Differential metabolomic signatures of declining renal function in Types 1 and 2 diabetes

Manca, Maria Laura

2021-10-01

Manca , M L , Solini , A , Haukka , J K , Sandholm , N , Forsblom , C , Groop , P H & Ferrannini , E 2021 , ' Differential metabolomic signatures of declining renal function in Types 1 and 2 diabetes ' , Nephrology Dialysis Transplantation , vol. 36 , no. 10 , pp. 1859-1866 . https://doi.org/10.1093/ndt/gfaa175

http://hdl.handle.net/10138/340442
https://doi.org/10.1093/ndt/gfaa175

unspecified
acceptedVersion

Differential metabolomic signatures of declining renal function in type 1 and type 2 diabetes

Maria Laura Manca,[1*] Anna Solini,[2*] Jani K Haukka,[3,4,5] Niina Sandholm,[3,4,5] Carol Forsblom,[3,4,5] Per-Henrik Groop,[3,4,5,6] and Ele Ferrannini[7]

equal contribution*

[1] Department of Clinical and Experimental Medicine, University of Pisa, Italy

[2] Department of Surgical, Medical, Molecular and Critical Area Pathology, University of Pisa, Italy

[3] Folkhälsan Institute of Genetics, Folkhälsan Research Center, Helsinki, Finland

[4] Abdominal Center Nephrology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

[5] Research Program for Clinical and Molecular Metabolism, Faculty of Medicine, University of Helsinki, Helsinki, Finland

[6] Department of Diabetes, Central Clinical School, Monash University, Melbourne, Victoria, Australia

[7] CNR Institute of Clinical Physiology, Pisa, Italy

Corresponding Author:

Ele Ferrannini

CNR Institute of Clinical Physiology, Pisa, Italy

Phone: +39050553235

Email: ferranni@ifc.cnr.it

ORCID ID: 0000-0002-1384-1584

Abstract

Background Chronic kidney disease (CKD) shows different clinical features in type 1 (T1D) and type 2 diabetes (T2D). Metabolomics have recently provided useful contribution to the identification of biomarkers of CKD progression in either form of the disease. However, no studies have so far compared plasma metabolomics between T1D and T2D in order to identify differential signatures of progression of eGFR decline.

Methods We used two large cohorts of T1D (from Finland) and T2D (from Italy) patients followed up to seven and three years, respectively. In both groups, progression was defined as the top quartile of yearly decline in eGFR. Pooled data from the two groups were analysed by univariate and bivariate random forest (RF), and confirmed by bivariate partial least squares (PLS) analysis, the response variables being type of diabetes and eGFR progression.

Results In progressors, yearly eGFR loss was significantly larger in T2D (-5.3[3.0] $ml \cdot min^{-1} \cdot 1.73m^{-2} \cdot yr^{-1}$) than T1D (-3.7[3.1], median[IQR], p=0.018). Out of several hundreds, bivariate RF extracted 22 metabolites associated with diabetes type (all higher in T1D than T2D except for 5-methylthioadenosine, pyruvate and ß-hydroxypyruvate.) and 13 molecules associated with eGFR progression (all higher in progressors than non-progressors except for sphyngomyelin). Three of the selected metabolites (histidylphenylalanine, leucylphenylalanine, tryptophylasparagine) showed a significant interaction between disease type and progression. Only 8 metabolites were common to both bivariate RF and PLS.

Conclusions Identification of metabolomic signatures of CKD progression is partially dependent on the statistical model. Dual analysis identified molecules specifically associated with progressive renal impairment in both T1D and T2D.

What is already known about this subject:

- Diabetic kidney disease impairs survival in type 1 and type 2 diabetes, and strategies to better identify, target, and treat people at risk of progressing to end-stage renal disease are an urgent need.

- Metabolomics is potentially useful for diagnosis, patient stratification and monitoring of therapeutic responses. A small group of metabolites as good markers of rapid GFR decline or progression toward albuminuria (or both) have been recently identified.

- No studies have, so far, compared metabolites as biomarkers of chronic renal disease in individuals with type 1 and type 2 diabetes.

What this study adds:

- In the attempt to identify differential metabolomic signatures of disease progression between type 1 or type 2 diabetes by using the same analytical platform and outcome variable, *i.e.*, the decline in eGFR, we found that the identification of associated metabolites partially depends on the statistical method of analysis.

- Out of several hundreds of metabolites, three (histidylphenylalanine, leucylphenylalanine, tryptophylasparagine) showed a significant interaction between disease type and progression of renal damage.

What impact this may have on practice or policy:

- To sharpen data interpretation, identification of metabolomic signatures of chronic kidney disease progression in diabetes should be validated by multiple statistical approaches.

Keywords: metabolomics; type 1 diabetes; type 2 diabetes; nephropathy; GFR

Introduction

Chronic kidney disease (CKD) is a major health burden in people with diabetes as its presence accelerates the physiologic, age-dependent decline in glomerular filtration rate (GFR). Renal impairment is also characterized by increased all-cause and cardiovascular (CVD) mortality and morbidity [1-3]. CVD mortality is at least 8-10-fold higher in patients with CKD than in the general population, and has not improved substantially in the past two decades. The alarming rise in the prevalence of both type 1 (T1D) and type 2 diabetes (T2D) amplifies the problem, calling for strategies to better identify, target, and treat people at risk of progressing to end-stage renal disease (ESRD) [4,5]. Despite the epidemiological relevance of the phenomenon and its large economic impact [6], limited therapeutic options are currently available to slow the progression of CKD in either T1D or T2D individuals; this is also due to the lack of suitable approaches for the early detection of renal damage. Age, gender, estimated GFR (eGFR) and albuminuria, the currently used markers to evaluate kidney function [7,8], lack sensitivity and specificity, and CKD often goes undetected until substantial renal injury has occurred. Therefore, a strong investigational effort has been made in the last decade in the search of molecules capable of more accurately identifying patients in whom clinical intervention is more likely to be effective.

Advances in laboratory technologies in the past decade have generated a rich panel of potential biomarkers of renal damage in diabetes [9,10]; randomized clinical trials are validating some of them, in the prospect of personalized medicine [11]. One of the most fruitful approaches is metabolomics, which can select and quantify small molecules linked with physiological homeostasis and gene-environment interactions and potentially useful for diagnosis, patient stratification, and monitoring of therapeutic responses. In the last five years, studies in T1D and T2D have recognized a relatively small group of metabolites as good

markers of rapid GFR decline or progression toward albuminuria (or both); some of these metabolites have received confirmation by multiple studies [12-14].

However, previous studies have used either proteinuria or GFR as the response variable [15,16], and markers have been screened in the plasma or urine [14,17]. In addition, no studies have attempted to distinguish CKD biomarkers for T1D *vs* T2D despite the well-established notion that CKD in T1D is predominantly diabetic nephropathy proper (with chronic hyperglycemia as the main pathogenetic factor) whereas CKD in T2D is a case mix of diabetic nephropathy, vascular and interstitial kidney disease [18]. This heterogeneity of approaches has prompted us to carry out the present longitudinal study, in which matrix (plasma), response variable (eGFR), and metabolomic platform were the same in relatively large groups of T1D and T2D patients. The specific question asked was: are there metabolites that reliably differentiate CKD progression in T1D *vs* T2D?

Subjects and Methods

Patients The study population consisted of a cohort of individuals with T1D and one of patients with T2D. The databases analyzed in the present manuscript have been previously used separately to explore the impact of albuminuria in type 2 or type 1 diabetes [13,19]. T1D patients were selected from the Finnish Diabetic Nephropathy Study Group (FinnDiane) [20]. T1D was defined as age at onset of diabetes below 40 years and insulin treatment initiated within one year of diagnosis. Data on recruitment (completed in 2004) and clinical characterization of patients have been presented in detail elsewhere [9,19]. Inclusion criteria were a normal albumin excretion rate <30 mg/day, an eGFR above 60 $ml \cdot min^{-1} \cdot 1.73m^{-2}$ at baseline, and available follow-up data on kidney status. Median follow up in these patients was 7.1 years [19]. The study protocol was approved by Ethics Committees of the Helsinki and Uusimaa Hospital districts as well as by the local ethics committees at each FinnDiane participating

center, and all patients signed a written informed consent. The study was performed in accordance with the Declaration of Helsinki.

Type 2 patients were recruited from two Italian outpatient clinics in 2011. Exclusion criteria were dialysis or transplantation, an eGFR below 30 $ml \cdot min^{-1} \cdot 1.73m^2$, liver disease and cancer. Follow up in these patients was 3 years [13]. The study protocol was approved by the respective Institutional Ethics Committees; patients signed an informed consent. All subjects volunteered to provide a fasting serum sample; other aliquots were collected for routine analyses. All samples were labeled with a blinded code and processed anonymously.

In each subject, the yearly change in eGFR (in $ml \cdot min^{-1} \cdot 1.73m^{-2} \cdot yr^{-1}$) was calculated as the difference in eGFR between follow up and baseline divided by the duration of follow up. Separately in each cohort, progression was defined as a change in eGFR falling into the top quartile of its distribution. The respective cutoff was an eGFR drop ≥ 7.0 $ml \cdot min^{-1} \cdot 1.73m^{-2} \cdot yr^{-1}$ in T2D patients, and one of 5.9 $ml \cdot min^{-1} \cdot 1.73m^{-2} \cdot yr^{-1}$ in T1D subjects.

Laboratory measurements Plasma glucose, serum creatinine and HbA_{1c} concentrations were measured by standard methods. eGFR was calculated by the CKD-EPI formula.

Sample Preparation Samples from both cohorts were stored at -20°C and processed according to previously published methods [9,19]. Briefly, preparation was conducted using an aqueous methanol extraction process to remove the protein fraction while allowing maximum recovery of small molecules. The extract was divided into four fractions: one for analysis by UPLC/MS/MS (positive mode), one for UPLC/MS/MS (negative mode), one for GC/MS, and one for backup. After removal of the organic solvent by a TurboVap® (Zymark), samples were frozen and dried under vacuum and then prepared for the appropriate instrument (UPLC/MS/MS or GC/MS).

Ultrahigh performance liquid chromatography/Mass Spectroscopy The LC/MS portion of the platform was based on a Waters ACQUITY ultra-performance liquid chromatography (UPLC) and a Thermo-Finnigan linear trap quadrupole (LTQ) mass spectrometer, consisting of

an electrospray ionization (ESI) source and linear ion-trap (LIT) mass analyzer. Dried sample extracts were reconstituted in acidic or basic LC-compatible solvents. One aliquot was analyzed using acidic positive ion optimised conditions and the other using basic negative ion optimised conditions in two independent injections using different columns. The MS analysis alternated between MS and data-dependent MS^2 scans using dynamic exclusion. Raw data files are archived and extracted as described below.

Gas chromatography/Mass Spectroscopy Samples destined for GC/MS analysis, samples were re-dried under vacuum desiccation and derivatized under dried nitrogen. Samples were analysed on a Thermo-Finnigan Trace DSQ fast-scanning single-quadrupole mass spectrometer using electron impact ionization. The instrument was daily tuned and calibrated for mass resolution and mass accuracy. The information output from the raw data files was automatically extracted as discussed below.

Data extraction and compound identification Raw data were extracted, peak-identified and processed using Metabolon's hardware and software [19]. Briefly, to improve QA/QC, extra samples were included for analyses every day. The samples were taken from a pool of well-studied and characterized human sera. Test samples were randomly distributed during the run, and QC samples were equally spaced between them. A selection QC compounds was carefully chosen not to interfere with the measurement of the test compounds, and added to every sample for chromatographic alignment. Instrument variability was determined by calculating the median relative standard deviation (RSD) for the standards that were added to each sample prior to injection into the mass spectrometers. Overall process variability was determined by calculating the median RSD for all endogenous metabolites (i.e., non-instrument standards) present in 100% of the pooled matrix samples. Compounds were identified by comparison to library entries of purified standards or recurrent unknown entities. Biochemical identification is based on three criteria: retention index within a narrow RI window of the proposed identification, nominal mass match to the library +/- 0.2 amu, and MS/MS forward and reverse

scores between the experimental data and authentic standards; these scores are based on a comparison of the ions present in the experimental spectrum to the ions present in the library spectrum. Identification of known chemical entities was based on comparison to library entries of purified standards. Missing values (if any) were imputed to the lowest measured value and metabolite data are scaled proportionately to a median of 1.

Statistical analysis Metabolite data are presented as median and [interquartile range, IQR]. Pooled data from the two patient groups were first analysed by random forest (RF) [21]. Using the R "VSURF" package, two separate univariate RF analyses were carried out, one with type of diabetes (T1D *vs* T2D) and the other with progression (progressors *vs* non-progressors) as the response variable; for either analysis, all measured metabolites were the explanatory variables. The Gini score and the mean decreases in classification accuracy were chosen to measure variable importance; a cutoff value of ≥1 was set for Gini score. Subsequently, using the R "random Forest SRC" package, a bivariate RF analysis was performed on both response variables, *i.e.,* diabetes type and progression, with all metabolites as explanatory variables.

As confirmatory analysis, data were also analyzed by bivariate partial least squares (PLS) in order to determine the robustness of RF in detecting the most important metabolites. The conventional VIP (variable importance in projection) score was used for ranking.

Groups (T1D and T2D cohorts) were compared using Wilcoxon test. For metabolites identified by the bivariate RF and/or PLS, group comparisons across diabetes type and progression were carried out by a 2-way ANOVA following log-transformation of the variables; correction for multiple testing was applied systematically. Multivariate logistic regression with a stepwise selection method was used to test the association of CKD progression – as the response variable – with top metabolites plus selected clinical characteristics; results are reported as odds ratio (OR) and 95% confidence intervals (95% CI). R and SPSS-IBM for Mac Os X software were used; the statistical significance threshold level was set at $p<0.05$.

Results

The clinical phenotype of the two patient cohorts was as expected: T1D patients were younger and leaner, with longer disease duration, lower BP and higher eGFR values at baseline (**Table S1**). Of note is that HbA_{1c} was significantly higher in progressors than non-progressors in both groups.

Changes in yearly eGFR were -0.7 [2.3] and -1.3 [3.3] $ml \cdot min^{-1} \cdot 1.73m^{-2} \cdot yr^{-1}$ in T1D and T2D, respectively (p=0.0043). In progressors, yearly eGFR loss was significantly larger in T2D than T1D, presumably on account of their higher age and BP and the presence of microalbuminuria in almost half of them. To account for differences in baseline eGFR between T1D and T2D, the yearly change in eGFR was also calculated as a percent change from baseline. As shown in Table 1, the differences between progressors and non-progressors across diabetes type were virtually identical for absolute and percent eGFR changes; 88.5% of subjects in the top quartile of absolute eGFR changes were also in the top quartile of percent eGFR changes (r^2 = 0.69, p<0.0001) (**Table 1**).

With diabetes type as the response variable, the univariate RF extracted 13 associated metabolites, all known molecules; of note, with the exception of 5-methylthioadenosine and ß-hydroxypyruvate, which were higher in T2D than T1D, all the other metabolites were higher in T1D than T2D (**Table S2**). With CKD progression as the response variable, the univariate RF detected 11 associated variables (8 known and 3 unknown), all of which – except for theophylline – were higher in progressors than non-progressors (**Table S3**). The metabolites identified by the two univariate RF's were different.

The bivariate RF analysis using both responses – diabetes type and progression – selected 35 metabolites, 32 known and 3 unknown, of which 20 (57%) were also detected by one or the other univariate RF (**Figure 1**). Signal levels and group comparisons of the known metabolites are given in **Table 2**.

The bivariate PLS, utilizing the same responses and a cutoff of 1.5 for VIP, screened 14 relevant molecules (13 known), of which 8 (57 %) were also identified by the bivariate RF; **Table 3** reports their concentrations across groups. Overall, 8 metabolites were common to both bivariate analyses (shaded rows in Tables 2&3).

When the bivariate random forest analysis was run on normoalbuminuric subjects only, there was a large overlap of 'hits' between the whole T2D cohort and the normoalbuminuric T2D subset (**Table S4**).

In a multivariate logistic regression model with CKD progression as the dependent variable and the 13 metabolites selected by the bivariate RF as the independent variables, age (OR=2.5; 95% CI:1.6-3.9 per 1 SD) and baseline eGFR (OR=3.0; 95% CI:1.9-4.7) – but not gender or HbA_{1c} – were also significant covariates in the combined cohorts along with plasma levels of glycerol, heptanoate, sphingomyelin, erythronate and pyroglutamylvaline (in that order). Running the same model in T2D data alone (T1D patients were all normoalbuminuric), female gender (OR=2.9; 95% CI:1.3-6.5) and the baseline albumin-to-creatinine ratio (OR=1.7; 95% CI:1.1-2.7 per 1 SD) – but not age or HbA_{1c} – were significant covariates. It is of interest that in these models including the clinical parameters type of diabetes was not a significant covariate.

Discussion

The primary aim of the present work was to identify differential metabolomic signatures of disease progression between patients with T1D or T2D by using the same analytical platform and outcome variable, *i.e.*, the decline in eGFR. However, the first finding emerging from the data is that the identification of associated metabolites is dependent on the statistical method of analysis: among the 4 sets of metabolites detected using different methods (2 univariate RF's, bivariate RF, bivariate PLS) there was a partial discrepancy both between the bivariate and the univariate RF's and between the bivariate RF and PLS. Furthermore, bivariate RF ranked 35 molecules while the PLS procedure ranked only 14. This is explained, at least in part, by the

sensitivity of the methods to select statistical predictors based on standard but user-defined cutoffs for the respective indices, *i.e.,* a Gini index ≥1 for RF and a VIP ≥1.5 for PLS. Different cutoffs would yield different numbers of predictors and overlap patterns. Both RF and PLS models scale well on large numbers of metabolite data compared to sample size (in the present work, 547 metabolites and 462 patients). RF has many interesting characteristics, for example, it is non-parametric, hard to over-train, robust to outliers and fast to train. PLS, being a dimension reduction method, is less liable than RF to collinearity [22].

Running data with both models, as was done here, increases confidence in the selected predictors. In our patient material, 12 out of the 13 biomarkers associated with T1D *vs* T2D in univariate RF were also detected by the bivariate RF, and 7 out of the 8 known metabolites associated with CKD progression in univariate RF were also detected by the bivariate RF. Even the overlap between bivariate RF and PLS was reasonable, as 8 out of 13 known metabolites were identified by both procedures.

Focusing on the plot of Gini indices from the bivariate RF (**Figure 1**), while some lipid molecules were selectively increased (2-hydroxyoctanoate, palmitic amide) or decreased (linoleamide) in T1D progressors, there appeared to be more metabolite differences by diabetes type than by CKD progression. This suggests that age, BMI and other differences (Table 1) between T1D and T2D at the time of sample collection have a greater impact on the metabolomic signature than the initial metabolic changes predicting loss of eGFR seven to three years later. This is indirectly confirmed by the logistic regression analyses testing the association of hit metabolites with CKD progression, where adding the main clinical variables (gender, age, BMI, HbA_{1c}, baseline eGFR and the albumin-to-creatinine ratio) was statistically equivalent to using type of diabetes as a lumped nominal variable. Nevertheless, bivariate RF also selected seven metabolites (erythyronate, gluconate, glycerol, heptanoate, hexanoylcarnitine, sphingomyelin, and xanthine) primarily associated with CKD progression but not with diabetes type (Table 2). Hexanoylcarnitine is a medium-chain acylcarnitine;

acylcarnitines are products of fatty acid oxidation that have been suggested to cause insulin resistance [23]. In a previous study of 123 mass spectrometry-derived serum metabolites measured in 409 individuals with T2D, several short acylcarnitines were elevated in participants with CKD [24]. Our findings support this observation and extend it to T1D. Furthermore, in our data the lower levels of sphingomyelin, which hydrolyzes to ceramide during the execution phase of apoptosis [25], may reflect ongoing apoptotic processes marked by elevated tissue ceramide levels [24]. Interestingly, among the other metabolites selected for CKD progression in the bivariate RF, glycerol is used to induce renal injury in rodent models of rhabdomyolysis [26]. Epidemiological studies found an association of triglycerides with diabetic nephropathy [27]; in addition, fenofibrate, alone or combined with statins, slowed down the decline in renal function [28], though these effects were not related to improvements in triglycerides or HDL-cholesterol levels. Cross-sectional analysis of the large RIACE cohort showed that hypertriglyceridemia is associated with renal, but not retinal, complications in subjects with T2D [29]. Raised plasma and urine xanthine levels can lead to a rare condition called xanthine nephropathy [30].

In the bivariate RF analysis, the signal intensity of three of the selected metabolites (histidylphenylalanine, leucylphenylalanine, tryptophylasparagine) showed a significant interaction between disease and CKD progression. Histidylphenylalanine, higher in T1D than T2D and higher in CKD progressors among T1D, is a competitive inhibitor of histidine decarboxylase, the enzyme responsible for histamine formation. Histamine has roles in allergy, inflammation and infection, and has also been linked with immune responses [31,32]. Recent work has shown that myeloid cells highly expressing histidine decarboxylase might promote foam cell formation and accelerate atherosclerosis [33]. Its lack prevents the onset of diabetes in the NOD mouse [34]; this is important in the light of the role of mast cells, which might infiltrate pancreatic islets in T1D [35]. Histidine decarboxylase has never before been connected with the known pathogenetic mechanisms of nephropathy, and therefore represents an interesting

target for further studies.

Another metabolite, leucylphenylalanine, showed the same pattern of changes as histidylphenylalanine. This is a neutral dipeptide, whose biological meaning is, so far as we know, obscure. Its distribution in biological fluid is large, and it has been reported in metabolomics in relation to the presence of obesity [36]. Intriguingly, these two metabolites are chemically similar (they both derive from phenylalanine). The kidney plays a major role in the uptake of phenylalanine and its hydroxylation and release as tyrosine. In CKD, the conversion of phenylalanine to tyrosine and tyrosine release from the kidney are both impaired; moreover, plasma tyrosine to phenylalanine ratio has been found to be reduced in advanced CKD [37]. In our data, the tyrosine-to-phenylalanine ratio also was lower in progressors than non-progressors (0.96 [0.28] *vs* 1.00 [0.27], $p<0.04$). As a mere speculation, the oxidation of these amino acids, forming nitrotyrosine and oxidation products of phenylalanine, may cause adverse metabolic or toxic effects in CKD patients.

The third discriminating metabolite is tryptophylasparagine, much higher in T1D, and again associated with CKD progression only in T1D. This is a dypeptide derived from protein breakdown, apparently never identified before in human tissues or fluids; its biological meaning is unknown.

Of interest is the finding of raised ß-hydroxypyruvate signal levels in T2D. This metabolite derives, at least in part, from serine metabolism, and has been found to be increased in proportion to hyperglycemia in T2D patients and to participate as a potential causative factor in the disease development by reducing the activity of the pancreatic enteric neuronal pathway [38]. In our T2D patients, ß-hydroxybutyrate was strongly correlated with fasting plasma glucose concentrations ($r=0.58$, $p<0.0001$).

Finally, in our series glycerol, urate, gluconate and heptanoate were consistently associated with the progression of renal damage. Urate is a well-established marker of diabetic and nondiabetic CKD [39,40]; however, its role as a determinant of renal impairment in T1D has

been recently questioned [41]. Gluconate is formed from gluconic acid and arises from glucose oxidation; it is not known to be involved in any pathological process.

Over the last five years, metabolomics as part of a systems biology approach [42,43] has made meaningful contributions to the knowledge in this field, with cross-sectional case-control studies and prospective cohort observations in T1D and T2D [12,14,44]. By this approach, acylcarnitines, acylglycines and metabolites related to tryptophan metabolism have emerged as discriminants in both types of disease [12,14,20]. Sharma *et al.* [17] reported differential expression of several urinary metabolites related to mitochondrial metabolism, suggesting a global impairment of mitochondrial activity in diabetes and CKD. To our knowledge, the present study is the first directly comparing prospective results obtained in Caucasian T1D and T2D individuals using an unbiased statistical definition of renal functional loss (upper quartile of eGFR changes over time). Additional strengths are the size of the cohorts and the prospective design. Furthermore, the T1D subjects were normoalbuminuric at baseline, and had well preserved renal function despite a long disease duration; thus, their metabolomic signature of eGFR loss likely reflected incipient nephropathy. However, patients at a more advanced stage of CKD may show a different cluster of associated metabolites as the pathology of renal damage evolves. Moreover, using urine metabolites or progression of proteinuria as the endpoint may also result in a different pattern of associations. In fact, the current analysis did not select the same metabolites as our previous analyses in the same cohorts of T2D [13] and T1D patients [19], because the outcome phenotype was progression to microalbuminuria in the previous reports and decline in renal function in the present report. Interestingly, among T1D subjects 30% of CKD progressors had not developed albuminuria at follow up despite similar eGFR decline (-3.5 [1.91] $ml \cdot min^{-1} \cdot 1.73m^{-2} \cdot yr^{-1}$) and follow-up time (7.6 [4.3] years) as the individuals who did progress to microalbuminuria (-3.9 [3.0] $ml \cdot min^{-1} \cdot 1.73m^{-2} \cdot yr^{-1}$ and 6.7 [3.0] years, respectively).

Our study has limitations. The two cohorts were not prospective in design or

contemporaneous. In addition, although the plasmas used for metabolomic analysis from both cohorts were never-thawed, and well frozen aliquots, confounding by calendar time, storage conditions, *etc*. is ineludible.

In conclusion, dual, unbiased analysis of screening metabolomics identified molecules (and associated pathways) specifically associated with progressive renal dysfunction common to T1D and T2D, thereby possibly representing 'pure' diabetic nephropathy. The findings can guide targeted metabolomics of those metabolites standing out as the most promising discriminants of functional decline in the kidneys of T1D and T2D patients.

Acknowledgements

We acknowledge FinnDiane participants and the doctors and nurses participating in the FinnDiane study, and thank the laboratory staff for their skillful work. We also wish to thank Giuseppe Pugliese (University of Roma La Sapienza) and Giuseppe Penno (University of Pisa) for the key help in forming the T2D cohort.

Conflict of Interest Statement

P-H. G. has received investigator-initiated research grants from Eli Lilly and Roche, is an advisory board member for AbbVie, Astellas, AstraZeneca, Boehringer Ingelheim, Cebix, Eli Lilly, Janssen, Medscape, Merck Sharp & Dohme, Mundipharma, Nestle, Novartis, Novo Nordisk and Sanofi; and has received lecture fees from AstraZeneca, Boehringer Ingelheim, Eli Lilly, Elo Water, Genzyme, Merck Sharp & Dohme, Medscape, Novartis, Novo Nordisk, PeerVoice and Sanofi. E.F. advises Boehringer Ingelheim, Nestlé and Sanofi, lectures for Boehringer Ingelheim, AstraZeneca, Eli Lilly, and MSD, and has research grant support by Boehringer Ingelheim, AstraZeneca and Janssen.

Authors' Contribution

M.L.M. analysed all data. A.S. contributed to design the study and to write the manuscript. J.H., N.S., C.F. and P-H. G. reviewed the manuscript and contributed to the discussion. E.F. designed the study, reviewed data analysis and wrote the manuscript. E.F. is the guarantor taking responsibility for the contents of the article.

Funding

This work was supported by Folkhälsan Research Foundation, the Wilhelm and Else Stockmann Foundation, the Liv och Hälsa Foundation, Helsinki University Central Hospital Research Funds (EVO), the Novo Nordisk Foundation (NNF OC0013659), Academy of Finland (299200, and 316664), and European Foundation for the Study of Diabetes (EFSD) Young Investigator Research Award funds. The sponsors were not involved in the design of the study; the collection, analysis, and interpretation of data; writing the report; or the decision to submit the report for publication.

References

1. Afkarian M, Sachs MC, Kestenbaum B, et al. Kidney disease and increased mortality risk in type 2 diabetes. *J Am Soc Nephrol* 2013; 24: 302-308

2. Gansevoort RT, Correa-Rotter R, Hemmelgarn BR, et al. Chronic kidney disease and cardiovascular risk: epidemiology, mechanisms, and prevention. *Lancet* 2013; 382: 339-352

3. Fox CS, Matsushita K, Woodward M et al.; Chronic Kidney Disease Prognosis Consortium. Associations of kidney disease measures with mortality and end-stage renal disease in individuals with and without diabetes: a meta-analysis. *Lancet* 2012; 380: 1662-1673

4. Rebholz CM, Inker LA, Chen Y, et al.; Chronic Kidney Disease Biomarkers Consortium. Risk of ESRD and mortality associated with change in filtration markers. *Am J Kidney Dis* 2017; 70: 551-560

5. Saulnier PJ, Gand E, Velho G, et al; SURDIAGENE Study Group. Association of circulating biomarkers (Adrenomedullin, TNFR1, and NT-proBNP) with renal function decline in patients with Type 2 Diabetes: A French Prospective Cohort. *Diabetes Care* 2017; 40: 367-374

6. Stenvinkel P. Chronic kidney disease: a public health priority and harbinger of premature cardiovascular disease. *J Intern Med* 2010; 268:456-467

7. Norris KC, Smoyer KE, Rolland C, Van der Vaart J, Grubb EB. Albuminuria, serum creatinine, and estimated glomerular filtration rate as predictors of cardio-renal outcomes in patients with type 2 diabetes mellitus and kidney disease: a systematic literature review. *BMC Nephrol* 2018; 19: 36

8. Tangri N, Grams ME, Levey AS, et al.; CKD Prognosis Consortium. Multinational assessment of accuracy of equations for predicting risk of kidney failure: a meta-analysis. *JAMA* 2016; 315: 164-174

9. Thorn LM, Forsblom C, Fagerudd J, et al.; FinnDiane Study Group. Metabolic syndrome in type 1 diabetes: association with diabetic nephropathy and glycemic control (the FinnDiane study). *Diabetes Care* 2005; 28: 2019-2024

10. Nassirpour R, Raj D, Townsend R, Argyropoulos C. MicroRNA biomarkers in clinical renal disease: from diabetic nephropathy renal transplantation and beyond. *Food Chem Toxicol* 2016; 98: 73-88

11. Lindhardt M, Persson F, Currie G, et al. Proteomic prediction and Renin angiotensin aldosterone system Inhibition prevention Of early diabetic nephRopathy in TYpe 2 diabetic

patients with normoalbuminuria (PRIORITY): essential study design and rationale of a randomised clinical multicentre trial. *BMJ Open* 2016; 6: e010310

12. Niewczas MA, Sirich TL, Mathew AV, et al. Uremic solutes and risk of end-stage renal disease in type 2 diabetes: metabolomic study. *Kidney Int* 2014; 85: 1214-1224

13. Solini A, Manca ML, Penno G, Pugliese G, Cobb JE, Ferrannini E. Prediction of declining renal function and albuminuria in patients with type 2 diabetes by metabolomics. *J Clin Endocrinol Metab* 2016; 101: 696-704

14. Niewczas MA, Mathew AV, Croall S, et al. Circulating modified metabolites and a risk of ESRD in patients with type 1 diabetes and chronic kidney disease. *Diabetes Care* 2017; 40: 383-390

15. Chen CJ, Liao WL, Chang CT, Lin YN, Tsai FJ. Identification of urinary metabolite biomarkers of type 2 diabetes nephropathy using an untargeted metabolomic approach. *J Proteome Res* 2018; 17: 3997-4007

16. Titan SM, Venturini G, Padilha K, et al. Metabolites related to eGFR: Evaluation of candidate molecules for GFR estimation using untargeted metabolomics. *Clin Chim Acta* 2019; 489: 242-248

17. Sharma K, Karl B, Mathew AV, et al. Metabolomics reveals signature of mitochondrial dysfunction in diabetic kidney disease. *J Am Soc Nephrol* 2013; 24: 1901-1912

18. Teng J, Dwyer KM, Hill P, et al. Spectrum of renal disease in diabetes. *Nephrology (Carlton)* 2014; 19: 528-536

19. Haukka JK, Sandholm N, Forsblom C, Cobb JE, Groop PH, Ferrannini E. Metabolomic profile predicts development of microalbuminuria in individuals with type 1 diabetes. *Sci Rep* 2018; 8: 13853

20. van der Kloet FM, Tempels FW, Ismail N, et al. Discovery of early-stage biomarkers for diabetic kidney disease using ms-based metabolomics (FinnDiane study). *Metabolomics* 2012; 8: 109-119

21. Breiman L. Random Forests. *Machine Learning* 2001*;* 45: 5-32

22. Palermo G, Piraino P, Zucht H-D. Performance of PLS regression coefficients in selecting variables for each response of a multivariate PLS for omics-type data. *Adv Appl Bioinform Chem* 2009; 2: 57-70

23. Schooneman MG, Vaz FM, Houten SM, Soeters MR. Acylcarnitines: reflecting or inflicting insulin resistance? *Diabetes* 2013; 62: 1-8

24. Liu JJ, Ghosh S, Kovalik JP, et al. Profiling of plasma metabolites suggests altered mitochondrial fuel usage and remodeling of sphingolipid metabolism in individuals with type 2 diabetes and kidney disease. *Kidney Int Rep* 2016; 2: 470-480

25. Green DR. Apoptosis and sphingomyelin hydrolysis. The flip side. *J Cell Biol* 2000; 150: F5-7

26. Kim JH, Lee SS, Jung MH, et al. N-acetylcysteine attenuates glycerol-induced acute kidney injury by regulating MAPKs and Bcl-2 family proteins. *Nephrol Dial Transplant* 2010; 25: 1435-1443

27. Retnakaran R, Cull CA, Thorne KI, Adler AI, Holman RR; UKPDS Study Group. Risk factors for renal dysfunction in type 2 diabetes: U.K. Prospective Diabetes Study 74. *Diabetes* 2006; 55: 1832-1839

28. Davis TM, Ting R, Best JD, et al. Effects of fenofibrate on renal function in patients with type 2 diabetes mellitus: the Fenofibrate Intervention and Event Lowering in Diabetes (FIELD) study. *Diabetologia* 2011; 54: 280-290

29. Penno G, Solini A, Zoppini G, et al; Renal Insufficiency And Cardiovascular Events (RIACE) Study Group. Hypertriglyceridemia is independently associated with renal, but not retinal complications in subjects with type 2 diabetes: a cross-sectional analysis of the Renal Insufficiency And Cardiovascular Events (RIACE) Italian Multicenter Study. *Plos One* 2015; 10: e0125512

30. LaRosa C, McMullen L, Bakdash S, et al. Acute renal failure from xanthine nephropathy during management of acute leukemia. *Pediatr Nephrol* 2007; 22: 132-135

31. Conti P, Caraffa A, Ronconi G, et al. Impact of mast cells in mucosal immunity of intestinal inflammation: Inhibitory effect of IL-37. *Eur J Pharmacol* 2018; 818: 294-299

32. Barcik W, Wawrzyniak M, Akdis CA, O'Mahony L. Immune regulation by histamine and histamine-secreting bacteria. *Curr Opin Immunol* 2017; 48: 108-113

33. Xu L, Cheng D, Huang Z, et al. Histamine promotes the differentiation of macrophages from CD11b(+) myeloid cells and formation of foam cells through a Stat6-dependent pathway. *Atherosclerosis* 2017; 263: 42-52

34. Alkan M, Machavoine F, Rignault R, Dam J, Dy M, Thieblemont N. Histidine decarboxylase deficiency prevents autoimmune diabetes in NOD mice. *J Diabetes Res* 2015; 965056: 2015

35. Martino L, Masini M, Bugliani M, et al. Mast cells infiltrate pancreatic islets in human type 1 diabetes. *Diabetologia* 2015; 58: 2554-2562

36. Wang C, Feng R, Sun D, Li Y, Bi X, Sun C. Metabolic profiling of urine in young obese men using ultra performance liquid chromatography and Q-TOF mass spectrometry (UPLC/Q-TOF MS). *J Chromatogr B Analyt Technol Biomed Life Sci* 2011; 879: 2871-2876

37. Duranton F, Lundin U, Gayrard N, et al. Plasma and urinary amino acid metabolomic profiling in patients with different levels of kidney function. *Clin J Am Soc Nephrol* 2014; 9: 37-45

38. Zhang S, Wang S, Puhl MD, et al. Global biochemical profiling identifies β-hydroxypyruvate as a potential mediator of type 2 diabetes in mice and humans. *Diabetes* 2015; 64: 1383-1394

39. Rincon-Choles H, Jolly SE, Arrigain S, et al. Impact of uric acid levels on kidney disease progression. *Am J Nephrol* 2017; 46: 315-322

40. Gu L, Huang L, Wu H, Lou Q, Bian R. Serum uric acid to creatinine ratio: a predictor of incident chronic kidney disease in type 2 diabetes mellitus patients with preserved kidney function. *Diab Vasc Dis Res* 2017; 14: 221-225

41. Ahola AJ, Sandholm N, Forsblom C, Harjutsalo V, Dahlström E, Groop PH; FinnDiane Study Group. The serum uric acid concentration is not causally linked to diabetic nephropathy in type 1 diabetes. *Kidney Int* 2017; 91: 1178-1185

42. Nicholson JK. Global systems biology, personalized medicine and molecular epidemiology. *Mol Syst Biol* 2006; 2: 52

43. Hocher B, Adamski J. Metabolomics for clinical use and research in chronic kidney disease. *Nat Rev Nephrol* 2017: 13: 269-284

44. Pena MJ, Lambers Heerspink HJ, Hellemons ME, et al. Urine and plasma metabolites predict the development of diabetic nephropathy in individuals with type 2 diabetes mellitus. *Diabet Med* 2014; 31: 1138-1147

Figure Legend

Figure 1 – Metabolites detected by bivariate Random Forest; red symbols identify metabolites also detected by PLS.

Table 1 – Clinical characteristics.*

	T1D		T2D		p^1	p^2	p^3
	Progressors (n=47)	**Non-Progr. (n=146)**	**Progressors (n=74)**	**Non-Progr. (n=195)**			
Sex (% Male)	60	40	54	66	-	ns	-
Age (years)	33 ± 12	34 ± 10	63 ± 7	62 ± 8	<0.0001	ns	ns
BMI (kg/m^2)	25.1 ± 3.7	25.1 ± 3.2	29.9 ± 5.6	28.7 ± 4.8	<0.0001	ns	ns
Diabetes duration (years)	20 ± 11	21 ± 9	15 ± 10	12 ± 9	<0.0001	ns	ns
HbA_{1c} (%)	9.3 ± 1.7	8.7 ± 1.5	7.4 ± 0.9	7.1 ± 0.8	<0.0001	0.0013	ns
Systolic BP (mmHg)	128 ± 13	129 ± 14	145 ± 21	140 ± 17	<0.0001	ns	ns
Diastolic BP (mmHg)	80 ± 9	78 ± 10	78 ± 9	77 ± 9	ns	ns	ns
Baseline eGFR ($ml \cdot min^{-1} \cdot 1.73m^{-2}$)	114 ± 19	108 ± 16	89 ± 17	84 ± 20	<0.0001	0.0109	ns
Change in eGFR ($ml \cdot min^{-1} \cdot 1.73m^{-2} \cdot yr^{-1}$)	-3.7 [3.1]	-0.3 [1.7]	-5.3 [3.0]	-0.7 [2.3]	0.0036	<0.0001	0.018
Change in eGFR ($\% \cdot yr^{-1}$)	-3.2 [2.9]	-0.3 [1.6]	-6.1 [3.5]	-0.9 [2.8]	<0.0001	<0.0001	0.0003
Normoalbuminuria at baseline (%)	100	100	53	69	<0.0001	-	0.028

* entries are mean ± SD or median [interquartile range]; p^1 = T1D *vs* T2D; p^2 = progressors *vs* non-progr.; p^3 = interaction of diabetes and progression.

Table 2 – Metabolites selected by bivariate random forest (RF) analysis.

Metabolite	T1D		T2D		p^1	p^2	p^3
	Progressor	***Non-Progr.***	***Progressor***	***Non-Progr.***			
1,5-anhydroglucitol	0.309 [0.775]	0.614 [0.439]	0.279 [0.538]	0.510 [0.721]	0.050	ns	ns
2aminoheptanoic acid	1.071 [0.537]	0.975 [0.500]	1.093 [0.663]	0.981 [0.524]	ns	ns	ns
2-hydroxyoctanoate	1.094 [0.502]	0.972 [0.482]	0.501 [0.537]	0.416 [0.584]	<0.001	ns	ns
2-hydroxystearate	1.068 [0.442]	0.985 [0.456]	1.035 [0.633]	0.998 [0.312]	ns	ns	ns
3-ethylphenylsulfate	1.176 [2.277]	0.941 [2.161]	0.259 [0.010]	0.257 [0.011]	<0.001	ns	ns
4-acetylphenylsulfate	0.996 [0.930]	1.000 [0.701]	0.246 [0.558]	0.247 [0.647]	<0.001	ns	ns
4-guanidinobutanoate	0.975 [0.279]	1.000 [0.346]	0.207 [0.010]	0.209 [0.012]	<0.001	ns	ns
5-methylthioadenosine	0.546 [1.207]	0.238 [0.967]	0.852 [1.170]	0.733 [0.720]	<0.001	0.036	ns
Acetylcarnitine	0.882 [3.935]	0.861 [1.990]	1.001 [0.366]	0.994 [0.324]	<0.001	ns	ns
Adipate	0.669 [1.083]	0.796 [1.087]	0.178 [0.361]	0.179 [0.020]	<0.001	ns	ns
α–CEHCglucuronide	1.031 [0.889]	0.948 [0.845]	0.130 [0.020]	0.132 [0.013]	<0.001	ns	ns
ß-hydroxypyruvate	0.829 [1.085]	0.648 [0.860]	1.003 [0.511]	0.915 [0.508]	<0.001	ns	ns
Erythronate	1.049 [0.273]	0.968 [0.232]	1.017 [0.314]	0.995 [0.329]	ns	0.023	ns
Gluconate	1.066 [0.684]	0.997 [0.533]	1.118 [0.415]	0.925 [0.491]	ns	0.003	ns
Glutamine	0.900 [0.973]	1.057 [1.521]	0.974 [0.255]	1.008 [0.244]	0.001	ns	ns
Glycerol	1.174 [0.674]	0.978 [0.445]	1.045 [0.428]	0.975 [0.328]	ns	0.001	ns
Heptanoate	1.062 [0.350]	0.992 [0.268]	1.041 [0.625]	0.922 [0.370]	ns	0.001	ns
Hexanoylcarnitine	1.193 [2.085]	0.993 [0.899]	0.998 [0.619]	0.996 [0.458]	ns	0.050	ns
Histidylalanine	1.205 [1.264]	0.992 [0.990]	0.248 [0.015]	0.244 [0.019]	<0.001	0.038	ns
Histidylphenylalanine	1.250 [1.982]	0.927 [1.458]	0.275 [0.070]	0.273 [0.015]	<0.001	0.025	0.042
Hyocholate	0.801 [0.829]	1.004 [1.171]	0.133 [0.761]	0.132 [0.779]	<0.001	ns	ns
Inosine	1.044 [1.413]	0.974 [1.524]	0.157 [0.807]	0.155 [0.521]	<0.001	ns	ns
Leucylphenylalanine	1.367 [1.280]	0.912 [1.445]	0.293 [0.020]	0.291 [0.018]	<0.001	0.004	0.007
Linoleamide	0.770 [9.746]	0.954 [4.264]	0.023 [1.103]	0.024 [0.472]	<0.001	ns	ns
Palmitic amide	1.000 [4.306]	0.967 [2.289]	0.089 [1.363]	0.046 [0.740]	<0.001	ns	ns
Pyroglutamylvaline	1.115 [1.019]	0.976 [0.789]	0.689 [0.537]	0.589 [0.510]	0.010	0.020	ns
Pyruvate	0.642 [1.057]	0.592 [0.827]	1.001 [0.445]	1.013 [0.215]	<0.001	ns	ns
Sphingomyelin	1.059 [0.490]	0.982 [0.338]	1.069 [0.436]	0.977 [0.385]	ns	0.001	ns
Theophylline	0.979 [0.802]	0.998 [0.933]	0.758 [0.978]	0.994 [1.028]	ns	ns	ns
Tryptophylasparagine	1.241 [1.915]	0.921 [1.794]	0.298 [0.016]	0.297 [0.018]	<0.001	0.005	0.003
Xanthine	1.036 [0,756]	1.001 [0.542]	1.091 [0.900]	0.951 [0.453]	ns	0.050	ns
Xanthosine	1.008 [0.640]	1.000 [0.415]	0.359 [0.040]	0.357 [0.030]	<0.001	ns	ns

* p^1 for T1D *vs* T2D; p^2 for progressors *vs* non-progressors; p^3 for the interaction of diabetes and progression.

Table 3 – Metabolites selected by bivariate PLS.*

	T1D		T2D		p^1	p^2	p^3
	Progressors	***Non-progr.***	***Progressors***	***Non-progr.***			
2-hydroxyoctanoate	1.094 (0.502)	0.972 (0.482)	0.501 (0.537)	0.416 (0.584)	<0.001	ns	ns
3-methylcatechol sulfate	0.762 (1.919)	0.912 (1.672)	0.375 (0.213)	0.376 (0.214)	ns	ns	ns
4-acetylphenylsulfate	0.996 (0.930)	1.000 (0.701)	0.246 (0.558)	0.247 (0.647)	<0.001	ns	ns
Adipate	0.669 (1.083)	0.796 (1.087)	0.178 (0.361)	0.179 (0.020)	<0.001	ns	ns
ß-hydroxypyruvate	0.829 (1.085)	0.648 (0.860)	1.003 (0.511)	0.915 (0.508)	<0.001	ns	ns
Hyocholate	0.801 (0.829)	1.004 (1.171)	0.133 (0.761)	0.132 (0.779)	<0.001	ns	ns
Palmitic amide	1.000 (4.306)	0.967 (2.289)	0.089 (1.363)	0.046 (0.740)	<0.001	ns	ns
Phenylalanylarginine	0.751 (2.223)	1.000 (3.152)	0.873 (0.655)	0.915 (0.497)	ns	ns	ns
Pyroglutamylvaline	1.115 (1.019)	0.976 (0.789)	0.689 (0.537)	0.589 (0.510)	0.010	0.020	ns
Pyruvate	0.642 (1.057)	0.592 (0.827)	1.001 (0.445)	1.013 (0.215)	<0.001	ns	ns
Sarcosine	0.532 (0.709)	0.576 (0.719)	0.856 (0.547)	0.983 (0.650)	<0.001	ns	ns
Serotonin (5HT)	1.068 (1.131)	1.000 (1.081)	0.355 (0.497)	0.357 (0.590)	<0.001	ns	ns
Stearamide	1.024 (1.238)	0.979 (1.035)	0.939 (1.033)	0.852 (0.739)	0.006	ns	ns

* p^1 for T1D *vs* T2D*;* p^2 for progressors *vs* non-progressors; p^3 for the interaction of diabetes and progression