of the 154 pathway gene signature was used to generate an ANG-TIE pathway activation score for each participant (54). The AddModuleScore function in R package Seurat (55; 56) was used to generate ANG-TIE pathway scores at the single cell cluster level. The pathway scores were also grouped into glomerular (podocytes, mesangial cells, vascular smooth muscle, and endothelial cells) and tubulointerstitial (proximal tubular, descending and ascending thin limb, think ascending limb, connecting tubular, intercalated and principal) cells based on cellular origin.

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Results

Circulating biomarkers associated with DKD progression

Fifty-eight C-PROBE participants were included in the discovery group (C-PROBE group A) and their baseline as well as follow-up characteristics are summarized in Table 1. Mean age at enrollment was 49.54 (13.74) years. Participants had impaired baseline kidney function, with mean eGFR of 49.54 (13.74) ml/min/1.73m^2, and median uACR 167.4 (IQR =683.09) mg/g. After 4.4 (IQR=3) years of follow-up, 28 of 58 patients reached the outcome.

Using the Cox proportional hazard model, 84 plasma proteins of 1301 measured on the SOMAScan platform were univariately associated with time to reaching the composite endpoint ($p<0.05$, Figure 2a, Supplementary Table S1). Epidermal growth factor receptor (EGFR), apolipoprotein A1 (APOA1), cathepsin V (CTSV), c-type lectin domain family 4 member M (CLEC4M) and TNF receptor superfamily member 1A (TNFRSF1A) were among the markers that showed the strongest associations with outcome. Many of the 84 proteins were intercorrelated, with Pearson correlation coefficients ranging from -0.63 to 0.96 (Figure S2a). Correlation matrix of those proteins ordered by hierarchical clustering showed distinct correlation patterns (Supplementary Figure S2b) suggesting that those proteins may represent different aspects of pathogenesis of DKD progression. The intercorrelated expression pattern suggested the need for a feature selection approach to reduce excessive multicollinearity.

Predictive biomarker panel for DKD progression

To reduce dimensionality and construct an actionable marker panel that could jointly predict DKD progression, a feature selection procedure (57-59) was applied to the 84 proteins (cross-

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

validation curve provided in Supplementary Figure S2c). A panel of 3 biomarkers (EGFR, CLEC4M and ANGPT2) was selected using this procedure. The markers were evaluated together with clinical variables (Figure 2b), with CLEC4M (Hazard ratio (HR)=0.076, p=0.005) and ANGPT2 (HR=3.39, p=0.02) remaining statistically significant in the multivariate Cox proportional hazard model. Addition of the biomarker panel significantly improved the prediction of DKD progression (LR test p =0.003) over clinical variables alone (model 0), with C statistic improving from 0.728 to 0.791 for the joint model including clinical variables and the selected biomarker panel (model 2, Table 2). Time-dependent ROC (receiver operating characteristic) curve for the Cox model (truncated at 5 years) also demonstrated that the lasso selected biomarker panel could improve the prediction when added to clinical variables, with an increase in AUC from 0.704 to 0.806 (Figure S3).

Validation of the additive value of the biomarker panel in an independent group of patients

To evaluate the performance of the 3-marker panel in an independent group of patients, concentrations of the identified biomarkers were measured by SOMAScan assay in plasma samples of the C-PROBE validation group B (N=68, Table 1). Group B included patients with transcriptomic data derived from kidney biopsies. In comparison with patients from Group A, patients from Group B were younger, had higher eGFR and uACR at baseline, and fewer patients were black. Mean eGFR at enrollment was 65.61 (34.43) ml/min per 1.73m2, and median uACR was 816.2 (1680.8) mg/g. Twenty-nine participants reached the outcome after a median follow-up of 3.6 years.

The three-marker panel showed significant improvement in prediction of progression to composite

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

endpoint with an increase of C statistic from 0.671 (model 0, clinical model) to 0.787 for model 2 (the joint clinical and biomarker model) with a significant LR test p value of 0.0004 (Table 2) in this validation group. Of the three markers, ANGPT2 remained statistically significant with HR of 3.59, namely one unit increase of ANGPT2 (in $\log_2$ scale) was significantly associated with 259% increased risk of progression given the other covariates withheld (p=0.006, Figure 2c).

Validation of ANGPT2 concentration using enzyme-linked immunosorbent assay (ELISA)

ANGPT2 stood out as a significant predictor in both the discovery (C-PROBE group A) and validation group (C-PROBE group B) after adjusting for age, sex, race, eGFR and uACR, and the other two biomarkers (Figures 2b and c). To confirm concentrations measured by SOMAScan with an independent technology, ANGPT2 in plasma samples from C-PROBE group A were measured using ELISA. The ANGPT2 concentrations measured using these two technologies were strongly correlated, with a correlation coefficient of 0.91 (p=2.58e-23, Figure 3a). The association of ANGPT2 with outcome was replicated (HR: 2.11 (1.30-3.41)). When stratified by ANGPT2 quartiles measured by ELISA, the risk of reaching the outcome increased by 8.04-fold (95% CI: 2.13-30.38, p=0.002) in patients from quartile 4 compared to quartile 1 (Table S2).

Validation of the association of ANGPT2 with composite endpoint in an external cohort

The association of ANGPT2 with composite outcome was further validated in an independent, external, multi-center cohort of patients with DKD, the Chinese Cohort Study of Chronic Kidney Disease (C-STRIDE). The baseline characteristics of the C-STRIDE patients (N=210, Table 1), including the means of age and eGFR, and median of uACR were similar to C-PROBE group A and B. Length of follow-up was shorter in patients from C-STRIDE. Higher baseline

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

concentrations of ANGPT2, measured using ELISA, were associated with increased risk of the composite endpoint, the estimated HR (95% CI) was 2.09 (1.40 to 3.12), adjusted for age, gender, eGFR, and ACR (Figure 3b). Thus, with one unit increase of ANGPT2 (in $\log_2$ scale) and the other covariates withheld, the hazard of disease progression would increase by 109%, in C-STRIDE.

ANGPT2-associated gene network in the kidney

The mechanistic role of circulating ANGPT2 on kidney disease progression were investigated by integrating transcriptomic data of C-PROBE group B participants. To understand the ANGPT2 downstream signaling cascade in the kidney and its association with outcome, we generated the 154 ANG-TIE signaling network gene list (Table S3, Figure 4a). Community clustering of the 154 genes through Humanbase kidney functional network analysis (53) resulted in three gene cluster modules (Figure S4) that contained key processes closely related to transmembrane receptor protein tyrosine kinase signaling (in both M1 and M2), regulation of cellular response to insulin stimulus and ephrin receptor signaling pathway (M1), superoxide metabolic process and positive regulation of cell motility (M2), and cellular response to oxidative stress and regulation of apoptotic process (M3).

Association of ANG-TIE activation score with kidney outcome

The ANG-TIE pathway activation scores were calculated for every C-PROBE group B participant by aggregating the expression levels of the 154 pathway genes, in the glomerular and tubulointerstitial compartments data. The association of ANG-TIE pathway activation scores with plasma ANGPT2 levels were evaluated in a compartment-specific manner. For patients with both glomerular transcriptomic data and plasma SOMAscan data (N=32), significant positive

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

correlation between plasma ANGPT2 levels and glomerular ANG-TIE activation scores (r=0.43, p=0.01, Figure 4b) were observed. No significant correlations were observed between tubulointerstitial ANG-TIE pathway activation scores and plasma ANGPT2 levels (N=25, r=0.16, p=0.82, Figure S5a).

The glomerular ANG-TIE pathway activation score was significantly higher in progressors (those who reached composite endpoint during follow-up) compared to non-progressors (the rest of C-PROBE group B) (p=0.02, Figure 4c); no significant differences were observed in tubulointerstitial ANG-TIE scores between these two groups (p=0.51, Figure S5b).

To test the generalizability of these findings, we downloaded the kidney cortex transcriptomic data from 28 patients with biopsy-proven diabetic nephropathy (DN) (43). Twenty-seven samples passed the quality control criteria and were divided into two groups (43), early (n=6) and advanced DN (n=21). Early DN was defined as uACR between 30 and 300 mg/g, eGFR <90 mL/min/1.73 m^2, whereas advanced DN was defined as uACR >300 mg/g, eGFR <90 mL/min/1.73 m^2. Significantly higher ANG-TIE pathway activation scores were observed in the advanced compared to early DN group (p=9.98e-7, Figure 4d) in this independent, external data set.

Compartment- and cell cluster-specific ANG-TIE signaling pathway activation in the kidney

To uncover the link between glomerular ANG-TIE scores, plasma ANGPT2 and disease outcome, we evaluated the mRNA expression of *TEK*, the receptor of ANGPT2, in micro-dissected glomeruli and tubulointerstitia from biopsies of 6 healthy donors using Nephroseq data from Lindenmeyer normal kidney tissue panel (48). *TEK* expression was significantly higher in

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

glomeruli compared to tubulointerstitia (Figure 4e).

Further, ScRNAseq data from kidney biopsies of 10 participants diagnosed with DKD were accessed through the KPMP kidney tissue atlas (Table S4, https://atlas.kpmp.org/repository) (47). The previously published scRNAseq data from 18 living donors (LDs) were the controls (18). A combined analysis of 56906 cells from these two data sets demonstrated that the TEK receptor was specifically expressed in endothelial cell clusters (Figure 4f, Figure S6a), and the average intensity of *TEK* is significantly higher in patients with DKD compared to living donors (Figure 4g, figure S6b and S6c).

ANG-TIE pathway scores were calculated at the single cell level. Comparison of the pathway scores grouped by cells from glomerular and tubulointerstitial compartments (glomeruli vs tubulointerstitia) indicated a more active ANG-TIE signaling in glomeruli compared to tubulointerstitia (Figure 4h, $p<2.2e-16$), consistent with a glomerulus-enriched TEK receptor expression pattern (Figures 4e).

About half the cells in the endothelial cell clusters (EC1 and EC2) in the DKD samples expressed *TEK* (*TEK* (+) cells). The level of ANG-TIE activation score was significantly higher in *TEK* (+) compared to *TEK* (-) endothelial cells in DKD (Figure 4i).

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Discussion

Multiple circulating protein biomarkers have been previously reported as strongly associated with progression of DKD (4-18). However, the mechanistic links of these molecules to cellular events in the kidney that would explain their association with kidney damage remain unclear. This study identified and validated a plasma marker panel, including ANGPT2, that is associated with long term clinical outcomes, but more importantly, provides insights into a plausible molecular signaling cascade underlying this association through multi-scalar data integration.

An aptamer-based proteome assay, enabled identification of plasma protein markers that are associated with DKD progression, including previously reported DKD biomarkers, such as β2-microglobulin (B2M) (60), tumor necrosis factor (TNF)-receptor superfamily members TNFRSF1A, TNFRSF1B, TNFRSF9 and TNFRSF4 (19; 61-63), CD27 antigen (CD27) (64; 65) CCL2/MCP-1 (66), and ANGPT2 (67). Novel candidate biomarkers involved in disease functions, such as apoptosis (diablo IAP-binding mitochondrial protein (DIABLO) (68), HtrA serine peptidase 2 (HTRA2) (69)); inflammation (interleukin 1 beta (IL1B) (70), interleukin 22 (IL22) (71)); and angiogenesis (fibroblast growth factor 2 (FGF2) (72), notch receptor 1 (NOTCH1) (73)) were also identified. Unbiased machine learning methods narrowed the list to a three-biomarker panel (EGFR, CLEC4M, ANGPT2) that improved prediction of the composite endpoint and amongst these, ANGPT2 persisted across three independent DKD groups.

ANGPT2 is a ligand for the endothelial TEK receptor tyrosine kinase (74). It inhibits TEK/TIE2 phosphorylation and induces pro-inflammatory and pro-fibrotic genes (74; 75). ANGPT2 is stored in and rapidly released upon stimulation from endothelial cell Weibel-Palade bodies (76). The

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

release of ANGPT2 into the bloodstream is triggered by hypoxia, inflammation and high glucose levels (77). ANGPT2 is crucial to the induction of inflammation and sensitizes endothelial cells to TNF-α (78). Increased serum or plasma levels of ANGPT2 were initially associated with cardiovascular disorders (74; 79-81) but also found elevated in DN (82). ANGPT2 has been associated with increased risk of kidney function decline in DKD (67) and CKD (83), and mortality in CKD (67; 83; 84). This study, for the first time, revealed that elevated circulating ANGPT2 functions through the TIE2 receptor in intra-renal endothelial cells and thereby activates a downstream pro-inflammatory and pro-fibrotic signaling cascade. We discovered that higher TEK receptor expression in glomeruli is associated with higher glomerular ANG-TIE signaling pathway scores and that the glomerular, not tubulointerstitial, ANG-TIE signaling pathway scores significantly correlated with circulating ANGPT2 levels, arguing for a critical role of glomerular endothelial cell signaling in mediating the association of circulating ANGPT2 with adverse DKD outcomes.

Our findings are consistent with studies showing an essential role for ANG-TIE signaling in the glomerular capillaries in both physiology and disease (85). Although ANGPT2 appears to act primarily as an autocrine antagonistic regulator (76), it could function as a paracrine regulator. For example, podocyte-specific ANGPT2 overexpression in pre-clinical models increased proteinuria and endothelial cell apoptosis (86), caused endothelial glycocalyx breakdown (87) and specifically stimulated endothelial expression of chemokines and adhesion molecules (75).

Potential therapeutics such as Faricimab (88) that manipulate ANGPT2 levels have provided promising results in cancer (89) and retinopathy (90; 91). ANGPT2 or ANG-TIE signaling have

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

yet to be targeted in DKD or CKD. Based on our findings, ANGPT2 or ANG-TIE signaling pathway could also offer therapeutic targets for DKD (92; 93).

The strengths of this study include the replication of the high throughput SOMAscan results with ANGPT2 concentration measured using ELISA, and validation of findings in the independent C-STRIDE cohort. The pathway activation score generated in this study can be used to determine the ANG-TIE pathway activity in kidney biopsies in future clinical studies and preclinical models of kidney disease. Further, this 154 gene set could be a starting point for identifying urinary predictive and pharmacodynamic biomarkers for ANG-TIE associated therapeutics.

Due to the limited sample size of the discovery group A and the large number of proteins measured, the association of circulating proteins with DKD progression was only significant at the nominal p value threshold, with statistical significance lost after multiple testing. This is a well-known limitation of small sample sizes in high-dimensional data analysis (94), that we addressed in several ways. Critically, we confirmed the significant association of ANGPT2 with composite endpoint in an independent group of patients from C-PROBE (group B) and further confirmed the finding in the C-STRIDE cohort, using two different assay platforms. Additionally, since biology is key for a robust biomarker, we validated the association of the ANG-TIE pathway activation score with outcome in an independent, publicly available data set. The verification of the initial discovery in multiple, independent, and external validation groups to reinforce the reliability of the discovery, potentially compensates for the limitations of multiple-testing associated with small sample size, frequently encountered in translational and preclinical research. Nevertheless, further validation using antibody-based immunoassays assays in large and independent study cohorts are

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

warranted to develop ANGPT2 or other candidate biomarkers, identified in this study, as clinical prognostic markers.

In summary, our integrative bioinformatics analysis identified and validated circulating ANGPT2 as a potential prognostic marker for DKD progression and revealed the link between plasma ANGPT2, glomerular ANG-TIE signaling and the DKD progression outcome of ESKD or 40% loss of eGFR. Our findings support further investigations into ANG-TIE signaling targeted therapies, currently being tested in other conditions, to prevent or ameliorate DKD progression.

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Acknowledgements

This work was supported by a grant from the University of Michigan Health System and Peking University Health Sciences Center Joint Institute for Translational and Clinical Research (No. BMU2017JI001). Support provided by the China Scholarship Council (201906370288) during a visit of JL to University of Michigan is acknowledged. This study was also supported, in part, by funding from NIH/NIDDK through the George M. O'Brien Michigan Kidney Translational Core Center, grant 2P30-DK-081943; the Integrated Systems Biology Approach to Diabetic Microvascular Complications grant, R24DK082841; P30DK89503; and JDRF Center for Excellence (5-COE-2019-861-S-B). The Kidney Precision Medicine Project (KPMP), UH3-DK-114907, is a multi-year project funded by the National Institute of Diabetes, Digestive, and Kidney Diseases (NIDDK) with the purpose of understanding and finding new ways to treat chronic kidney disease (CKD) and acute kidney injury (AKI). See Supplemental Acknowledgments for consortium details. C-STRIDE is supported by a grant from China International Medical Foundation-Renal Anemia Fund, the grants from the National Natural Science Found (No. 82070748, 82090020 and 82090021). Dr. Vasquez is supported through funds from the National Institutes of Health, 5UH3DK114870-05. The authors thank Drs. Yuee Wang and Virginia Vega-Warner for their technical support.

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Conflict of Interest Statement

Dr. M. Kretzler reports grants from NIH/NIDDK in support of this manuscript. Grants and contracts outside the submitted work through the University of Michigan with NIH, Chan Zuckerberg Initiative, JDRF, AstraZeneca, NovoNordisk, Eli Lilly, Gilead, Goldfinch Bio, Janssen, Boehringer-Ingelheim, Moderna, European Union Innovative Medicine Initiative, Certa, Chinook, amfAR, Angion, RenalytixAI, Travere, Regeneron, IONIS, consulting fees through the University of Michigan from Astellas, Poxel, Janssen and UCB. In addition, Drs. Kretzler, Ju and Nair have a patent PCT/EP2014/073413 "Biomarkers and methods for progression prediction for chronic kidney disease" licensed. Dr. Rosas reports grants from NIH/NIDDK in support of this manuscript, grants, and contracts through the Joslin Diabetes Center with NIDDK, Bayer, and AstraZeneca.

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

References

1. Alicic RZ, Rooney MT, Tuttle KR: Diabetic Kidney Disease Challenges, Progress, and Possibilities. Clin J Am Soc Nephrol 2017;12:2032-2045

2. Pena MJ, Mischak H, Heerspink HJL: Proteomics for prediction of disease progression and response to therapy in diabetic kidney disease. Diabetologia 2016;59:1819-1831

3. Snider JT, Sullivan J, van Eijndhoven E, Hansen MK, Bellosillo N, Neslusan C, O'Brien E, Riley R, Seabury S, Kasiske BL: Lifetime benefits of early detection and treatment of diabetic kidney disease. PLoS One 2019;14

4. Colhoun HM, Marcovecchio ML: Biomarkers of diabetic kidney disease. Diabetologia 2018;61:996-1011

5. MacIsaac RJ, Jerums G: Diabetic kidney disease with and without albuminuria. Curr Opin Nephrol Hypertens 2011;20:246-257

6. Yamamoto CM, Murakami T, Oakes ML, Mitsuhashi M, Kelly C, Henry RR, Sharma K: Uromodulin mRNA from Urinary Extracellular Vesicles Correlate to Kidney Function Decline in Type 2 Diabetes Mellitus. Am J Nephrol 2018;47:283-291

7. Cardenas-Gonzalez M, Srivastava A, Pavkovic M, Bijol V, Rennke HG, Stillman IE, Zhang X, Parikh S, Rovin BH, Afkarian M, de Boer IH, Himmelfarb J, Waikar SS, Vaidya VS: Identification, Confirmation, and Replication of Novel Urinary MicroRNA Biomarkers in Lupus Nephritis and Diabetic Nephropathy. Clin Chem 2017;63:1515-1526

8. Kwan B, Fuhrer T, Zhang J, Darshi M, Van Espen B, Montemayor D, de Boer IH, Dobre M, Hsu CY, Kelly TN, Raj DS, Rao PS, Saraf SL, Scialla J, Waikar SS, Sharma K, Natarajan L: Metabolomic Markers of Kidney Function Decline in Patients With Diabetes: Evidence From the Chronic Renal Insufficiency Cohort (CRIC) Study. Am J Kidney Dis 2020;76:511-520

9. Saulnier PJ, Darshi M, Wheelock KM, Looker HC, Fufaa GD, Knowler WC, Weil EJ, Tanamas SK, Lemley KV, Saito R, Natarajan L, Nelson RG, Sharma K: Urine metabolites are associated with glomerular lesions in type 2 diabetes. Metabolomics : Official journal of the Metabolomic Society 2018;14:84

10. Niewczas MA, Mathew AV, Croall S, Byun J, Major M, Sabisetti VS, Smiles A, Bonventre JV, Pennathur S, Krolewski AS: Circulating Modified Metabolites and a Risk of ESRD in Patients With Type 1 Diabetes and Chronic Kidney Disease. Diabetes Care 2017;40:383-390

11. Sas KM, Kayampilly P, Byun J, Nair V, Hinder LM, Hur J, Zhang H, Lin C, Qi NR, Michailidis G, Groop PH, Nelson RG, Darshi M, Sharma K, Schelling JR, Sedor JR, Pop-Busui R, Weinberg JM, Soleimanpour SA, Abcouwer SF, Gardner TW, Burant CF, Feldman EL, Kretzler M, Brosius FC, 3rd, Pennathur S: Tissue-specific metabolic reprogramming drives nutrient flux in diabetic complications. JCI Insight 2016;1:e86976

12. Afshinnia F, Rajendiran TM, He C, Byun J, Montemayor D, Darshi M, Tumova J, Kim J,

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Limonte CP, Miller RG, Costacou T, Orchard TJ, Ahluwalia TS, Rossing P, Snell-Bergeon JK, de Boer IH, Natarajan L, Michailidis G, Sharma K, Pennathur S: Circulating Free Fatty Acid and Phospholipid Signature Predicts Early Rapid Kidney Function Decline in Patients With Type 1 Diabetes. Diabetes Care 2021;44:2098-2106

13. Afshinnia F, Nair V, Lin J, Rajendiran TM, Soni T, Byun J, Sharma K, Fort PE, Gardner TW, Looker HC, Nelson RG, Brosius FC, Feldman EL, Michailidis G, Kretzler M, Pennathur S: Increased lipogenesis and impaired beta-oxidation predict type 2 diabetic kidney disease progression in American Indians. JCI Insight 2019;4

14. Murakoshi M, Gohda T, Suzuki Y: Circulating Tumor Necrosis Factor Receptors: A Potential Biomarker for the Progression of Diabetic Kidney Disease. Int J Mol Sci 2020;21

15. Coca SG, Nadkarni GN, Huang Y, Moledina DG, Rao V, Zhang J, Ferket B, Crowley ST, Fried LF, Parikh CR: Plasma Biomarkers and Kidney Function Decline in Early and Established Diabetic Kidney Disease. J Am Soc Nephrol 2017;28:2786-2793

16. Rotbain Curovic V, Theilade S, Winther SA, Tofte N, Eugen-Olsen J, Persson F, Hansen TW, Jeppesen J, Rossing P: Soluble Urokinase Plasminogen Activator Receptor Predicts Cardiovascular Events, Kidney Function Decline, and Mortality in Patients With Type 1 Diabetes. Diabetes Care 2019;42:1112-1119

17. Lindhardt M, Persson F, Zürbig P, Stalmach A, Mischak H, de Zeeuw D, Lambers Heerspink H, Klein R, Orchard T, Porta M, Fuller J, Bilous R, Chaturvedi N, Parving HH, Rossing P: Urinary proteomics predict onset of microalbuminuria in normoalbuminuric type 2 diabetic patients, a sub-study of the DIRECT-Protect 2 study. Nephrol Dial Transplant 2017;32:1866-1873

18. Menez S, Ju WJ, Menon R, Moledina DG, Philbrook HT, McArthur E, Jia YQ, Obeid W, Mansour SG, Koyner JK, Shlipak MG, Coca SG, Garg AX, Kellum JA, Bomback AS, Kretzler M, Parikh CR, Bi TRI, Project KPM: Urinary EGF and MCP-1 and risk of CKD after cardiac surgery. Jci Insight 2021;6

19. Niewczas MA, Pavkov ME, Skupien J, Smiles A, Dom ZIM, Wilson JM, Park J, Nair V, Schlafly A, Saulnier PJ, Satake E, Simeone CA, Shah H, Qiu CX, Looker HC, Fiorina P, Ware CF, Sun JK, Doria A, Kretzler M, Susztak K, Duffin KL, Nelson RG, Krolewski AS: A signature of circulating inflammatory proteins and development of end-stage renal disease in diabetes. Nat Med 2019;25:805-+

20. Madaio MP: Renal biopsy. Kidney Int 1990;38:529-543

21. Henger A, Schmid H, Kretzler M: Gene expression analysis of human renal biopsies: recent developments towards molecular diagnosis of kidney disease. Curr Opin Nephrol Hypertens 2004;13:313-318

22. Lindenmeyer MT, Kretzler M: Renal biopsy-driven molecular target identification in glomerular disease. Pflugers Arch 2017;469:1021-1028

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

23. Hodgin JB, Nair V, Zhang H, Randolph A, Harris RC, Nelson RG, Weil EJ, Cavalcoli JD, Patel JM, Brosius FC, 3rd, Kretzler M: Identification of cross-species shared transcriptional networks of diabetic nephropathy in human and mouse glomeruli. Diabetes 2013;62:299-308

24. Nair V, Komorowsky CV, Weil EJ, Yee B, Hodgin J, Harder JL, Godfrey B, Ju W, Boustany-Kari CM, Schwarz M, Lemley KV, Nelson PJ, Nelson RG, Kretzler M: A molecular morphometric approach to diabetic kidney disease can link structure to function and outcome. Kidney Int 2018;93:439-449

25. Berthier CC, Zhang H, Schin M, Henger A, Nelson RG, Yee B, Boucherot A, Neusser MA, Cohen CD, Carter-Su C, Argetsinger LS, Rastaldi MP, Brosius FC, Kretzler M: Enhanced expression of Janus kinase-signal transducer and activator of transcription pathway members in human diabetic nephropathy. Diabetes 2009;58:469-477

26. Menon R, Otto EA, Hoover P, Eddy S, Mariani L, Godfrey B, Berthier CC, Eichinger F, Subramanian L, Harder J, Ju W, Nair V, Larkina M, Naik AS, Luo J, Jain S, Sealfon R, Troyanskaya O, Hacohen N, Hodgin JB, Kretzler M, Kpmp K: Single cell transcriptomics identifies focal segmental glomerulosclerosis remission endothelial biomarker. JCI Insight 2020;5

27. Chung JJ, Goldstein L, Chen YJ, Lee J, Webster JD, Roose-Girma M, Paudyal SC, Modrusan Z, Dey A, Shaw AS: Single-Cell Transcriptome Profiling of the Kidney Glomerulus Identifies Key Cell Types and Reactions to Injury. J Am Soc Nephrol 2020;31:2341-2354

28. Ju W, Nair V, Smith S, Zhu L, Shedden K, Song PXK, Mariani LH, Eichinger FH, Berthier CC, Randolph A, Lai JY, Zhou Y, Hawkins JJ, Bitzer M, Sampson MG, Thier M, Solier C, Duran-Pacheco GC, Duchateau-Nguyen G, Essioux L, Schott B, Formentini I, Magnone MC, Bobadilla M, Cohen CD, Bagnasco SM, Barisoni L, Lv J, Zhang H, Wang HY, Brosius FC, Gadegbeku CA, Kretzler M, Ercb CPN, Consortium PK-I: Tissue transcriptome-driven identification of epidermal growth factor as a chronic kidney disease biomarker. Sci Transl Med 2015;7:316ra193

29. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, 3rd, Feldman HI, Kusek JW, Eggers P, Van Lente F, Greene T, Coresh J, Ckd EPI: A new equation to estimate glomerular filtration rate. Ann Intern Med 2009;150:604-612

30. Gao B, Zhang L, Wang H, Zhao M: Chinese cohort study of chronic kidney disease: design and methods. Chin Med J (Engl) 2014;127:2180-2185

31. Ma Y-C, Zuo L, Chen J-H, Luo Q, Yu X-Q, Li Y, Xu J-S, Huang S-M, Wang L-N, Huang W, Wang M, Xu G-B, Wang H-Y: Modified Glomerular Filtration Rate Estimating Equation for Chinese Patients with Chronic Kidney Disease. J Am Soc Nephrol 2006;17:2937

32. Christensson A, Ash JA, DeLisle RK, Gaspar FW, Ostroff R, Grubb A, Lindstrom V, Bruun L, Williams SA: The Impact of the Glomerular Filtration Rate on the Human Plasma Proteome. Proteomics Clinical Applications 2018;12

33. Gold L, Ayers D, Bertino J, Bock C, Bock A, Brody EN, Carter J, Dalby AB, Eaton BE,

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Fitzwater T, Flather D, Forbes A, Foreman T, Fowler C, Gawande B, Goss M, Gunn M, Gupta S, Halladay D, Heil J, Heilig J, Hicke B, Husar G, Janjic N, Jarvis T, Jennings S, Katilius E, Keeney TR, Kim N, Koch TH, Kraemer S, Kroiss L, Le N, Levine D, Lindsey W, Lollo B, Mayfield W, Mehan M, Mehler R, Nelson SK, Nelson M, Nieuwlandt D, Nikrad M, Ochsner U, Ostroff RM, Otis M, Parker T, Pietrasiewicz S, Resnicow DI, Rohloff J, Sanders G, Sattin S, Schneider D, Singer B, Stanton M, Sterkel A, Stewart A, Stratford S, Vaught JD, Vrkljan M, Walker JJ, Watrobka M, Waugh S, Weiss A, Wilcox SK, Wolfson A, Wolk SK, Zhang C, Zichi D: Aptamer-based multiplexed proteomic technology for biomarker discovery. PLoS One 2010;5:e15004

34. Tibshirani R: The lasso method for variable selection in the cox model. Stat Med 1997;16:385-395

35. Uno H, Cai TX, Pencina MJ, D'Agostino RB, Wei LJ: On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. Stat Med 2011;30:1105-1117

36. Blanche P, Dartigues JF, Jacqmin-Gadda H: Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. Stat Med 2013;32:5381-5397

37. Simon N, Friedman J, Hastie T, Tibshirani R: Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. J Stat Softw 2011;39:1-13

38. Friedman J, Hastie T, Tibshirani R: Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw 2010;33:1-22

39. Cohen CD, Frach K, Schlondorff D, Kretzler M: Quantitative gene expression analysis in renal biopsies: a novel protocol for a high-throughput multicenter application. Kidney Int 2002;61:133-140

40. Dai M, Wang P, Boyd AD, Kostov G, Athey B, Jones EG, Bunney WE, Myers RM, Speed TP, Akil H, Watson SJ, Meng F: Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. Nucleic Acids Res 2005;33:e175

41. Lockstone HE: Exon array data analysis using Affymetrix power tools and R statistical software. Brief Bioinform 2011;12:634-644

42. Edgar R, Domrachev M, Lash AE: Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 2002;30:207-210

43. Fan Y, Yi Z, D'Agati VD, Sun Z, Zhong F, Zhang W, Wen J, Zhou T, Li Z, He L, Zhang Q, Lee K, He JC, Wang N: Comparison of Kidney Transcriptomic Profiles of Early and Advanced Diabetic Nephropathy Reveals Potential New Mechanisms for Disease Progression. Diabetes 2019;68:2301-2314

44. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR: STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15-21

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

45. Anders S, Pyl PT, Huber W: HTSeq-a Python framework to work with high-throughput sequencing data. Bioinformatics 2015;31:166-169

46. Law CW, Chen YS, Shi W, Smyth GK: voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol 2014;15

47. Menon R, Otto EA, Sealfon R, Nair V, Wong AK, Theesfeld CL, Chen X, Wang Y, Boppana AS, Luo J, Yang Y, Kasson PM, Schaub JA, Berthier CC, Eddy S, Lienczewski CC, Godfrey B, Dagenais SL, Sohaney R, Hartman J, Fermin D, Subramanian L, Looker HC, Harder JL, Mariani LH, Hodgin JB, Sexton JZ, Wobus CE, Naik AS, Nelson RG, Troyanskaya OG, Kretzler M: SARS-CoV-2 receptor networks in diabetic and COVID-19-associated kidney disease. Kidney Int 2020;98:1502-1518

48. Lindenmeyer MT, Eichinger F, Sen K, Anders HJ, Edenhofer I, Mattinzoli D, Kretzler M, Rastaldi MP, Cohen CD: Systematic Analysis of a Novel Human Renal Glomerulus-Enriched Gene Expression Dataset. Plos One 2010;5

49. Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, Hannay T, Buetow KH: PID: the Pathway Interaction Database. Nucleic Acids Res 2009;37:D674-679

50. Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, Sidiropoulos K, Cook J, Gillespie M, Haw R, Loney F, May B, Milacic M, Rothfels K, Sevilla C, Shamovsky V, Shorser S, Varusai T, Weiser J, Wu G, Stein L, Hermjakob H, D'Eustachio P: The reactome pathway knowledgebase. Nucleic Acids Res 2020;48:D498-D503

51. Khan AA, Sandhya VK, Singh P, Parthasarathy D, Kumar A, Advani J, Gattu R, Ranjit DV, Vaidyanathan R, Mathur PP, Prasad TS, Mac Gabhann F, Pandey A, Raju R, Gowda H: Signaling Network Map of Endothelial TEK Tyrosine Kinase. J Signal Transduct 2014;2014:173026

52. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498-2504

53. Greene CS, Krishnan A, Wong AK, Ricciotti E, Zelaya RA, Himmelstein DS, Zhang R, Hartmann BM, Zaslavsky E, Sealfon SC, Chasman DI, FitzGerald GA, Dolinski K, Grosser T, Troyanskaya OG: Understanding multicellular function and disease with human tissue-specific networks. Nat Genet 2015;47:569-576

54. Tao J, Mariani L, Eddy S, Maecker H, Kambham N, Mehta K, Hartman J, Wang W, Kretzler M, Lafayette RA: JAK-STAT Activity in Peripheral Blood Cells and Kidney Tissue in IgA Nephropathy. Clin J Am Soc Nephrol 2020;

55. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd, Hao Y, Stoeckius M, Smibert P, Satija R: Comprehensive Integration of Single-Cell Data. Cell 2019;177:1888-1902.e1821

56. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, 2nd, Treacy D, Trombetta JJ, Rotem A,

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Rodman C, Lian C, Murphy G, Fallahi-Sichani M, Dutton-Regester K, Lin JR, Cohen O, Shah P, Lu D, Genshaft AS, Hughes TK, Ziegler CG, Kazer SW, Gaillard A, Kolb KE, Villani AC, Johannessen CM, Andreev AY, Van Allen EM, Bertagnolli M, Sorger PK, Sullivan RJ, Flaherty KT, Frederick DT, Jane-Valbuena J, Yoon CH, Rozenblatt-Rosen O, Shalek AK, Regev A, Garraway LA: Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science 2016;352:189-196

57. Krstajic D, Buturovic LJ, Leahy DE, Thomas S: Cross-validation pitfalls when selecting and assessing regression and classification models. J Cheminform 2014;6

58. De Groote MA, Sterling DG, Hraha T, Russell TM, Green LS, Wall K, Kraemer S, Ostroff R, Janjic N, Ochsner UA: Discovery and Validation of a Six-Marker Serum Protein Signature for the Diagnosis of Active Pulmonary Tuberculosis. J Clin Microbiol 2017;55:3057-3071

59. Ganz P, Heidecker B, Hveem K, Jonasson C, Kato S, Segal MR, Sterling DG, Williams SA: Development and Validation of a Protein-Based Risk Score for Cardiovascular Outcomes Among Patients With Stable Coronary Heart Disease. JAMA 2016;315:2532-2541

60. Kim MK, Yun KJ, Chun HJ, Jang EH, Han KD, Park YM, Baek KH, Song KH, Cha BY, Park CS, Kwon HS: Clinical utility of serum beta-2-microglobulin as a predictor of diabetic complications in patients with type 2 diabetes without renal impairment. Diabetes Metab 2014;40:459-465

61. Ihara K, Skupien J, Krolewski B, Md Dom ZI, O'Neil K, Satake E, Kobayashi H, Rashidi NM, Niewczas MA, Krolewski AS: 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;

62. Niewczas MA, Gohda T, Skupien J, Smiles AM, Walker WH, Rosetti F, Cullere X, Eckfeldt JH, Doria A, Mayadas TN, Warram JH, Krolewski AS: Circulating TNF Receptors 1 and 2 Predict ESRD in Type 2 Diabetes. J Am Soc Nephrol 2012;23:507-515

63. Gohda T, Niewczas MA, Ficociello LH, Walker WH, Skupien J, Rosetti F, Cullere X, Johnson AC, Crabtree G, Smiles AM, Mayadas TN, Warram JH, Krolewski AS: Circulating TNF receptors 1 and 2 predict stage 3 CKD in type 1 diabetes. J Am Soc Nephrol 2012;23:516-524

64. Colombo M, Valo E, McGurnaghan SJ, Sandholm N, Blackbourn LAK, Dalton RN, Dunger D, Groop PH, McKeigue PM, Forsblom C, Colhoun HM, FinnDiane Study G, the Scottish Diabetes Research Network Type 1 Bioresource C: Biomarker panels associated with progression of renal disease in type 1 diabetes. Diabetologia 2019;62:1616-1627

65. Colombo M, McGurnaghan SJ, Blackbourn LAK, Dalton RN, Dunger D, Bell S, Petrie JR, Green F, MacRury S, McKnight JA, Chalmers J, Collier A, McKeigue PM, Colhoun HM, Scottish Diabetes Research Network Type 1 Bioresource I: Comparison of serum and urinary biomarker panels with albumin/creatinine ratio in the prediction of renal function decline in type 1 diabetes. Diabetologia 2020;

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

66. Satirapoj B, Dispan R, Radinahamed P, Kitiyakara C: Urinary epidermal growth factor, monocyte chemoattractant protein-1 or their ratio as predictors for rapid loss of renal function in type 2 diabetic patients with diabetic kidney disease. BMC Nephrol 2018;19

67. Tsai YC, Lee CS, Chiu YW, Lee JJ, Lee SC, Hsu YL, Kuo MC: Angiopoietin-2, Renal Deterioration, Major Adverse Cardiovascular Events and All-Cause Mortality in Patients with Diabetic Nephropathy. Kidney Blood Press Res 2018;43:545-554

68. Justo P, Sanz A, Lorz C, Gomez-Garre D, Mezzano S, Egido J, Ortiz A: Expression of Smac/Diablo in tubular epithelial cells and during acute renal failure. Kidney Int 2003;64:S52-S56

69. Kim J, Kim DS, Park MJ, Cho HJ, Zervos AS, Bonventre JV, Park KM: Omi/HtrA2 protease is associated with tubular cell apoptosis and fibrosis induced by unilateral ureteral obstruction. American Journal of Physiology-Renal Physiology 2010;298:F1332-F1340

70. Anders HJ: Of Inflammasomes and Alarmins: IL-1 beta and IL-1 alpha in Kidney Disease. J Am Soc Nephrol 2016;27:2564-2575

71. Wang SF, Li YB, Fan JJ, Zhang XY, Luan JY, Bian Q, Ding T, Wang YC, Wang ZY, Song P, Cui DX, Mei XB, Ju DW: Interleukin-22 ameliorated renal injury and fibrosis in diabetic nephropathy through inhibition of NLRP3 inflammasome activation. Cell Death Dis 2017;8

72. Murakami M, Simons M: Fibroblast growth factor regulation of neovascularization. Curr Opin Hematol 2008;15:215-220

73. Siekmann AF, Lawson ND: Notch Signalling and the Regulation of Angiogenesis. Cell Adhes Migr 2007;1

74. Akwii RG, Sajib MS, Zahra FT, Mikelis CM: Role of Angiopoietin-2 in Vascular Physiology and Pathophysiology. Cells 2019;8

75. Chang FC, Chiang WC, Tsai MH, Chou YH, Pan SY, Chang YT, Yeh PY, Chen YT, Chiang CK, Chen YM, Chu TS, Wu KD, Lin SL: Angiopoietin-2-induced arterial stiffness in CKD. J Am Soc Nephrol 2014;25:1198-1209

76. Fiedler U, Scharpfenecker M, Koidl S, Hegen A, Grunow V, Schmidt JM, Kriz W, Thurston G, Augustin HG: The Tie-2 ligand angiopoietin-2 is stored in and rapidly released upon stimulation from endothelial cell Weibel-Paladebodies. Blood 2004;103:4150-4156

77. He FF, Zhang D, Chen Q, Zhao Y, Wu L, Li ZQ, Zhang C, Jiang ZH, Wang YM: Angiopoietin-Tie signaling in kidney diseases: an updated review. FEBS Lett 2019;593:2706-2715

78. Fiedler U, Reiss Y, Scharpfenecker M, Grunow V, Koidl S, Thurston G, Gale NW, Witzenrath M, Rosseau S, Suttorp N, Sobke A, Herrmann M, Preissner KT, Vajkoczy P, Augustin HG: Angiopoietin-2 sensitizes endothelial cells to TNF-α and has a crucial role in the induction of inflammation. Nat Med 2006;12:235-239

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

79. Lee KW, Lip GY, Blann AD: Plasma angiopoietin-1, angiopoietin-2, angiopoietin receptor tie-2, and vascular endothelial growth factor levels in acute coronary syndromes. Circulation 2004;110:2355-2360

80. Patel JV, Lim HS, Varughese GI, Hughes EA, Lip GY: Angiopoietin-2 levels as a biomarker of cardiovascular risk in patients with hypertension. Ann Med 2008;40:215-222

81. Chong AY, Caine GJ, Freestone B, Blann AD, Lip GY: Plasma angiopoietin-1, angiopoietin-2, and angiopoietin receptor tie-2 levels in congestive heart failure. J Am Coll Cardiol 2004;43:423-428

82. Khalaf N, Helmy H, Labib H, Fahmy I, El Hamid MA, Moemen L: Role of Angiopoietins and Tie-2 in Diabetic Retinopathy. Electronic physician 2017;9:5031-5035

83. Tsai YC, Chiu YW, Tsai JC, Kuo HT, Lee SC, Hung CC, Lin MY, Hwang SJ, Kuo MC, Chen HC: Association of angiopoietin-2 with renal outcome in chronic kidney disease. PLoS One 2014;9:e108862

84. David S, John SG, Jefferies HJ, Sigrist MK, Kumpers P, Kielstein JT, Haller H, McIntyre CW: Angiopoietin-2 levels predict mortality in CKD patients. Nephrol Dial Transplant 2012;27:1867-1872

85. Gnudi L: Angiopoietins and diabetic nephropathy. Diabetologia 2016;59:1616-1620

86. Davis B, Cas AD, Long DA, White KE, Hayward A, Ku CH, Woolf AS, Bilous R, Viberti G, Gnudi L: Podocyte-specific expression of angiopoietin-2 causes proteinuria and apoptosis of glomerular endothelia. J Am Soc Nephrol 2007;18:2320-2329

87. Lukasz A, Hillgruber C, Oberleithner H, Kusche-Vihrog K, Pavenstadt H, Rovas A, Hesse B, Goerge T, Kumpers P: Endothelial glycocalyx breakdown is mediated by angiopoietin-2. Cardiovasc Res 2017;113:671-680

88. Sahni J, Patel SS, Dugel PU, Khanani AM, Jhaveri CD, Wykoff CC, Hershberger VS, Pauly-Evers M, Sadikhov S, Szczesny P, Schwab D, Nogoceke E, Osborne A, Weikert R, Fauser S: Simultaneous Inhibition of Angiopoietin-2 and Vascular Endothelial Growth Factor-A with Faricimab in Diabetic Macular Edema: BOULEVARD Phase 2 Randomized Trial. Ophthalmology 2019;126:1155-1170

89. Martin-Liberal J, Hollebecque A, Aftimos P, Jungels C, Martin-Romano P, Rodon J, Kremer JD, Zhang W, Bendell J: First-in-human, dose-escalation, phase 1 study of anti-angiopoietin-2 LY3127804 as monotherapy and in combination with ramucirumab in patients with advanced solid tumours. Br J Cancer 2020;123:1235-1243

90. Khanani AM, Patel SS, Ferrone PJ, Osborne A, Sahni J, Grzeschik S, Basu K, Ehrlich JS, Haskova Z, Dugel PU: Efficacy of Every Four Monthly and Quarterly Dosing of Faricimab vs Ranibizumab in Neovascular Age-Related Macular Degeneration: The STAIRWAY Phase 2 Randomized Clinical Trial. JAMA ophthalmology 2020;138:964-972

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

91. Khan M, Aziz AA, Shafi NA, Abbas T, Khanani AM: Targeting Angiopoietin in Retinal Vascular Diseases: A Literature Review and Summary of Clinical Trials Involving Faricimab. Cells 2020;9

92. Dei Cas A, Gnudi L: VEGF and angiopoietins in diabetic glomerulopathy: how far for a new treatment? Metabolism 2012;61:1666-1673

93. Gnudi L, Benedetti S, Woolf AS, Long DA: Vascular growth factors play critical roles in kidney glomeruli. Clin Sci (Lond) 2015;129:1225-1236

94. Konietschke F, Schwab K, Pauly M: Small sample sizes: A big data problem in high-dimensional data analysis. Stat Methods Med Res 2021;30:687-701

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Table 1. Basic demographics and clinical characteristics of participating patients from C-PROBE and C-STRIDE.

	C-PROBE Discovery group A (n=58)	C-PROBE Validation group B (n=68)	C-STRIDE (n=210)
At baseline			
Age (years)	59.90 (11.70)	43.57 (15.35)	58.52 (10.61)
Female (%)	33 (56.9)	41 (60.3)	80 (47.3)
Race black (%)	28 (48.3)	13 (19.1)	NA
eGFR (ml/min per 1.73m2)	49.54 (13.74)	65.61 (34.43)	57.22 (30.62)
uACR (mg/g)	167.4 [683.09]	816.2 [1680.8]	595.0 [1364.9]
During follow-up			
Length of follow-up(years)	4.4 (3.0)	3.6 (2.8)	1.9 (2.3)
Numbers reaching outcome (%)	28 (48.3)	29 (49.2)	74 (35.2)

Data are shown as mean (SD) or median [interquartile range, IQR] for continuous variable or n (proportions) for categorical variable.
Outcome was defined as time to composite event of developing end stage kidney disease (ESKD) or loss >40% of the baseline eGFR, whichever occurs first.
Data were complete except uACR missing for 1, 4 and 7 individuals in C-PROBE group A, C-PROBE group B, and C-STRIDE, respectively. Time of follow-up is missing for 5 people in C-PROBE group B.
Abbreviations: eGFR- estimated glomerular filtration rate; uACR- urinary albumin-to-creatinine ratio; NA – not applicable.

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Table 2. Incremental value of the biomarker panel in prediction of composite endpoint of ESKD or 40% reduction in eGFR.

	C statistic (CI)	delta C statistic (CI)[a]	P value (LR test)[a]
Group A (discovery)			
Model 0	0.728 (0.583-0.873)		
Model 1	0.773 (0.671- 0.874)		
Model 2	0.791 (0.662,0.920)	0.063 (-0.031-0.157)	0.003
Group B (validation)			
Model 0	0.671 (0.531-0.811)		
Model 1	0.703 (0.596 -0.810)		
Model 2	0.787 (0.666-0.907)	0.116 (-0.002-0.234)	0.0004

Model 0 covariates: Age, gender, race, eGFR and uACR
Model 1 covariates: three biomarkers EGFR, CLEC4M and ANGPT2
Model 2 covariates: Age, gender, race, eGFR, uACR and three biomarkers EGFR, CLEC4M and ANGPT2.
[a]delta C statistic and P value are the results comparing model 2 and model 0.
Abbreviations: CI-confidence interval; LR test- likelihood ratio test.

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Supplementary Materials:

1. Supplementary acknowledgement – KPMP
2. Supplementary methods
3. Supplementary references
4. Supplementary tables:
   - Table S1. Proteins associated with DKD progression.
   - Table S2. Association of plasma ANGPT2 (ELISA) with risk of composite outcome
   - Table S3. ANG-TIE signaling pathway related genes from three database
   - Table S4. Basic clinical data of DKD patients included in the ScRNAseq analysis
5. Supplementary Figures: S1 to S6
   - Figure S1. Curation of an unbiased ANG-TIE signaling gene set.
   - Figure S2. Correlation matrix between the univariate significant plasma proteins and the lasso cross validation curve.
   - Figure S3. Time-dependent ROC curve truncated at 5 years for clinical model and the joint clinical and biomarker model.
   - Figure S4. Functional characterization of the ANG-TIE signaling network in the kidney.
   - Figure S5. Association of tubular ANG-TIE signaling pathway score with plasma ANGPT2 level and kidney outcome (n=25).
   - Figure S6. TEK gene expression in KPMP single cell data: DKD and living kidney donors (LD).

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

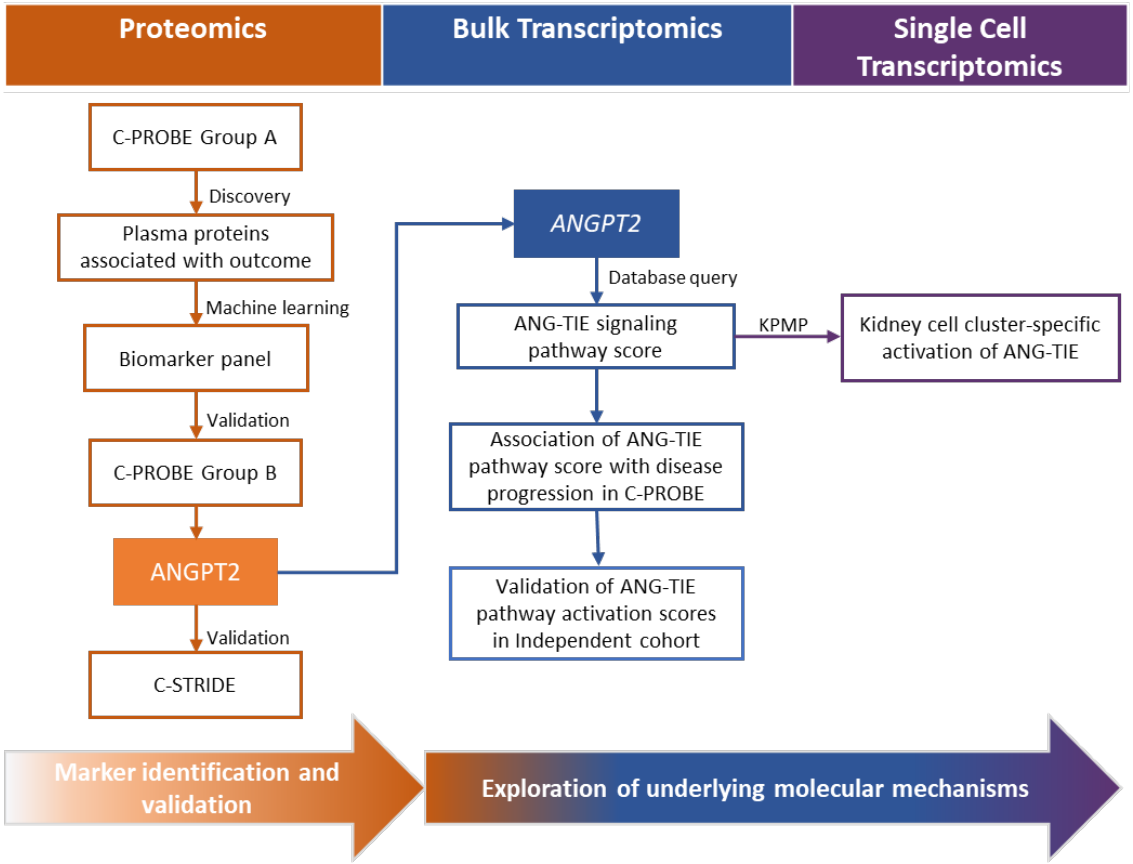


Figure 1. Overall study workflow.

SOMAScan plasma proteomics data from C-PROBE group A was used to identify protein markers associated with DKD progression. A three-marker panel that jointly predicted DKD progression was constructed using a machine learning method and validated in plasma proteomics data from C-PROBE group B. Of the three markers, ANGPT2 was further validated using ELISA in the C-STRIDE cohort. To understand the potential biological mechanisms underlying the association of ANGPT2 with kidney outcomes, an ANG-TIE signaling cascade was curated and quantified using bulk transcriptomic data. The associations of kidney ANG-TIE pathway activation scores with circulating ANGPT2 levels and kidney outcomes were evaluated, and the latter validated in an external cohort of patients with DKD (GSE142025). Cell type-specific pathway activation was determined in key kidney cell types using single cell RNAseq.

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

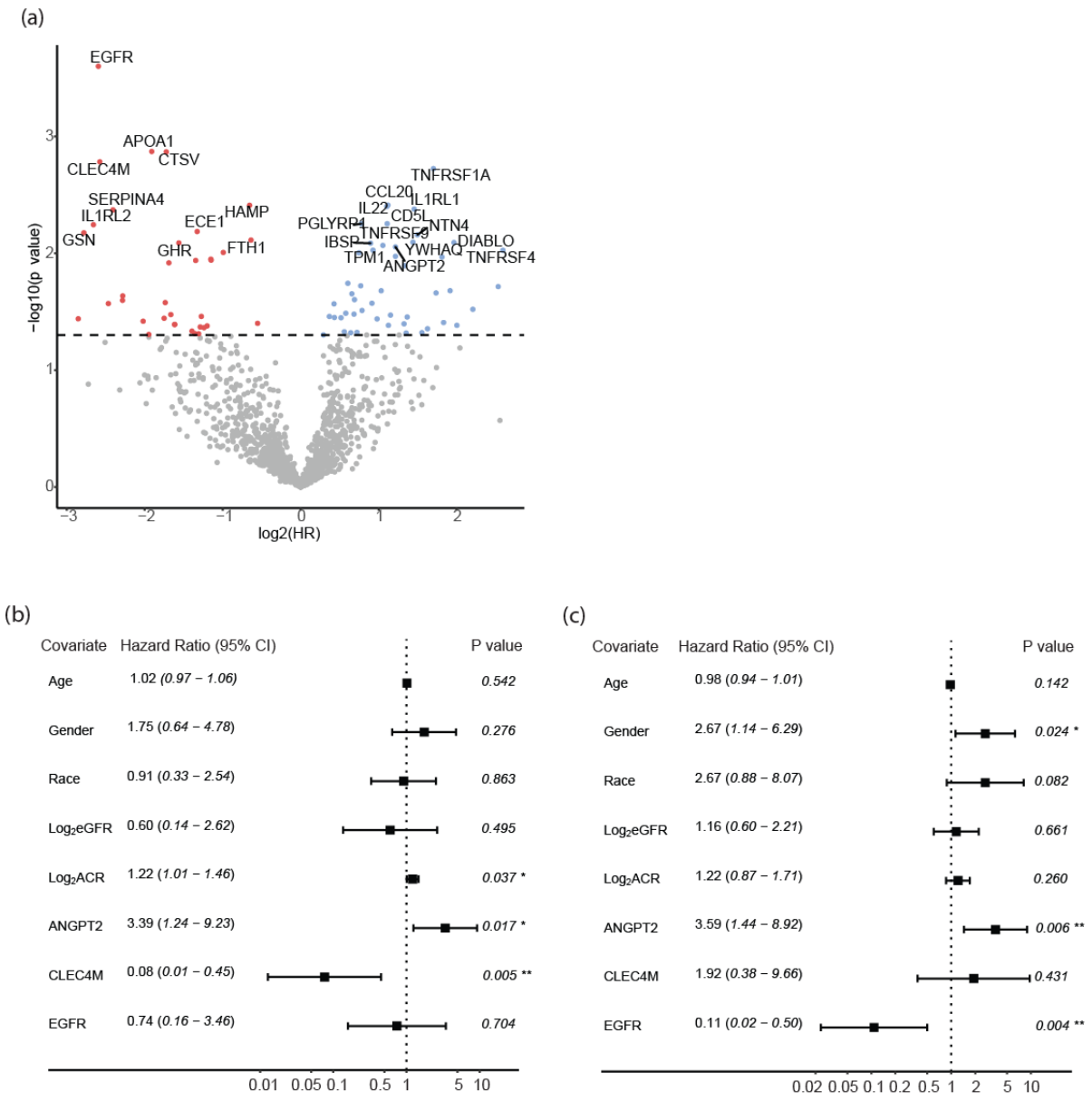


Figure 2. Identification and validation of the biomarker panel for DKD progression.

(a) Volcano plot demonstrating the associations of circulating proteins with risk of progression to the composite outcome. Hazard ratio are in $\log_2$ scale (x-axis) and statistical significance of associations are in $\log_{10}$ scale (y axis). Forest plot of the multivariate Cox proportional hazard model in C-PROBE discovery Group A (N=58) (b) and validation group B (N=68) (c).

*- p<0.05, **- p<0.01

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

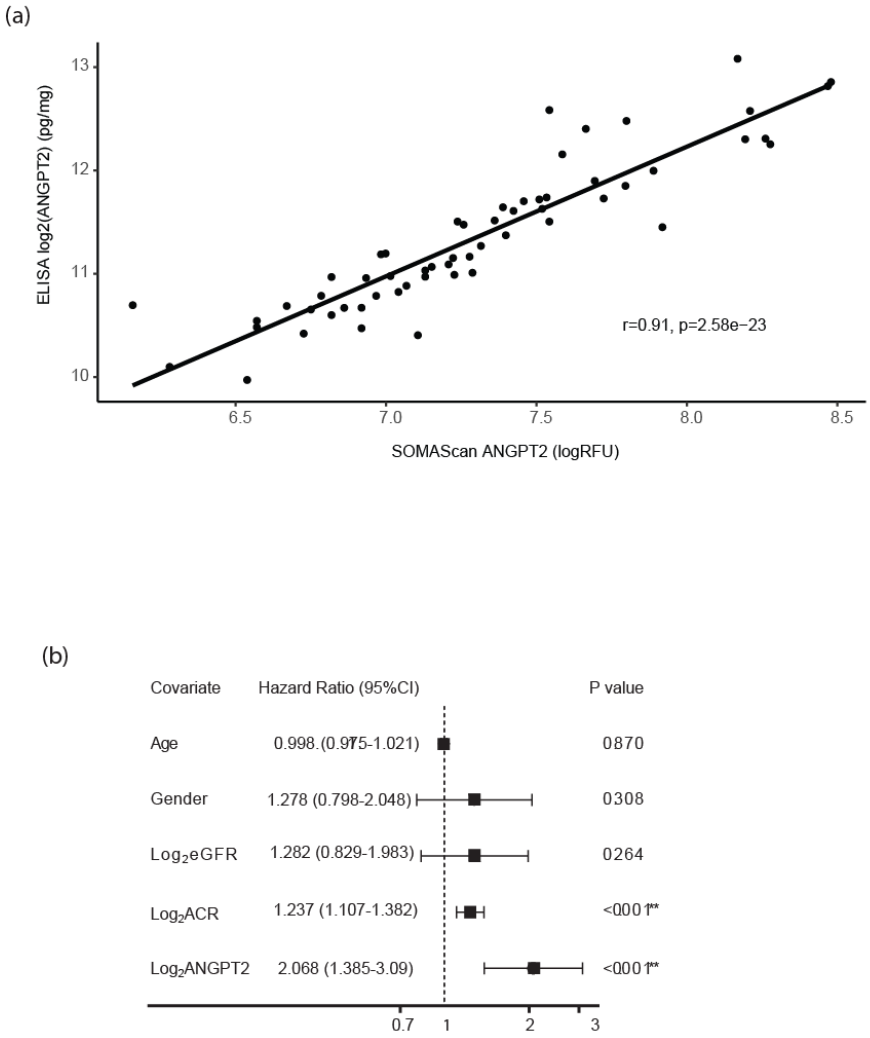
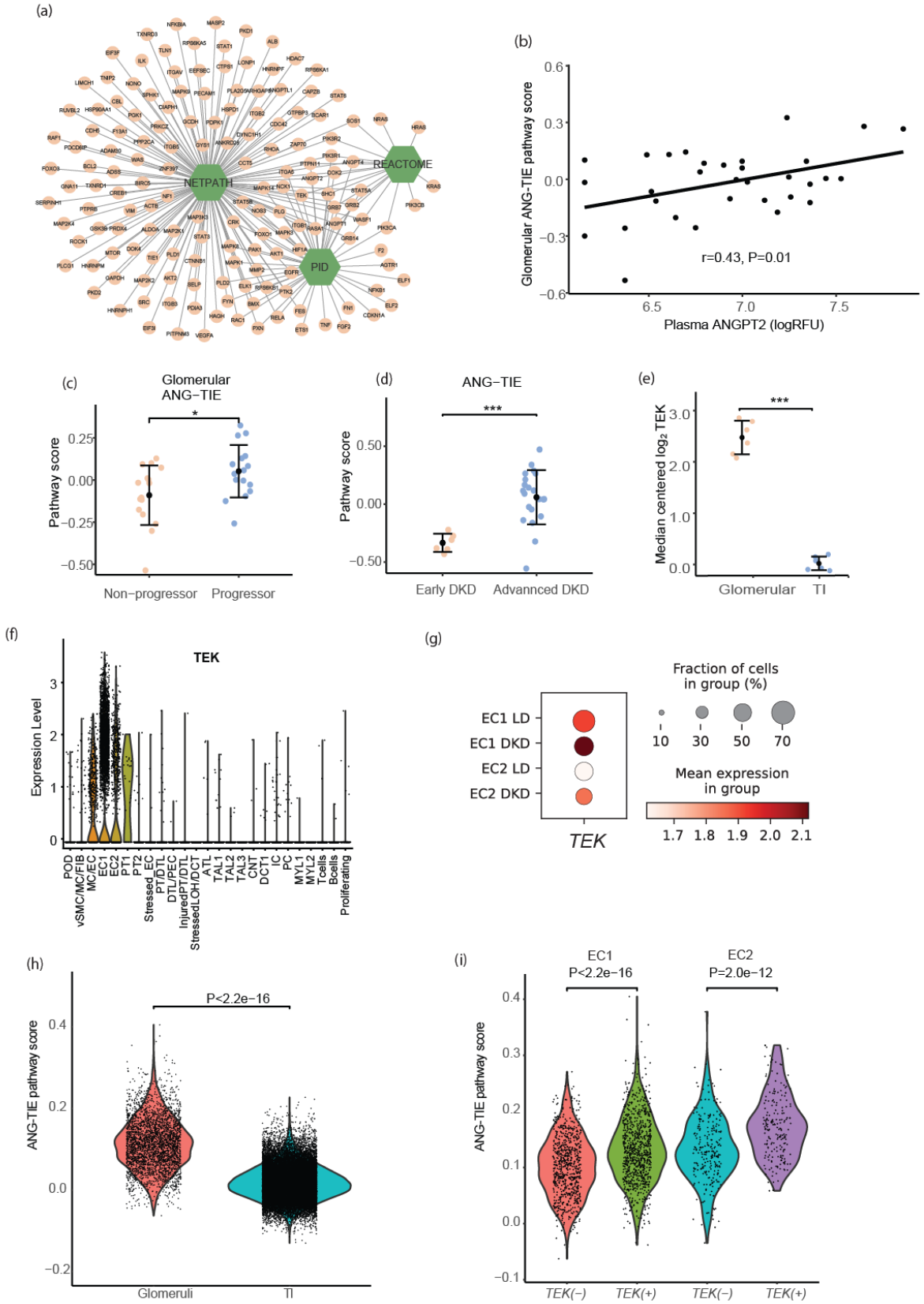


Figure 3. Validation of the significant association of plasma ANGPT2 with the risk of composite endpoint in C-STRIDE cohort.

(a) ANGPT2 concentrations measured by two different assay platforms, SOMAscan and ELISA, are significantly correlated with each other in patients from C-PROBE cohort A; (b) Forest plot showing significant association of ANGPT2 with composite endpoint in multivariate Cox proportional hazard model in DKD patients from C-STRIDE cohort where ANGPT2 concentrations were measured using ELISA. *-p<0.05, **-p<0.01

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Figure 4. ANG-TIE signaling pathway in micro-dissected transcriptome and KPMP scRNAseq data.

(a) Network visualization of the literature curated ANG-TIE signaling cascade network. (b) Association of plasma ANGPT2 with glomerular ANG-TIE signaling pathway score (n=32). (c) The glomerular ANG-TIE signaling pathway score is significantly higher in progressors compared to non-progressors in C-PROBE cohort B (n=32). (d) The ANG-TIE signaling pathway score is significantly higher in patients with advanced DKD compared to early DKD from GSE142025 dataset (n=28). Definition of early and advanced DKD is adapted from the original paper and outlined in the methods section. (e) *TEK* gene expression is significantly higher in glomeruli than in tubulointerstitia (Data derived from Lindenmeyer normal kidney tissue panel in Nephroseq). (f) ScRNAseq analysis of a combined analysis of 56906 cells from KPMP datasets demonstrate *TEK* gene expression is enriched in endothelial cells (EC), followed by MC/EC cluster, sporadic expression was also observed in PT1 cells. (g) Dot plot of *TEK* expression (higher expression indicated by increase in color intensity of the dots) in patients with DKD compared to living donors. (h) Violin plots showing ANG-TIE pathway activation scores are significantly higher in glomerular cells compared to tubulointerstitial cells of kidney biopsy samples from patients with DKD. (i) Violin plots showing the ANG-TIE pathway activation scores are significantly higher in cells expressing higher *TEK* (*TEK*-H) cells compared to cells expressing low level of *TEK* (*TEK*-L), in both endothelial cell clusters of DKD patients. For figure c, d and e, student t test was used and for panels h and i, Mann–Whitney U test was used to compare the difference between groups.

Abbreviations: TI-tubulointerstitial, ATL-ascending thin loop of Henle, CNT-connecting tubule, DCT-distal convoluted tubule, DTL-descending loop of Henle, EC-endothelial cell, FIB-fibroblast, IC-intercalated cell, LOH-loop of Henle, MC-mesangial cell, MYL-myeloid cell, PC-

medRxiv preprint doi: https://doi.org/10.1101/2021.12.21.21267377; this version posted January 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

principal cell, PEC-parietal epithelial cell, POD-podocyte, PT-proximal tubular epithelial cell, TAL-thick ascending loop of Henle, VSMC-vascular smooth muscle cell.