Identification of biomarkers for diabetic kidney disease progression via Proteonano™: a nanobinder-enabled deep plasma proteomics platform

Ban Zhao[1,2], Yonghao Zhang[3], Xiehua Ouyang[3], Le Shen[3,4], Yi Wang[3], Hao Wu[3], Yonghui Mao[1,2]

1. Department of Nephrology, Beijing Hospital, Beijing, China, 2.National Center for Genrontology, Institute of Geriatric Medicine, Chinese Academy of Medical Sciences, Beijing, China, 3. Nanomics Biotechnology Co., Ltd., Hangzhou, Zhejiang, China, 4.Department of Surgery, The University of Chicago, Chicago, USA

Nanomics
The Next-Gen Proteomics Company

北京医院 Beijing Hospital

ABSTRACT

Introduction: Diabetes is a global health challenge. One of the major complications of diabetes is diabetic kidney disease (DKD), which can progress to end-stage kidney disease that requires kidney transplant and blood dialysis. However, only a fraction of diabetes patients develops DKD, and within the DKD population, disease progression is also heterogeneous. Thus, it is imperative to identify patients with a higher likelihood of developing DKD and a higher risk for DKD progression at an early stage. However, conventional blood biomarker-based methodologies are incapable of predicting DKD progression with high sensitivity and specificity. To overcome such challenges, we developed a deep, untargeted plasma proteome profiling technology (Proteonano™ platform) to facilitate the discovery of protein biomarkers associated with DKD progression.

Methods: Proteonano™ technology was developed as an affinity-selective mass spectrometry platform, including the usage of nanoparticle-based affinity protein binders (nanobinders) to enrich low abundance proteins and the employment of an automated pre-treatment workstation for parallel sample preparation. Patients were serially enrolled under approved ethical review. Serum samples were collected at the time of enrollment, kidney functions were assessed for up to five years, and patients were stratified as progressors and non-progressors. Serum samples were processed through the Proteonano™ Lite pipeline and analyzed by a ThermoFisher Exlporis 480 mass spectrometer with FAIMS module at data-independent acquisition mode. Raw data was analyzed using DIA-NN, normalized, and further processed using a customized biostatistics and bioinformatic pipeline.

Results: Serum samples from 67 serially enrolled DKD patients with known clinical outcomes (DKD progressors, n = 38; non-progressors, n = 29) were used for Proteonano™ assisted untargeted proteomic analysis. 1393 protein groups were identified, with 951±11 (AVG±SE) protein groups identified in each sample. 1185 of them were mapped to the human plasma protein project protein catalog. Concentrations of these proteins spanned eight orders of magnitude, with lowest protein concentration of 3.0 pg/mL. Differential protein expression analysis showed 45 proteins were differentially expressed between DKD progressors and non-progressors. The most upregulated serum protein was von Willebrand factor (VWF), and the most downregulated protein was Per-Arnt-Sim kinase (PASK). Receiver operating curve (ROC) analysis showed the best performing single protein that can differentiate non-progressors and progressors was VWF, which was not superior to a widely used clinical test, urine albumin to serum creatinine ratio (UACR). Thus, multivariate analyses were subsequently conducted. Eight feature selection methods, including least absolute shrinkage and selection operator (LASSO) and random forest (RF), were employed, and the top features selected in each method were subjected to Akaiki information criteria (AIC) based model selection. Each of the AIC selected models had a ROC-AUC value higher than 0.89. Features of the best-performing model were selected by the RF method, with five proteins in the panel (ROC-AUC=0.97, 5-95 % confidence interval: 0.94-1.00), superior to the discriminative power of UACR (ROC-AUC=0.87). The discriminative power of this model was only marginally improved by including UACR as a variate (ROC-AUC=0.99, 5-95 % confidence interval: 0.97-1.00). These results demonstrate Proteonano™ nanobinder assisted sample preprocessing effectively improves untargeted proteomic analysis of blood samples, which allowed identification of a novel protein panel that effectively identifies patients with increased risk for DKD progression.

Novel aspect: Proteonano™ platform significantly improves untargeted proteomics based biomaker discovery, enabling identification of a biomarker panel that accurately predicts DKD progression.

METHOD

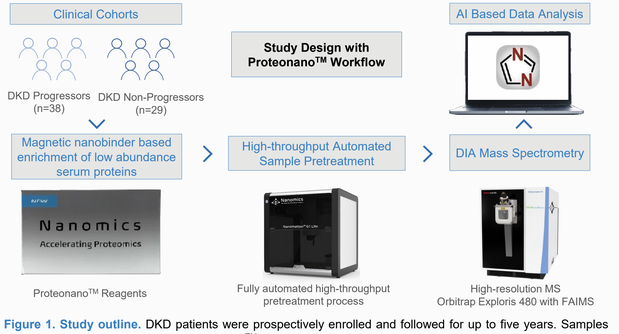

Figure 1. Study outline. DKD patients were prospectively enrolled and followed for up to five years. Samples were pre-processed by using Proteonano™ magnetic nanobinders by using an automatic work flow, and subjected to mass spectrometry (Orbitrap Exploris 480 with FAIMS, DIA data acquisition mode).

RESULTS

Proteonano™ enables deep proteomic analysis of DKD patient serum samples

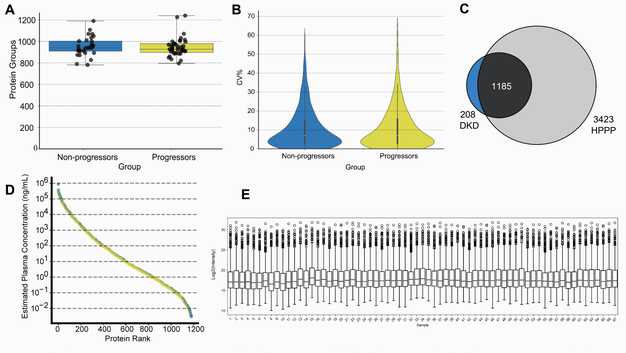

Figure 2. Sample preprocessing and LC-MS analysis. (A) Protein groups detected in serum samples from DKD patients with or without disease progression. (B). Distribution of coefficients of variation (CV) for detected protein groups in DKD non-progressors and progressors. (C) Overlap between identified protein groups in our DKD patient cohort (blue) and and those reported by Human Plasma Proteome Project (HPPP, grey). (D) Distribution of protein abundance of identified protein groups in DKD patient cohort. Protein concentrations referenced to the HPPP database. (E) Distribution of protein abundance in each of the DKD patient within the cohort prior to data normalization.

Differentially expressed porteins between DKD progressors and non-progressors

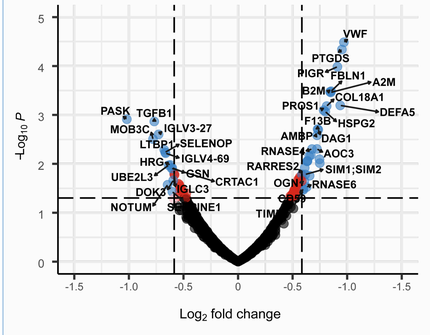

Figure 3. Differentially expressed proteins in DKD patients with and without disease progression. A total of 45 differentially expressed protein groups were identified.

Identification of pathway differences between DKD progressors and non-progressors

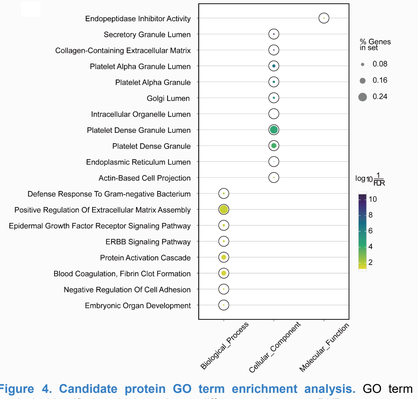

Figure 4. Candidate protein GO term enrichment analysis. GO term analysis identified multiple pathway differences between DKD progressors and non-progressors.

Functional associations among differentially expressed proteins

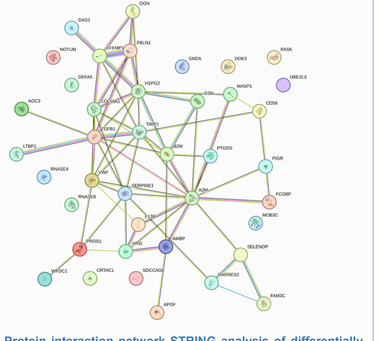

Figure 5. Protein interaction network STRING analysis of differentially expressed proteins. Different line colors between proteins represent types of associations. The length of lines represent the confidence of the interaction associations; shorter distances indicate more reliable predictions.

Predictive power of each differentially expressed protein group

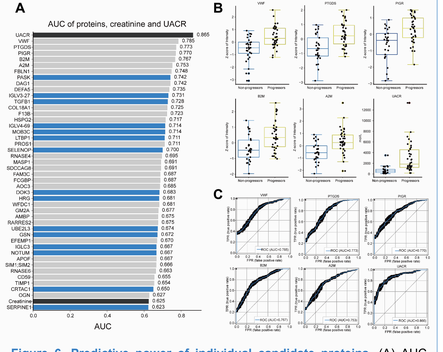

Figure 6. Predictive power of individual candidate proteins. (A) AUC values of receiver operation curves for baseline urine albumin-creatinine ratio (UACR), serum creatinine, and 45 differentially expression proteins. (B) Relative abundance of top 5 proteins with highest AUC values and UACR differentiating progressors and non-progressors. (C) ROC curves of top 5 proteins with highest AUC values and UACR. AUC values are shown on the bottom of each panel.

Identification of effective biomarker combinations to predict DKD progression

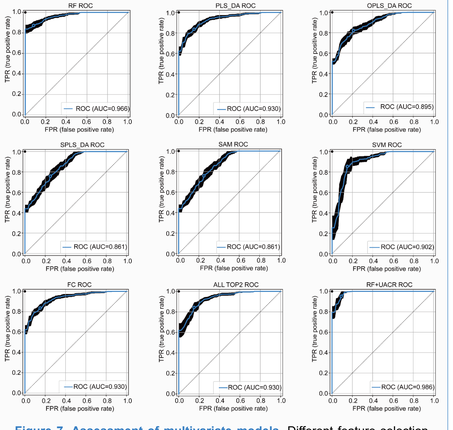

Figure 7. Assessment of multivariate models. Different feature selection methods were used to identify features that could impact predictive power for DKD progression. Akaike information criteria (AIC) was used to identify best combinations of features selected by each method, and effectiveness for each model was assessed by ROC method. Random forest method selected five features have best performance, with a small increase in predictive power when UACR value was incorporated in the analysis.

CONCLUSION AND DISCUSSION

- Utilization of Proteonano™ platform allowed deep profiling of serum proteomics in diabetic kidney disease patient samples.
- This method enabled identification of 45 differentially expressed protein groups between DKD progressors and non-progressors.
- While individual proteins have limited predictive power relative to currently used predictive biomarker (UACR), multivariate analysis of detected proteins allowed us to identify algorithms that have AUC values higher than 0.95 which combines 5 biomarkers.
- Subsequent validation studies are needed to assess the effectiveness of multivariate biomarker panel for DKD progression prediction.

ACKNOWLEDGEMENTS

This work was supported by grants from the Beijing Hospital (IIT-2021-09-03, 121-2016008), and internal fundings from Nanomics Biotech. X. Gao, J. Fang and Z. Deng (Nanomics Biotech) provided technical assistance to the study. YZ, XO, LS, YW, and HW are employees of Nanomics Biotechnology.