A Panel of Novel Biomarkers Representing Different Disease Pathways Improves Prediction of Renal Function Decline in Type 2 Diabetes

Michelle J. Pena[1], Andreas Heinzel[2], Georg Heinze[3], Alaa Alkhalaf[4,5], Stephan J. L. Bakker[6], Tri Q. Nguyen[7], Roel Goldschmeding[7], Henk J. G. Bilo[4,6], Paul Perco[2], Bernd Mayer[2], Dick de Zeeuw[1], Hiddo J. Lambers Heerspink[1] *

1 Department of Clinical Pharmacy & Pharmacology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands, 2 emergentec biodevelopment GmbH, Vienna, Austria, 3 Center For Medical Statistics, Informatics, And Intelligent Systems, Medical University of Vienna, Vienna, Austria, 4 Diabetes Centre, Isala Clinics, Zwolle, The Netherlands, 5 Department of Gastroenterology and Hepatology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands, 6 Department of Internal Medicine, Division of Nephrology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands, 7 Department of Pathology, University Medical Center Utrecht, Utrecht, The Netherlands

* h.j.lambers.heerspink@umcg.nl

Abstract

Objective

We aimed to identify a novel panel of biomarkers predicting renal function decline in type 2 diabetes, using biomarkers representing different disease pathways speculated to contribute to the progression of diabetic nephropathy.

Research Design and Methods

A systematic data integration approach was used to select biomarkers representing different disease pathways. Twenty-eight biomarkers were measured in 82 patients seen at an outpatient diabetes center in The Netherlands. Median follow-up was 4.0 years. We compared the cross-validated explained variation (R^2) of two models to predict eGFR decline, one including only established risk markers, the other adding a novel panel of biomarkers. Least absolute shrinkage and selection operator (LASSO) was used for model estimation. The C-index was calculated to assess improvement in prediction of accelerated eGFR decline defined as <-3.0 mL/min/1.73m^2/year.

Results

Patients' average age was 63.5 years and baseline eGFR was 77.9 mL/min/1.73m^2. The average rate of eGFR decline was -2.0 ± 4.7 mL/min/1.73m^2/year. When modeled on top of established risk markers, the biomarker panel including matrix metallopeptidases, tyrosine kinase, podocin, CTGF, TNF-receptor-1, sclerostin, CCL2, YKL-40, and NT-proCNP

improved the explained variability of eGFR decline (R^2 increase from 37.7% to 54.6%; p=0.018) and improved prediction of accelerated eGFR decline (C-index increase from 0.835 to 0.896; p=0.008).

Conclusions

A novel panel of biomarkers representing different pathways of renal disease progression including inflammation, fibrosis, angiogenesis, and endothelial function improved prediction of eGFR decline on top of established risk markers in type 2 diabetes. These results need to be confirmed in a large prospective cohort.

Introduction

The growing prevalence of type 2 diabetes is a great global health problem. Type 2 diabetes is the leading cause of chronic kidney disease (CKD) in the United States and is associated with high cardiovascular risk [1, 2]. Optimizing treatment has been shown to improve life expectancy, reduce costs, and lower the risk of death in patients with type 2 diabetes [3, 4]. Despite important progress in improving therapy, many patients are still at risk for renal disease.

Early identification of patients with type 2 diabetes at risk for progressive renal function loss during the early stages of disease may lead to better patient outcomes. In clinical practice, estimated glomerular filtration rate (eGFR) and albuminuria are used to assess renal function when gold-standard measured GFR is not feasible or practical. The search for novel biomarkers that improve risk prediction models on top of established risk markers has been a priority of many researchers for many years. Various studies have assessed the performance of single biomarkers representing a single, disease-associated pathway to predict progression of renal function loss in type 2 diabetes [5, 6]. However, because type 2 diabetes is a multifactorial disease, several pathways involving pro-inflammatory, pro-fibrotic, and angiogenic processes, among others, are activated during the course of the disease [7]. Given the complexity of the multiple pathophysiological processes involved in progression of type 2 diabetes together with the intra-individual variability of biomarkers, it is questionable if a single biomarker may possess useful diagnostic and prognostic power. Alternatively, a combination of biomarkers that capture different pathways of renal damage may provide a more realistic picture of a patient's actual pathophysiological status and hence may yield better assessment of disease prognosis performance.

Therefore, we aimed to identify a novel panel of biomarkers representing different disease pathways that are speculated to contribute to the progression of renal disease in type 2 diabetes, and to evaluate their combined predictive performance of accelerated renal function decline.

Research Design and Methods

Patients and methods

This observation cohort study was performed in Caucasian patients from Zwolle, The Netherlands, who participated in the PREvention of DIabetic ComplicaTIONS (PREDICTIONS) study [8]. Patients aged 35–75 with type 2 diabetes with a documented duration of ≥5 years were eligible for the PREDICTIONS study. Type 2 diabetes was defined according to World Health Organization criteria [9]. A total of 82 patients were recruited in 2007–2008 and followed for a median of 4.0 [1st, 3rd quartile 3.7 to 4.4] years. Follow-up information on urinary

albumin:creatinine ratio (UACR), serum creatinine, cholesterol, and glycated hemoglobin (HbA$_{1c}$) was obtained from electronic patient files from visits to the outpatient diabetes clinic during their annual visit to the diabetes specialist.

Ethics Statement

The PREDICTIONS study was approved by the ethical review boards of the medical ethics committees of the Isala Clinics in Zwolle and of the University Medical Center in Groningen, The Netherlands and was conducted in accordance with the guidelines of the Declaration of Helsinki. All patients gave written, informed consent.

Selection of biomarkers, sample collection, preparation, and measurement

Twenty-eight biomarkers were selected for testing using three distinct approaches, namely a literature review [10], identification of molecular processes and pathways [7], and ranking of consolidated Omics signatures [11]. A complete list of biomarkers is presented in Table 1, and the biomarker selection procedure is described in S1 Appendix.

Fasting serum and plasma samples were stored at -80°C. All samples were stored for 4–5 years and did not undergo any freeze-thaw cycles. Biomarkers were assayed on baseline samples by enzyme-linked immunosorbent assay (ELISA) or multiplex assay by Biomarker Design Forschungs GmbH (BDF), in Vienna, Austria, except for connective tissue growth factor (CTGF). CTGF was measured using specific antibodies (FibroGen Inc., San Francisco, USA) directed against distinct epitopes in the amino-terminal fragment of CTGF, as described previously [12]. All assays were used according to manufacturer's instructions. A complete list of assays, and information on stability and determination of limits of detection are available in S2 Appendix. All biomarker analyses were performed blinded, and the results were then reported back to the study center for analysis.

Statistical analysis

Analyses were performed with SAS software (version 9.2; SAS Institute, Cary, NC) and R version 3.0.2 [13] using the packages mice and glmnet [14, 15]. Data are presented as mean (standard deviation) or median [1st, 3rd quartile] for skewed variables. Graphical techniques were used to detect outliers. The natural logarithm of UACR and the binary logarithm of all biomarkers were used to normalize their distributions. Log transformed variables were used in all regression analysis. Values below the detection limit were set to the detection limit. Variables with missing values were multiply imputed using chained equations [16]. Five of the twenty-eight biomarkers had values with >10% missing or >25% below the detection limit were not used in analysis. Details on our implementation of multiple imputation can be found in S3 Appendix. All p-values were two-tailed, and values < 0.05 were considered statistically significant.

The outcome of interest was eGFR decline, defined as the within-patient annual eGFR slope. EGFR decline was calculated using a minimum of 3 serum creatinine measurements during follow-up by fitting a straight line through the eGFR values using linear regression. The eGFR value at each time-point was estimated using the 4-variable Modification of Diet in Renal Disease (MDRD) Study Equation [17].

Statistical modeling consisted of several steps. First, established risk markers were selected as best predictors of eGFR decline using least absolute shrinkage and selection operator (LASSO) selection [18]. The LASSO is advantageous for small samples sizes because it places restrictions on the absolute sizes of the regression coefficients in the model while optimally selecting the subset of variables that best predicts the outcome. This restriction also controls for

Table 1. Concentrations of biomarkers* and univariate and multivariable associations of single biomarkers with eGFR decline.

Pathway Biomarker	Concentrations	Univariate association			Multivariable association†		
	Median [1st, 3rd quartile]	β	95% CI	p-value	β	95% CI	p-value
Inflammation							
Monocyte chemoattractant protein-1 (CCL2) (pg/mL)	316.2 [258.3, 386.4]	-1.1	-3.6, 1.5	0.41	0.1	-2.0, 2.3	0.89
Tumor necrosis factor receptor-1 (TNFR1) (ng/mL)	3.8 [3.1, 6.7]	-3.2	-5.2, -1.2	**<0.01**	-2.1	-4.0, -0.3	**0.03**
Tumor necrosis factor receptor-2 (TNFR2) (pg/mL)	317.6 [248.3, 475.0]	-2.4	-3.8, -0.9	**<0.01**	-2.2	-4.6, 0.3	0.08
Chitinase 3-like 1 (YKL-40) (ng/mL)	36.5 [21.1, 87.1]	-1.1	-1.9, -0.2	**0.01**	-0.5	-1.3, 0.3	0.20
Chemokine (C-X-C motif) 1 (CXCL1) (pg/mL)	80.0 [71.1, 94.7]	1.6	-1.7, 4.8	0.33	-0.2	-3.0, 2.6	0.88
Chemokine (C-X-C motif) 10 (CXCL10) (pg/mL)	80.6 [58.1, 121.7]	-0.5	-1.7, 0.6	0.35	-0.5	-1.4, 0.5	0.35
Fibrosis							
Connective tissue growth factor (CTGF) (nmol/L)	1.0 [0.8, 1.5]	-3.4	-7.8, 0.3	0.07	-3.8	-7.7, 0.1	0.05
Matrix metallopeptidase 1 (MMP1) (pg/mL)	767.4 [470.0, 1315.5]	-0.4	-1.4, 0.7	0.49	0.2	-0.8, 1.2	0.72
Matrix metallopeptidase 2 (MMP2) (ng/mL)	38.3 [36.1, 40.0]	9.6	0.4, 18.7	**0.04**	5.6	-1.9, 13.1	0.14
Matrix metallopeptidase 7 (MMP7) (ng/mL)	1.6 [0.6, 3.0]	-1.4	-2.1, -0.7	**<0.01**	-0.8	-1.5, -0.04	**0.04**
Matrix metallopeptidase 8 (MMP8) (ng/mL)	2.4 [1.5, 4.8]	0.1	-0.8, 0.7	0.88	0.3	-0.3, 0.9	0.34
Matrix metallopeptidase 13 (MMP13) (pg/mL)	120.0 [108.5, 142.4]	-1.2	-3.0, 0.7	0.23	-0.7	-2.4, 1.0	0.40
Podocin (NPHS2) (ng/mL)	0.9 [0.3, 1.2]	-2.9	-5.0, -0.7	**0.01**	-0.9	-3.0, 1.3	0.44
Leptin (LEP) (ng/mL)	15.9 [10.1, 33.3]	0.2	-0.6, 0.9	0.66	-0.3	-1.1, 0.6	0.54
Angiogenesis							
Endostatin (Frag. COL18A1) (pmol/L)	7.6 [6.3, 9.7]	-2.6	-5.0, -0.3	**0.03**	-1.3	-4.4, 1.7	0.39
Tyrosine kinase (TEK) (pg/mL)	666.9 [315.7, 1275.8]	-0.6	-1.6, 0.3	0.18	-0.9	-1.7, -0.1	**0.03**
Vascular endothelial growth factor-A (VEGF-A) (pg/mL)	66.7 [30.6, 155.9]	-0.5	-1.1, 0.2	0.13	-0.2	-0.8, 0.4	0.49
Hepatocyte growth factor (HGF) (pg/mL)	65.9 [35.0, 120.3]	-0.7	-1.5, 0.2	0.11	-0.2	-0.9, 0.5	0.61
Endothelial Dysfunction							
Amino terminal pro C-type natriuretic peptide (NT-proCNP)(pmol/L)	2.9 [2.3, 4.3]	-1.3	-2.9, 0.3	0.12	-0.7	-2.4, 0.9	0.38
Mineral metabolism							
Fibroblast growth factor 23 (FGF23) (pmol/L)	4.0 [2.6, 5.5]	-0.8	-2.1, 0.5	0.24	-0.2	-1.3, 1.0	0.78
Sclerostin (SOST) (pmol/L)	42.7 [33.4, 52.2]	-0.1	-2.4, 2.3	0.96	0.3	-1.6, 2.3	0.75
Lipid metabolism							

(Continued)

Table 1. (*Continued*)

Pathway	Biomarker	Concentrations Median [1st, 3rd quartile]	Univariate association β	95% CI	p-value	Multivariable association† β	95% CI	p-value
Zinc-binding alpha-2-glycoprotein 1 (AZGP1) (ng/mL)		13.4 [9.0, 20.2]	-1.1	-2.5, 0.4	0.14	-0.6	-1.8, 0.7	0.37
Glomerular damage								
Growth hormone 1 (GH1) (pg/mL)		330.8 [53.4, 994.2]	-0.2	-0.7, 0.3	0.38	-0.3	-0.7, 0.1	0.14

*Concentrations of nephrin (NPHS1), neuropilin-1 (NRP1), interleukin-1 alpha (IL1A), interleukin-1 beta (IL1B), and epidermal growth factor (EGF) were missing in 10% of observations or undetectable in >25% of observations, and these biomarkers were therefore not used in analysis.

†Adjusted for established risk markers: baseline UACR, current vs. never smoker, sex, systolic blood pressure, use of oral diabetic medication, diastolic blood pressure, and baseline eGFR.

multicollinearity. LASSO involves the estimation of a tuning parameter controlling the amount of restriction, which was optimized by minimizing the leave-one-out cross-validated mean squared error of prediction. The established risk markers listed in Table 2 were considered as potential predictors of eGFR decline. All established risk markers were first included in a multivariable model using LASSO regression. The best predictors of eGFR decline were then identified from the multivariable model and are reported in the results section. Second, univariate linear models were fit for each of the novel biomarkers to assess a single biomarker association with eGFR decline. Third, multivariable models were then fit by linear regression with single novel biomarkers adjusting for the selected established risk markers. Fourth, a multivariable

Table 2. Baseline characteristics in patients with type 2 diabetes (n = 82).

Risk marker	Baseline values
Age (years)	63.5 ± 9.4
Male Gender (%)	44 (53.7)
Current smoker (%)	8 (9.6)
Body mass index (kg/m^2)	32.4 ± 6.3
Systolic blood pressure (mmHg)	135.2 ± 16.3
Diastolic blood pressure (mmHg)	72.7 ± 10.5
Duration of diabetes (years)	15.7 ± 7.3
Baseline laboratory measurements	
UACR (mg/mmol)	1.2 [0.5, 57.7]
Serum creatinine (μmol/L)	88.4 ± 33.5
eGFR (mL/min/1.73m^2)	77.9 ± 22.6
HDL Cholesterol (mmol/L)	1.3 ± 0.4
LDL Cholesterol(mmol/L)	2.0 ± 0.6
HbA$_{1c}$ (%)	7.7 ± 1.3
Medication use	
RAAS* (%)	27 (42.9)
Insulin* (%)	58 (92.1)
Oral diabetic medication* (%)	35 (55.6)

Data are reported as mean ± standard deviation or number (percent) or median [1st, 3rd quartile].

*Data available for n = 63.

model including the selected established risk markers and all biomarkers was fit using the LASSO selection in order to find the best subset of predictors.

Bootstrap validation was performed to determine the validity of the model to assess the ability of the biomarker panels to predict renal function decline. The bootstrap (N = 1000) was used to evaluate selection probabilities of each biomarker, and to construct 95% confidence intervals and two-sided *p*-values for the regression coefficients by the percentile method. A global *p*-value testing the global null hypothesis of no added value of the biomarkers was constructed by counting the number of bootstrap resamples in which the multivariable biomarker model led to a smaller cross-validated mean squared error (MSE) than a model based on the established risk markers alone. In a simple bootstrap validation, LASSO models were fit to the 1000 bootstrap resamples, each time optimizing the cross-validated MSE as described above. These models were then applied to the original data without modification. The resulting MSE was calculated by averaging the squared average difference between the original outcome and the predicted outcome for each patient. This was done for models only considering the established risk markers, and for models considering clinical and biomarker predictors. From the MSEs, R^2 measures were finally derived in order to determine whether the biomarkers significantly improved prediction.

The added value of the biomarker panel was also evaluated using the discriminative index (C-index) by dichotomizing the observed outcome variables into accelerated or non-accelerated renal function decline (eGFR decline <-3 or >-3 mL/min/1.73m^2/year, respectively) and comparing this with predicted probabilities of eGFR decline (see S3 Appendix). The C-index was also calculated using the simple bootstrap validation scheme, and the differences in the C-index between a model of only established risk markers and a model of established risk markers plus biomarkers were assessed. The threshold of -3 mL/min/1.73m^2 was based on prior studies and its concurrence with the high quartile of eGFR decline [19, 20].

Results

Baseline characteristics and association with eGFR decline

Baseline characteristics are presented in Table 2. The average age of the cohort was 63.5 (SD 9.4) years and 53.7% were male. Type 2 diabetes was well established in the study population with average diabetes duration of 15.7 (SD 7.3) years. Renal function was relatively preserved in the cohort with an average eGFR of 77.9 (SD 22.6) mL/min/1.73m^2 at baseline. Median UACR was 10.6 [1st, 3rd quartile: 4.42, 510.1] mg/g. The average rate of eGFR decline over the median of 4.0 [1st, 3rd quartile: 3.7, 4.4] years of follow-up was -2.1 (SD 4.5) mL/min/1.73m^2/year.

The following best predictors of eGFR decline were selected from the LASSO selection: baseline UACR, current vs. never smoker, sex, systolic and diastolic blood pressure, use of oral diabetic medication, and baseline eGFR (S1 Fig).

Biomarker concentrations and associations with eGFR decline

Baseline biomarker concentrations and univariate associations of the single biomarkers with eGFR decline are reported in Table 1. Higher concentrations of the individual biomarkers matrix metallopeptidases 2 (MMP2) (*p* = 0.04), matrix metallopeptidases 7 (MMP7) (*p* < 0.01), chitinase 3-like 1 (YKL-40) (*p* = 0.01), tumor necrosis factor receptor-1 (TNFR1) (*p* < 0.01), podocin (NPHS2) (*p* = 0.01), and endostatin (frag.COL18A1) (*p* = 0.03) were significantly associated with eGFR decline. When single biomarkers were modeled adjusting for established risk markers, MMP7, tyrosine kinase (TEK), and TNFR1 were independently associated with eGFR decline (Table 1). For every two-fold increase in the log concentration of MMP7, TEK,

or TNFR1, a corresponding decrease of eGFR of 0.77 ($p = 0.04$), 0.90 ($p = 0.02$), and -2.1 ($p = 0.03$) mL/min/1.73m^2/year, respectively, was observed.

When these three biomarkers were modeled on top of the established risk markers, they did not improve the explained variability (R^2) of eGFR decline (35.7% compared to 37.7% of the reference model; $p = 0.988$). The three biomarkers also did not increase the C-index for prediction of accelerated renal function decline (0.860 compared to 0.835 of the reference model; $p = 0.262$).

Selection of optimal combination of established risk markers and biomarkers

Although most individual biomarkers were not found to be independently associated with eGFR decline, we hypothesized that the combination of biomarkers representing different disease pathways may improve prediction of eGFR decline. In a multivariable LASSO selection, the optimal model for prediction of eGFR decline was achieved after inclusion of 19 variables (Fig 1). The model included a subset of 13 novel biomarkers representing fibrosis, angiogenesis, inflammation, mineral metabolism, and endothelial function that, when added to the established risk markers, more accurately predicted the rate of eGFR decline (Table 3). The explained variability of the model (R^2) markedly increased from 37.7% to 54.6% ($p = 0.018$) and predicted a higher probability of accelerated renal function decline (Fig 2). There was also a
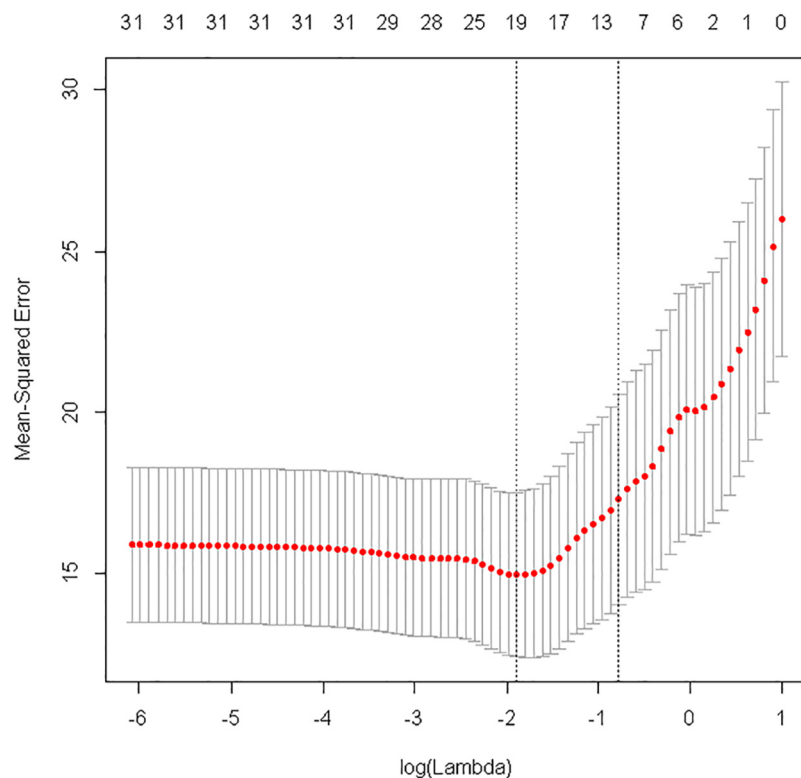

Fig 1. LASSO selection of optimal model of established risk markers and biomarkers: cross validated mean squared error (Y-axis; red bullets; MSE) vs. amount of restriction (X-axis; log(Lambda)). Vertical bars refer to standard errors across the 82 cross-validations. The best cross-validated MSE was obtained after inclusion of 19 variables (step 31), which included baseline UACR, MMP7, current vs. never smoker, sex, TEK, MMP2, systolic blood pressure, baseline eGFR, TNFR1, NPHS2, CTGF, use of oral diabetic medication, YKL-40, MMP1, MMP13, MMP8, SOST, CCL2, and NT-proCNP.

doi:10.1371/journal.pone.0120995.g001

Table 3. Optimal model of established risk markers and biomarkers, results from LASSO selection and bootstrap resampling (N = 1000).

Variable	mean β	95% CI*	p	Selection probability[†]
Baseline UACR	-0.509	-0.834, -0.159	0.002	0.999
Systolic blood pressure	0.049	0.010, 0.085	0.012	0.994
MMP2	7.382	0.010, 0.085	0.018	0.993
TEK	-0.793	-1.416, -0.139	0.018	0.993
Baseline eGFR	-0.072	-0.130, -0.014	0.026	0.987
CTGF	-5.911	-10.358, -0.913	0.026	0.987
MMP7	-0.540	-1.191, 0.0	0.078	0.966
Current vs. never smoker	-1.593	-3.905, 0.0	0.144	0.943
MMP8	0.472	0.0, 1.036	0.134	0.935
NPHS2	-1.509	-3.667, 0.0	0.206	0.908
MMP1	0.392	-0.081, 1.051	0.298	0.897
TNFR1	-1.618	-4.037, 0.0	0.228	0.889
SOST	0.983	-0.014, 2.556	0.278	0.888
Oral diabetic medication	-1.060	-2.673, 0.0	0.274	0.884
MMP13	-0.363	-1.835, 1.020	0.798	0.820
Sex	0.792	-0.905, 2.814	0.592	0.785
CCL2	0.461	-1.228, 2.672	0.854	0.781
YKL-40	-0.405	-1.358, 0.019	0.518	0.771
NT-proCNP	0.756	-0.002, 2.452	0.568	0.742

*95% confidence interval, estimated from the 2.5[th] and 97.5[th] percentiles of the bootstrap distribution.
[†]The relative frequency of the marker being included in the model across 1000 bootstrap resamples.

doi:10.1371/journal.pone.0120995.t003

significant improvement in the C-index of the optimal model for prediction of accelerated renal function decline (0.896 compared to 0.835 of the reference model; p = 0.008) (Fig 3).

To investigate the importance of each of the predictors in the optimal model, we omitted, one by one, variables from the full model. If a variable was omitted from the model, the other predictor variables could be selected instead. Only the omission of UACR or systolic blood pressure resulted in relevant inclusions of other novel biomarkers (S1 Table).

Discussion

In this study, we established that a combination of different biomarkers representing different pathways that are speculated to be involved in the progression of renal disease improves prediction of eGFR decline. Although some biomarkers were not independently associated with eGFR decline, when combined into a multi-biomarker model, the combination of biomarkers improved renal risk stratification, suggesting that these biomarkers may possess synergistic effects in predicting renal function loss.

Diabetic kidney disease is characterized by the functional impairment and structural remodeling of the kidney and is linked to the changes in the kidney. Diabetic nephropathy is well characterized by glomerular hypertrophy and hyperfiltration, inflammation of glomeruli and tubuliointerstitial regions, and reduction of cell number by apoptosis and accumulation of extracellular matrix (ECM). Each of the biomarkers selected in the optimal model has been associated with one of these pathophysiological processes involved in diabetic nephropathy.

First, chronic inflammation has long been identified in the pathogenesis of type 2 diabetes and progression of diabetic nephropathy, and inflammation is well represented by the biomarkers included in the optimal model. Tumor necrosis factor alpha is a key mediator of
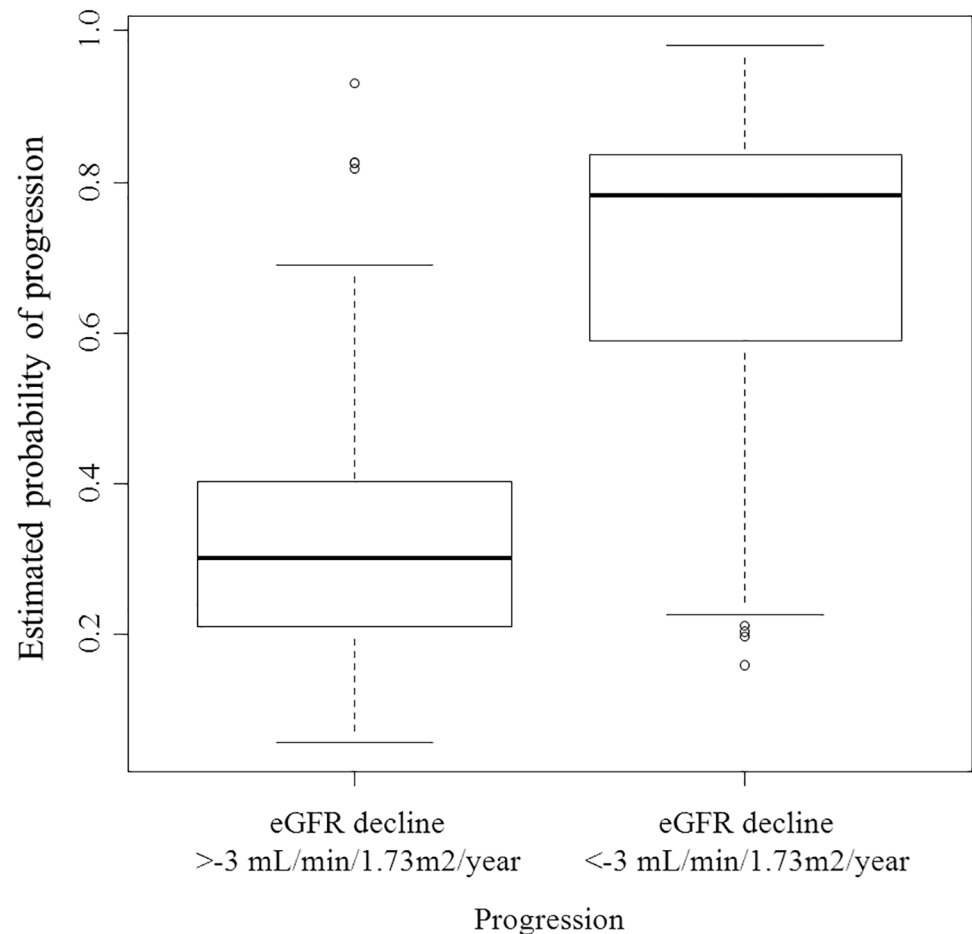
Fig 2. Predicted probability of accelerated renal function decline (eGFR decline <-3 or >-3 mL/min/ 1.73m² /year) in patients with type 2 diabetes.

inflammation and plays a role in apoptosis. It mediates its signal via two distinct receptors, TNFR1 and TNFR2. Circulating forms of both TNF receptors were recently shown to predict ESRD in type 2 diabetes[6]. Monocyte chemoattractant protein-1 (CCL2), another marker of inflammation, is a potent C-C chemokine for monocyte/macrophages and T cells. Increased amounts of CCL2 have been detected in renal biopsies and urine from patients with diabetic nephropathy [21], and CCL2 has been shown to be a marker of late stage diabetic nephropathy [22]. Currently there are a couple of clinical trials ongoing that target CCL2 receptor as a means to delay progression of diabetic nephropathy (www.clinicaltrials.gov identifier NCT01712061, NCT01752985). Results of these studies will provide more insight whether CCL2 is a causal factor or consequence of renal function loss. Additionally, YKL-40, a proinflammatory marker, has been identified as an independent factor associated with albuminuria in early stage nephropathy in type 2 diabetes and might have a useful role as a noninvasive marker for the early diabetic nephropathy detection [5, 23]. High YKL-40 levels have been shown to predict mortality in patients with type 2 diabetes [24]. Future mechanistic studies exploring the interplay between different inflammatory markers will help determine which markers are causal factors or consequences in the progression of diabetic kidney disease.

Second, the optimal model included several biomarkers linked to pro-fibrotic processes. Fibrosis, resulting from expansion and change in composition of ECM in the kidney, is a well-known
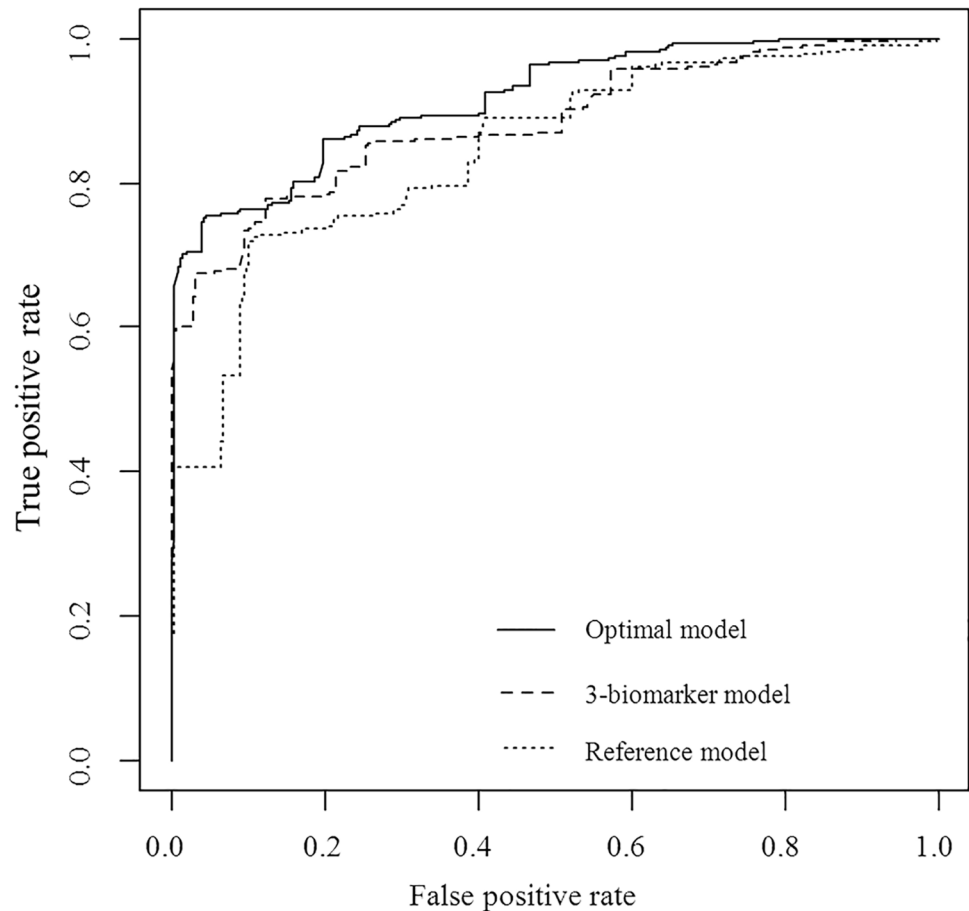
Fig 3. C-index for prediction of accelerated renal function decline (eGFR decline <-3 or >-3 mL/min/1.73m²/year) for a) established risk markers (reference model: baseline UACR, current vs. never smoker, sex, systolic and diastolic blood pressure, use of oral diabetic medication, and baseline eGFR) (C-index = 0.835), b) 3-biomarker model (MMP7, TEK, and TNFR1 on top of reference model) (C-index = 0.835; p = 0.262 compared to reference model), and c) Optimal model (baseline UACR, MMP7, current vs. never smoker, sex, TEK, MMP2, systolic blood pressure, baseline eGFR, TNFR1, NPHS2, CTGF, use of oral diabetic medication, YKL-40, MMP1, MMP13, MMP8, SOST, CCL2, and NT-proCNP) (C-index = 0.896; p = 0.008 compared to reference model).

doi:10.1371/journal.pone.0120995.g003

pathologic feature of diabetic complications. Altered expression of matrix metalloproteinases (MMPs) have been implicated in the progression of diabetic nephropathy by affecting the breakdown and turnover of ECM. In mice, the overexpression of MMP-9 has been shown to induce podocyte dedifferentiation, interrupt podocyte cell integrity, and promote podocyte monolayer permeability to albumin and extracellular matrix protein synthesis [25]. In humans, serum MMP7 has been shown to be increased in diabetic renal disease and diabetic diastolic dysfunction [26]. In support of this, our study showed that higher concentrations of MMP7 were independently associated with eGFR decline. CTGF is another well investigated pro-fibrotic biomarker that was included in the optimal model. CTGF, which is upregulated in diabetic nephropathy and contributes to extracellular matrix accumulation, has been associated with both early and late stage diabetic nephropathy [12, 22]. Down-regulation of CTGF and vascular endothelial growth factor-A (VEGF-A) in diabetic nephropathy is speculated to be a result of podocyte loss [27]. Our data, in conjunction with data from literature, support the importance of fibrotic pathways in the initiation and progression of diabetic kidney disease.

Third, we included a marker representing angiogenesis. Angiogenesis is the formation of new blood vessels from pre-existing vasculature. Neovascularization has been implicated in the genesis of diverse diabetic complications such as retinopathy, impaired wound healing, neuropathy, and diabetic nephropathy. In both physiological and pathological angiogenesis, tyrosine kinase (TEK) plays a key role. TEK is principally expressed in endothelial cells and inhibits vascular permeability and tightens preexisting vessels [28]. Additionally, TEK plays a critical role in the angiogenesis of endothelial cells via binding to angiopoietin [29].

Finally the model included a marker representing endothelial function. Endothelial dysfunction is considered an initial step of the atherosclerotic process because diabetes substantially impairs vasodilating properties of the endothelium which leads to impaired vasodilation and ultimately endothelial dysfunction [30]. C-type natriuretic peptide (CNP), a member of the natriuretic peptide family, is produced in vascular endothelium. Our study implies that natriuretic (NT)-proCNP, the N-terminal fragment of the C-type natriuretic peptide precursor, contributes to prediction of eGFR decline. NT-proCNP has been shown to be associated with arterial stiffness, endothelial dysfunction, and early atherosclerosis [31], however the link of NT-proCNP to type 2 diabetes and nephropathy is still under investigation.

In our study, most biomarkers were not able to individually predict eGFR decline after adjustment for established risk markers, and the model of 3 biomarkers did not statistically improve prediction. Rather, the optimal model of 13 biomarkers yielded best and significant improvements in the C-index. Advancing laboratory techniques allowing simultaneous measurement of many biomarkers are becoming more and more realistic in clinical practice. Whether the biomarkers identified are either involved in the causal pathway contributing to CKD progression, or are markers of its risk, or are merely the end-product of existing pathological processes, remains an important and unresolved question that requires further exploration. A future study on etiology to examine the causal relationship between these biomarkers as risk factors of renal disease would be appropriate, and issues of confounding could then be addressed. Testing for confounding was beyond the scope of this prediction study; however, we were able to investigate the importance of each of the predictors in the optimal model. Baseline UACR was found to have the largest impact on eGFR decline, and only the omission of baseline UACR or systolic blood pressure allowed inclusions of other novel biomarkers into the model. The combination of multiple biomarkers in the final, optimal model appears to be more accurate in risk stratification for accelerated renal function decline in patients with type 2 diabetes.

There are some studies in literature that use a multi-biomarker approach for risk prediction in CKD. A recent study showed that the combination of a panel of biomarkers including inflammation, fibrosis, and cardiac stretch and injury improved prediction of death in a Canadian CKD cohort; however, this study was conducted in a cohort with different CKD etiology [32]. Additionally, in another study of multiple protein biomarkers, 17 urinary and 7 plasma biomarkers were evaluated to predict progression. C-terminal FGF-23 and VEGF-A were found to be associated with the end point independent of urine albumin/creatinine. In that study many biomarkers were tested one by one, but did not use a combined biomarker approach to predict renal disease progression [33]. Furthermore, a panel of multiple urinary cytokines was found to predict rapid renal function decline in overt diabetic nephropathy [34]. However, that study included a heterogeneous population of patients with both type 1 and type 2 diabetes. Finally, in a *post-hoc* study from the IRMA-2 trial showed that multiple biomarkers of endothelial dysfunction and possibly inflammation were predictors of progression to diabetic nephropathy in patients with type 2 diabetes and microalbuminuria independent of traditional risk markers [35].

Advances in high throughput analytical methods has fueled novel biomarker discovery. Two such platforms, namely proteomics and metabolomics, have shown promise in multi-bio-marker discovery for the diabetic CKD. A urinary peptide classifier, consisting of 273 defined urinary peptides, was recently discovered as a good classifier in patients with CKD [36] and validated in an independent cohort as a predictor of albuminuria progression in patients with type 2 diabetes [37]. Furthermore, a panel of 13 metabolites linked with mitochondrial metabo-lism was significantly reduced in CKD patients with diabetes compared to healthy controls [38], and the combination of plasma metabolites butenoylcarnitine and histidine, and urine metabolites glutamine, tyrosine, and hexoses were able to predict the progression from micro- to macroalbuminuria in patients with type 2 diabetes [39].

Interestingly in our study, HbA_{1c} and duration of diabetes were not strong predictors of eGFR decline, whereas albuminuria was identified as the strongest predictor. The exclusion of HbA_{1c} and duration of diabetes from the reference model may be due to small variations in these parameters within this population. Regarding albuminuria, there is evidence that demonstrates albuminuria as a strong risk predictor of renal function loss in patients with type 2 diabetes [40–43]. Moreover, experimental data show that increased albumin exposure to the tubuli causes tubulo-interstitial damage through activation of pro-inflammatory medi-ators, which leads to a progressive decline in glomerular and tubular function, ultimately cul-minating in end-stage renal disease [44, 45]. Our data on albuminuria as a strong predictor of eGFR decline are in line with this and highlight the importance of screening for high albu-minuria to identify individuals at risk of progressive renal function loss. At the same time, it may be interesting to explore the predictive ability of urine biomarkers alongside albumin-uria for renal disease progression as urine is considered quite a suitable substrate to measure biomarkers linked to kidney disease due to the practical advantages of collecting urine com-pared to blood samples. Since our study measured biomarkers in blood, we are unable to speculate if urine biomarkers, or the combination of both blood and urine markers, would yield similar predictive capabilities.

There are strengths and limitations to this study. A clear strength is the use of a multi-marker, multi-pathway approach for identifying and testing biomarkers in a population of patients with type 2 diabetes over approximately 4 years of follow-up. The clear limitation is the measurement of multiple biomarkers in a small sample size. However, as advancing labo-ratory techniques generate larger amounts of data, methods of data analysis to accommodate "big data" with smaller sample sizes are needed. The rigorous statistical method of the LASSO regression allowed for modeling many biomarkers in the small sample size, and mul-tiple imputation was used to avoid truncating observations due to missing data. The true pre-dictive capacity of the model could have been overestimated due to the prediction model being developed and tested in the same sample, and we do agree that external validation is necessary. In the absence of external validation, we performed internal bootstrap validation in an attempt to minimize this limitation [46]. GFR was estimated using a serum creatinine-based equation instead of by direct measurement, which may have contributed to misclassifi-cation bias. However, this could have only resulted in an underestimation of the strength of the reported associations. We chose to omit five biomarkers from our analysis due to many missing or below LOD values. While the exclusion of these biomarkers from our analysis may have resulted in an underrepresentation of pathways, the omission of biomarkers could have only underestimated the predictive ability of the biomarker panel. Additional limita-tions include the lack of information concerning insulin use, diet, and renin-angiotensin-aldosterone system medication type and dose, which clearly represent unmeasured con-founders in our study.

In conclusion, novel biomarkers may provide deeper understanding into the pathophysiology of CKD or diabetic nephropathy but identification of progression-associated molecular pathways via biomarkers as proxy may also help to identify novel therapeutic targets. We identified a novel panel of biomarkers representing different pathways of renal damage, including inflammation, fibrosis, angiogenesis, and endothelial function. This combined panel improved prediction of accelerated renal function decline in patients with type 2 diabetes on top of established risk markers. The results of this study need to be validated in a large, prospective cohort to validate and assess its applicability in a broad type 2 diabetes population.

Supporting Information

S1 Appendix. Selection of biomarkers.
(DOC)

S2 Appendix. Biomarker assay information.
(DOC)

S3 Appendix. Extended statistical methods.
(DOC)

S1 Fig. LASSO selection of established risk markers.
(DOC)

S1 Table. Impact of omitting variables from the model.
(DOC)

Acknowledgments

Author Contributions

Conceived and designed the experiments: MJP HJLH AA SJLB HJGB DdZ. Performed the experiments: MJP AH GH HJLH. Analyzed the data: MJP AH GH PP BM HJLH. Contributed reagents/materials/analysis tools: AA SJLB HJGB TQN RG. Wrote the paper: MJP AH GH AA SJLB TQN RG HJGB PP BM DdZ HJLH.

References

1. United States Renal Data System, 2014 Annual Data Report: Epidemiology of Kidney Disease in the United States. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2014. Available: http://www.usrds.org/adr.aspx.

2. Gregg EW, Li Y, Wang J, Burrows NR, Ali MK, Rolka D, et al. Changes in diabetes-related complications in the United States, 1990–2010. N Engl J Med. 2014; 370: 1514–1523. doi: 10.1056/NEJMoa1310799 PMID: 24738668

3. Tseng CL, Kern EF, Miller DR, Tiwari A, Maney M, Rajan J, et al. Survival benefit of nephrologic care in patients with diabetes mellitus and chronic kidney disease. Arch Intern Med. 2008; 168: 55–62. doi: 10.1001/archinternmed.2007.9 PMID: 18195196

4. Palmer AJ, Annemans L, Roze S, Lamotte M, Lapuerta P, Chen R, et al. Cost-effectiveness of early irbesartan treatment versus control (standard antihypertensive medications excluding ACE inhibitors,

other angiotensin-2 receptor antagonists, and dihydropyridine calcium channel blockers) or late irbesartan treatment in patients with type 2 diabetes, hypertension, and renal disease. Diabetes Care. 2004; 27: 1897–1903. PMID: 15277414

5. Rondbjerg AK, Omerovic E, Vestergaard H. YKL-40 levels are independently associated with albuminuria in type 2 diabetes. Cardiovasc Diabetol. 2011; 10: 54-2840-10-54.

6. Niewczas MA, Gohda T, Skupien J, Smiles AM, Walker WH, Rosetti F, et al. Circulating TNF receptors 1 and 2 predict ESRD in type 2 diabetes. J Am Soc Nephrol. 2012; 23: 507–515. doi: 10.1681/ASN.2011060627 PMID: 22266663

7. Fechete R, Heinzel A, Perco P, Monks K, Sollner J, Stelzer G, et al. Mapping of molecular pathways, biomarkers and drug targets for diabetic nephropathy. Proteomics Clin Appl. 2011; 5: 354–366. doi: 10.1002/prca.201000136 PMID: 21491608

8. Alkhalaf A, Zurbig P, Bakker SJ, Bilo HJ, Cerna M, Fischer C, et al. Multicentric validation of proteomic biomarkers in urine specific for diabetic nephropathy. PLoS One. 2010; 5: e13421. doi: 10.1371/journal.pone.0013421 PMID: 20975990

9. World Health Organization. Dept. of Noncommunicable Disease Surveillance. Definition, diagnosis and classification of diabetes mellitus and its complications: report of a WHO consultation. Part 1, Diagnosis and classification of diabetes mellitus. Geneva, World Health Organization. 1999. Available: http://apps.who.int/iris/handle/10665/66040.

10. Hellemons ME, Kerschbaum J, Bakker SJ, Neuwirt H, Mayer B, Mayer G, et al. Validity of biomarkers predicting onset or progression of nephropathy in patients with type 2 diabetes: a systematic review. Diabet Med. 2012; 29: 567–577. doi: 10.1111/j.1464-5491.2011.03437.x PMID: 21913962

11. Stelzer G, Dalah I, Stein TI, Satanower Y, Rosen N, Nativ N, et al. In-silico human genomics with GeneCards. Hum Genomics. 2011; 5: 709–717. PMID: 22155609

12. Nguyen TQ, Tarnow L, Jorsal A, Oliver N, Roestenberg P, Ito Y, et al. Plasma connective tissue growth factor is an independent predictor of end-stage renal disease and mortality in type 1 diabetic nephropathy. Diabetes Care. 2008; 31: 1177–1182. doi: 10.2337/dc07-2469 PMID: 18344285

13. R core team. R: A language and environment for statistical computing. R foundation for statistical computing. Vienna, Austria. 2013. Available: http://www.R-project.org/. doi: 10.3758/s13428-013-0330-5 PMID: 23519455

14. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. J Stat Softw. 2011; 45:3: 1–67.

15. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010; 33: 1–22. PMID: 20808728

16. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med. 2011; 30: 377–399. doi: 10.1002/sim.4067 PMID: 21225900

17. Levey AS, Bosch JP, Lewis JB, Greene T, Rogers N, Roth D. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group. Ann Intern Med. 1999; 130: 461–470. PMID: 10075613

18. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Series B Stat Methodol. 1996; 58: 267–288.

19. Eriksen BO, Ingebretsen OC. The progression of chronic kidney disease: a 10-year population-based study of the effects of gender and age. Kidney Int. 2006; 69: 375–382. PMID: 16408129

20. Shlipak MG, Katz R, Kestenbaum B, Fried LF, Newman AB, Siscovick DS, et al. Rate of kidney function decline in older adults: a comparison using creatinine and cystatin C. Am J Nephrol, 2009; 30: 171–178. doi: 10.1159/000212381 PMID: 19349699

21. Morii T, Fujita H, Narita T, Shimotomai T, Fujishima H, Yoshioka N, et al. Association of monocyte chemoattractant protein-1 with renal tubular damage in diabetic nephropathy. J Diabetes Complications. 2003; 17: 11–15. PMID: 12623163

22. Tam FW, Riser BL, Meeran K, Rambow J, Pusey CD, Frankel AH. Urinary monocyte chemoattractant protein-1 (MCP-1) and connective tissue growth factor (CCN2) as prognostic markers for progression of diabetic nephropathy. Cytokine. 2009; 47: 37–42. doi: 10.1016/j.cyto.2009.04.001 PMID: 19409809

23. Lee JH, Kim SS, Kim IJ, Song SH, Kim YK, In Kim J, et al. Clinical implication of plasma and urine YKL-40, as a proinflammatory biomarker, on early stage of nephropathy in type 2 diabetic patients. J Diabetes Complications, 2012; 26: 308–312. doi: 10.1016/j.jdiacomp.2012.04.012 PMID: 22705282

24. Persson F, Rathcke CN, Gall MA, Parving HH, Vestergaard H, Rossing P. High YKL-40 levels predict mortality in patients with type 2 diabetes. Diabetes Res Clin Pract. 2012; 96: 84–89. doi: 10.1016/j.diabres.2011.12.008 PMID: 22209020

25. Li SY, Huang PH, Yang AH, Tarng DC, Yang WC, Lin CC, et al. Matrix metalloproteinase-9 deficiency attenuates diabetic nephropathy by modulation of podocyte functions and dedifferentiation. Kidney Int. 2014; 86:358–69. doi: 10.1038/ki.2014.67 PMID: 24670409

26. Ban CR, Twigg SM, Franjic B, Brooks BA, Celermajer D, Yue DK, et al. Serum MMP-7 is increased in diabetic renal disease and diabetic diastolic dysfunction. Diabetes Res Clin Pract. 2010; 87: 335–341. doi: 10.1016/j.diabres.2010.01.004 PMID: 20096949

27. Baelde HJ, Eikmans M, Lappin DW, Doran PP, Hohenadel D, Brinkkoetter PT, et al. Reduction of VEGF-A and CTGF expression in diabetic nephropathy is associated with podocyte loss. Kidney Int. 2007; 71: 637–645. PMID: 17264876

28. Thurston G, Rudge JS, Ioffe E, Zhou H, Ross L, Croll SD, et al. Angiopoietin-1 protects the adult vasculature against plasma leakage. Nat Med. 2000; 6: 460–463. PMID: 10742156

29. Ward NL, Dumont DJ. The angiopoietins and Tie2/Tek: adding to the complexity of cardiovascular development. Semin Cell Dev Biol. 2002; 13: 19–27. PMID: 11969368

30. Avogaro A, Fadini GP, Gallo A, Pagnin E, de Kreutzenberg S. Endothelial dysfunction in type 2 diabetes mellitus. Nutr Metab Cardiovasc Dis. 2006; 16 Suppl 1: S39–45. PMID: 16530129

31. Vlachopoulos C, Ioakeimidis N, Terentes-Printzios D, Aznaouridis K, Baou K, Bratsas A, et al. Amino-terminal pro-C-type natriuretic peptide is associated with arterial stiffness, endothelial function and early atherosclerosis. Atherosclerosis. 2010; 211: 649–655. doi: 10.1016/j.atherosclerosis.2010.03.020 PMID: 20399429

32. Levin A, Rigatto C, Barrett B, Madore F, Muirhead N, Holmes D, et al. Biomarkers of inflammation, fibrosis, cardiac stretch and injury predict death but not renal replacement therapy at 1 year in a canadian chronic kidney disease cohort. Nephrol Dial Transplant. 2014; 29: 1037–1047. doi: 10.1093/ndt/gft479 PMID: 24371297

33. Agarwal R, Duffin KL, Laska DA, Voelker JR, Breyer MD, Mitchell PG. A prospective study of multiple protein biomarkers to predict progression in diabetic chronic kidney disease. Nephrol Dial Transplant. 2014; 29:2293–302. doi: 10.1093/ndt/gfu255 PMID: 25085239

34. Verhave JC, Bouchard J, Goupil R, Pichette V, Brachemi S, Madore F, et al. Clinical value of inflammatory urinary biomarkers in overt diabetic nephropathy: a prospective study. Diabetes Res Clin Pract. 2013; 101: 333–340. doi: 10.1016/j.diabres.2013.07.006 PMID: 23880038

35. Persson F, Rossing P, Hovind P, Stehouwer CD, Schalkwijk CG, Tarnow L, et al. Endothelial dysfunction and inflammation predict development of diabetic nephropathy in the irbesartan in patients with type 2 diabetes and microalbuminuria (IRMA 2) study. Scand J Clin Lab Invest. 2008; 68: 731–738. doi: 10.1080/00365510802187226 PMID: 18609080

36. Good DM, Zurbig P, Argiles A, Bauer HW, Behrens G, Coon JJ, et al. Naturally occurring human urinary peptides for use in diagnosis of chronic kidney disease. Mol Cell Proteomics. 2010; 9: 2424–2437. doi: 10.1074/mcp.M110.001917 PMID: 20616184

37. Roscioni SS, de Zeeuw D, Hellemons ME, Mischak H, Zurbig P, Bakker SJ, et al. A urinary peptide biomarker set predicts worsening of albuminuria in type 2 diabetes mellitus. Diabetologia. 2013; 56: 259–267. doi: 10.1007/s00125-012-2755-2 PMID: 23086559

38. Sharma K, Karl B, Mathew AV, Gangoiti JA, Wassel CL, Saito R, et al. Metabolomics reveals signature of mitochondrial dysfunction in diabetic kidney disease. J Am Soc Nephrol. 2013; 24: 1901–1912. doi: 10.1681/ASN.2013020126 PMID: 23949796

39. Pena MJ, Lambers Heerspink HJ, Hellemons ME, Friedrich T, Dallmann G, Lajer M, et al. Urine and plasma metabolites predict the development of diabetic nephropathy in individuals with type 2 diabetes mellitus. Diabet Med. 2014; 31:1138–47. doi: 10.1111/dme.12447 PMID: 24661264

40. Ninomiya T, Perkovic V, de Galan BE, Zoungas S, Pillai A, Jardine M, et al. Albuminuria and kidney function independently predict cardiovascular and renal outcomes in diabetes. J Am Soc Nephrol. 2009; 20: 1813–1821. doi: 10.1681/ASN.2008121270 PMID: 19443635

41. Babazono T, Nyumura I, Toya K, Hayashi T, Ohta M, Suzuki K, et al. Higher levels of urinary albumin excretion within the normal range predict faster decline in glomerular filtration rate in diabetic patients. Diabetes Care. 2009; 32: 1518–1520. doi: 10.2337/dc08-2151 PMID: 19435960

42. Hellemons ME, Persson F, Bakker SJ, Rossing P, Parving HH, de Zeeuw D, et al. Initial angiotensin receptor blockade-induced decrease in albuminuria is associated with long-term renal outcome in type 2 diabetic patients with microalbuminuria: a post hoc analysis of the IRMA-2 trial. Diabetes Care. 2011; 34: 2078–2083. doi: 10.2337/dc11-0324 PMID: 21788629

43. Roscioni SS, Lambers Heerspink HJ, de Zeeuw D. Microalbuminuria: Target for renoprotective therapy PRO. Kidney Int. 2014; 86:40–9. doi: 10.1038/ki.2013.490 PMID: 24759154

44. Remuzzi G, Bertani T. Is glomerulosclerosis a consequence of altered glomerular permeability to macromolecules? Kidney Int. 1990; 38: 384–394. PMID: 2232481

45. Abbate M, Zoja C, Remuzzi G. How does proteinuria cause progressive renal damage? J Am Soc Nephrol. 2006; 17: 2974–2984. PMID: 17035611

46. Steyerberg EW. Validation of prediction models. In: Gail M, Krickeberg K, Samet J, Tsiatis A, Wong W, editors. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. New York: Springer US; 2009. pp. 299–311.