Journal of proteome research

Cite This: J. Proteome Res. 2019, 18, 1796−1805

Article

pubs.acs.org/jpr

A Novel Metabolic Signature To Predict the Requirement of Dialysis or Renal Transplantation in Patients with Chronic Kidney Disease

Helena U. Zacharias,[†] Michael Altenbuchinger,[‡] Ulla T. Schultheiss,[∥,#] Claudia Samol,[§] Fruzsina Kotsis,[∥,#] Inga Poguntke,[∥] Peggy Sekula,[∥] Jan Krumsiek,[†,⊥] Anna Köttgen,[∥] Rainer Spang,[‡] Peter J. Oefner,[§] and Wolfram Gronwald*[,§]

[†]Institute of Computational Biology, Helmholtz Zentrum München, Neuherberg 85764, Germany

[‡]Statistical Bioinformatics, Institute of Functional Genomics, and [§]Institute of Functional Genomics, University of Regensburg, Regensburg 93053, Germany

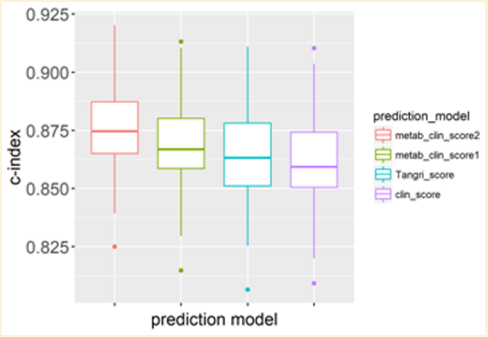
[∥]Institute of Genetic Epidemiology, Department of Biometry, Epidemiology, and Medical Bioinformatics, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg 79106, Germany

[⊥]Institute for Computational Biomedicine, Englander Institute for Precision Medicine, Department of Physiology and Biophysics, Weill Cornell Medicine, New York, New York 10065, United States

[#]Renal Division, Department of Medicine IV, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg 79106, Germany

Supporting Information

ABSTRACT: Identification of chronic kidney disease patients at risk of progressing to end-stage renal disease (ESRD) is essential for treatment decision-making and clinical trial design. Here, we explored whether proton nuclear magnetic resonance (NMR) spectroscopy of blood plasma improves the currently best performing kidney failure risk equation, the so-called Tangri score. Our study cohort comprised 4640 participants from the German Chronic Kidney Disease (GCKD) study, of whom 185 (3.99%) progressed over a mean observation time of 3.70 ± 0.88 years to ESRD requiring either dialysis or transplantation. The original four-variable Tangri risk equation yielded a *C* statistic of 0.863 (95% CI, 0.831−0.900). Upon inclusion of NMR features by state-of-the-art machine learning methods, the *C* statistic improved to 0.875 (95% CI, 0.850−0.911), thereby outperforming the Tangri score in 94 out of 100 subsampling rounds. Of the 24 NMR features included in the model, creatinine, high-density lipoprotein, valine, acetyl groups of glycoproteins, and Ca^{2+}-EDTA carried the highest weights. In conclusion, proton NMR-based plasma fingerprinting improved markedly the detection of patients at risk of developing ESRD, thus enabling enhanced patient treatment.



KEYWORDS: *kidney failure risk equation, metabolomics, chronic kidney disease*

INTRODUCTION

Chronic Kidney Disease (CKD) represents one of the largest burdens on the world's health system.[1] The prevalence of CKD is 10−15% in the general adult population across both low- and high-income countries.[2] CKD is staged according to the estimated glomerular filtration rate (eGFR) and the amount of albuminuria into five major CKD categories ranging from G1 (normal eGFR of >90 mL/min/1.73 m^2, but evidence of kidney disease) to G5 (eGFR of <15 mL/min/1.73 m^2), and three albuminuria categories A1 (<30 mg/24 h) to A3 (>300 mg/24 h).[3,4] Ultimately, CKD can progress to end-stage renal disease (ESRD), requiring dialysis or transplantation.[5] In addition, CKD is a risk factor for all-cause and cardiovascular mortality,[6] acute kidney injury (AKI), cognitive decline, anemia, mineral and bone disorders, and fractures.[1,3,5,7]

The timely discrimination of patients with an increased risk of ESRD might facilitate effective intervention. So-called kidney failure risk equations, which are based on demographic and laboratory data, have shown quite excellent performance in predicting ESRD.[8−11] The best performing equations so far include either age, sex, eGFR, and urine albumin-to-creatinine ratio (UACR) only (four-variable equation), or additionally serum levels of calcium, phosphate, bicarbonate, and albumin (eight-variable equation).[8] Their prediction accuracies have been evaluated in a large meta-analysis of 31 cohorts, comprising in total 721 357 patients with CKD stages 3−5 across 30 countries spanning 4 continents.[9] Overall, the four- and eight-variable equations performed fairly similar in this

Received: December 21, 2018
Published: February 28, 2019

DOI: 10.1021/acs.jproteome.8b00983
J. Proteome Res. 2019, 18, 1796−1805

Table 1. Baseline Characteristics of the Investigated GCKD Cohort[a]

total number of participants: 4640				
health parameters	mean	SD	min	max
age (years)	60.19	11.9	18	76
serum albumin (g/L)	38.32	4.27	10.1	59.53
total cholesterol (mg/dL)	211.23	52.4	72.26	771.58
HDL cholesterol (mg/dL)	52.05	18.14	7.27	152.2
LDL cholesterol (mg/dL)	118.29	43.46	4.67	601.78
serum creatinine (mg/dL)	1.51	0.47	0.45	7.01
HbA1c (%)	6.33	1.02	4.6	16.3
HbA1c (mmol/mol)	45.74	11.18	27	154.2
eGFR (CKD-EPI; mL/min/1.73 m^2)	49.44	18.2	8	136
systolic blood pressure (mmHg)	139.61	20.3	77	232
diastolic blood pressure (mmHg)	79.35	11.73	42	151
pulse (bpm)	70.42	12.09	29	135
waste−hip ratio	0.94	0.09	0.64	1.3
BMI (kg/m^2)	29.74	5.93	15.5	69.7
follow-up time (days)	1351.4	322.91	1	1925
	median	IQR	min	max
CRP (mg/L)	2.25	3.96	0.07	216.64
triglycerides (mg/dL)	168.50	121.76	33.3	1714.9
UACR (mg/g)	50.24	373.60	0.84	15783.59
gender	no.		%	
male	2799		60.32	
smoking	no.		%	
former smoker	1991		42.91	
nonsmoker	1879		40.50	
smoker	755		16.27	
unknown	15		0.32	
proteinuria	no.		%	
<30 mg/g creatinine	2248		48.45	
30−300 mg/g creatinine	1321		28.47	
>300 mg/g creatinine	1071		23.08	
outcome	no.		%	
kidney failure events, total	185		3.99	
dialysis	177		3.81	
transplantation	6		0.13	
unspecified	2		0.04	

[a]Abbreviations: HDL, high-density lipoprotein; LDL, low-density lipoprotein; eGFR, estimated glomerular filtration rate; HbA1c, glycated hemoglobin; BMI, body mass index; IQR, interquartile range; CRP, C-reactive protein; UACR, urine albumin-to-creatinine ratio.

meta-analysis, with a pooled *C* statistic, that is, a measure of goodness of fit for binary outcomes in a logistic regression model, of 0.90 (95% CI, 0.89−0.92) at 2 years, and 0.88 (95% CI, 0.86−0.90) at 5 years for the four-variable equation over a median follow-up period of 4 years. A subsequent attempt to improve the eight-variable equation by a dynamic model that included aside from baseline age the latest-available-estimate of eGFR and the latest-available-measurements of serum albumin, phosphorus, calcium, and bicarbonate yielded only an incremental discrimination improvement of 0.73%.[12] Hence, there is a continued need to find additional biomarkers capable of improving the discrimination of CKD patients at high risk of developing ESRD.

Nuclear magnetic resonance (NMR) spectroscopy is, besides hyphenated mass spectrometry, the preferred analytical method in metabolomics,[13] as it allows the nondestructive, simultaneous detection and absolute quantification of free metabolites and other compounds in biological fluids with little sample pretreatment.[14] Several metabolomic studies have already shown the ability of NMR spectroscopy to detect novel biomarkers for such kidney diseases as rhabdomyolysis-related renal tubular damage,[15] nephrotoxicity,[16] pediatric nephrourological diseases,[17] autosomal dominant polycystic kidney disease,[18] CKD in the general population,[19,20] acute kidney injury after cardiac surgery,[21,22] and IgA nephropathy.[23] Here, we employed one-dimensional proton NMR spectroscopy to baseline blood plasma specimens from the German Chronic Kidney Disease (GCKD) cohort, a 10-year prospective observational study of 5217 patients with CKD of various etiologies,[2,7] to test whether NMR metabolite fingerprinting might improve the identification of CKD patients at risk of developing ESRD.

MATERIALS AND METHODS

Patient Selection

With the approval of the local ethics committees of the RWTH Aachen University, Aachen, Germany, the Charité-University-Medicine, Berlin, Germany, the Friedrich-Alexander University, Erlangen, Germany (coordination center), the Albert-Ludwigs-University, Freiburg, Germany, the Friedrich-Schiller

DOI: 10.1021/acs.jproteome.8b00983
J. Proteome Res. 2019, 18, 1796−1805

University, Jena, Germany, the Hannover Medical School, Hannover, Germany, the medical faculty, Ruprecht-Karls University, Heidelberg, Germany, the medical faculty, Ludwig-Maximilians-University, Munich, Germany, and the Julius-Maximilians-University, Würzburg, Germany in total, 4640 of the 5217 GCKD patients originally enrolled were included after receipt of written confirmed consent. Specifically, all patients provided written informed consent that their data and plasma, serum, and urine specimens will be analyzed for research purposes. The study was registered in the national registry for clinical studies (DRKS 00003971) and carried out in accordance with relevant guidelines and regulations. Reasons for exclusion from the present study were insufficient amounts of plasma for ^{1}H NMR spectroscopy and missing clinical data. Original inclusion criteria had been an estimated glomerular filtration rate (eGFR) of 30−60 mL/min/1.73 m^2 or overt proteinuria with an eGFR of >60 mL/min/1.73 m^2 as determined by the treating nephrologist prior to inclusion. Excluded were patients of non-Caucasian ancestry, a history of solid organ or bone marrow transplantation, cancer within the last 24 months, and severe heart failure.[2,7] Further baseline clinical and demographic characteristics are given in Table 1. Baseline EDTA-plasma specimens had been collected in dedicated vacutainer tubes (Sarstedt, Nümbrecht, Germany) from each patient, aliquoted, and stored at −80 °C until measurement. Within a mean observation time of 3.70 ± 0.88 years (up to the fourth yearly follow up visit, GCKD FU4), 185 out of the 4640 patients (3.99%) developed ESRD. Dialysis was initiated in 177 patients, 6 received a kidney transplant, and 2 suffered from ESRD without further specification. In the case that a patient experienced more than one event during the observation period, only the earliest event was taken into account. Patients, who died of any cause, got censored at the date of death. In addition, for a competing risk analysis, we censored at the time point of a cardiovascular event. The latter comprised coronary revascularization procedures, nonfatal nonhemorrhagic strokes, and nonfatal myocardial infarctions. Note that all fatal events had been already censored. A flowchart illustrating the patient selection process as well as the end-point abstraction for the two different analysis procedures is given in Figure S1.

Clinical Variables

Clinical variables employed for risk prediction included age, sex, eGFR, and UACR. Baseline serum creatinine was measured by means of the Roche Creatinine Plus enzymatic creatinine assay, and GFR values were estimated using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) creatinine equation.[24]

NMR Spectroscopy and Data Preprocessing

Details regarding NMR spectroscopy and data preprocessing are given in the Supporting Information. In short, a 1D ^{1}H NMR spectrum was acquired for each of the 4640 participants. For further analysis, the spectral region from 9.5−0.5 ppm was evenly split into 900 buckets or features of 0.01 ppm width. Note that a bucket may contain the NMR signals of one or more metabolites. In case that several metabolites contributed to a bucket, we explicitly specified major and minor contributions. In addition, the use of EDTA plasma allowed the measurement of Ca^{2+} due to its complexation with EDTA, yielding NMR signals specific for Ca^{2+}-EDTA.

Multivariate Data Analysis

Aside from the original 5-year, four-variable risk score model of Tangri et al.[9] for non-American Caucasians, which includes as predictive variables age, sex, eGFR, and the natural log transformed value of UACR, we generated three different Cox proportional hazards (PH) models to compute individual risk scores. To that end, a resampling approach was used, splitting the available cohort of 4640 patients at random 100 times into training and test cohorts of 3093 and 1547 patients, respectively, to allow for an unbiased validation of the resulting models. The first Cox PH regression model, hereafter referred to as clin_score, employed again only age, sex, eGFR, and UACR. Yet in contrast to the Cox PH model used by Tangri et al., both eGFR and UACR were $\log_2$-transformed. Further details regarding data transformation are given in the Supporting Information. For the two subsequent models that included features from the ^{1}H NMR plasma spectra, we excluded all buckets containing only noise or heavily overlapping signals by manual inspection and thereby reduced the number of buckets to 127 (see Figure S2). Next, to obtain relatively sparse models that contain only NMR and/or clinical features impacting classification performance, we trained a LASSO Cox PH model employing the *R*-package glmnet[25] to benefit from the inherent feature selection of LASSO regression.[26] By this, it is also ensured that the resulting models are not prone to overfitting. The penalization parameter λ, which regulates the trade-off between optimal classification performance and model sparsity, was determined by the minimal partial likelihood deviance estimated in an inner 5-fold cross-validation. In the first model metab_clin_score1, each NMR feature was penalized with an equal penalty weight of 1, whereas age, gender, eGFR, and UACR received a penalty weight equal to 0, and thus were always included in the model. In the second model metab_clin_score2, both NMR and clinical as well as demographic features were equally penalized with penalty weights of 1. Therefore, their corresponding regression coefficients were all subjected to the inherent shrinkage and selection procedure of LASSO regression.

The performance of the estimated risk scores in distinguishing patients in subsequent need of dialysis or transplantation was assessed by the concordance statistic *C*.[27] The *C* statistic or index is equal to the area under a ROC curve and represents the probability that, for a pair of randomly chosen patients, the patient with the higher risk prediction will experience an event before the other patient. Thus, a *C* statistic of 1.0 indicates the perfect discrimination of two groups, whereas a *C* statistic of 0.5 indicates a model that does not perform better than chance at predicting membership to a group. The *C* statistic was computed with the *R*-package *survcomp*.[28]

RESULTS

Cohort Description

We investigated 4640 patients from the GCKD cohort, whose baseline clinical characteristics are listed in Table 1. They comprised a subset of the German Chronic Kidney Disease (GCKD) study cohort, as patients with missing NMR, clinical chemistry, and demographic data, as well as missing time to renal event information, were excluded. At baseline, the respective mean ± SD values for age and eGFR (based on the CKD-EPI equation) were 60.2 ± 12.0 years and 49.44 ± 18.20 mL/min/1.73 m^2, respectively, and the median and

DOI: 10.1021/acs.jproteome.8b00983
J. Proteome Res. 2019, 18, 1796−1805

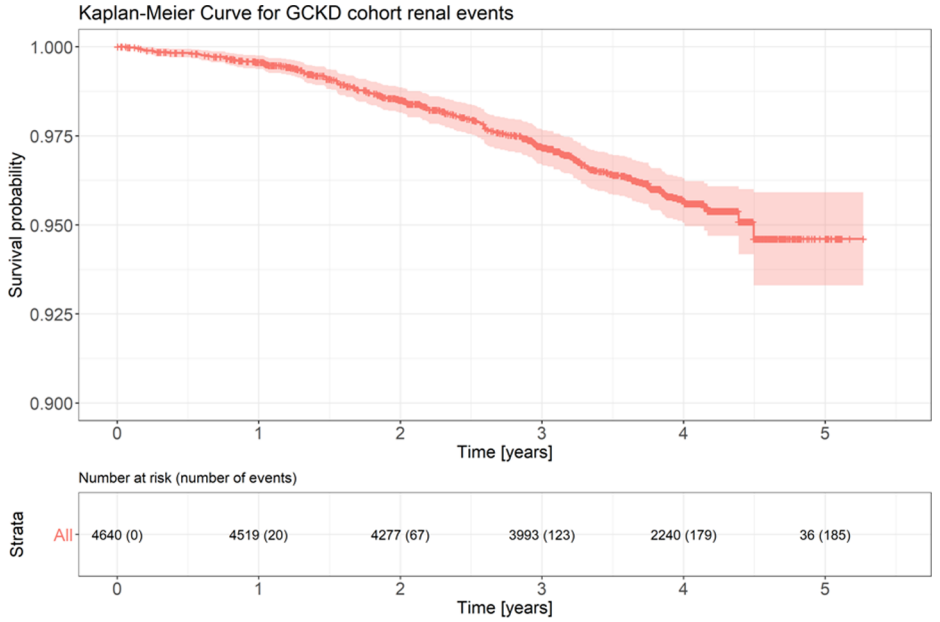


Figure 1. Kaplan−Meier curve of all 4640 patients. The *x*-axis depicts the observation time in years, the *y*-axis the probability of event-free survival at a given time-point. The lower table gives the number of individuals at risk, that is, the number of patients still included in the study at the given time-point, as well as the number of cumulative events in brackets for each follow-up year.

interquartile range (IQR) of UACR were 50.24 mg/g (IQR = 373.60 mg/g). Of the investigated patients, 60.2% were male. Of the 4640 patients, 185 (3.99%) developed ESRD over the mean observation period of 3.70 ± 0.88 years. The corresponding Kaplan−Meier-curve, together with the number of individuals at risk and the number of events per year in brackets, is shown in Figure 1. The event-free survival probabilities after 1, 2, 3, and 4 years were 99.6%, 98.5%, 97.2%, and 95.7%, respectively. Note that the Kaplan−Meier-curve is displayed for slightly more than 5 years corresponding to the maximum observation period.

Multivariate Analysis

We trained three different Cox PH models on the GCKD cohort and compared their performance in discriminating patients at risk of ESRD to that of the original Tangri score by means of the *C* statistic. Because the distributions of both eGFR and UACR values were heavily right skewed, see Figure S3, both were $\log_2$ transformed prior to inclusion in our Cox PH models. The first model included the four variables of the Tangri score. As evident from Figure 2, its median *C* statistic of 0.859 (95% CI, 0.826−0.900) over 100 subsampling rounds was slightly inferior to that of the original 5-year, four-variable Tangri score with a median *C* statistic of 0.863 (95% CI, 0.831−0.900). Improvements over the Tangri score, however, were observed upon inclusion of NMR features using a LASSO Cox PH model (Figure 2). Penalizing only the NMR features but not age, sex, eGFR, and UACR (referred to as metab_clin_score1) yielded a median *C* statistic of 0.867 (95% CI, 0.837−0.910) over 100 subsampling rounds, while the inclusion of both the NMR features and the clinical variables in the penalty term (metab_clin_score2) resulted in the best performing model with a median *C* statistic of 0.875 (95% CI, 0.850−0.911).

Next, we analyzed the distribution of differences in *C* statistic between the Tangri score and the three new models (Figure 3). As compared to the Tangri score, the best performing metab_clin_score2 model scored higher *C*

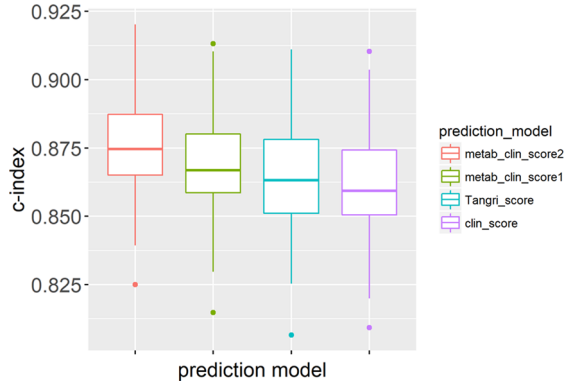


Figure 2. Comparison of the predictive performances of the four different ESRD risk scores according to concordance statistics. The four boxplots show from left to right the *C* statistics of 100 subsampling rounds each for the following risk models: (1) LASSO Cox PH model that penalized the 127 selected NMR features and the four clinical variables age, gender, eGFR, and UACR, with an equal weight of 1 in the l_1-norm of the penalty term (metab_clin_score2), (2) LASSO Cox PH model that penalized only the 127 NMR features with an equal weight of 1 in the l_1-norm of the penalty term, but not the four clinical variables (metab_clin_score1), (3) Cox PH model built on the four clinical variables only (clin_score), and (4) the four-variable Tangri score at 5 years for non-North Americans as proposed by Tangri et al.[9] For the subsampling rounds, except for the application of the original four-variable Tangri risk equation, which did not involve a training step, the complete data set was divided at random 100 times into a training and test set of 3093 and 1547 patients, respectively. The *C* statistic was computed for the test sets only and displayed in boxplots, where the midlines represent the median and the upper and lower limits of the boxes the third and first quartiles (75th and 25th percentiles), respectively.

statistics in 94 out of 100 subsampling rounds, with a median improvement of 0.011 (95% CI −0.001−0.028). The metab_clin_score1 model, in contrast, yielded a higher *C* statistic in only 76 of the 100 subsampling rounds (median

DOI: 10.1021/acs.jproteome.8b00983
J. Proteome Res. 2019, 18, 1796−1805

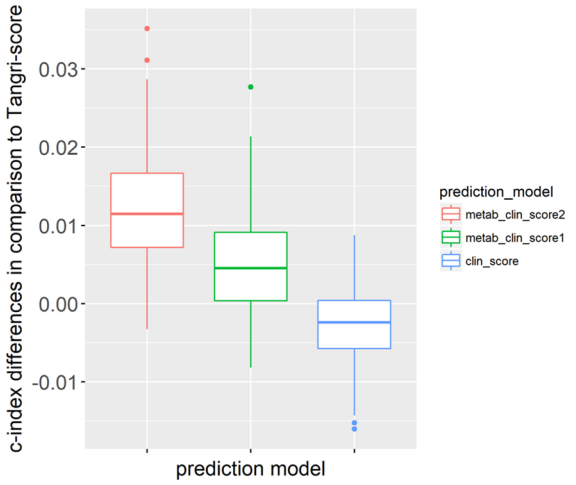


Figure 3. Boxplots comparing the predictive performances of the three Cox proportional hazards models generated here to that of the 5-year, four-variable Tangri score based on differences in the concordance statistic (*C* statistic). The midlines of the boxplots give the median differences in *C* statistic, while the upper and lower limits of the boxes represent the third and first quartiles (75th and 25th percentiles), respectively. The three models, to which the Tangri score was compared, are from left to right as follows: (1) LASSO Cox PH model that penalized the 127 selected NMR features and the four clinical variables age, gender, eGFR, and UACR, with an equal weight of 1 in the l_1-norm of the penalty term (metab_clin_score2), (2) LASSO Cox PH model that penalized only the 127 NMR features with an equal weight of 1 in the l_1-norm of the penalty term, but not the four clinical variables (metab_clin_score1), and (3) Cox PH model built on the four clinical variables only (clin_score). For this, the complete data set was divided into a training set, comprising 2/3 of the total number of 4640 samples, and a test set for the evaluation of *c*-indices, comprising the remaining 1/3 of the complete cohort. The whole procedure was repeated 100 times, each starting with a different subsampling for training and test data to assess the variability of the results.

improvement of 0.005 (95% CI −0.007−0.017)). Finally, the clin_score model outperformed the Tangri score in only 29% of the subsampling rounds (median decline of −0.002 (95% CI −0.014−0.006)).

Because of the used resampling approach, each split of the data into test and training data yielded individual models. To consolidate the best performing metab_clin_score2 model, the complete cohort of 4640 samples was used for LASSO Cox PH model estimation. The final model contained 24 NMR features and the clinical variables age, eGFR, and UACR. A list of all features with regression coefficients unequal to zero and their molecular identities is given in Table 2. The metabolic features having the largest absolute β-coefficients, and thus contributing the most to the predictive model, were creatinine (β = +1.25), high-density lipoprotein (HDL) (β = −0.91), valine (β = −0.76), acetyl groups of glycoproteins (β = +0.41), and Ca^{2+}-EDTA (β = −0.38), with positive and negative signs indicating ESRD risk increasing and decreasing features, respectively. As the patients in this study were treated with numerous medications, we carefully checked all NMR features of the final signature for contributions from drugs. Only for one feature, the feature located at 7.385 ppm and assigned to phenylalanine, did we find for some patients taking pain killers on a regular basis in a few cases minute contributions from paracetamol-glucuronide. However, these contributions were much smaller than the main signal of phenylalanine. Therefore,

Table 2. NMR and Clinical Features with LASSO Regression Coefficients Unequal to Zero in the Final metab_clin_score2 Model Trained on the Complete Data Set

NMR bucket [ppm]	LASSO coefficient[a]	bucket assignments[b]
4.065	1.253	creatinine
0.835	−0.907	HDL
0.995	−0.763	valine (major), lipids (minor)
3.045	0.629	creatinine (equal), creatine (equal), albumin lysyl (minor)
2.045	0.406	acetyl groups of glycoproteins (major), other acetyl groups (minor)
2.575	−0.382	Ca^{2+}-EDTA
1.925	−0.326	acetate
3.275	0.273	TMAO (major), D-glucose (minor), betaine (minor)
0.865	0.237	LDL (major), HDL (minor)
3.305	0.187	myo-inositol (equal), others (equal)
2.645	0.186	citrate
7.385	−0.172	phenylalanine
4.125	0.154	lactate (major), proline (minor)
1.485	−0.152	alanine
3.705	0.142	D-glucose
3.775	0.139	D-glucose (equal), glutamine (equal)
3.825	0.103	D-glucose
3.895	0.082	D-glucose
7.625	0.057	1-methylhistidine
7.685	0.056	pyridoxine, weak signal
8.125	0.051	small peptide, weak signal
2.375	−0.038	proline (equal), +glutamate (equal), others (minor)
7.415	−0.031	phenylalanine
3.865	0.014	D-glucose
eGFR (ckdepi)	−1.019	
UACR	0.251	
age	−0.009	

[a]LASSO coefficients denote the weights that individual NMR buckets and clinical variables carry in the final metab_clin_score2 model. [b]In case that NMR signals of more than one metabolite correspond to a given bucket, major and minor contributions are explicitly denoted.

it is safe to conclude that signals from medication have not obscured the differentiation between patients progressing to ESRD and those who did not. Of the clinical parameters, only eGFR, UACR, and age yielded β-coefficients unequal to zero, whereas gender was not included. Note that LASSO will select in the presence of collinear features only one of these features.

We further analyzed correlations between the Tangri score and the metab_clin_score2 (Figure 4). To this end, the results obtained for the model with the closest to median predictive performance were selected from the resampling approach. In most instances, scores agreed quite well with an overall Pearson correlation coefficient of r = 0.85. However, as can be seen from the upper left quadrant, there are several cases of renal events that were only identified upon inclusion of NMR features in the predictive model. This clearly shows that additional predictive information is provided by the metabolite data. However, because metab_clin_score2 also yielded more false positives, the overall improvement in predictive performance remained moderate.

Finally, the performance analysis was repeated taking nonfatal cardiovascular diseases as competing events into

DOI: 10.1021/acs.jproteome.8b00983
J. Proteome Res. 2019, 18, 1796−1805

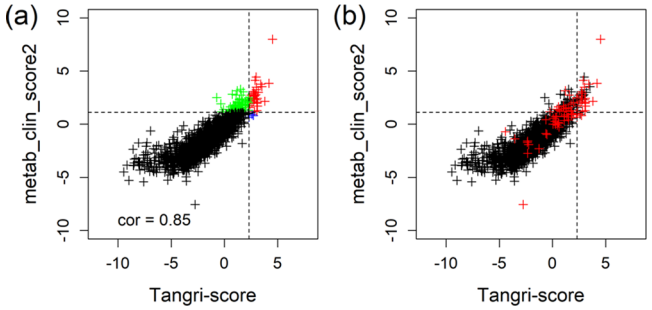


Figure 4. Correlation between the Tangri score (*x*-axis) and the metab_clin_score2 (*y*-axis) of each patient and the agreement between predicted and true ESRD events. (a) The dashed lines set classification boundaries for patients that had been predicted by both the Tangri and the metab_clin_score2 (red, upper right quadrant) to suffer a subsequent ESRD event, or by only the Tangri (blue, lower right quadrant) or the metab_clin_score2 (green upper left quadrant). The classification boundaries (thresholds) were set in a way to obtain an optimal separation of the respective survival curves of the two groups separated by the threshold. Patients who had not been predicted by either score to suffer an ESRD event are plotted in black. The Pearson correlation coefficient is given in the lower left corner. (b) Patients who had actually progressed to ESRD are plotted in red, while all other patients are plotted in black.

account. For the Tangri score, the LASSO Cox PH model, the metab_clin_score1, and the metab_clin_score2 median *C* statistics of 0.867, 0.864, 0.875, and 0.882 were obtained, respectively (Figure S4A and B). Comparison of Figures 3 and 4 with Figure S4A and B shows that cardiovascular events did not have a major effect on the metab_clin_score2 model performance. The body mass index (BMI) is one of the most common risk factors in cardiovascular disease. Therefore, we repeated the analyses where cardiovascular diseases were considered as competing events by including the BMI as an additional predictor (Figure S5A and B). A slight improvement in *C* statistics can be observed for both the metab_clin_bmi_score1 and the metab_clin_bmi_score2 in comparison to the metab_clin_score1 and the metab_clin_score2 with an improvement of 0.0017 in both cases.

DISCUSSION

End-stage renal disease, that is, the complete loss of renal function, is a significant outcome of chronic kidney disease. To survive, patients entering this final stage of CKD are in need of a renal replacement therapy, either dialysis or kidney transplantation. These therapies cause both severe impacts on a patient's quality of life and high costs for the health system. Thus, the early detection of CKD patients possibly progressing to ESRD is of utmost importance to either prevent or at least delay the complete loss of kidney function.

The four-variable Cox PH model published by Tangri et al. in 2011,[8] which took age, gender, eGFR, and the natural log transformed value of UACR into account, has shown great power in discriminating CKD patients at risk of progressing to ESRD. In a multinational assessment of 721 357 patients from 31 cohorts with CKD stages 3−5, the four-variable risk equation has yielded an overall accuracy of 90% and 88% in predicting risk of kidney failure at 2 and 5 years, respectively, with *C* statistics ranging from 0.79 to 0.99 for 2 years and 0.77 to 0.96 for 5 years.[9] With a median *C* statistic of 0.863 at 5 years, the impressive predictive performance of the four-variable Tangri score also held true for the 4640 participants of the German Chronic Kidney Disease (GCKD) study, despite its moderate kidney failure incidence of 10.77 per 1000 patient-years, and the fact that the present study included only 10.15% patients with CKD stages 4 and 5 as compared to the 33.2% of the original development cohort employed by Tangri et al.[8] Furthermore, the present study included 461 (9.9%) proteinuric patients (30 < 300 mg/L; 261 > 300 mg/L) with CKD stages 1 and 2 that had not been included in the original training of the Tangri score.[8] Nevertheless, there is an ongoing interest in further improvement of the four-variable risk equation, which has been neither met by the additional inclusion of baseline serum levels of calcium, phosphate, bicarbonate, and albumin in an expanded eight-variable Cox PH model,[8] or only very incrementally by a dynamic Cox PH model that included age, eGFR, as well as serum albumin, phosphorus, calcium, and bicarbonate as time-dependent covariates, upon which the C statistic improved to 0.91 as compared to 0.90 for the baseline visit eight-variable model.[12] However, given the already excellent performance of the four-variable Tangri score, any improvements will be rather incremental. Therefore, the average improvement of 1.27% in discrimination achieved over 100 subsampling rounds, in which the metab_clin_score2 model outperformed the four-variable Tangri score 94 times, can be still considered a significant advance despite its fairly incremental nature.

Given its reliability, minimum need for sample pretreatment other than the addition of buffer, deuterium oxide, and an internal standard for referencing and absolute quantification, as well as its near universal detection capability, ^{1}H NMR spectroscopy is a particularly attractive technique for large-scale studies requiring the analysis of thousands of serum, plasma, and urine specimens.[13,14] In this context, for the suppression of signals arising from macromolecules present in the sample matrix, often Carr−Purcell−Meiboom−Gill (CPMG) NMR pulse sequences will be used. Although it is possible to semiautomatically identify and quantify metabolites in ^{1}H NMR spectra, it is nowadays customary to use first machine learning algorithms to select from the totality of available features those that, for instance, predict best in a Cox PH model the outcome of disease, before identifying the compound(s) constituting a given predictive feature. Although the goal is to select as few predictors as possible from a high-dimensional data set, to facilitate their targeted analysis, a trade-off has to be found between model sparsity and optimal classification performance, which in the case of LASSO is determined by the penalization parameter λ. Here, the best performing model, metab_clin_score2, contained ultimately 24 out of 127 NMR features in addition to eGFR, UACR, and age. Further, on the basis of their chemical structure, compounds may be represented by multiple signals in an NMR spectrum. Generally, the LASSO should select only one of the signals of a compound, as they are highly correlated. However, given the complexity of plasma, signal overlap is expected, and, therefore, more than one bucket may be assigned to a compound, as it was the case for glucose (see Table 2). On the basis of the available metabolite reference libraries, it is possible to estimate the contributions of different compounds to a given bucket. However, this is not always feasible due to either the incompleteness of libraries or the contribution of functional groups from macromolecules whose signals have not been completely suppressed by the applied CPMG pulse sequence.

DOI: 10.1021/acs.jproteome.8b00983
J. Proteome Res. 2019, 18, 1796−1805

The feature that carried the largest weight in the ultimate metab_clin_score2 model was creatinine, the blood concentration of which is used routinely together with other readily available clinical parameters such as sex, age, and race, that function as surrogates for an individual's muscle mass, to estimate eGFR.[29] Interestingly, our metab_clin_score2 model includes both plasma creatinine measured by NMR spectroscopy and eGFR based on serum creatinine measured by an enzymatic assay. In absolute terms, plasma creatinine (β = 1.25) and eGFR (β = −1.02) have similar values in our model, except for the negative sign of the β-coefficient of eGFR, which is inversely proportional to serum creatinine. One possible explanation for the inclusion of both in the present model might be that creatinine is known to bind to albumin.[30] As demonstrated recently,[31] however, ^{1}H CPMG-NMR will determine only the unbound fraction of creatinine in plasma, the magnitude of which depends on the concentration of albumin and, possibly, also on the presence of other ligands competing with creatinine for albumin binding. In the present case, the ^{1}H −CPMG-NMR measured plasma concentration may likely serve as an indirect measure of plasma protein content, which is decreased in patients with proteinuria.

High-density lipoprotein (HDL) and low-density lipoprotein (LDL) were negatively and positively associated with the incidence of ESRD in our model, respectively. An inverse association between HDL cholesterol concentration and kidney function as well as risk for eGFR < 60 mL/min/1.73m^2 has been reported recently.[32] As far as LDL cholesterol is concerned, there is clear evidence that it contributes to atherosclerotic cardiovascular disease (ASCVD). In more than 30 randomized trials, reduction of plasma LDL has been found to reduce the risk of ASCVD.[33] Cardiovascular disease (CVD) and kidney disease are closely interrelated, and the disease of one organ causes dysfunction of the other.[34] Consequently, it is not unexpected that high levels of LDL are positively associated with an increased risk for ESRD.

We also found plasma levels of the branched-chain amino acid valine to be negatively associated with the risk for ESRD. This is in accordance with a recent study that found decreasing serum levels of valine, as well as alanine, which is also among the metabolites included in the ultimate metab_clin_score2 model, and tyrosine to be associated with decreasing kidney function.[35]

Ca^{2+}-EDTA was also negatively correlated with the risk of experiencing ESRD in our model. Interestingly, serum calcium is part of the Tangri 8-variable risk score, where higher levels of serum calcium are associated with a lower risk for ESRD in concordance with our findings. Additionally, in another study, an association between lower calcium levels and progressive CKD was reported, although it did not hold after adjustment for eGFR. Therefore, it was suggested that serum calcium is only a surrogate marker for GFR but not an independent risk factor for CKD progression.[36]

We found a positive association between myo-inositol and increased ESRD risk. Myo-inositol functions as an organic osmolyte that enables renal cells to adapt to hyperosmotic environments. Raised blood levels in patients with CKD have been previously noticed by Clements et al.[37] Additionally, Sekula et al.[38] have observed a highly significant inverse correlation between serum myo-inositol-levels and eGFR. They reported that myo-inositol concentrations were associated with eGFR but not incident CKD, suggesting that its association with reduced kidney function is a reflection of the altered ability of the kidney to degrade or excrete myo-inositol, but not a marker causally involved in the pathogenesis of CKD.

Diabetic kidney disease (DKD) is one of the most frequent complications of both type I and type II diabetes. It is the leading cause of ESRD, accounting for approximately 50% of the cases in the developed world.[39] The current study included 1630 diabetic patients, in 686 of whom diabetic nephropathy was the leading renal disease. Previously, Colhoun et al. reported that fasting plasma glucose levels in type II diabetes patients were associated with ESRD independent of other risk factors.[40] Here, higher blood glucose concentrations were also associated with the risk of adverse renal events.

Another interesting feature at 3.275 ppm represents mostly trimethylamine *N*-oxide (TMAO), a gut microbial-dependent metabolite of dietary choline, phosphatidylcholine, and L-carnitine. TMAO has been linked before in CKD patients to both a higher risk for progressive renal fibrosis and functional impairment, as well as an elevated mortality risk.[41] Further, in the Framingham Heart Study, TMAO was significantly associated with incident CKD.[42] Moreover, using a mixed graphical model, we have recently observed an association between baseline plasma TMAO and both cardiac arrhythmia and cardiac infarction in the GCKD study.[43]

There have been other, albeit smaller studies reporting novel biomarkers for progression to CKD or ESRD. For example, liquid chromatography−mass spectrometry-based profiling of 217 metabolites in plasma specimens from 1434 participants in the Framingham Heart Study yielded a panel of five metabolites, kynurenic aid, xanthosine, 5-hydroxyindoleacetic acid, kynurenine, and citrulline, the plasma levels of which led on top of established CKD risk factors to a significant improvement in the discrimination of individuals at risk of developing CKD (eGFR < 60 mL/min/1.73 m^2) as compared to a model taking only the established risk factor into account, with an increase of the C statistic from 0.77 to 0.83.[42] Using a combination of flow injection and liquid chromatography coupled to tandem mass spectrometry for the determination of the baseline serum concentrations of 140 metabolites in 1104 participants from the Cooperative Health Research in the Region of Augsburg Study, Goek et al. found instead the kynurenine-to-tryptophan ratio to be associated with a significantly higher incidence of CKD at follow-up visits.[44] For the identification of novel biomarkers associated with the progression to ESRD, Niewczas et al. employed a combination of gas and liquid chromatography−mass spectrometry for the comprehensive profiling of metabolites in baseline plasma specimens from 40 CKD patients each with type 2 diabetes that had either progressed to ESRD during 8−12 years of follow-up or had remained alive without ESRD.[45] Using logistic regression analysis, they found, for example, *p*-cresol sulfate, phenylacetylglutamine, myo-inositol, pseudouridine, and urate, to be significantly associated with the risk of progression to ESRD. Another nested case-control study, which used liquid chromatography−mass spectrometry to profile a total of 160 polar metabolites in baseline plasma specimens from 200 individuals each from the Chronic Renal Insufficiency Cohort (CRIC) Study with declining and stable eGFR, respectively, reported a composite score of the natural log transformed, standardized, and then summed signal intensities threonine, methionine, and arginine to be significantly associated with rapid CKD progression in a logistic regression model adjusted for age, sex, race, hypertension, systolic and diastolic blood pressure, diabetes, eGFR,

DOI: 10.1021/acs.jproteome.8b00983
J. Proteome Res. 2019, 18, 1796−1805

and 24-h urine protein.[46] In summary, these studies show that, depending on the cohort and metabolomics approach chosen, different metabolites may be identified that putatively improve risk prediction.

A limitation of the present study, as well as those mentioned above, is that its findings cannot be generalized, as the discriminatory power of a given risk equation may quite vary across different cohorts as exemplified by the multinational assessment of the accuracy of the four-variable risk equation developed by Tangri et al.[8] While ^{1}H CPMG-NMR allows not only the detection of metabolites but also functional groups such as acetyl groups found on macromolecules, the technology is still not widely available due to instrument cost. Further, assessment of the accuracy of the present risk equation using methodology other than NMR is hampered by the fact that the NMR features included in the equation are likely to represent more than just the identified compounds list in Table 2, although the contribution of the unknown compounds to the feature integrals may be too small to exert a significant effect on the performance of the model. Further, while ^{1}H CPMG-NMR offers in contrast to hyphenated mass spectrometric methods the advantage that serum or plasma samples may be analyzed directly without prior protein precipitation, one has to keep in mind that in case of protein binding ^{1}H CPMG-NMR will not measure the total but rather the unbound concentration of a metabolite provided that the metabolite is in slow exchange between its bound and free state.[31] Consequently, plasma may have to be ultrafiltered prior to analysis. However, one should keep in mind nonspecific binding of analytes to the plastic components or the ultrafiltration membrane of the filtration device, or in rare instances even the conversion of analytes as reported for glutamine, which has been observed to convert to pyroglutamic acid through cyclization during ultrafiltration.[47]

CONCLUSIONS

It remains to be elucidated whether a dynamic model including features of the latest available rather than the baseline NMR plasma fingerprint will yield a further improvement in predictive power. Nevertheless, the current study demonstrates the value of NMR-based metabolomics in combination with traditional clinical markers for the timely discrimination of CKD patients at risk of progressing rapidly to ESRD.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.8b00983.

Details regarding NMR spectroscopy; details regarding data preprocessing; details regarding multivariate data analysis; Figure S1, flowchart of patient selection procedure and end-point determination; Figure S2, selection of buckets for further analysis; Figure S3, distributions of eGFR and UACR values before and after $\log_2$ transformation; Figure S4, comparison of risk scores upon consideration of cardiovascular events as competing events; and Figure S5, comparison of risk scores upon consideration of cardiovascular events as competing events with BMI as additional predictor variable (PDF)

AUTHOR INFORMATION

Corresponding Author

*Phone: +49-(0)941-943-5015. Fax: +49-(0)941-943-5020. E-mail: wolfram.gronwald@klinik.uni-regensburg.de.

ORCID

Wolfram Gronwald: 0000-0003-3646-0060

Notes

The authors declare no competing financial interest.

Data Availability: NMR spectra are available via the publicly accessible MetaboLights database https://www.ebi.ac.uk/metabolights/accession ID MTBLS798. Patients provided written informed consent for their data to be shared within the scope of scientific collaborations. The authors should therefore be contacted with collaboration requests.

ACKNOWLEDGMENTS

This work was supported in part by the German Federal Ministry of Education and Research (BMBF grant no. 01 ER 0821) and the German Research Foundation (SFB1350 grant no. 387509280). We thank Dr. Vera Krane for critical reading of the manuscript. GCKD investigators are University of Erlangen-Nürnberg: Kai-Uwe Eckardt, Heike Meiselbach, Markus Schneider, Thomas Dienemann, Hans-Ulrich Prokosch, Barbara Bärthlein, Andreas Beck, Thomas Ganslandt, André Reis, Arif B. Ekici, Susanne Avendaño, Dinah Becker-Grosspitsch, Ulrike Alberth-Schmidt, Birgit Hausknecht, Rita Zitzmann, Anke Weigel; University of Freiburg: Gerd Walz, Anna Köttgen, Ulla Schultheiß, Fruzsina Kotsis, Simone Meder, Erna Mitsch, Ursula Reinhard; RWTH Aachen University: Jürgen Floege, Georg Schlieper, Turgay Saritas, Sabine Ernst, Nicole Beaujean; Charité, University Medicine Berlin: Elke Schaeffner, Seema Baid-Agrawal, Kerstin Theisen; Hannover Medical School: Hermann Haller, Jan Menne; University of Heidelberg: Martin Zeier, Claudia Sommerer, Rebecca Woitke; University of Jena: Gunter Wolf, Martin Busch, Rainer Fuß; Ludwig-Maximilians University of München: Thomas Sitter, Claudia Blank; University of Würzburg: Christoph Wanner, Vera Krane, Antje Börner-Klein, Britta Bauer; Medical University of Innsbruck, Division of Genetic Epidemiology: Florian Kronenberg, Julia Raschenberger, Barbara Kollerits, Lukas Forer, Sebastian Schönherr, Hansi Weissensteiner; University of Regensburg, Institute of Functional Genomics: Peter Oefner, Wolfram Gronwald, Helena Zacharias; and Department of Medical Biometry, Informatics and Epidemiology (IMBIE), University Hospital of Bonn: Matthias Schmid, Jennifer Nadal.

REFERENCES

(1) Jha, V.; Garcia-Garcia, G.; Iseki, K.; Li, Z.; Naicker, S.; Plattner, B.; Saran, R.; Wang, A. Y.-M.; Yang, C.-W. Chronic kidney disease: global dimension and perspectives. *Lancet* **2013**, *382*, 260−272.

(2) Titze, S.; Schmid, M.; Köttgen, A.; Busch, M.; Floege, J.; Wanner, C.; Kronenberg, F.; Eckardt, K.-U. Disease burden and risk profile in referred patients with moderate chronic kidney disease: composition of the German Chronic Kidney Disease (GCKD) cohort. *Nephrol., Dial., Transplant.* **2015**, *30*, 441−451.

(3) KDIGO workgroup. KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. *Kidney Int. Suppl.* **2013**, *3*, 1−150.

(4) Eckardt, K.-U.; Coresh, J.; Devuyst, O.; Johnson, R. J.; Köttgen, A.; Levey, A. S.; Levin, A. Evolving importance of kidney disease.

DOI: 10.1021/acs.jproteome.8b00983
J. Proteome Res. 2019, 18, 1796−1805

From subspecialty to global health burden. *Lancet* **2013**, *382*, 158−169.
(5) Kuhlmann, U.; Walb, D.; Luft, F. C. *Nephrologie. Pathophysiologie - Klinik - Nierenersatzverfahren*, 4th ed.; Thieme: Stuttgart, 2003.
(6) Thomas, B.; Matsushita, K.; Abate, K. H.; Al-Aly, Z.; Ärnlöv, J.; Asayama, K.; Atkins, R.; Badawi, A.; Ballew, S. H.; Banerjee, A.; Barregård, L.; Barrett-Connor, E.; Basu, S.; Bello, A. K.; Bensenor, I.; Bergstrom, J.; Bikbov, B.; Blosser, C.; Brenner, H.; Carrero, J.-J.; Chadban, S.; Cirillo, M.; Cortinovis, M.; Courville, K.; Dandona, L.; Dandona, R.; Estep, K.; Fernandes, J.; Fischer, F.; Fox, C.; Gansevoort, R. T.; Gona, P. N.; Gutierrez, O. M.; Hamidi, S.; Hanson, S. W.; Himmelfarb, J.; Jassal, S. K.; Jee, S. H.; Jha, V.; Jimenez-Corona, A.; Jonas, J. B.; Kengne, A. P.; Khader, Y.; Khang, Y.-H.; Kim, Y. J.; Klein, B.; Klein, R.; Kokubo, Y.; Kolte, D.; Lee, K.; Levey, A. S.; Li, Y.; Lotufo, P.; El Razek, Hassan Magdy Abd; Mendoza, W.; Metoki, H.; Mok, Y.; Muraki, I.; Muntner, P. M.; Noda, H.; Ohkubo, T.; Ortiz, A.; Perico, N.; Polkinghorne, K.; Al-Radaddi, R.; Remuzzi, G.; Roth, G.; Rothenbacher, D.; Satoh, M.; Saum, K.-U.; Sawhney, M.; Schöttker, B.; Shankar, A.; Shlipak, M.; Silva, D. A. S.; Toyoshima, H.; Ukwaja, K.; Umesawa, M.; Vollset, S. E.; Warnock, D. G.; Werdecker, A.; Yamagishi, K.; Yano, Y.; Yonemoto, N.; Zaki, M. E. S.; Naghavi, M.; Forouzanfar, M. H.; Murray, C. J. L.; Coresh, J.; Vos, T. Global Cardiovascular and Renal Outcomes of Reduced GFR. *J. Am. Soc. Nephrol.* **2017**, *28*, 2167−2179.
(7) Eckardt, K.-U.; Barthlein, B.; Baid-Agrawal, S.; Beck, A.; Busch, M.; Eitner, F.; Ekici, A. B.; Floege, J.; Gefeller, O.; Haller, H.; Hilge, R.; Hilgers, K. F.; Kielstein, J. T.; Krane, V.; Kottgen, A.; Kronenberg, F.; Oefner, P.; Prokosch, H.-U.; Reis, A.; Schmid, M.; Schaeffner, E.; Schultheiss, U. T.; Seuchter, S. A.; Sitter, T.; Sommerer, C.; Walz, G.; Wanner, C.; Wolf, G.; Zeier, M.; Titze, S. The German Chronic Kidney Disease (GCKD) study: design and methods. *Nephrol., Dial., Transplant.* **2012**, *27*, 1454−1460.
(8) Tangri, N.; Stevens, L. A.; Griffith, J.; Tighiouart, H.; Djurdjev, O.; Naimark, D.; Levin, A.; Levey, A. S. A predictive model for progression of chronic kidney disease to kidney failure. *JAMA* **2011**, *305*, 1553−1559.
(9) Tangri, N.; Grams, M. E.; Levey, A. S.; Coresh, J.; Appel, L. J.; Astor, B. C.; Chodick, G.; Collins, A. J.; Djurdjev, O.; Elley, C. R.; Evans, M.; Garg, A. X.; Hallan, S. I.; Inker, L. A.; Ito, S.; Jee, S. H.; Kovesdy, C. P.; Kronenberg, F.; Heerspink, H. J. L.; Marks, A.; Nadkarni, G. N.; Navaneethan, S. D.; Nelson, R. G.; Titze, S.; Sarnak, M. J.; Stengel, B.; Woodward, M.; Iseki, K. Multinational Assessment of Accuracy of Equations for Predicting Risk of Kidney Failure: A Meta-analysis. *JAMA* **2016**, *315*, 164−174.
(10) Grams, M. E.; Li, L.; Greene, T. H.; Tin, A.; Sang, Y.; Kao, W. H. L.; Lipkowitz, M. S.; Wright, J. T.; Chang, A. R.; Astor, B. C.; Appel, L. J. Estimating time to ESRD using kidney failure risk equations: results from the African American Study of Kidney Disease and Hypertension (AASK). *Am. J. Kidney Dis.* **2015**, *65*, 394−402.
(11) Peeters, M. J.; van Zuilen, A. D.; van den Brand, Jan A J G; Bots, M. L.; Blankestijn, P. J.; Wetzels, J. F. M. Validation of the kidney failure risk equation in European CKD patients. *Nephrol., Dial., Transplant.* **2013**, *28*, 1773−1779.
(12) Tangri, N.; Inker, L. A.; Hiebert, B.; Wong, J.; Naimark, D.; Kent, D.; Levey, A. S. A Dynamic Predictive Model for Progression of CKD. *Am. J. Kidney Dis.* **2017**, *69*, 514−520.
(13) Tzoulaki, I.; Ebbels, T. M. D.; Valdes, A.; Elliott, P.; Ioannidis, J. P. A. Design and analysis of metabolomics studies in epidemiologic research: a primer on -omic technologies. *Am. J. Epidemiol.* **2014**, *180*, 129−139.
(14) Nicholson, J. K.; Lindon, J. C. Systems biology. Metabonomics. *Nature* **2008**, *455*, 1054−1056.
(15) Bairaktari, E.; Seferiadis, K.; Liamis, G.; Psihogios, N.; Tsolas, O.; Elisaf, M. Rhabdomyolysis-related renal tubular damage studied by proton nuclear magnetic resonance spectroscopy of urine. *Clin. Chem.* **2002**, *48*, 1106−1109.
(16) Sieber, M.; Hoffmann, D.; Adler, M.; Vaidya, V. S.; Clement, M.; Bonventre, J. V.; Zidek, N.; Rached, E.; Amberg, A.; Callanan, J. J.; Dekant, W.; Mally, A. Comparative analysis of novel noninvasive renal biomarkers and metabonomic changes in a rat model of gentamicin nephrotoxicity. *Toxicol. Sci.* **2009**, *109*, 336−349.
(17) Atzori, L.; Antonucci, R.; Barberini, L.; Locci, E.; Cesare Marincola, F.; Scano, P.; Cortesi, P.; Agostiniani, R.; Weljie, A.; Lai, A.; Fanos, V. 1H NMR-based metabolic profiling of urine from children with nephrouropathies. *Front. Biosci., Elite Ed.* **2010**, *2*, 725−732.
(18) Gronwald, W.; Klein, M. S.; Zeltner, R.; Schulze, B. D.; Reinhold, S. W.; Deutschmann, M.; Immervoll, A. K.; Böger, C. A.; Banas, B.; Eckardt, K. U.; Oefner, P. J. Detection of Autosomal Polycystic Kidney Disease Using NMR Spectroscopic Fingerprints of Urine. *Kidney Int.* **2011**, *79*, 1244−1253.
(19) Qi, S.; Ouyang, X.; Wang, L.; Peng, W.; Wen, J.; Dai, Y. A pilot metabolic profiling study in serum of patients with chronic kidney disease based on 1H-NMR- spectroscopy. *Clin. Transl. Sci.* **2012**, *5*, 379−385.
(20) Posada-Ayala, M.; Zubiri, I.; Martin-Lorenzo, M.; Sanz-Maroto, A.; Molero, D.; Gonzalez-Calero, L.; Fernandez-Fernandez, B.; La Cuesta, F. de; Laborde, C. M.; Barderas, M. G.; Ortiz, A.; Vivanco, F.; Alvarez-Llamas, G. Identification of a urine metabolomic signature in patients with advanced-stage chronic kidney disease. *Kidney Int.* **2014**, *85*, 103−111.
(21) Zacharias, H. U.; Schley, G.; Hochrein, J.; Klein, M. S.; Köberle, C.; Eckardt, K. U.; Willam, C.; Oefner, P. J.; Gronwald, W. Analysis of Human Urine Reveals Metabolic Changes Related to the Development of Acute Kidney Injury Following Cardiac Surgery. *Metabolomics* **2013**, *9*, 697−707.
(22) Zacharias, H. U.; Hochrein, J.; Vogl, F. C.; Schley, G.; Mayer, F.; Jeleazcov, C.; Eckardt, K.-U.; Willam, C.; Oefner, P. J.; Gronwald, W. Identification of Plasma Metabolites Prognostic of Acute Kidney Injury after Cardiac Surgery with Cardiopulmonary Bypass. *J. Proteome Res.* **2015**, *14*, 2897−2905.
(23) Kalantari, S.; Nafar, M.; Samavat, S.; Parvin, M. 1H NMR-based metabolomics study for identifying urinary biomarkers and perturbed metabolic pathways associated with severity of IgA nephropathy: a pilot study. *Magn. Reson. Chem.* **2017**, *55*, 693−699.
(24) Levey, A. S.; Stevens, L. A.; Schmid, C. H.; Zhang, Y.; Castro, A. F.; Feldman, H. I.; Kusek, J. W.; Eggers, P.; van Lente, F.; Greene, T.; Coresh, J. A New Equation to Estimate Glomerular Filtration Rate. *Ann. Intern. Med.* **2009**, *150*, 604.
(25) Friedman, J.; Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **2010**, *33*, 1−22.
(26) Tibshirani, R. The LASSO Method for Variable Selection in the COX Model. *Stat. Med.* **1997**, *16*, 385−395.
(27) Uno, H.; Cai, T.; Pencina, M. J.; D'Agostino, R. B.; Wei, L. J. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat. Med.* **2011**, *30*, 1105−1117.
(28) Schröder, M. S.; Culhane, A. C.; Quackenbush, J.; Haibe-Kains, B. survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. *Bioinformatics* **2011**, *27*, 3206−3208.
(29) Stevens, L. A.; Levey, A. S. Measured GFR as a confirmatory test for estimated GFR. *J. Am. Soc. Nephrol.* **2009**, *20*, 2305−2313.
(30) Varshney, A.; Rehan, M.; Subbarao, N.; Rabbani, G.; Khan, R. H. Elimination of endogenous toxin, creatinine from blood plasma depends on albumin conformation: site specific uremic toxicity & impaired drug binding. *PLoS One* **2011**, *6*, e17230.
(31) Wallmeier, J.; Samol, C.; Ellmann, L.; Zacharias, H. U.; Vogl, F. C.; Garcia, M.; Dettmer, K.; Oefner, P. J.; Gronwald, W. Quantification of Metabolites by NMR Spectroscopy in the Presence of Protein. *J. Proteome Res.* **2017**, *16*, 1784−1796.
(32) Lanktree, M. B.; Thériault, S.; Walsh, M.; Paré, G. HDL Cholesterol, LDL Cholesterol, and Triglycerides as Risk Factors for CKD: A Mendelian Randomization Study. *Am. J. Kidney Dis.* **2018**, *71*, 166−172.
(33) Ference, B. A.; Ginsberg, H. N.; Graham, I.; Ray, K. K.; Packard, C. J.; Bruckert, E.; Hegele, R. A.; Krauss, R. M.; Raal, F. J.;

DOI: 10.1021/acs.jproteome.8b00983
J. Proteome Res. 2019, 18, 1796−1805

Schunkert, H.; Watts, G. F.; Borén, J.; Fazio, S.; Horton, J. D.; Masana, L.; Nicholls, S. J.; Nordestgaard, B. G.; van de Sluis, B.; Taskinen, M.-R.; Tokgözoglu, L.; Landmesser, U.; Laufs, U.; Wiklund, O.; Stock, J. K.; Chapman, M. J.; Catapano, A. L. Low-Density Lipoproteins Cause Atherosclerotic Cardiovascular Disease. 1. Evidence from Genetic, Epidemiologic, and Clinical Studies. A Consensus Statement from the European Atherosclerosis Society Consensus Panel. *Eur. Heart J.* **2017**, *38*, 2459−2472.

(34) Liu, M.; Li, X.-C.; Lu, L.; Cao, Y.; Sun, R.-R.; Chen, S.; Zhang, P.-Y. Cardiovascular disease and its relationship with chronic kidney disease. *Eur. Rev. Med. Pharmaco.* **2014**, *18*, 2918−2926.

(35) Tizianello, A.; Deferrari, G.; Garibotto, G.; Robaudo, C.; Lutman, M.; Passerone, G.; Bruzzone, M. Branched-chain amino acid metabolism in chronic renal failure. *Kidney Int. Suppl.* **1983**, *16*, S17−22.

(36) Schwarz, S.; Trivedi, B. K.; Kalantar-Zadeh, K.; Kovesdy, C. P. Association of disorders in mineral metabolism with progression of chronic kidney disease. *Clin. J. Am. Soc. Nephrol.* **2006**, *1*, 825−831.

(37) Clements, R.; Dejesus, P. V.; Winegrad, A. Raised Plasma-Myoinositol Levels in Uraemia and Experimental Neuropathy. *Lancet* **1973**, *301*, 1137−1141.

(38) Sekula, P.; Goek, O.-N.; Quaye, L.; Barrios, C.; Levey, A. S.; Römisch-Margl, W.; Menni, C.; Yet, I.; Gieger, C.; Inker, L. A.; Adamski, J.; Gronwald, W.; Illig, T.; Dettmer, K.; Krumsiek, J.; Oefner, P. J.; Valdes, A. M.; Meisinger, C.; Coresh, J.; Spector, T. D.; Mohney, R. P.; Suhre, K.; Kastenmüller, G.; Köttgen, A. A Metabolome-Wide Association Study of Kidney Function and Disease in the General Population. *J. Am. Soc. Nephrol.* **2016**, *27*, 1175−1188.

(39) Tuttle, K. R.; Bakris, G. L.; Bilous, R. W.; Chiang, J. L.; Boer, I. H. de; Goldstein- Fuchs, J.; Hirsch, I. B.; Kalantar-Zadeh, K.; Narva, A. S.; Navaneethan, S. D.; Neumiller, J. J.; Patel, U. D.; Ratner, R. E.; Whaley-Connell, A. T.; Molitch, M. E. Diabetic kidney disease: a report from an ADA Consensus Conference. *Diabetes Care* **2014**, *37*, 2864−2883.

(40) Colhoun, H. M.; Lee, E. T.; Bennett, P. H.; Lu, M.; Keen, H.; Wang, S. L.; Stevens, L. K.; Fuller, J. H. Risk factors for renal failure: the WHO Mulinational Study of Vascular Disease in Diabetes. *Diabetologia* **2001**, *44* (Suppl 2), S46−53.

(41) Tang, W. H. W.; Wang, Z.; Kennedy, D. J.; Wu, Y.; Buffa, J. A.; Agatisa-Boyle, B.; Li, X. S.; Levison, B. S.; Hazen, S. L. Gut microbiota-dependent trimethylamine N- oxide (TMAO) pathway contributes to both development of renal insufficiency and mortality risk in chronic kidney disease. *Circ. Res.* **2015**, *116*, 448−455.

(42) Rhee, E. P.; Clish, C. B.; Ghorbani, A.; Larson, M. G.; Elmariah, S.; McCabe, E.; Yang, Q.; Cheng, S.; Pierce, K.; Deik, A.; Souza, A. L.; Farrell, L.; Domos, C.; Yeh, R. W.; Palacios, I.; Rosenfield, K.; Vasan, R. S.; Florez, J. C.; Wang, T. J.; Fox, C. S.; Gerszten, R. E. A combined epidemiologic and metabolomic approach improves CKD prediction. *J. Am. Soc. Nephrol.* **2013**, *24*, 1330−1338.

(43) Zacharias, H. U.; Altenbuchinger, M.; Solbrig, S.; Schäfer, A.; Büyüközkan, M.; Schultheiss, U. T.; Kotsis, F.; Köttgen, A.; Krumsiek, J.; Theis, F. J.; Spang, R.; Oefner, P. J.; Gronwald, W. Fully integrative data analysis of NMR metabolic fingerprints with comprehensive patient data: a case report based on the German Chronic Kidney Disease (GCKD) study, arXiv, 2018; 1810.04281.

(44) Goek, O.-N.; Prehn, C.; Sekula, P.; Römisch-Margl, W.; Döring, A.; Gieger, C.; Heier, M.; Koenig, W.; Wang-Sattler, R.; Illig, T.; Suhre, K.; Adamski, J.; Köttgen, A.; Meisinger, C. Metabolites associate with kidney function decline and incident chronic kidney disease in the general population. *Nephrol., Dial., Transplant.* **2013**, *28*, 2131−2138.

(45) Niewczas, M. A.; Sirich, T. L.; Mathew, A. V.; Skupien, J.; Mohney, R. P.; Warram, J. H.; Smiles, A.; Huang, X.; Walker, W.; Byun, J.; Karoly, E. D.; Kensicki, E. M.; Berry, G. T.; Bonventre, J. V.; Pennathur, S.; Meyer, T. W.; Krolewski, A. S. Uremic solutes and risk of end-stage renal disease in type 2 diabetes: metabolomic study. *Kidney Int.* **2014**, *85*, 1214−1224.

(46) Rhee, E. P.; Clish, C. B.; Wenger, J.; Roy, J.; Elmariah, S.; Pierce, K. A.; Bullock, K.; Anderson, A. H.; Gerszten, R. E.; Feldman, H. I. Metabolomics of Chronic Kidney Disease Progression: A Case-Control Analysis in the Chronic Renal Insufficiency Cohort Study. *Am. J. Nephrol.* **2016**, *43*, 366−374.

(47) Nagana Gowda, G. A.; Gowda, Y. N.; Raftery, D. Massive glutamine cyclization to pyroglutamic acid in human serum discovered using NMR spectroscopy. *Anal. Chem.* **2015**, *87*, 3800−3805.

DOI: 10.1021/acs.jproteome.8b00983
J. Proteome Res. 2019, 18, 1796−1805