



# Systems Biology–Derived Biomarkers to Predict Progression of Renal Function Decline in Type 2 Diabetes Mellitus

DOI: 10.2337/dc16-2202

Gert Mayer,<sup>1</sup> Hiddo J.L. Heerspink,<sup>2</sup> Constantin Aschauer,<sup>3</sup> Andreas Heinzel,<sup>4</sup> Georg Heinze,<sup>5</sup> Alexander Kainz,<sup>3</sup> Judith Sunzenauer,<sup>3</sup> Paul Perco,<sup>4</sup> Dick de Zeeuw,<sup>2</sup> Peter Rossing,<sup>6</sup> Michelle Pena,<sup>2</sup> and Rainer Oberbauer,<sup>3</sup> for the SYSKID Consortium

## OBJECTIVE

Chronic kidney disease (CKD) in diabetes has a complex molecular and likely multifaceted pathophysiology. We aimed to validate a panel of biomarkers identified using a systems biology approach to predict the individual decline of estimated glomerular filtration rate (eGFR) in a large group of patients with type 2 diabetes mellitus and CKD at various stages.

## RESEARCH DESIGN AND METHODS

We used publicly available “omics” data to develop a molecular process model of CKD in diabetes and identified a representative parsimonious set of nine molecular biomarkers: chitinase 3-like protein 1, growth hormone 1, hepatocyte growth factor, matrix metalloproteinase (MMP) 2, MMP7, MMP8, MMP13, tyrosine kinase, and tumor necrosis factor receptor-1. These biomarkers were measured in baseline serum samples from 1,765 patients recruited into two large clinical trials. eGFR decline was predicted based on molecular markers, clinical risk factors (including baseline eGFR and albuminuria), and both combined, and these predictions were evaluated using mixed linear regression models for longitudinal data.

## RESULTS

The variability of annual eGFR loss explained by the biomarkers, indicated by the adjusted  $R^2$  value, was 15% and 34% for patients with eGFR  $\geq 60$  and  $< 60$  mL/min/1.73 m<sup>2</sup>, respectively; variability explained by clinical predictors was 20% and 31%, respectively. A combination of molecular and clinical predictors increased the adjusted  $R^2$  to 35% and 64%, respectively. Calibration analysis of marker models showed significant (all  $P < 0.0001$ ) but largely irrelevant deviations from optimal calibration (calibration-in-the-large:  $-1.125$  and  $0.95$ ; calibration slopes:  $1.07$  and  $1.13$  in the two groups, respectively).

## CONCLUSIONS

A small set of serum protein biomarkers identified using a systems biology approach, combined with clinical variables, enhances the prediction of renal function loss over a wide range of baseline eGFR values in patients with type 2 diabetes mellitus and CKD.

<sup>1</sup>Department of Internal Medicine IV (Nephrology and Hypertension), Medical University of Innsbruck, Innsbruck, Austria

<sup>2</sup>Department of Clinical Pharmacy and Pharmacology, University of Groningen, University Medical Center Groningen, Groningen, Netherlands

<sup>3</sup>Department of Nephrology, Medical University of Vienna, Vienna, Austria

<sup>4</sup>emergentec biodevelopment GmbH, Vienna, Austria

<sup>5</sup>Center for Medical Statistics, Informatics and Intelligent Systems (CeMSIIS), Section for Clinical Biometrics, Medical University of Vienna, Vienna, Austria

<sup>6</sup>Steno Diabetes Center, Gentofte, University of Copenhagen, Copenhagen, Denmark

Corresponding author: Rainer Oberbauer, rainer.oberbauer@meduniwien.ac.at.

Received 13 October 2016 and accepted 20 December 2016.

Clinical trial reg. nos. NCT00252694 and NCT00130312, clinicaltrials.gov.

This article contains Supplementary Data online at <http://care.diabetesjournals.org/lookup/suppl/doi:10.2337/dc16-2202/-/DC1>.

G.M. and H.J.L.H. contributed equally to this study.

© 2017 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered. More information is available at <http://www.diabetesjournals.org/content/license>.

Chronic kidney disease (CKD) in patients with diabetes is a global public health burden (1). The two best-validated biomarkers, estimated glomerular filtration rate (eGFR) and albuminuria, are currently used to predict renal prognosis in and enrich clinical trials with high-risk patients. However, the accuracy and precision of predicting renal prognosis is far from optimal, likely because of the complex pathophysiology of CKD in diabetes (2,3). As a consequence, it is difficult to select high-risk populations for clinical trials, which reduces their statistical power and limits the ability to power them to show efficacy of treatment. Another point of concern is that renal function progressively declines in many patients, even when using state-of-the-art therapies. Current treatments address “uniformly active” pathologies such as hyperglycemia or elevated blood pressure, but trials with new drugs interfering with processes active only in subgroups often fail because the “unselected” population recruited is not ideally suited for the targeted therapy. Thus there is a great clinical need to improve the prediction of prognosis for individuals or subgroups, which may be accomplished using novel biomarkers. Novel biomarkers should ideally reflect pathophysiology and thus guide targeted therapy in well-selected cohorts. It is obvious that these goals can only be achieved if biomarker panels capture multiple pathways. Given the complex molecular pathophysiology of CKD in patients with diabetes, it is not surprising that the addition of single biomarkers to risk prediction models does not improve their performance in the general population of patients with CKD (4,5).

Recent advancements in “omics” technologies, followed by innovations in computational sciences, have improved the ability to integrate high-dimensional data using a systems biology approach. This approach offers new ways to describe the pathophysiology of complex diseases. Techniques have been developed that allow us to refine the accuracy of individual predictions based on molecular pathophysiology, account for different stage-specific dominant processes, and uncover novel therapeutic targets/leads. Some recent studies used such a workflow in small cohorts of patients with later stages of CKD of various etiologies. Ju and colleagues

(6) identified epidermal growth factor as a prognostic biomarker in 164 patients, and the test characteristics were validated in two, albeit smaller, cohorts; another group used higher-order descriptive urinary peptide classifiers to predict CKD progression in patients with diabetes (7). A review of these and other studies of prognostic molecular biomarkers in patients with CKD and diabetes was published recently (8).

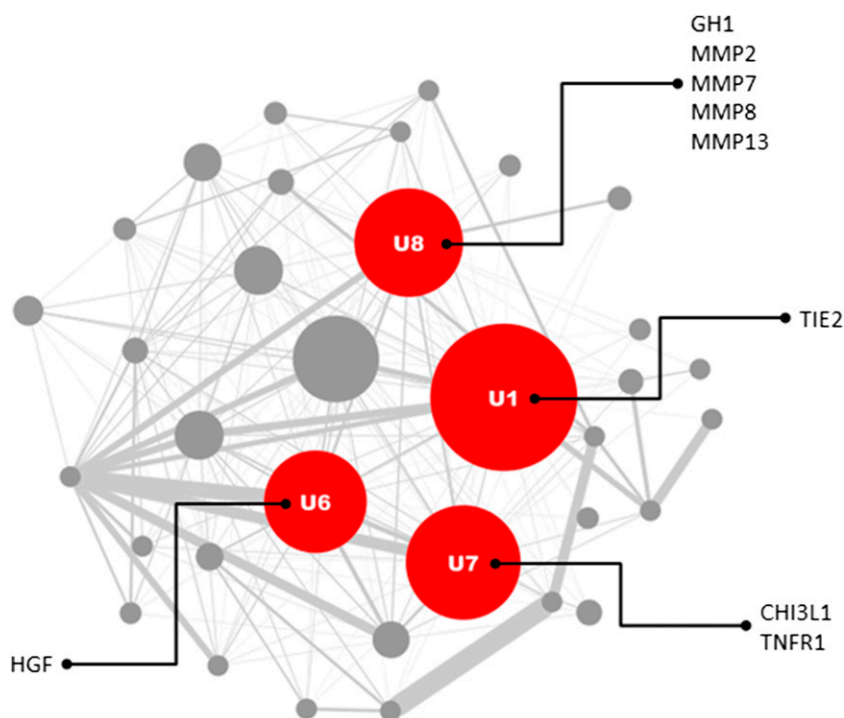
In the EU FP-7–funded project SysKid (Systems Biology toward Novel Chronic Kidney Disease Diagnosis and Treatment), our group used publicly available and proprietary experimental data to construct a molecular process model of CKD in diabetes using advanced systems biology techniques (9,10). For *in silico*–derived molecular processes, we proposed protein biomarkers representing different processes and tested the prognostic power of some of these markers to predict the progression of CKD in an initial small discovery study (11). In the study presented here, we validated these markers in 1,765 patients with type 2 diabetes mellitus with incident, early, and later stages of CKD using multiplexed Luminex analysis of serum

samples from two prospective interventional studies.

## RESEARCH DESIGN AND METHODS

### Biomarker Identification and Selection

Biomarker candidates were selected on the basis of a molecular process model representation of CKD in diabetes (3). In brief, large-scale omics profiling was combined with scientific literature mining to derive a set of 2,466 protein-coding genes (PCGs) deregulated in CKD in diabetes. This feature set was mapped onto a hybrid protein interaction network holding more than 16,000 PCGs as nodes and more than 600,000 connecting edges reflecting their interactions, as described by Fechete et al. (12). The resulting subgraph was further segmented into 34 internally highly connected clusters of PCGs using MCODE (13). These processes covered 688 PCGs; the size of each ranged from 3 to 128 nodes (3) (Fig. 1). Sixteen markers that best represented the four largest processes, holding a total of 398 PCGs, were selected and measured in the discovery study (11). Gene ontologies (GOs) of biological processes enriched in PCGs of these four clusters included terms such as regulation



**Figure 1**—*In silico*–derived molecular model of CKD in diabetes. Each node represents a process unit; node size correlates to the number of assigned proteins. The nine selected marker candidates are assigned to the four largest process units (U1, U6, U7, and U8).

of cell proliferation, epithelial-to-mesenchymal transition, regulation of apoptosis, regulation of inflammatory response, collagen metabolic process, and response to hypoxia, among others. Vascular endothelial growth factor-A needed to be removed from further analysis because of minimal value spread, and epidermal growth factor and interleukin-1  $\beta$  were removed because of a lack of data points below the lower assay detection limit. With the remaining 13 marker candidates and available clinical parameters, a multi-variable model was designed, resulting from least absolute shrinkage and selection operator selection and bootstrap resampling for loss of eGFR. Finally, a subset of nine molecular markers significantly improved the prediction of accelerated eGFR decline.

The average patient age in the discovery study was 63.5 years and was comparable between the discovery and the validation cohorts. Renal function, however, was more preserved in the discovery study, with an average baseline eGFR of 77.9 mL/min/1.73 m<sup>2</sup>, compared with average baseline eGFR values of 33.5 and 69.4 mL/min/1.73 m<sup>2</sup> in the SMACRO and DIRECT studies, respectively. The average HbA<sub>1c</sub> value (7.7%) in the discovery cohort was slightly lower than those in the SMACRO (7.9%) and the DIRECT (8.2%) studies, and the average BMI of 32.4 kg/m<sup>2</sup> in the discovery cohort was between the average BMI values in the DIRECT (29.5 kg/m<sup>2</sup>) and the SMACRO (33.1 kg/m<sup>2</sup>) studies. The average rate of eGFR decline in the discovery study was  $-2.0 \pm 4.7$  mL/min/1.73m<sup>2</sup>/year over a median of 4 years.

### Biomarker Protein Measurement

The marker proteins (chitinase 3-like protein 1 [CHI3L1], growth hormone 1 [GH1], hepatocyte growth factor [HGF], matrix metalloproteinase [MMP] 2, MMP7, MMP8, MMP13, tyrosine kinase [TIE2], and tumor necrosis factor receptor-1 [TNFR1]) in baseline serum samples from two longitudinal studies including participants with diabetes and different stages of CKD were measured using standard multiplexed Luminex technology.

A Human Premixed Multi-Analyte Kit (catalog no. LXSAH-10; R&D Systems, Minneapolis, MN) with a custom assay panel of the markers based on Luminex xMAP technology was used to prepare

and analyze samples. We performed all measurements using a Luminex 200 machine (Bio-Rad, Hercules, CA) with xPONENT software, according to the manufacturer's instructions. Samples were diluted 1:2 with Calibrator Diluent RD6-52. Samples were washed with a vacuum manifold (Millipore, Darmstadt, Germany) and an ATMOS C 261 Aspirator (ATMOS MedizinTechnik, Lenzkirch, Germany). A Luminex 200 Performance Verification Kit containing an xMAP control (catalog no. LX200CONK25; Bio-Rad) was used as the control and a Luminex 200 Calibration Kit with an xMAP calibrator (catalog no. LX200CALK25; Bio-Rad) as the calibrator. The standard curves for each cytokine were generated using the premixed lyophilized standards provided in the kits.

In a 50- $\mu$ L sample, 50 events per bead were analyzed at a flow rate of 60  $\mu$ L/min. Minimum events were set to zero, and the median fluorescence was measured. Doublet discriminator gates for two respective measurements were set at 7,500 and 15,500.

### Study Design, Data Sources, and Selection of Study Participants

To capture a wide range of CKD stages, we analyzed samples from two prospective interventional studies of patients with type 2 diabetes mellitus. The DIRECT study (770 available samples included, 364 from patients in the active treatment group receiving candesartan) included normoalbuminuric, normotensive, or treated hypertensive patients with mild to moderately severe retinopathy (14). In addition, we analyzed 995 available samples (489 from patients receiving active therapy) from the sulodexide macroalbuminuria (SMACRO) trial, which evaluated the renoprotective effects of sulodexide (>900 mg/day) in patients with renal impairment and significant proteinuria (15). The participants from these two studies were grouped according to their baseline eGFR ( $\geq 60$  and  $< 60$  mL/min/1.73 m<sup>2</sup>); the characteristics of the two populations are shown in Table 1.

### Statistical Analysis

Characteristics of patients by eGFR study group were described using mean and SD for normally distributed variables, and frequency and percentage for categorical variables, respectively. We used the Student *t* test to compare continuous variables and the

$\chi^2$  test or Fisher exact test to compare categorical variables between groups.

To establish an eGFR slope prediction formula, we estimated two separate random-coefficient linear mixed models—one for patients with eGFR  $\geq 60$  mL/min/1.73 m<sup>2</sup> at baseline, and one for those with eGFR  $< 60$  mL/min/1.73 m<sup>2</sup> at baseline. We included only patients with at least 3 months of follow-up, and in each of the mixed models we used the repeated measurements of eGFR, including the baseline value, as the dependent variable, and clinical predictors and protein markers as fixed effects. The rationale for including the baseline eGFR in the dependent variable is that it is assumed to be subject to the same random variation as later values, and therefore should not be treated as a fixed value. It also provides useful information about random, intraindividual variation. Interactions of all variables with time of measurement were also included. The random part of the model consisted of a patient-specific random intercept and a random slope, imposing no restrictions on the structure of their covariance. We eliminated protein marker effects if this increased the Akaike information criterion, yielding a model with optimal predictive performance. This corresponds to selection at an  $\alpha$  level of 0.157. In this elimination, we kept the hierarchy of the main effect and the interaction of the markers with time by first eliminating insignificant interactions and then evaluating main effects only if corresponding interactions had already been dropped from the model. Clinical variables were not eliminated because they were selected based on background knowledge. The clinical covariables were sex, age, smoking status, log urinary albumin-to-creatinine ratio, BMI, total cholesterol, mean arterial pressure, and HbA<sub>1c</sub>.

The models were then described by reporting regression coefficients and associated 95% CIs for estimating the eGFR "intercept" and "slope," that is, the change per year (Tables 2 and 3). To improve predictions, available baseline values of eGFR were incorporated into predictions. We computed adjusted *R*<sup>2</sup> values to estimate the explained variation of the model (i.e., unadjusted *R*<sup>2</sup> values were multiplied by  $[N - K - 1]/[N - 1]$ , where *N* and *K* denote the number of patients and the number of fixed

**Table 1—Baseline characteristics of participants in the two subpopulations defined by baseline eGFR <60 (n = 1143) and ≥60 mL/min/1.73 m<sup>2</sup> (n = 622)**

	Patients, n			Values*		P value
	eGFR <60 mL/min/1.73 m <sup>2</sup>	eGFR ≥60 mL/min/1.73 m <sup>2</sup>	Patients with data missing, n (%)	eGFR <60 mL/min 1.73 m <sup>2</sup>	eGFR ≥60 mL/min 1.73 m <sup>2</sup>	
Age (years)	975	622	168 (9.52)	63.1 (8.8)	56.6 (7.4)	<0.001
Female sex	977	622	166 (9.41)	312 (31.9%)	298 (46.3%)	<0.001
Smoking status at baseline (never/previous or current)	861	629	284 (16.09)	460/401	467/153	<0.001
BMI (kg/m <sup>2</sup> )	863	622	280 (15.86)	32.5 (15.4)	29.5 (4.6)	<0.001
Blood pressure at baseline (mmHg)						
Systolic	960	622	183 (10.37)	138 (13.8)	134 (13.7)	<0.001
Diastolic	960	622	183 (10.37)	74 (9.6)	78 (7.2)	<0.001
Heart rate (bpm)	774	7	984 (55.75)	70 (12)	76 (10)	0.151
HbA <sub>1c</sub> (%)	1,081	618	66 (3.74)	8.0 (1.5)	8.2 (1.5)	0.017
Serum glucose (mmol/L)	956	7	802 (45.44)	8.7 (3.7)	9.1 (4.8)	0.776
Serum cholesterol (mmol/L)	1,114	620	31 (1.76)	4.7 (1.2)	5.4 (1.1)	<0.001
Serum creatinine (μmol/L)	1,117	622	26 (1.47)	176 (58.3)	60 (15.5)	<0.001
Baseline MDRD eGFR (mL/min/1.73 m <sup>2</sup> )	952	622	191 (10.82)	36.5 (11.4)	73.3 (9.0)	<0.001
Baseline UACR (mg/g)	1,108	622	35 (1.98)	1566 (1623)	32 (187)	<0.001

\*Values are mean (SD) or n (%). UACR, urinary albumin-to-creatinine ratio.

[main and interaction] effects in the model, respectively). To assess the importance of each predictor, we also decomposed the adjusted  $R^2$  values into contributions from the group of clinical variables and the group of protein markers by averaging the adjusted  $R^2$  values resulting from models with one group only and the reduction in  $R^2$  when a group was completely omitted from the full model. The same was computed separately for each protein marker, scaling the resulting values to add up to the total adjusted  $R^2$  contribution of the protein markers.

All analyses were done on a complete-case basis. A  $P$  value less than 0.05 was considered statistically significant, and all reported  $P$  values are two-sided. We used SAS 9.4 TS 1M2 for Windows (SAS Institute, Inc., Cary, NC) for all analyses.

## RESULTS

### Biomarker Selection

Nine markers remained significantly associated with the outcome of interest and were used in the analysis in this study: CHI3L1, GH1, HGF, MMP2, MMP7, MMP8, MMP13, TIE2, and TNFR1.

We observed a wide spread of the nine biomarker values in the baseline serum samples of the two studies; a box-and-whisker plot of the concentrations is provided in Supplementary Fig. 1A and B. The calculated decrease of eGFR per year per included patient is depicted in Supplementary Fig. 2A and B.

### Performance of the Marker Panel in Predicting eGFR Slopes

Tables 2 and 3 show selected regression coefficients for estimating eGFR slope by demographic and clinical variables and protein markers. To compare the importance of the markers in predicting eGFR, the contribution of each marker to the adjusted  $R^2$  value, as well as that of all markers and of all clinical variables together to the adjusted  $R^2$  value, were computed and are provided in Table 4. The regression coefficients of some markers were different in the two study groups, with different baseline eGFRs and occasionally even with opposite signs, and some of the markers were selected in one group only (TIE2, TNFR1), supporting the necessity of stage-specific dynamic marker panel modeling. Nonetheless, the biomarkers contributed 15.2% and 33.4% to the explained variances in patients with baseline eGFR levels ≥60 and <60 mL/min/1.73 m<sup>2</sup>, respectively. Further, 19.6% and 30.6% of explained variance were contributed by clinical and demographic parameters, respectively.

**Table 2—Predictors of eGFR slope per year for patients with eGFR ≥60 mL/min/1.73 m<sup>2</sup> at baseline**

Variable	Coefficient	95% CI	P value
Constant	1.5722	−3.2473 to 6.3917	0.5219
Female sex	−0.3543	−0.7563 to 0.04760	0.0840
Age (per year)	−0.2758	−0.3738 to −0.1778	0.3621
Current or former smoker	0.3130	−0.1045 to 0.7306	0.1416
Total cholesterol (mmol/L)	0.1286	−0.02665 to 0.2838	0.1044
BMI (kg/m <sup>2</sup> )	−0.03714	−0.07862 to 0.004335	0.0792
Log <sub>2</sub> UACR (mg/g)	−0.1846	−0.3638 to −0.00536	0.0435
Mean arterial pressure (mmHg)	0.000487	−0.02104 to 0.02202	0.9646
HbA <sub>1c</sub> (%)	−0.1162	−0.2355 to 0.002958	0.0560
Log <sub>2</sub> MMP2 (pg/mL)	0.2242	−0.01261 to 0.4610	0.0635
Log <sub>2</sub> MMP8 (pg/mL)	−0.1725	−0.3163 to −0.02883	0.0186
Log <sub>2</sub> MMP7 (pg/mL)	−0.2245	−0.4737 to 0.02471	0.0774
Log <sub>2</sub> GH1 1(pg/mL)	−0.07095	−0.1673 to 0.02544	0.1490

UACR, urinary albumin-to-creatinine ratio.

**Table 3—Predictors of eGFR slope per year for patients with eGFR <60 mL/min/1.73 m<sup>2</sup> at baseline**

Variable	Coefficient	95% CI	P value
Constant	16.3718	4.37–28.37	0.0076
Female sex	−0.1631	−0.8252 to 0.4989	0.6289
Age (per year)	0.01165	−0.02756 to 0.05086	0.5601
Current or former smoker	1.0000	0.3887–1.6114	0.0014
Total cholesterol (mmol/L)	−0.2004	−0.4187 to 0.01799	0.0721
BMI (kg/m <sup>2</sup> )	0.002429	−0.05085 to 0.05571	0.9288
Log <sub>2</sub> UACR (mg/g)	−0.4268	−0.5689 to −0.2847	<0.0001
Mean arterial pressure (mmHg)	−0.02697	−0.05768 to 0.003744	0.0852
HbA <sub>1c</sub> (%)	0.06177	−0.1234 to 0.2470	0.5131
Log <sub>2</sub> MMP2 (pg/mL)	−0.8673	−1.4858 to −0.2488	0.0060
Log <sub>2</sub> Tie2 (pg/mL)	−0.3853	−0.6866 to −0.08405	0.0122
Log <sub>2</sub> MMP7 (pg/mL)	−0.3595	−0.8228 to 0.1038	0.1282
Log <sub>2</sub> MMP13 (pg/mL)	−0.6628	−1.4872 to 0.1617	0.1151
Log <sub>2</sub> TNFR1 (pg/mL)	0.9630	0.2966–1.6294	0.0046

UACR, urinary albumin-to-creatinine ratio.

The mean (2.5th, 97.5th percentiles) predicted trajectory of eGFR over time is illustrated in Fig. 2, and the distributions of observed values are superimposed in a box-and-whisker plot. The calibration of the model—that is, the assessment of the goodness of fit of the predicted to the observed eGFR values for the two groups—is shown in Fig. 3A and B. Given the high intraindividual variability in the progression of eGFR over time, our models were well calibrated and not overfitted (calibration

slopes of observed vs. predicted values with baseline eGFR levels  $\geq 60$  and  $< 60$  mL/min/1.73 m<sup>2</sup>: 1.069 [ $P = 0.012$  for test against a perfect calibration of 1] and 1.137 [ $P < 0.0001$ ], respectively), suggesting that the in silico biomarker selection was successful and, in particular, models were not overfitted (calibration slopes  $< 1$ ). Calibration-in-the-large, that is, the overall unbiasedness of the eGFR predictions, was adequate as well, and deviations from optimal calibration, although statistically significant, were

clinically irrelevant (−1.125 and 0.956, respectively; both  $P < 0.0001$  against perfect unbiasedness).

## CONCLUSIONS

To our knowledge, this study validates for the first time that a systems biology approach to omics data can identify a combination of serum biomarkers (in this case, nine) that are able to predict the decline of eGFR in a large population of patients with type 2 diabetes mellitus across a wide range of glomerular filtration rates. The biomarker models for eGFR  $\geq 60$  and  $< 60$  mL/min/1.73 m<sup>2</sup> performed similarly to those including established clinical risk factors, and a combination of both improved the accuracy, indicating that additional information is provided by the markers.

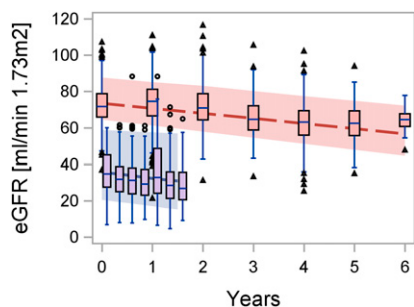
We used the molecular interaction network model of CKD in diabetes that was developed during the FP-7–funded project SysKid (10). The network consists of various processes characterized by PCGs with a high-grade interaction and goes beyond conventional molecular analysis tools such as Gene Ontology or PANTHER, as the complexity of interactions between pathways is also taken into account. Our systems biology analysis provided the rationale to focus on only a subset of PCGs reported to be deregulated in CKD (688 of 2,466). The in silico processes generated directly reflect the pathophysiology of CKD and guided the selection of the biomarker panel, which was tested for its prognostic potential.

We hypothesized that the progression of CKD in diabetes is complex, with high interindividual variability, and over time, even the high intraindividual variability of the molecular processes is involved. First, we conclude that the biomarkers selected are able to predict the prognosis, which demonstrates that the representative processes are of pathophysiological importance. Next, we were able to show that these processes are stage specific (some biomarkers predict prognosis in early, others in late disease) by testing the biomarker in a group of patients with a wide range of renal dysfunction at baseline. Finally, the wide spread of baseline eGFR values is an indication of considerable interindividual variability of CKD pathophysiology in diabetes mellitus. Given this complexity, our model approach seems superior to

**Table 4—Decomposition of adjusted R<sup>2</sup> values into contributions of clinical variables and protein markers**

Protein marker	Importance in model (scaled contribution to adjusted R <sup>2</sup> )	
	Model for eGFR $\geq 60$ ml/min/1.73m <sup>2</sup> at baseline	Model for eGFR $< 60$ ml/min/1.73m <sup>2</sup> at baseline
Clinical variables only	29.2%	55.9%
Protein markers only	24.7%	58.7%
Total adjusted R <sup>2</sup>	34.8%	64.0%
Decomposed into contributions of:		
Clinical variables	19.6%	30.6%
Protein markers	15.2%	33.4%
Decomposed into contributions of:		
MMP2	3.7%	2.3%
TIE2	n.s.	1.5%
CHI3L1	n.s.	n.s.
MMP8	3.7%	n.s.
MMP7	3.7%	9.6%
HGF	n.s.	n.s.
MMP13	n.s.	3.3%
TNFR1	n.s.	16.6%
GH	4.0%	n.s.

n.s., Not selected by Akaike information criterion.



**Figure 2**—Predicted and observed eGFR values over the study duration. Predicted slopes are indicated by solid lines with 95% CIs, and actual (measured) eGFR values are presented as box plots in the eGFR  $\geq 60$  and  $< 60$  mL/min/1.73 m<sup>2</sup> groups. (A high-quality color representation of this figure is available in the online issue.)

many conventional risk factor analyses. Even though eGFR and age are valuable biomarkers for risk assessment, our panel is directly linked to the pathophysiology of the disease and thus not only predicts prognosis but also could allow the identification of new treatment targets and, importantly, guide targeted therapy.

Some biomarkers included in the panel have already been investigated by others. A group from the Joslin Diabetes Center found that TNFR1 serum concentration predicted the incidence of end-stage renal disease and eGFR loss to values  $< 60$  mL/min/1.73 m<sup>2</sup> in patients with type 2 and type 1 diabetes, and preserved renal function (eGFR  $> 60$  mL/min/1.73 m<sup>2</sup> and no or mild proteinuria) (16,17). Those authors did not find a correlation with other glomerular or tubular biomarkers or clinical

variables. TNFR1 data related to eGFR loss in advanced CKD stages and patients with proteinuria are lacking. In our study, with no further adjustment, the effect of TNFR1 goes in the same direction as in the study by Niewczas et al. (16), whereas adjusting for clinical variables and for biomarkers turns this effect in the opposite direction. This is a result of confounding with those other variables. Niewczas et al., for example, adjusted for clinical predictors and did not adjust for other biomarkers. Thus there is no contradiction between the studies since the sign of the slope depends on the covariables in the model.

An additional likely explanation for this intriguing finding is the competing risk of death in patients with elevated TNFR1, as was recently shown in patients with type 2 diabetes and advanced proteinuric CKD (18). Thraikill et al. (19) found MMP2 to be dysregulated in plasma and in particular in urine of patients with type 1 diabetes. Based on the latter result, those authors suggested that MMP2 excretion might directly mirror intrarenal pathology. Among the other matrix metalloproteinases, only urinary MMP9 (20) and serum MMP7 (21) were investigated previously (not MMP7, MMP8, or MMP13).

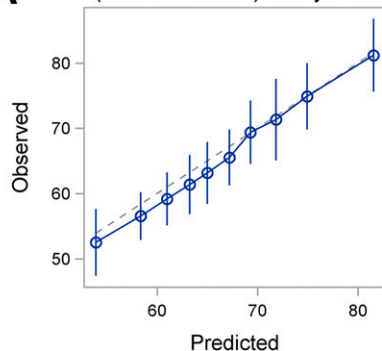
Several experimental and clinical studies showed that HGF, which is also part of our panel, is associated with renoprotection and may even serve as a therapeutic target in patients with diabetes and CKD (22,23). In this context, even though HGF was not included in the prediction models, a strong pathophysiological

relevance should not be excluded. GH1, which was negatively associated with eGFR slope (and thus faster progression) in our analysis, induced transforming growth factor- $\beta$  and accelerated podocyte depletion and proteinuria in an experimental model of diabetic nephropathy (24). CHI3L1 is a member of the glycosyl hydrolase 18 family and is involved in stress response. We found CHI3L1 to be a strong contributor to eGFR loss in patients with higher baseline GFR (i.e.,  $\geq 60$  mL/min/1.73 m<sup>2</sup>), which is in line with a study by Żurawska-Płaksej et al. (25).

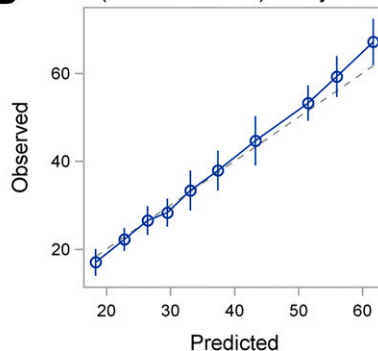
In contrast to the targeted approaches mentioned above, our parsimonious biomarker panel representing molecular features deregulated in CKD covers a wide range of the critically involved molecular pathways and might therefore better allow patient subclassification (3,5,12). In addition, the processes can also be directly linked to clinical phenotypes. For example TIE2, which is associated with disease progression, represents process unit 1. Using GO term analysis, PCGs in process unit 1 preferentially are associated with regulation of cell proliferation, cell signaling, and cell surface receptor-linked signal transduction, among other processes, and one could speculate that medications interfering with these pathways might be tested in CKD. HGF as a proxy for process unit 6 is assigned to the enriched GO terms epithelial-to-mesenchymal transition, antiapoptosis, and cell proliferation. TNFR1 as a candidate from process unit 7 is assigned to the enriched GO biological process regulation of inflammatory response. The set of matrix metalloproteinases (MMP2, MMP7, MMP8, and MMP13) from process unit 8 are involved in the collagen metabolic process and thus are linked to fibrotic processes, one of the hallmarks of CKD in patients with diabetes.

Certainly, our approach has some limitations; these were, however, carefully addressed by several sensitivity analyses. Omission of some of the markers in the panel always led to inferior model performance, and the addition of other random markers, which by chance were available for some samples, did reduce the adjusted  $R^2$ . Furthermore, all patients in this study had type 2 diabetes mellitus, and thus results may not be generalizable to other populations.

**A** eGFR (ml/min/1.73m<sup>2</sup>) at 3 years



**B** eGFR (ml/min/1.73m<sup>2</sup>) at 1 year



**Figure 3**—Calibration plots showing mean ( $\pm$ SD) observed vs. predicted eGFR values per decile of predicted eGFR values. A: Calibration of eGFR at 3 years in patients with baseline eGFR  $\geq 60$  mL/min/1.73 m<sup>2</sup> (slope = 1.069;  $P = 0.012$  for test against a perfect calibration of 1). B: Calibration of eGFR at 1 year in patients with baseline eGFR  $< 60$  mL/min/1.73 m<sup>2</sup> (slope = 1.138;  $P < 0.0001$  for test against a perfect calibration of 1). (A high-quality color representation of this figure is available in the online issue.)

We were able to derive a representative set of parsimonious biomarkers from an *in silico* model of CKD in diabetes mellitus elaborated from omics data via sophisticated bioinformatics. The biomarker panel was able to predict renal function loss in a large group of patients with type 2 diabetes mellitus and a wide range of baseline eGFR values.

### Summary

We conclude that a small set of biomarkers identified by a systems biology approach enhances the prediction of renal function loss using standard clinical variables in patients with type 2 diabetes mellitus with a wide range of baseline eGFR.

**Funding.** Funding for DIRECT and SMACRO was provided by AstraZeneca and Takeda Pharmaceuticals, and Keryx Biopharmaceuticals, respectively. Funding for this study was provided by SysKid, an EU FP7 research project (SysKid HEALTH-F2-2009-241544).

**Duality of Interest.** No potential conflicts of interest relevant to this article were reported.

The funding sources had no role in the design, conduct, or analysis of the study.

**Author Contributions.** G.M., H.J.L.H., D.d.Z., P.R., and R.O. designed the study. C.A., A.H., and P.P. performed the systems biology analysis. G.H., A.K., and M.P. conducted the statistical analysis. A.K. and J.S. designed the Luminex assays and measured the samples. All authors contributed substantially to the analysis of data and preparation of the manuscript. R.O. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

### References

- Eckardt KU, Coresh J, Devuyst O, et al. Evolving importance of kidney disease: from subspecialty to global health burden. *Lancet* 2013;382:158–169
- Hallan SI, Ritz E, Lydersen S, Romundstad S, Kvenild K, Orth SR. Combining GFR and albuminuria to classify CKD improves prediction of ESRD. *J Am Soc Nephrol* 2009;20:1069–1077
- Heinzel A, Mühlberger I, Stelzer G, et al. Molecular disease presentation in diabetic nephropathy. *Nephrol Dial Transplant* 2015;30(Suppl. 4):iv17–iv25
- Parving HH, Persson F, Rossing P. Microalbuminuria: a parameter that has changed diabetes care. *Diabetes Res Clin Pract* 2015;107:1–8
- Heinzel A, Mühlberger I, Fehete R, Mayer B, Perco P. Functional molecular units for guiding biomarker panel design. *Methods Mol Biol* 2014;1159:109–133
- Ju W, Nair V, Smith S, et al.; ERCB, C-PROBE, NEPTUNE, and PKU-IgAN Consortium. Tissue transcriptome-driven identification of epidermal growth factor as a chronic kidney disease biomarker. *Sci Transl Med* 2015;7:316ra193
- Schanstra JP, Zúrbig P, Alkhalaf A, et al. Diagnosis and prediction of CKD progression by assessment of urinary peptides. *J Am Soc Nephrol* 2015;26:1999–2010
- Pena MJ, de Zeeuw D, Mischak H, et al. Prognostic clinical and molecular biomarkers of renal disease in type 2 diabetes. *Nephrol Dial Transplant* 2015;30(Suppl. 4):iv86–iv95
- Lambers Heerspink HJ, Oberbauer R, Perco P, et al. Drugs meeting the molecular basis of diabetic kidney disease: bridging from molecular mechanism to personalized medicine. *Nephrol Dial Transplant* 2015;30(Suppl. 4):iv105–iv112
- Mayer P, Mayer B, Mayer G. Systems biology: building a useful model from multiple markers and profiles. *Nephrol Dial Transplant* 2012;27:3995–4002
- Pena MJ, Heinzel A, Heinze G, et al. A panel of novel biomarkers representing different disease pathways improves prediction of renal function decline in type 2 diabetes. *PLoS One* 2015;10:e0120995
- Fehete R, Heinzel A, Soellner J, Perco P, Lukas A, Mayer B. Using information content for expanding human protein coding gene interaction networks. *J Comput Sci Syst Biol* 2013;6:073–082
- Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 2003;4:2
- Sjølie AK, Klein R, Porta M, et al.; DIRECT Programme Study Group. Effect of candesartan on progression and regression of retinopathy in type 2 diabetes (DIRECT-Protect 2): a randomised placebo-controlled trial. *Lancet* 2008;372:1385–1393
- Packham DK, Wolfe R, Reutens AT, et al.; Collaborative Study Group. Sulodexide fails to demonstrate renoprotection in overt type 2 diabetic nephropathy. *J Am Soc Nephrol* 2012;23:123–130
- Niewczas MA, Gohda T, Skupien J, et al. Circulating TNF receptors 1 and 2 predict ESRD in type 2 diabetes. *J Am Soc Nephrol* 2012;23:507–515
- Nowak N, Skupien J, Niewczas MA, et al. Increased plasma kidney injury molecule-1 suggests early progressive renal decline in non-proteinuric patients with type 1 diabetes. *Kidney Int* 2016;89:459–467
- Saulnier PJ, Gand E, Ragot S, et al.; SURDIAGENE Study Group. Association of serum concentration of TNFR1 with all-cause mortality in patients with type 2 diabetes and chronic kidney disease: follow-up of the SURDIAGENE Cohort. *Diabetes Care* 2014;37:1425–1431
- Thraikill KM, Bunn RC, Moreau CS, et al. Matrix metalloproteinase-2 dysregulation in type 1 diabetes. *Diabetes Care* 2007;30:2321–2326
- Li SY, Huang PH, Yang AH, et al. Matrix metalloproteinase-9 deficiency attenuates diabetic nephropathy by modulation of podocyte functions and dedifferentiation. *Kidney Int* 2014;86:358–369
- Ban CR, Twigg SM, Franjic B, et al. Serum MMP-7 is increased in diabetic renal disease and diabetic diastolic dysfunction. *Diabetes Res Clin Pract* 2010;87:335–341
- Flaquer M, Franquesa M, Vidal A, et al. Hepatocyte growth factor gene therapy enhances infiltration of macrophages and may induce kidney repair in db/db mice as a model of diabetes. *Diabetologia* 2012;55:2059–2068
- Kusunoki H, Taniyama Y, Azuma J, et al. Telmisartan exerts renoprotective actions via peroxisome proliferator-activated receptor- $\gamma$ /hepatocyte growth factor pathway independent of angiotensin II type 1 receptor blockade. *Hypertension* 2012;59:308–316
- Chitra PS, Swathi T, Sahay R, Reddy GB, Menon RK, Kumar PA. Growth hormone induces transforming growth factor- $\beta$ -induced protein in podocytes: implications for podocyte depletion and proteinuria. *J Cell Biochem* 2015;116:1947–1956
- Żurawska-Płaksej E, Ługowska A, Hetmańczyk K, Knapik-Kordecka M, Adamiec P, Piwowar A. Proteins from the 18 glycosyl hydrolase family are associated with kidney dysfunction in patients with diabetes type 2. *Biomarkers* 2015;20:52–57