THE UNIVERSITY *of* EDINBURGH

Risk prediction modelling for renal decline in patients with Type 2 Diabetes

Justina Krasauskaite

Doctor of Philosophy

Precision Medicine Doctoral Training Programme

University of Edinburgh

March, 2024

Abstract

Background

The continuous growth in prevalence of type 2 diabetes mellitus (T2DM), currently reaching epidemic proportions, along with the demographic shift towards an ageing population, means that the incidence of diabetes-related co-morbidities is also rising rapidly. One of the most common of such health conditions is chronic kidney disease (CKD), affecting approximately a third of people with T2DM. Both CKD and T2DM are long-term conditions, which are heterogeneous and complex. Affected individuals face a high risk of progression and development of further complications, which can negatively impact on quality of life as well as increased healthcare costs. One issue is that the early stages of CKD are symptomless, thus early kidney function decline that eventually progresses to CKD is often undetected.

Aims

To explore metabolomic profiling as a novel means to identify prognostic markers associated with the risk of declining renal function in people with T2DM and to undertake prediction modelling incorporating such markers.

Methods

(1) Scoping review: I systematically searched the primary literature (up to February, 2024) for observational studies in humans that investigated metabolomic profiles in blood and urine using high-throughput analysis to identify biomarkers associated with renal function decline.

(2) Secondary data analysis: Using data from the Edinburgh Type 2 Diabetes Study (ET2DS), a prospective cohort study of 1,066 men and women aged between 60 and 75 years in 2006/7, I undertook both cross-sectional and prospective analyses.

Main outcome measures: i) Cross-sectional analysis: CKD diagnosis based on serum creatinine-based estimated glomerular filtration rate (eGFR) < 60 ml min^{-1} (1.73 $m)^{-2}$ on at least two occasions, three months apart, during the two-year period up to and including baseline. ii) Prospective analysis: CKD onset defined as eGFR <60 ml min^{-1} (1.73 $m)^{-2}$ on at least two occasions, three months apart during follow-up and representing at least a 25% change from baseline eGFR. Additional outcomes for prospective analyses were the annual rate of change in eGFR as percentage and rapid decliners, defined as participants with annual change in eGFR of at least -5%.

Metabolomics panel: The metabolomics panel, applied at baseline to ET2DS participants, was developed by Nightingale Health Ltd. The metabolomic markers (n=149 concentrations of individual metabolites and n=79 derived ratios) were quantified from blood serum using high-throughput proton nuclear magnetic resonance spectroscopy platform. Metabolite levels were log-transformed and standardised (mean= 0, standard deviation = 1) prior analyses.

Prediction study: For prediction of 5-year CKD, 100 repeats of Cox least absolute shrinkage and selection operator (LASSO) were used to select metabolites as components for a multi-marker score. The predictive performance of the metabolites-based score (MetS) was assessed in relation to a reference score, a published risk prediction model for incident CKD consisting of clinical risk factors.

Results

The scoping review of published literature revealed an array of metabolites which have been found previously to be associated with renal function decline. Most studies used univariable statistical methods with adjustment for clinical variables and multiple testing correction to evaluate associations.

At ET2DS baseline, mean eGFR was 77.5 (ml min^{-1} (1.73 m)$^{-2}$, N=1058) and 216 individuals had evidence of CKD. Of those without CKD, 155 developed CKD over a median 6.7-year follow-up [inter-quartile range: 6.4, 7.1]. In cross-sectional analyses using baseline data, 88 metabolites were significantly associated with eGFR [βs (95% CI) ranged from -3.95 (-4.92, -2.92) to -1.19 (-2.25, -0.13) for negative associations and 3.53 (2.44, 4.62) to 1.17 (0.12, 2.23) for positive associations; all P_{FDR}< 0.05]. Very low-density lipoproteins (VLDL), triglycerides, amino acids, glycoprotein acetyls, and fatty acids showed inverse associations with eGFR, while cholesterol and phospholipids in high density lipoproteins (HDL) exhibited positive associations. Following prospective analysis, it was found that isoleucine, apolipoprotein A1 and total cholines were not only associated with baseline kidney measures, but also showed stable, nominally significant association with incident CKD and rapid decline.

In risk prediction analysis, 18 metabolites relating to lipoproteins, glycolysis, ketone bodies, inflammation and amino acids were selected as components for the multi-marker score, MetS, which showed positive association with incident CKD over 5 years after adjustment for traditional CKD risk factors [hazard ratio 2.21, [1.71, 2.84],

P=8.62*10^{-10}]. The independent validation of the published risk prediction model that was later used as the reference model, revealed c-statistic was 0.80 (95%CI 0.76, 0.84) which increased modestly to 0.84 (95%CI 0.80, 0.87) for the reference model combined with the MetS. Compared with the reference model, the net reclassification index and integrated discrimination index were 0.61 (95%CI 0.30, 0.81) and 0.06 (95%CI 0.02, 0.11, p=0.01) respectively, indicating modest improvement in reclassification of events and non-events.

Conclusion

This study revealed widespread changes within the metabolomic profile of CKD in people with T2DM, particularly in lipoproteins and their lipid compounds. I identified a smaller number of individual metabolites which specifically associated with incident CKD and decline in kidney function. Further investigation in additional studies will be needed to confirm the longitudinal findings and explore if the metabolic signals at baseline may be informative of mechanisms underlying the development of CKD. Furthermore, I found a sparse set of metabolites that could potentially enhance predictive accuracy for CKD risk in individuals with T2DM beyond the conventional risk factors. Additional research, including external validation, will be required before making definitive conclusions which may ultimately lead to clinically-relevant recommendations for the prediction and management of CKD in T2DM.

Lay summary

In recent years, the number of people with type 2 diabetes has been steadily increasing, and this condition is becoming more widespread, especially as our population ages. Unfortunately, type 2 diabetes often comes with other health problems, and one of the most common issues is chronic kidney disease (CKD), which affects about a third of individuals with type 2 diabetes. Both type 2 diabetes and kidney disease are long-term and complex conditions that can worsen over time, affecting quality of life and increasing healthcare costs. A significant challenge is that kidney problems in diabetes can develop silently without any noticeable symptoms in the early stages. This means that the decline in kidney function can go unnoticed until it becomes a more serious issue.

The goal of this research is to use a new approach called metabolomic profiling to find metabolites (markers found in blood) that can predict the risk of declining kidney

function in people with type 2 diabetes. To do this, I looked at existing scientific literature and also analysed data from a group of people with type 2 diabetes in the Edinburgh Type 2 Diabetes Study (ET2DS). I wanted to see if adding specific metabolites to traditional risk prediction models could improve their ability to predict kidney problems in people with type 2 diabetes.

Here are the key findings:

- I reviewed previous studies and found that certain metabolites (small molecules in the body) were linked to kidney function decline. These metabolites included amino acids and lipids.
- In the ET2DS cohort, I examined the metabolomic profiles of over a thousand participants. I discovered that several metabolites were associated with the participants' baseline kidney function. Some metabolites, were linked to lower kidney function, while others, were associated with better kidney function.
- I then created a multi-marker score using 18 metabolites related to various metabolic processes. This score was found to be positively associated with the development of CKD over five years, even after accounting for traditional CKD risk factors. This means that these metabolites could help predict the risk of CKD in people with type 2 diabetes.
- However, it is important to emphasise that more research is needed to validate these findings in other studies before making any definitive recommendations.

In summary, this study discovered significant changes in the metabolomic profile of CKD in people with type 2 diabetes, particularly in lipoproteins and their related lipid compounds. I identified specific metabolites that could be used to predict the risk of kidney problems, but more research is needed to be certain. I hope that in the future, these findings can help improve the early detection of kidney problems in people with type 2 diabetes.

Contents

List of tables

List of figures

List of abbreviations

AUC	Area under the curve
AASK	African American Study of Kidney Disease and Hypertension
ARIC	Atherosclerosis risk in communities
BMI	Body mass index
CKD	Chronic kidney disease
CVD	Cardiovascular disease
CI	Confidence interval
CRIC	Chronic Renal Insufficiency Cohort
DBP	Diastolic blood pressure
DKD	Diabetic kidney disease
ESRD	End-stage renal disease
ET2DS	Edinburgh Type 2 Diabetes Study
eGFR	Estimated glomerular filtration rate
FDR	False discovery rate
FinnDIane	Finnish Diabetic Nephropathy study

FWER	Family wise error rate
GCKD	German Chronic Kidney Disease
GFR	Glomerular filtration rate
HbA1c	Glycated haemoglobin
HDL	High density lipoprotein
HTN	Hypertension
HPLC	High-performance liquid chromatography
HR	Hazzard ratio
IDI	Integrated discrimination improvement
IDL	Intermediate density lipoprotein
IDMS	Isotope dilution mass spectrometry
IQR	Interquartile range
KDIGO	Kidney Disease: Improving Global Outcomes
KFRE	Kidney Failure Risk Equation
KIM-1	Kidney injury molecule 1
KORA	Cooperative Health Research in the Region of Augsburg
LASSO	Least absolute shrinkage and selection operator
LDL	Low density lipoprotein
LDR	Lothian Diabetes Register
MetS	Metabolite-based risk score
NHANES III	The Third National Health and Nutrition Examination Survey
NMR	Nuclear magnetic resonance
NRI	Net reclassification improvement
PREVEND	Prevention of REnal and Vascular End-stage Disease
SD	Standard deviation
SBP	Systolic blood pressure
SDRNT1BIO	Scottish Diabetes Research Network Type 1 Bioresource
T1DM	Type 1 diabetes mellitus
T2DM	Type 2 diabetes mellitus
TNFR1	tumor necrosis factor receptor 1
TNFR2	tumor necrosis factor receptor 2
uACR	Urinary albumin to creatinine ratio
VLDL	Very low-density lipoprotein

Acknowledgements

While this thesis represents the culmination of my own efforts, its completion would not have been possible without the invaluable contributions of numerous individuals whom I wish to acknowledge with my sincere gratitude. Foremost, I extend my appreciation to my esteemed supervisors, Prof Jackie Price, Dr Bryan Conway, and Prof Christopher Weir. Their unwavering support, constructive feedback, and insightful guidance have been indispensable throughout my doctoral journey. Their dedication to nurturing my growth as a researcher has been beyond measure, and I am grateful for their mentorship. I am also deeply indebted to the research team and participants of the Edinburgh Type 2 Diabetes Study. Without their collaboration and commitment, this work would not have been conceivable.

My heartfelt thanks extends to the remarkable peers with whom I have had the privilege to work alongside. In particular, I am grateful to Zhe and Wardah, whose friendship has been a source of inspiration and camaraderie. Our shared experiences and insightful conversations, often held in Room 666 at Teviot, have enriched both my academic and personal growth.

To those closest to my heart, I owe a debt of gratitude that words cannot adequately express. To George Wheatley, my steadfast partner, your unwavering love and encouragement over the past decade have provided me with the courage and determination to embark on this academic journey. To my mom, Daiva, and my sister, Evelina, your words of encouragement have been a source of strength. I am also deeply appreciative of the support extended to me by my parents-in-law, Mark and Elaine, who provided reassurance and stability. Last but certainly not least, I must express my thanks to my loyal feline companion, Sherman. Throughout the challenges of working from home during the pandemic and beyond, your companionship has brought me solace.

To each and every individual mentioned, and to those whose names I may have inadvertently omitted, please accept my apreciation for your support, encouragement, and belief in me throughout the realization of this thesis.

Declaration

I affirm that this thesis is entirely my original work. The content presented in this document has not been used in pursuit of any other academic degree or professional certification.

The data from the Edinburgh Type 2 Diabetes Study (baseline and follow-up), had already been collected prior to the commencement of my PhD studies. The data was subjected to cleaning and, in some instances, adjustments, which were made by previous PhD students. I took on the role of deriving specific variables used in this thesis. All research concepts and statistical analyses involving this data were carried out by me.

Signed: Justina Krasauskaite

Date: 22/03/2024

Publications and presentations

This section contains a list of publications and conference presentations relating to the work carried out during my PhD project. The abstracts of conference presentations can be found in Appendix A.

Publications:

Krasauskaite, J., Conway, B., Weir, C., Huang, Z. and Price, J., 2024. Exploration of metabolomic markers associated with declining kidney function in people with type 2 diabetes mellitus. Journal of the Endocrine Society, 8(1).

Huang, Z., Klaric, L., Krasauskaite, J., Khalid, W., Strachan, M.W., Wilson, J.F. and Price, J.F., 2023. Combining serum metabolomic profiles with traditional risk factors improves 10-year cardiovascular risk prediction in people with type 2 diabetes. European Journal of Preventive Cardiology, 30(12) .

Huang, Z., Klaric, L., Krasauskaite, J., McLachlan, S., Strachan, M.W., Wilson, J.F. and Price, J.F., 2022. Serum metabolomic profiles associated with subclinical and clinical cardiovascular phenotypes in people with type 2 diabetes. Cardiovascular diabetology, 21(1).

Conference presentations:

Associations between metabolomic profiles and kidney function decline in older people with Type 2 diabetes: the Edinburgh Type 2 Diabetes Study. J Krasauskaite, B R Conway, C J Weir, Z Huang, and J F Price. EDNSG - European Diabetic Nephropathy Study Group, May, 2020

Identification of markers for predicting the onset of chronic kidney disease in older people with type 2 diabetes by metabolomic profiling: Edinburgh Type 2 Diabetes Study. J Krasauskaite, B R Conway, C J Weir, Z Huang, and J F Price. *Diabetologia* 65 (SUPPL 1), S22-S23. European Association for the Study of Diabetes (EASD) Annual Meeting September, 2022 (Stockholm, Sweden)

Associations between metabolomic profiles and chronic kidney disease in older people with type 2 diabetes: The Edinburgh Type 2 Diabetes Study (ET2DS). J Krasauskaite, B R Conway, C J Weir, Z Huang, and J F Price. Diabetes UK Professional Conference 2022 (Virtual)

Chapter 1. Introduction: chronic kidney disease, type 2 diabetes mellitus and biomarkers

This chapter presents the rationale for this thesis, including the importance of studying chronic kidney disease (CKD) particularly in people diagnosed with type 2 diabetes mellitus (T2DM). The first part of the chapter provides an overview of CKD including a description of clinical measures relevant to kidney function evaluation and description of diagnosis and screening in nephrology to provide context for the adapted definitions in this thesis. Additionally, it describes the relationship between CKD and diabetes and delves into the consequences of CKD, including the risk of cardiovascular disease, and renal failure. The second part of the chapter focuses on the concept of risk prediction modelling and provides background on its applications in the field of nephrology, including an overview of available risk prediction models for new onset CKD. The final part of the chapter introduces the advancement of analytical tools that have enabled high-throughput analysis of metabolites in biological systems, leading to the establishment of metabolomics as a field within biomarker research. This section highlights the importance of metabolomics for improving prediction of kidney-related outcomes, as well as enhancing our understanding of the disease aetiology.

1.1 Chronic Kidney Disease

1.1.1 Global burden of chronic kidney disease

Chronic kidney disease (CKD) has emerged as a significant contributor to morbidity and mortality in the 21st century. By 2017, the global prevalence of CKD was estimated to be approximately 11-13% (Hill et al., 2016), affecting an estimated 843.6 million individuals (Jager et al., 2019). It is expected that the burden of CKD will continue to increase in the coming years, primarily due to the global ageing population and the rising incidence of CKD accelerators such as hypertension, obesity, and T2DM (Glassock et al., 2017, Darlington et al., 2021). For example, the projections in the United Kingdom (UK) anticipate a 1% increase in CKD prevalence by 2025, affecting 14% of the adult population (Garcia Sanchez et al., 2021). These forecasts also suggest a shift in the composition of CKD patients, with a projected 7% increase (relative to the total CKD population) in the more advanced stages compared to the overall CKD population in the UK.

Since CKD is often diagnosed based on estimated GFR (eGFR), factors affecting eGFR, particularly age, significantly impact prevalence estimates. As individuals age, eGFR declines due to age-related muscle mass loss, which can mask reductions in kidney function even when serum creatinine levels remain stable. A meta-analysis by Hill et al. showed a linear increase in CKD prevalence with age, ranging from 13.7% in those aged 30- 40 to 27.9% in individuals over 70. Similarly, U.S. data from 2015-2016 found CKD stages 1- 4 present in 5.6% of people aged 20–39 and 44% in those over 70. While the clinical significance of early-stage CKD (stage 3a) caused solely by aging is widely debated, its high prevalence among older individuals underscores the importance of monitoring the rate of decline. In other words, if an older person's kidney function declines very slowly, they are unlikely to reach severely low kidney function before natural death, unlike those whose kidney function deteriorates more rapidly over a short period. Furthermore, regions with ageing population and higher death rates attributed to CKD, adjusted for age and population, showed that CKD caused by diabetes, and to a lesser extent by hypertension and other causes, was more common than expected based on population growth and aging alone (Xie et al., 2018).

Global burden of disease study revealed that in 2017, approximately 1.2 million deaths were attributed to CKD-related conditions (Collaboration, 2020). Consequently, CKD has become the 12th leading cause of death out of 133 major health conditions, surpassing the number of deaths caused by HIV or tuberculosis (Collaborators, 2018). The most comprehensive analysis of trends in the burden of CKD stages G3–G5 across global regions and countries comes from the Global Burden of Disease study, which analysed data from 1990 to 2016 (Xie et al., 2018). The results showed that the leading cause of CKD-related mortality was CKD due to diabetes mellitus, followed by CKD caused by hypertension and glomerulonephritis. Overall, CKD is a highly prevalent condition that is increasingly recognized as a global public health concern (Collaboration, 2020).

1.1.2 Definition of chronic kidney disease

In general, CKD is a long-term condition characterised by a gradual decline in kidney function (Go et al., 2004), which reduces kidney ability to fulfil its vital function of filtering blood (kidney function illustrated in figure 1-1) . The onset of the disease involves a spectrum of functional and structural changes. More than a dozen pathophysiological mechanisms underlying CKD have been suggested, including

hyperactivity of the renin–angiotensin-aldosterone system, endothelial dysfunction, dyslipidaemia, modification of the purinergic system, activation of myofibroblasts and inflammation (Gajjala et al., 2015).

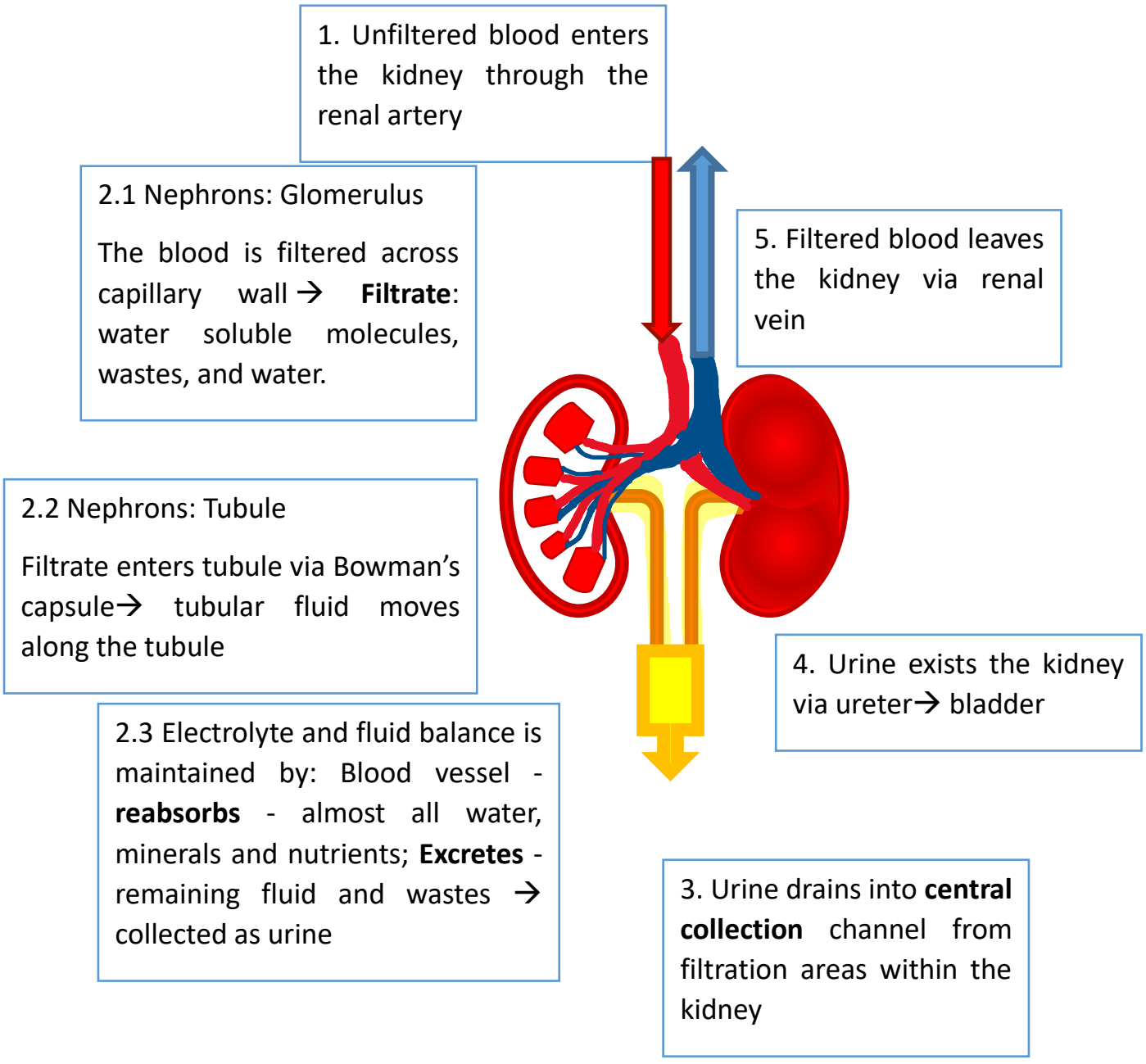


Figure 1-1 Illustration of kidney function to remove waste and extra water from the blood via nephrons*

*Nephrons are structural components of the kidney which consist of glomerulus, described as clusters of capillaries, which act as filtration barriers and tubules, which are ducts that run along the blood vessels and allows reabsorption and excretion of fluids.

The National Kidney Foundation's Kidney Disease Quality Outcome Initiative (NKF-KDOQI) formulated the initial conceptual framework for CKD in 2002 (Levey et al., 2002), encompassing key elements to describe the condition, including the definition, staging, expected outcomes, treatment modalities, and factors that increase the risk of CKD development, progression, and complications. Subsequently, these concepts were endorsed by KDIGO (Kidney Disease: Improving Global Outcomes) on a global basis in 2005 (Levey et al., 2005) and were refined to incorporate the latest evidence in 2012 (Levin et al., 2013, Levin and Stevens, 2014). The medical and public health communities recognised these guidelines and applied the approach for CKD prevention and management in the United States (Levey et al., 2009a), Europe (Verbeke et al., 2014) and the UK (Carville et al., 2014).

In brief, the conceptual model of CKD defines kidney failure or end-stage renal disease (ESRD) as the terminal stage of CKD, establishing connections with preceding stages (Levey et al., 2009b). According to this concept, ESRD is preceded by a decline in the glomerular filtration rate (GFR), which, in turn, is preceded by kidney damage. The progression of CKD typically spans a considerable duration, commencing with a prolonged latent period, during which kidney decline is symptomless and may remain undetected. The later stages that follow are accompanied with emergence of symptoms, driven by complications stemming from reduced kidney function (e.g., hypertension, uremic symptoms, anaemia) (Bello et al., 2017). Hence, it should be feasible to identify and diagnose CKD before it progresses to ESRD by screening for markers of kidney function and/or damage. This model underscores the gradual nature of CKD, suggesting that the pace of progression varies among individuals and may depend on the presence of comorbidities (such as prevalent cardiovascular disease (CVD) or diabetes mellitus).

Importantly, not all patients with CKD progress to the advanced stages, thus, a diagnosis of CKD does not guarantee the eventual development of kidney failure. Interventions administered in the earlier stages may serve to decelerate the kidney function decline or even avert the progression. This highlights the importance of predicting the risk of new-onset CKD and identify those at greatest risk of progressive decline in kidney function in order to enable timely implementation of lifestyle and therapeutic interventions that aim to prevent further decline and complications.

1.1.3 Staging of chronic kidney disease

The definition of CKD used in this thesis is based on the updated guidelines developed by KDIGO (Levin et al., 2013). In these guidelines, CKD is a term used to describe heterogeneous disorders that irreversibly affect kidney structure and/or function for more than three months. It is defined as decreased kidney function determined using measures or estimates of GFR and/or the presence of kidney damage markers such as albuminuria (excess protein in the urine) or the presence of structural changes, such as polycystic kidneys. The disease development and progression are usually classified into five stages based on the level of kidney function and/or kidney damage, as shown in table 1.1 and described below (Levey et al., 2002, Levey et al., 2003). The staging approach allows healthcare providers to assess the severity of the condition and determine appropriate treatment and management strategies.

Table 1-1 Kidney Disease Outcomes Quality Initiative (K/DOQI). Guideline used for the classification of standardised stages for CKD (Levey et al., 2003).

Stage	Description
G1	Normal GFR (> 90 mL/min/1.73 m²) with other evidence of chronic kidney damage
G2	Mild impairment: GFR 60 to 89 mL/min/1.73 m² with other evidence of chronic kidney damage
G3	Moderate impairment: GFR 30 to 59 mL/min/1.73 m²
G4	Severe impairment: GFR 15 to 29 mL/min/1.73 m²
G5	ESRD: GFR < 15 mL/min/1.73 m²
A1	Normal to mildly increased ACR: < 3 mg/mmol / <30 mg/g
A2	Moderately increased ACR: 3 - 30 mg/mmol /30 to 300 mg/g
A3	Severely increased ACR: > 30 mg/mmol/ exceeding 300 mg/g

Abbreviations: GFR, glomerular filtration rate; mg/g, milligrams per gram of creatinine; mg/mmol, milligrams per millimolar of creatinine.

The level of kidney damage indicated by albuminuria may play a role in CKD, which is defined as the presence of excess albumin (protein) in the urine (Levin et al., 2013). The stages based on albuminuria start with normal albuminuria, which is characterised by a normal level of albumin in the urine. Kidney function is usually normal at this stage. The stage A1 or mildly increased albuminuria, indicates a modest increase of albumin in the urine. It is an early sign of kidney damage and is often a precursor to more severe CKD, especially when accompanied with reduction in GFR. The penultimate stage (A2) indicates severely increased albuminuria, which signifies a significant increase in albumin in the urine. Kidney damage is more pronounced, but

generally patients would not notice symptoms due to this level of albuminuria alone. The final stage, A3 is characterised by extremely high levels of albumin in the urine, often exceeding 3,500 mg/g. It is a sign of severe injury to the glomerular filtration barrier and may be associated with conditions like nephrotic syndrome, which includes a triad of proteinuria (the loss of significant amounts of protein in the urine), low serum albumin and the presence of swelling. This classification is often used in conjunction with the GFR measurements to assess the severity of CKD and guide treatment.

The CKD stages based on GFR, which is the main indicator of the disease severity. During early stage G1, there may be evidence of kidney damage, such as the albuminuria or abnormalities in kidney imaging. However, the GFR remains relatively normal (just below the normal range of 100-120 mL/min/1.73m2), indicating that the kidneys are still functioning well. Stage G2 is associated with a slight reduction in GFR, but the kidneys are still functioning effectively, and many people in this stage would not experience noticeable symptoms. As kidney continues to decline and reaches a moderate reduction or stage G3 (split into two sub-categories - a and b), which is associated with an increased risk of complications, such as high blood pressure and anaemia. Severe reduction in GFR is classed as stage G4, which is associated with a significant decline in kidney function. Individuals in this stage often require medical intervention and may be referred to a nephrologist for specialised care. Complications such as development of CVD are more likely (Go et al., 2004). The final stage G5, indicates the most severe form of CKD, also known as ESRD or kidney failure. Due to severely compromised kidney function, the kidneys can no longer effectively filter waste and excess fluids from the body. Patients with CKD stage G5 often require dialysis or a kidney transplant to survive.

As discussed before, the kidney decline is irreversible, so prophylactic intervention using preventative strategies, before significant kidney decline and emergence of complication is advantageous. Thus, the stage G3a is the point of interest for the purpose of this thesis as it marks the onset of symptomatic disease and predicting this stage, would enable early interventions that prevent or reduce the rate of the progression of kidney decline, which ultimately results in disease progression. In addition, stage A1 also marks the onset of disease according to the KDIGO guidelines, so this definition will also be included when defining the CKD status in some of the analyses carried out for the purpose of this thesis.

1.1.4 Clinical and economic burden of chronic kidney disease

Once the GFR drops below 60 ml min^{-1} (1.73 $m)^{-2}$ (CKD stage G3a), the deterioration of the kidneys is considered advanced and often coincides with increased risk of metabolic complications as well as cardiovascular-related outcomes, including heart failure, lower extremity amputations and mortality (Levey et al., 2003, Levey et al., 2005, Yang et al., 2011, Levin and Stevens, 2014, Matsushita et al., 2022). Moreover, decreased kidney function is a major risk factor for hospitalisations (Go et al., 2004, Gansevoort et al., 2013), cognitive dysfunction (Etgen et al., 2012), as well as a poor quality of life (Yapa et al., 2023). As a result, CKD diagnosis is accompanied by high clinical burden and healthcare costs.

The annual cost of CKD management for the National Health Service (NHS) in England (between 2009-2010) was approximately £1.45 billion, with 50% used for the treatment of ESRD, while the remaining funds were spent on primary care costs, such as hypertension treatment (Kerr et al., 2012). More recently, an evaluation of healthcare resource use in patients in the UK showed that the mean annual per-patient costs (between 2008-2020) varied from £4966 (CKD stage A1) to £9196 (CKD stage A3) and from £4997 (CKD stage G2) to £7595 (CKD stage G5) (Pollock et al., 2022). Costs for each healthcare resource generally rose with increasing CKD severity, as a result of additional hospitalisations, outpatient visits, critical care, general practice visits, emergency room visits, and ambulance use. Furthermore, individuals with CKD who also have comorbid T2DM and/or CVD incurred even higher expenses than those with CKD alone, again primarily due to higher hospitalisation rates and outpatient visits. On the whole, the average healthcare costs, regardless of the CKD stage, amounted to £6,149 per person per year for individuals diagnosed with both CKD and T2DM, compared to £5,076 for those with CKD only. Considering the projected rise in the prevalence of T2DM (Saeedi et al., 2019), it is expected that the occurrence of CKD and associated co-morbidities will further increase the burden on the healthcare system. This health economic data further highlights the importance of predicting early kidney function decline, as it allows for the introduction of interventions that may slow the decline and, in turn, reduce risk of complications which are associated with higher healthcare costs.

The prevalence of ESRD is also on the rise, particularly among individuals with diabetes mellitus, with up to half of people with diabetes-related CKD being affected

(Tuttle et al., 2014, McCullough et al., 2019, Cheng et al., 2021). Managing patients in the final stages of CKD places an additional significant financial burden on the NHS. In 2017, National Institute of Health and Care Excellence (NICE) estimated that each patient costs £30,591 per year for dialysis (NationalGuidelineCentre, 2017), and an initial cost of over £27,000 for kidney transplantation (NHS, 2021). Notably, the expenses associated with dialysis alone exceed six times the total healthcare costs for individuals in earlier CKD stages. These substantial costs highlight the advantages to the economy and healthcare system of early detection of CKD and proactive management, aimed at delaying the progression to the advanced stages.

In addition to the significant healthcare burden, CKD has a substantial impact on the quality of life and daily well-being, as described by four key themes (Yapa et al., 2023):

i) The first theme relates to the physiological experience of deteriorating kidney function. Patients become acutely aware of changes in their bodies, such as fatigue and during the more advanced stages the loss of appetite, swelling and shortness of breath. They also experience shifts in mental well-being, often accompanied by anxiety or depression, stemming from the CKD diagnosis and the physical symptoms that hinder their ability to perform everyday tasks.

ii) The second theme, titled as 'Changes in Everyday Life,' suggested that CKD prevents patients from engaging in daily physical activities and participating in social events. In some cases, these changes can have a direct impact on family dynamics, including employment loss and financial hardship, as well as the need for frequent travel to renal clinics.

iii) 'Adapting to a Different Everyday' is the third theme, which highlights the necessary adjustments in daily life required to confront the illness and its associated treatments. The life-limiting nature of CKD compels patients to identify strategies for managing the challenges they encounter in daily life.

iv) The final theme explores how decisions regarding CKD-related treatments are influenced by future planning, and vice versa. While some patients believe that dialysis would help them achieve specific life goals, such as continuing to work or living until their children grow up, even if it comes at the cost of disrupting their daily routines. Others prefer conservative care,

which may result in earlier mortality but without the need for dialysis, which may be seen as an “unnecessary burden and source of suffering.”

The substantial healthcare cost, clinical burden and the negative influence on everyday life, underscores the increasing complexity of care required as CKD progresses. The repercussions of CKD on both individuals and broader society, accentuate the importance of early detection facilitating CKD prevention and timely intervention, a process contingent upon a comprehensive understanding of the key risk factors.

1.1.5 Risk factors for chronic kidney disease

Risk factors are characteristics linked to an increased likelihood of experiencing unfavourable outcomes. Previous epidemiological studies have identified numerous risk factors linked to CKD, categorizing them into initiating and perpetuating elements. Initiating factors contribute to the commencement of nephron loss, including advanced age, male sex, or diabetes, while perpetuating factors, such as increased proteinuria, drive the ongoing progression of the disease (McClellan and Flanders, 2003, Menon et al., 2005, Taal and Brenner, 2006, Levey et al., 2009b, Tsai et al., 2016). Levey et al. (2009b) categorized some examples of CKD risk factors associated with the development and progression of CKD, which are summarised in Table 1-2 (Menon et al., 2005, Levey et al., 2009b, Kazancioglu, 2013).

Table 1-2. Risk factors for development and progression of CKD.

CKD phase	Definition	Examples of risk factors
Development	Increase susceptibility to kidney damage	Family history of CKD, ethnicity (African-American), older age, reduced kidney mass, hyperfiltration states; genetic component* lower socioeconomic status*, elevated uric acid*
	Directly initiate kidney damage	Diabetes, hypertension/high blood pressure, obesity, higher BMI, dyslipidaemia,

		autoimmune diseases, infections, kidney stones, acute kidney injury *
Progression	Worsen kidney damage or accelerate GFR decline	Increased proteinuria; nephrotoxins* (e.g. alcohol, recreational drugs, analgesic drugs, heavy metals), smoking*

Note, many risk factors may be involved in both development and progression, so for consistency, they are listed in the first category where they would appear. Abbreviations: BMI, body mass index; CKD, chronic kidney disease; ESRD, end stage renal disease; GFR, glomerular filtration rate. Examples of risk factors marked with * symbol were taken from Kazancioglu (2013) and the remainder were taken from (Levey et al., 2009b).

A comprehensive understanding of CKD risk factors enhances the potential for preventing disease onset or its advancement to later stages. Recognizing the substantial care burden posed by ESRD and dialysis patients, recent attention has increasingly focused on early intervention and modifying risk factors for individuals with CKD. Whilst some 'risk factors', such as demographic characteristics (e.g., age and male sex) and presence of co-morbidities (e.g., diabetes), may be useful to identify higher risk populations (which can then be targeted for screening and preventive actions), potentially causal risk factors, which are also modifiable (e.g., hypertension, obesity) are those most amenable to preventive measures and/or treatment of CKD. Some key risk factors associated with CKD development are described in more detail below.

Diabetes mellitus

The primary underlying disease associated with CKD is diabetes mellitus, particularly in high and middle-income countries (Tsai et al., 2016). It remains unclear whether CKD is directly caused by diabetes, or stems from diabetes-related microvascular issues, such as albuminuria (Tsai et al., 2016). The impact of diabetes mellitus on kidney function and CKD development is well-documented (de Boer et al., 2020). Individuals with diabetes face an increased risk of developing CKD in comparison to those without a diabetes diagnosis. For example, between 2017 and 2021 in the United States (US), the prevalence of CKD stages 3 - 4 was 21.4% among people diagnosed with diabetes, compared to 12.1% in people with pre-diabetes and 4.4% in people without diabetes (Centers for Disease Control and Prevention, 2021). A global

meta-analysis consisting of 82 studies further affirmed that diabetes was significantly associated with CKD (Hill et al., 2016).

Hypertension

Another major primary disease associated with CKD is hypertension. Among hypertensive adults in the US (between 2017 and 2021), the prevalence of CKD was notably high at 28.3%, in contrast to non-hypertensive individuals with a lower prevalence of 8.9% (Centers for Disease Control and Prevention, 2021). The association between hypertension and CKD prevalence was reaffirmed in a comprehensive meta-analysis, encompassing 75 global studies (Hill et al., 2016).

Hypertension and CKD exhibit a closely interlinked pathophysiologic relationship, wherein sustained hypertension can precipitate deteriorating kidney function, and conversely, progressive decline in kidney function can contribute to compromised blood pressure regulation (Bidani and Griffin, 2004, Brantsma et al., 2006, Kestenbaum et al., 2008). In addition, the co-existence of the two conditions increases the risk of both, cardiovascular morbidity and mortality (Gansevoort et al., 2013). For individuals with eGFR below 60 ml min^{-1} (1.73 $m)^{-2}$, the risk of cardiovascular-related mortality surpasses the risk of progression to ESRD (eGFR < 15 ml min^{-1} (1.73 $m)^{-2}$) (Keith et al., 2004, Foley et al., 2005). Thus, from a therapeutic standpoint, the reduction of blood pressure can attenuate the decline in eGFR, postpone the progression to ESRD, and reduce the incidence of CVD in these patients (Cheung et al., 2017).

Obesity

Obesity, characterised by a body mass index (BMI) exceeding 30 kg/m^2, was associated with increased morbidity and mortality (Sarma et al., 2021). In particular, it is considered a well-established risk factor for developing T2DM, hypertension and CVD. Furthermore, the presence of multiple risk factors, such as obesity and T2DM, can synergistically elevate the risk of CKD onset and progression (Foster et al., 2008, Lu et al., 2015, Sarma et al., 2021).

Some studies have shown a strong association between obesity and CKD progression to ESRD (Ejerblad et al., 2006, Hsu et al., 2006), suggesting that excess body weight is one of the major modifiable risk factors for accelerated kidney function

decline. In addition, a systematic review and meta-analysis showed that both BMI (as a continuous variable) and obese status were significantly associated with the new-onset of low eGFR and albuminuria (Garofalo et al., 2017). Interestingly, overweight status showed no significant effect on the same outcomes, indicating a non-linear relationship between BMI and the risk of kidney function decline. These findings could hold major clinical significance when assessing the risk of new-onset CKD and implementation of preventative measures, particularly given the growing global prevalence of obesity (Hay and Collaborators, 2017) and CKD (Jager et al., 2019), both of which are recognised as major epidemic diseases.

Other risk factors

Non-traditional CKD risk factors such as exposure to nephrotoxins, kidney stones, foetal and maternal factors, infections, and acute kidney injury() events are increasingly gaining recognition as significant threats to kidney health (Kazancioglu, 2013). For example, in low and middle-income countries, CKD is linked to infectious diseases, glomerulonephritis, and the improper use of medications, including traditional remedies with potential nephrotoxins (Jha et al., 2013, Stanifer et al., 2017). Also, low birthweight, typically defined as less than 2.5 kg due to preterm birth or intrauterine growth restriction, is associated with CKD later in life. Globally, around 10% of births are preterm, and approximately 15% are low birthweight, putting millions of children at risk for CKD later in life (Charlton et al., 2014, Khalsa et al., 2016).

The wide range of risk factors that are associated with CKD development and progression highlights the complex nature of this condition. The growing prevalence of high-risk populations, such as rising prevalence of people with diabetes mellitus, presents additional challenges for disease prevention and management strategies. Nonetheless, the modifiable nature of many risk factors (e.g., obesity and hypertension) emphasises the possibility of disease prevention through early interventions and highlights the importance of understanding the disease presentation in these high-risk populations.

1.2 Chronic kidney disease and type 2 diabetes mellitus

While the association between CKD and T2DM is well-established (as discussed in section 1.1.5) (Levin et al., 2013, de Boer et al., 2020), the pathophysiology of CKD in people with T2DM is complex, since a cluster of cardiovascular risk factors, such as

obesity, hypertension, and dyslipidaemia, may contribute to disease development and progression. The intricate interplay between these health issues is of great concern to healthcare professionals and researchers, as managing one condition often involves addressing the other. The coexistence of CKD and T2DM can significantly impact patient outcomes, necessitating a comprehensive understanding and targeted approaches for prevention and management.

Although the management of diabetes has witnessed improvements that mitigate the risk of CKD development among these individuals, the combination of increased life expectancy and the growing incidence of both type 2 and type 1 diabetes, has led to a rising global prevalence of CKD (Hovind et al., 2003, Kianmehr et al., 2022). Considering that T2DM is responsible for most diabetes cases (over 85%) (Forouhi and Wareham, 2014) and its prevalence continues to rise (Khan et al., 2020), the increasing incidence of CKD is also primarily linked to the growing numbers of individuals with T2DM. Interestingly, the prevalence of CKD among patients with diabetes exhibits significant global variation, as indicated by the evidence from a narrative review (Koye et al., 2018). Estimates range from 27.1% in Shanghai, China, to as high as 83.7% in Tanzania (Guo et al., 2016, Mpondo et al., 2016). While UK was found in the middle of the two extremes, with 42.3% of individuals with T2DM documented as having CKD (Hill et al., 2014).

The definitions of CKD among studies differ, which may in part explain the discrepancies seen in the prevalence. For example, multiple Asian studies defined CKD diagnosis based on a single spot urinary albumin to creatinine ratio (uACR) measurement, and consequently reported a lower prevalence of CKD in people with diabetes, compared to similar studies elsewhere that used eGFR to define cases (Koye et al., 2018). Other factors that may cause variation in prevalence may be attributed to differences in diabetes prevalence across countries (Khan et al., 2020, Sun et al., 2022), as well as local factors, including differences in access to nephrology care, genetic predisposition, public health policies, and environmental exposures (Cheng et al., 2021). Regardless of inconsistencies in the definitions and differences between global populations, it is evident that CKD is highly prevalent among people with T2DM across the globe (Koye et al., 2018, Cheng et al., 2021).

1.2.1 Presentation of chronic kidney disease in type 2 diabetes mellitus

Originally, a screening strategy for CKD in individuals with diabetes was established in parallel with the aetiology of diabetic nephropathy (Mogensen et al., 1983). While diabetic nephropathy and CKD are both conditions that affect the kidneys, they differ in their causes, progression, and clinical features. Diabetic nephropathy is a specific kidney condition caused directly by diabetes. It refers to the damage to the filtering units (glomeruli) of the kidneys and occurs as a result of long-term high blood sugar levels in people with diabetes, particularly uncontrolled Type 1 or Type 2 diabetes. Diabetic nephropathy is essentially a subset of CKD that occurs only due to diabetes. In contrast, CKD is a general term that refers to any long-term, progressive loss of kidney function, regardless of the underlying cause. In people with diabetes, CKD can be caused by diabetic nephropathy, but it may also arise from other factors such as hypertension (which often coexists with diabetes), glomerulonephritis, or other non-diabetes-related kidney conditions.

Diabetic nephropathy is usually a direct consequence of diabetes, the hallmark of which is nodular glomerulosclerosis (Anders et al., 2018). The classic stages of glomerulopathy (glomerular injury in the kidneys and histopathological feature of diabetic nephropathy) begins with the development of albuminuria (an increase in uACR >30 mg-300 mg), followed by progressive decline in kidney function (eGFR <60 ml min^{-1} (1.73 $m)^{-2}$) and ultimately ESRD. Thus, traditionally, increased albuminuria has been widely accepted as a predictor for the initiation of a decline in renal function and has been used for screening, early diagnosis, and the management of CKD in people with diabetes.

However, growing evidence has demonstrated that declining kidney function in diabetes is not a straightforward pathogenic process, undermining the classic definition. Studies have suggested that diabetic nephropathy is caused by microvascular complications (i.e., affects small blood vessels) that predominantly affect young and lean individuals with T1DM (Tervaert et al., 2010). Furthermore, Penno et al. (2012) demonstrated that the majority of patients with diabetic retinopathy had an increased albuminuria, suggesting that both are microvascular complications and that diabetic retinopathy is a key risk factor for diabetic nephropathy. In contrast, a study found that people with diabetes who developed significantly reduced GFR indicative of CKD, but without albuminuria, had more advanced renal dysfunction, and

lower haemoglobin, compared to patients with increased albuminuria (Laranjinha et al., 2016). While albuminuria is still recognised as the classic hallmark of diabetic nephropathy, non-albuminuric CKD is more commonly observed in T2DM than in T1DM (National Kidney Foundation, 2012).

Indeed, it appears that individuals with T2DM who do not follow the natural evolution of diabetic nephropathy, but have renal impairment (classed as CKD stage G3 or above) are highly prevalent (Halimi, 2012, Laranjinha et al., 2016). In fact, the UK Prospective Diabetes Study cohort exemplified this phenomenon, wherein 50% of patients displaying renal dysfunction did not exhibit concurrent albuminuria (Retnakaran et al., 2006). Similarly, a National Health and Nutrition Examination Survey (NHANES)–III study (based in US) found that 46% of individuals with T2DM and reduced GFR did not have increased albuminuria (Afkarian et al., 2013). The Edinburgh Type 2 Diabetes Study (ET2DS) cohort provided an extreme example, where 101 out of 119 patients who experienced a significant decline in their renal function did not have preceding albuminuria (Jenks et al., 2017). The ET2DS and NHANES III studies, among others, provide examples where albuminuria is not a predictor of kidney function decline in people with T2DM, highlighting the need for additional predictors. The prevalence of these patients could be attributed to the fact that the non-albuminuric phenotype in CKD is fairly common among older individuals with longer diabetes duration, as observed consistently in the studies described above. Therefore, albuminuria lacks precision in defining the predominant form of kidney dysfunction in individuals diagnosed with T2DM, rendering uACR as only modestly predictive of subsequent renal decline in this population.

Consequently, international KDIGO guidelines classify individuals with both CKD and diabetes mellitus (or DKD) based on clinical measurements of uACR and eGFR, whereas the term diabetic nephropathy was reserved exclusively for the histological diagnosis of glomerular alterations in a biopsy setting (de Boer et al., 2020). However, kidney biopsies are typically omitted in people with diabetes due to restricted treatment alternatives, so the histopathological features causing kidney dysfunction are often unknown. As a result, people with T2DM and renal decline are diagnosed with CKD, which encompasses a wide range of underlying causes and presents with heterogeneous course of disease development.

Recognising CKD among individuals with diabetes holds significance for enhancing clinical outcomes. An illustrative example is the application of a population-based screening strategy (Kidney disease Program) established by Indian Health Service specifically for Native American and Alaskan Native populations, as they are recognized for their elevated diabetes and kidney disease prevalence (Narva, 2018). This screening approach enabled a more effective implementation of combined management of CKD and diabetes within primary care settings, which resulted in a notable 54% reduction of incident ESRD between 2000 and 2016 (Doshi and Friedman, 2017, Narva, 2018). Similarly, in Pima Indian population, there was a notable reduction in albuminuria cases (Pavkov et al., 2009), and patients presented with slower rate of kidney decline and fewer incident ESRD cases, which at least in part, may be attributed to increased usage of antihypertensive medications and glycaemic control at earlier stages of CKD development (Pavkov et al., 2006). Evidently, screening and diagnosis of CKD assume paramount importance in the initiation of therapeutic interventions, aiming to prevent or delay associated complications.

1.2.2 Risk factors for chronic kidney disease in type 2 diabetes mellitus

Individuals with T2DM are considered a high-risk population and CKD is recognized as a common complication of diabetes, which is also known as diabetic kidney disease (DKD) (Chu et al., 2021). Thus, annual screening for CKD is recommended upon the diagnosis of T2DM (de Boer et al., 2020). The identification of risk factors associated with DKD is imperative for the implementation of targeted preventive measures or interventions to decelerate CKD progression. Notably, many risk factors overlap between CKD and DKD, such as hypertension and age, both of which increase the risk of kidney function decline in people with T2DM or without (Jitraknatee et al., 2020, Swartling et al., 2021). Whereas, poor glycaemic control (i.e., management of blood glucose levels) is a well-recognised factor of diabetes-related complications, including the risk of CKD (Guo et al., 2016, Russo et al., 2018, Jitraknatee et al., 2020). Our understanding of kidney disease in people with T2DM was particularly shaped by studies of Pima Indians (a native American Tribe based in Arizona), which has one of the highest incidences of T2DM affecting over half of those aged 35 and above (Knowler et al., 1983). Early studies of this community highlighted kidney disease as a common complication of T2DM (Nelson et al., 1988, Nelson et al., 1993) and aided

characterisation of many risk factors associated with kidney decline specifically in people with T2DM, such as low birth weight (Nelson et al., 1998) and genetic determinants (Pettitt et al., 1990).

In general, risk factors for CKD can be categorized into non-modifiable and modifiable factors (Saran et al., 2015, Thomas et al., 2015, Fiorentino et al., 2017, Hoogeveen et al., 2017, Tonneijck et al., 2017, Esmeijer et al., 2018, McCullough et al., 2019, Esmeijer et al., 2019). Non-modifiable risk factors encompass genetic predisposition, male sex, advanced age, familial history of T2DM or CKD, insulin resistance, and ethnicity (e.g., Black, Hispanic, American Indian, Asian) and low birth weight; Conversely, modifiable risk factors include obesity, metabolic syndrome, suboptimal glycaemic control, hypertension, acute kidney injury, cigarette smoking, dyslipidaemia, sedentary lifestyle and elevated salt consumption. Some key examples of risk factors are described in more detail below:

Older age

The increasing prevalence of T2DM is largely attributed to demographic shifts, characterized by economic development, improved living standards, and healthcare, leading to longer life expectancy (Boyle et al., 2010, Khan et al., 2020). As societies age, with a notable rise in the elderly population, the incidence of T2DM is expected to surge (Khan et al., 2020). The aging population, especially those over 65 years old, faces a higher likelihood of T2DM and associated comorbidities, such as hypertension and atherosclerotic vascular disease. The prevalence of renal impairment is also significant in this age group, with approximately 20% of T2DM patients over 65 experiencing reduced kidney function (Parving et al., 2006, Thomas et al., 2006b, Thomas et al., 2016). As the global population continues to age, the intersection of T2DM and renal impairment is anticipated to rise, posing challenges for healthcare management and outcomes.

Blood glucose and hypertension

The escalating prevalence of CKD in individuals with T2DM is also influenced by changes in risk factors, particularly glucose control and hypertension. Over the past three decades, there has been notable progress in managing these factors, driven by increased awareness of their impact on clinical outcomes in trials and emphasized in treatment guidelines (Imperatore et al., 2004, Hoerger et al., 2008, Tseng et al., 2012).

Effective control of glucose levels and blood pressure has been shown to significantly prevent albuminuria in T2DM (Derby et al., 1989, Schmitz et al., 1994, Tanaka et al., 1998, Boussageon et al., 2011, Emdin et al., 2015), although the evidence regarding other renal outcomes is less conclusive (Boussageon et al., 2011). The mandated early and intensive management of CKD risk factors in individuals with or at risk of T2DM may have contributed to a modest reduction in albuminuria prevalence, but has not shown a clear reduction in renal impairment. Some researchers suggest that the growing number of T2DM patients with renal impairment, but without albuminuria could be attributed to improved diabetes mellitus treatment for glycaemic control (Thomas et al., 2016).

Dyslipidaemia

Dyslipidaemia is a significant risk factor for CKD in individuals with T2DM. While mean total cholesterol levels have decreased in T2DM patients (Imperatore et al., 2004), the prevalence of dyslipidaemia, characterised by elevated triglycerides, apolipoprotein B, or low high density lipoprotein (HDL) cholesterol, is on the rise (Thomas et al., 2006a, Kaysen, 2006, Penno et al., 2015). This dyslipidaemia pattern is linked to the development and severity of CKD, particularly in individuals with insulin resistance and adiposity (de Vries et al., 2014). The correlation between dyslipidaemia and renal function suggests that excess free fat and renal lipid accumulation may hasten kidney damage associated with T2DM. Treatment with peroxisome proliferator-activator receptor α agonists, such as fenofibrate, which target triglyceride-rich very low density lipoprotein (VLDL) particles, has been associated with a reduction in albuminuria (Jun et al., 2012). However, the reno-protective effects of statins, which primarily modify low density lipoprotein (LDL) cholesterol, remain unclear in T2DM patients (Haynes et al., 2014, Griffin et al., 2011), and some studies even suggest potential detrimental effects on renal function associated with potent statin use (de Zeeuw et al., 2015).

Obesity

The prevalence and severity of obesity are on the rise in individuals with T2DM and the general population (as discussed in section 1.1.5). The average BMI in T2DM patients has increased over the years, with a notable shift towards morbid obesity (BMI >40 kg/m2) (de Boer et al., 2011). Many individuals with T2DM have a history of long-term obesity, which not only contributes to the development of diabetes but also exerts

marked effects on the kidneys. Long-term obesity influences intra-glomerular hemodynamic, increases sympathetic activity, induces hypertension, triggers systemic inflammation, leads to endothelial dysfunction, and contributes to visceral adiposity-related compression (Thomas et al., 2015), all of which are factors that heighten the risk of CKD in T2DM. Conversely, weight loss is linked to a reduced incidence of CKD in T2DM, as evidenced by randomized controlled trials like Look-Ahead, which showed significant reductions in albuminuria and modest improvements in renal function with lifestyle-induced weight loss (Look, 2014). Bariatric surgery has also been associated with a reduced incidence of CKD in obese individuals with T2DM (Carlsson et al., 2015).

1.2.3 Complications of chronic kidney disease in type 2 diabetes mellitus

The development of CKD in individuals with T2DM predicts an increased risk of ESRD and other morbidities such as CVD, retinopathy, neuropathy, foot ulcers, and amputations, which amplify the risk of premature mortality, often occurring before reaching the ESRD stage (McCullough et al., 2008, Afkarian et al., 2013, Jha et al., 2013). Renal decline not only predicts the risk of hospitalisations, but also contributes to the escalating burden on public healthcare (Go et al., 2004). Consequently, CKD stands out as one of the most prevalent, burdensome, and costly long-term complications associated with diabetes (Sun et al., 2022). By 2040, CKD is projected to be the 5th highest cause of years of life lost globally (Foreman et al., 2018)

In particular, the individuals with T2DM and declining kidney function are at a heightened risk of experiencing cardiovascular complications (e.g., myocardial infarction, stroke, ischemia, arrhythmia, and heart failure) (Holman et al., 2008, Gansevoort et al., 2013). In fact, the risk of CVD in individuals with diabetes and CKD is firmly established and can be viewed as a "CVD risk-equivalent." This means that the incidence of CVD in people with diabetes and CKD, even without evident CVD, is as significant as in individuals without diabetes who have prevalent CVD (de Boer et al., 2020). Hence, CVD is one of the major causes of mortality in patients suffering with T2DM and CKD (Afkarian et al., 2013).

Given that CVD is a major complication of DKD, the KDIGO guidelines recommend a comprehensive and multifaceted treatment strategy for these individuals to improve both kidney and cardiovascular outcomes (Group, 2022). This strategy includes

lifestyle adjustments (such as smoking cessation, diet and exercise) along with first-line drug therapies supported by robust evidence for enhancing both kidney and cardiovascular outcomes. The goal is to improve clinical outcomes regardless of its effects on intermediary targets. For instance, as a result of consistent demonstration of kidney and cardiovascular benefits across GFR categories with the use of sodium-glucose co-transporter type 2 inhibitors (Bakris et al., 2020, Packer et al., 2020, Chertow et al., 2021), it was recommended to prioritise them as a primary treatment for individuals with DKD, irrespective of glycaemia levels. Therefore, sodium-glucose co-transporter type 2 inhibitors are considered a cornerstone in the pharmacologic therapy for individuals suffering with CKD and T2DM. Additional medications with demonstrated kidney and heart protective properties may be also incorporated based on assessments of residual risk, including controlling lipid levels and blood pressure, both of which are the major drivers of CVD (National Kidney Foundation, 2012). For example, the use of antihypertensive agents for blood pressure control, especially angiotensin receptor blockades demonstrated reno-protective effects in terms of slowing down kidney function decline, lowering albuminuria and reduction of the ESRD risk in Angiotensin II Antagonist Losartan Study of people diagnosed with T2DM and renal dysfunction (Brenner et al., 2001, Bakris et al., 2003, de Zeeuw et al., 2004). Likewise, in the Irbesartan in Patients with Type 2 Diabetes and Microalbuminuria Study trial, the administration of 300 mg of irbesartan once daily demonstrated a significant 70% reduction in the risk of diabetic nephropathy progression and a 38% decrease in microalbuminuria among hypertensive individuals with T2DM (Parving et al., 2001).

Notably, the evidence for DKD management is rapidly evolving as seen in the original KDIGO guidelines published in 2020 (Disease, 2020), which were soon updated in 2022 (Group, 2022). The brief duration between guideline updates reflects the growing availability of new-evidence that enables swift progress in the management of diabetes and CKD. Theoretically, specialised healthcare services designed to address all aspects of care for individuals with diabetes-related CKD could offer comprehensive care, targeting complications and potentially leading to improved clinical outcomes by slowing down the progression. However, given the current structure of services, providing an all-encompassing approach to care for the extensive diabetes population affected by CKD is not feasible. Therefore, the focus should be on disease screening

and developing methods to predict the onset of disease early, enabling the implementation of early preventative and therapeutic interventions that slow down the progression and development of complications.

1.3 Novel biomarkers of chronic kidney disease

Clinical screening programs routinely measure kidney function in high-risk groups, such as individuals with diabetes, using well-established biomarkers like albuminuria and eGFR. However, eGFR alone is often not specific enough, as it is influenced by factors unrelated to kidney health, such as age and muscle mass (Stevens et al., 2006). Consequently, these traditional biomarkers typically become elevated only after substantial kidney damage or significant loss of filtration capacity has occurred (Levin and Stevens, 2014). As a result, these markers reflect renal injury after a significant amount of kidney function has already been lost. Detecting CKD at an early stage remains a critical yet unmet medical need, as early detection of kidney decline would allow for timely, targeted interventions. This is crucial not only for predicting and preventing CKD onset but also for improving patient outcomes, including survival and reducing comorbidities like cardiovascular disease.

In addition to identifying various risk factors that enable screening and prevention, there has been a significant emphasis on researching novel preclinical markers associated with diverse disease pathways. Non-invasive biomarkers (e.g. measurable in blood and urine) are preferable as kidney biopsy confers a risk of bleeding and hence is not performed in most people (Lees et al., 2017). Recent studies have found several preclinical markers associated with incident CKD, including the high risk genetic variants in *APOL1*, gene encoding for Apolipoprotein L1 (Parsa et al., 2013), urinary epithelial growth factor (Ju et al., 2015, Betz et al., 2016), urinary and circulating levels of uromodulin (Kottgen et al., 2010, Leiherer et al., 2018, Steubl et al., 2019), and circulating levels of soluble urokinase-type plasminogen activator receptor (Hayek et al., 2016, Schaefer et al., 2017). The overarching goal here is to identify biomarkers that go beyond traditional measures of kidney health, specifically glomerular dysfunction (serum creatinine) and glomerular injury (albuminuria) to aid early and more accurate prediction of CKD development.

Establishing sensitive biomarkers for predicting clinical outcomes is crucial for risk stratification, which provides direct benefits in clinical research and beyond. For example, identifying high-risk individuals contributes to more efficient clinical trials, by reducing sample size and follow-up duration without compromising statistical power, as successfully implemented in heart failure and oncology trials (Ibrahim et al., 2016, Hu and Dignam, 2019). In nephrology, data from the Chronic Renal Insufficiency Cohort (CRIC), suggested that employing tumour necrosis factor receptor 2 (TNFR2) levels ≥75th percentile to identify high-risk individuals can lead to a 50% reduction in the required sample size for detecting a 20% reduction in DKD progression over a 5-year period (Coca et al., 2017). Furthermore, the integration of such biomarkers enhances our understanding of biological pathways and their responses to therapeutic interventions, as previously exemplified by sodium-glucose co-transporter type 2 inhibitors, which have demonstrated the ability to reduce tumour necrosis factor receptor 1 (TNFR1), TNFR2, and kidney injury molecule-1 (KIM-1) levels in people diagnosed with DKD (Heerspink et al., 2019, Sen et al., 2021)

Several studies have focused on evaluating the effectiveness of individual biomarkers, each representing a single disease-associated pathway, in predicting the progression of renal function decline in the general population and in patients with T2DM (Liu et al., 2022). For example, plasma TNFR1 was the most studied biomarker found in blood and has been associated with the inflammatory processes affecting endothelial cells in the kidneys. Statistically significant positive correlations between TNRF1 and incident CKD as well as progression to ESRD have been consistently demonstrated across 33 independent studies (Liu et al., 2022). Likewise, KIM-1, the predominantly studied urinary biomarker indicative of tubular damage in the kidneys, exhibited a consistent association with the development and progression of CKD across various populations (Liu et al., 2022). However, Liu et al. (2022) highlighted that in clinical studies these biomarkers demonstrate differing strengths or weakened association with CKD outcomes after adjustment for confounding factors. Thus, currently preclinical markers lack evidence that prove their effectiveness in detecting the new-onset of CKD or prediction of kidney function decline. The shortage of robust findings could be attributed, at least partially, to diverse study designs, insufficient follow-up intervals, or variations in biomarker platforms used among studies. Consequently, the

dependability of preclinical markers for autonomously predicting kidney events across different CKD groups and offering significant clinical insights remains uncertain.

On the other hand, the development of CKD is a complex, multifactorial process characterized by the activation of various pathways, including pro-inflammatory, pro-fibrotic, and angiogenic processes, among others (Gajjala et al., 2015). Given the intricacy of these multiple pathophysiological processes involved in the development of CKD, along with the inherent variability of biomarkers within individuals, the utility of a single biomarker in terms of diagnostic and prognostic power is a matter of doubt. Alternatively, a combination of biomarkers also known as multi-marker panels, may offer a more accurate reflection of the true pathophysiological status in particular patient groups, and consequently, provide a superior assessment of disease prediction (Colhoun and Marcovecchio, 2018, Bidin et al., 2019).

Advancements in analytical methodologies, gave rise to high-throughput and high-content omics, which enabled insights into biological molecules involved in multiple pathways of disease. For example: i) next-generation sequencing enabled genomics, which allowed comprehensive study of genetic variation (Hu et al., 2021) ii) high resolution nuclear magnetic resonance analytics enabled metabolomics, which empowered the level of information gathered about the structure and dynamics of molecules in a sample (Vignoli et al., 2019) iii) mass spectrometry, enabled proteomic analysis, which facilitated the identification and quantification of many proteins in a given sample. Applying these technologies in large-scale population studies not only improved our understanding of molecular processes involved in diseases such as CKD, but also enabled the development of novel multi-marker panels to improve prediction of disease development. For instance, CKD273 is a proteomic classifier based on 273 urinary peptides, which successfully distinguished between CKD (resulting from various aetiologies) and healthy individuals (Good et al., 2010). The same CKD273 classifier was effective at differentiating between CKD stages and predicting which individuals are set to develop rapidly declining kidney function in both general population and people diagnosed with diabetes (Zurbig et al., 2012, Schanstra et al., 2015).

Evidently, novel biomarker research focused on predicting the onset or progression of CKD is an area of interest. The analytical advances facilitated opportunities for gaining

better understanding of molecular pathways involved in CKD development, which offered insights into elevated risk profiles of disease. Numerous studies have demonstrated that single biomarkers as well as multi-marker panels have shown potential in predicting CKD outcomes, warranting further research in this area.

1.4 Metabolomics

1.4.1 Omics

Omics technologies enabled large-scale, comprehensive and systematic exploration of the molecular components within a specific class of biological molecule, which transformed molecular research by providing extensive data that can be used to uncover patterns, associations, and potential biomarkers. For example, the completion of the Human Genome Project in 2003 was made possible by genomic sequencing technologies, which analysed almost the entire genetic code of humans and allowed an investigation of the human genome as a whole (International Human Genome Sequencing, 2004). This was one of the largest scientific collaborations that accelerated the utility of genomic technologies in research and commercial settings, aiding our understanding of many diseases. In particular, the rise of genome-wide association studies (GWAS) led to the discovery of many associations between genetic features and pre-disease conditions (Kruglyak, 2008). However, the genetic code is not a complete picture of human health and thus cannot answer all of our questions about disease and environmental health. To understand the phenotype, we need to measure the combined effect of the genome, lifestyle, and environment on the function of biological systems. Other branches of omics, downstream of genomics, such as proteomics (study of proteins expressed in a cell, tissue or organism) and metabolomics (study of small molecules within biological system) provide further insight into phenotypes related to health and disease.

1.4.2 Metabolomic profiling

Metabolomics is one of the emerging analytical fields in molecular epidemiology. The terms "metabolome" and "metabolomics" were first introduced into the scientific literature by Steve Oliver and colleagues in 1998 (Oliver et al., 1998). Although metabolomics is regarded as the younger branch of the omics, metabolites and metabolism have been studied for more than 100 years. Metabolites are low-molecular weight molecules that provide cells with energy, structural components, and the

necessary materials to facilitate the creation of larger molecules like DNA or proteins. Metabolites are present within the cells and in the environment surrounding the cells and tissues found in fluids such as blood and urine. The metabolome is the entire collection of metabolites in a biological system, which can provide insights into ongoing or completed biological processes (Nicholson and Lindon, 2008). The metabolome is downstream of the genome and provides an amplified and dynamic measure of the cellular changes resulting from the processes involved in the genome, transcriptome, proteome, lifestyle, and environment (Holmes et al., 2008a, Bictash et al., 2010).

Metabolomic profiling involves the identification and measurement of metabolites present in a biological sample (e.g., blood serum), focusing on low-molecular-weight compounds (Nicholson and Lindon, 2008, Bictash et al., 2010). These analyses have the capacity to identify numerous metabolites from a single sample, ranging from sugars (hexoses) and amino acids to dipeptides, lipids, and organic acids (Holmes et al., 2008b). Currently, there are more than 217,920 chemically and structurally diverse compounds described in the Human Metabolome Database (Wishart et al., 2022).

Analysing the metabolome requires sophisticated separation and detection methods. One of the major driving forces in the rapid development of the metabolomics field is the improvement in analytical technologies. Metabolomic measurements can be performed using either mass spectrometry or proton nuclear magnetic resonance spectroscopy. The nuclear magnetic resonance distinguishes and quantifies various compounds by recording their resonance frequencies, which correspond to the energy released during rapid changes in an external magnetic field. Conversely, mass spectrometry measures the mass-to-charge ratio of ionized analytes to achieve the same goal. The nuclear magnetic resonance offers several advantages, including its non-invasive and non-destructive nature, minimal sample preparation requirements, and ease of use for structural elucidation. However, it has some limitations; nuclear magnetic resonance can identify a more limited number of metabolites and is generally less sensitive when compared to mass spectrometry (Nicholson and Lindon, 2008)

The datasets generated by metabolomics are complex, necessitating robust, reproducible methods for data processing and analysis. Comprehensive statistical analysis is essential for biological interpretation of metabolomic profiles as it enables identification of metabolites with abnormal levels, which may be indicative of changes

in metabolism in the context of disease. Statistical analysis is an important aspect in the metabolomics field and requires careful consideration (Blaise et al., 2009, De Livera et al., 2013, Barnes et al., 2016, Chen et al., 2022, Anwardeen et al., 2023).

1.4.3 Statistical approaches in metabolomics

Typically, statistical analysis of metabolomic data involves data pre-treatment process, such as normalization and scaling. Normalization adjusts the complete profile of each sample by a factor, ensuring effective comparison of profiles from diverse samples (Craig et al., 2006). This data transformation eliminates variation associated with systematic effects, such as total sample mass and aligns the measurements with the absolute abundance of analytes in samples. Scaling modifies each variable by a factor to accommodate distinct statistical characteristics, such as variations in metabolite abundance (van den Berg et al., 2006). For example, unit-variance scaling, also referred to as standardization, involves dividing each variable by its standard deviation, ensuring that each variable attains a standard deviation of one.

After the data pre-treatment, the dataset may undergo exploration, typically encompassing the identification of variables capable of classifying and/or predicting the outcome associated with a given sample. Multivariate statistical methods are well-suited for exploring the association of variables with outcomes, either individually or collectively. These methods consider the variation across all variables and samples simultaneously, revealing multiparametric signatures that elucidate the interconnections among samples and variables. The fundamental principle behind multivariate techniques is that similar samples cluster together, while dissimilar ones are positioned farther apart in this space. Various approaches, such as linear projections, machine learning, and Bayesian statistics, can be employed for this type of analysis.

Unsupervised and supervised analyses can be conducted to investigate discrimination between sample groups or relationships with continuous outcomes. Unsupervised methods seek latent patterns in data without prior knowledge of outcomes. For example, principal component analysis, examines the dataset to identify directions of greatest variance in multivariate space and reduce its dimensionality (Carey et al., 1975). Supervised approaches facilitate outcome prediction, such as disease classes, based on input data (metabolic variables). These analyses specifically examine the

relationship between the X data matrix (spectral variables e.g., metabolomic measurements) and outcomes Y (sample classes or other non-spectral metadata like age and sex), offering explicit information on the strength of the association between X and Y. However, there is a risk of overfitting (capturing noise instead of genuine structure), which can be mitigated by employing an independent test set.

In contrast, univariable approaches, which consider each metabolite independently are commonly used for analysis that aims to select candidate biomarkers. Univariate analysis methods focus on significance testing to assess the statistical significance of observed associations (Blaise et al., 2009). However, to address the increased risk of false positives due to multiple testing, correction of significance threshold is essential. Various procedures, such as the Bonferroni correction for controlling family-wise error rate (FWER) (Bland and Altman, 1995), or methods like Benjamini-Hochberg corrections for controlling false discovery rate (FDR) (Benjamini and Hochberg, 1995), can be employed to manage these corrections. The FWER method represents the proportion of variables incorrectly identified as significant by a statistical method among those with a true null hypothesis (no effect). On the other hand, the FDR measures the proportion of variables declared as significant that actually correspond to a null hypothesis. For instance, in a scenario with 1,000 tested variables, where 40 truly differentiate classes, if 30 are deemed significant at a specific P-value cut off, but only 25 are true positives, there are five false positives, resulting in an FWER of 0.5%. Similarly, with 30 positives, but 5 false positives, the FDR is 16.7%. While the FDR approach tends to be less stringent in controlling false positives than the FWER, it can offer higher statistical power.

In general, the interpretation of multivariate models is inherently more complex than that of univariate models. However, univariate models lack consideration for correlations, dependencies, or interactions between variables. Thus, the choice of specific analysis approach should be guided by overarching goal of the study, and in many instances multiple approaches can be used to answer different aspects of the research question.

1.4.4 Metabolomic profiling in chronic kidney disease and diabetes

Considering that metabolomics has the potential to capture and quantify hundreds or even thousands of circulating metabolites across multiple pathways in a single

biological sample, it may enhance the resolution of our understanding of complex diseases (e.g., CKD and diabetes) at the molecular level. For example, in CRIC study involving 1,001 participants with diabetes exhibiting various stages of kidney disease, Kwan et al. (2020) found that elevated levels of 3-hydroxyisobutyrate in urine were linked to a more rapid decline in GFR and a shorter time to the onset of ESRD. Interestingly, 3-hydroxyisobutyrate is an intermediate metabolic product of valine, a compound that has demonstrated positive associations with T2DM via mechanisms that promote insulin resistance (McCormack et al., 2013, Halama et al., 2016). More, specifically, it has been suggested that 3-hydroxyisobutyrate contributes to insulin resistance in skeletal muscle by enhancing the uptake of fatty acids from endothelial cells (Jang et al., 2016). This demonstrates metabolomics as a potentially useful tool for identifying novel biomarkers related to diabetes, which may also improve the risk prediction of renal decline in these patients.

In addition, metabolomic profiling provides an opportunity to study hundreds of potential novel biomarkers simultaneously, free of hypotheses, rather than focusing on a single biomarker. Previous studies suggested that the combination of metabolomic profiles with traditional risk factors could provide better performance for predicting incident CKD risk in general population (Rhee et al., 2013) and risk of progression to ESRD in individuals with T2DM (Solini et al., 2016) as well as general population (Zacharias et al., 2019). In all three studies, the multi-marker panel based on metabolomic profiles led to improved discriminative ability and reclassification. However, to my best knowledge, there are no studies that explored the use of metabolomics based multi-marker panels in predicting incident CKD specifically in individuals with T2DM.

1.5 Risk Prediction Modelling

1.5.1 Prognosis and risk prediction

Risk prediction and prognosis or the anticipation of future outcomes plays a crucial role in guiding healthcare decisions and enhancing health-related results. While the concept of risk prediction and prognosis predates modern medicine, its significance has evolved over time. In the seventh century BC, prognosis and prediction was a key medical skill, emphasizing understanding, explaining, and forecasting outcomes based on cumulative observations of patients with similar symptoms (Chauffard,

1913). During this pre-scientific era, prognosis was more about the characteristics of the patient rather than specific diseases (Christakis, 1997). The dominance of prognosis persisted for centuries, but by the mid-nineteenth century, advancements in understanding biological organisms and diagnostic techniques, shifted the focus to diagnosis in healthcare. Prognosis took a back seat, becoming more associated with the characteristics of the disease itself (Christakis, 1997). In the twentieth century, early clinical epidemiology introduced a broader perspective, acknowledging the impact of patient characteristics on outcomes, which directly challenged the assumption that diagnosis and treatment alone could reliably predict and prognosticate (Hutchison, 1928).

In the current era, there is a renewed emphasis on prognosis and risk prediction as a vital tool for achieving optimal healthcare outcomes. Biomedical advancements, including molecular biology and omics technologies, have provided insights into the cellular level of disease development. This progress has a potential to better our abilities in risk prediction, disease prognosis and treatment responses among patients. The wealth of comprehensive patient information available in various databases presents both opportunities and challenges for research concerned with risk prediction.

1.5.2 Risk prediction models and scores

The goal of risk prediction research is to determine which patient features are relevant to the outcomes, how to select and use them for prediction, and ultimately guide personalized healthcare decisions. The key objective of risk prediction is to summarise, explain, and predict outcomes in clinically relevant populations, which supports decision making in healthcare. The four types of risk prediction research were summarised by the Prognosis Research Strategy (PROGRESS) framework (Hemingway et al., 2013, Riley et al., 2013, Steyerberg et al., 2013). In brief, the framework encompasses the four key types of research that investigates questions related to risk prediction, starting with studying the overall outcomes observed in individuals with a specific health condition (Type I). This is followed by the identification of risk factors associated with the outcome (Type II) and combining these factors to predict outcomes, i.e., prognostic/risk prediction models (Type III). The final type of studies are concerned with identifying factors that predict treatment effect (Type IV).

The primary emphasis of this thesis centres on predicting the risk of kidney decline, which falls within the research purposes described in PROGRESS framework Type III. The prognostic model research (Type III) aims to evaluate the utility of combining multiple prognostic factors, termed as predictors, to develop a prognostic model, also referred to as a risk prediction model. Other aims of this type of research include the evaluation of the predictive performance of an existing model in a new population, updating existing models with new predictors, and examining the impact of a risk prediction model when comparing health-related outcomes (e.g., treatment decision or clinical management of disease) are assisted by the model versus without.

A risk prediction model or prognostic tool is a mathematical or statistical calculation that can translate individual risk factor data (individual values of predictors measured at the starting point) into an estimated probability that the individual will experience a health-related outcome in the future (Tangri et al., 2011, Hippisley-Cox et al., 2017, Nelson et al., 2019). More recently, artificial intelligence algorithms have also been used for development of new models and tools (Kanegae et al., 2020). The use of these models facilitates the movement towards health stratification and personalised medicine, which is guided by the individual profiles of predictors or clinical risk factors (Kent and Hayward, 2007, Hingorani et al., 2013). The approach that employs risk prediction modelling assists healthcare professionals by allowing clearer communication of risk to the patient. Furthermore, risk prediction models may improve research study design by enabling stratified randomisation strategies to create a better balance in the prognosis for intervention and control groups (Roozenbeek et al., 2009).

1.5.3 Risk Prediction of incident chronic kidney disease

Currently, there are no conventional risk prediction models recommended for use in clinical setting to predict the risk of incident CKD. Nonetheless, a number of risk prediction models for incident CKD have been developed in the research setting for use in the general population and specifically in people with diabetes, which were reviewed previously (Echouffo-Tcheugui and Kengne, 2012, Collins et al., 2013, Fraccaro et al., 2016, Ramspek et al., 2020, Slieker et al., 2021, van Rijn et al., 2021). The interest in risk prediction of kidney dysfunction has been increasing as indicated by the growing number of studies aiming to develop risk prediction models. The earliest review by Echouffo-Tcheugui and Kengne (2012) identified 26 publications reporting on a total of 47 models, which included 30 models for the prediction of new-

onset CKD and 17 models for CKD progression. Whereas the latest review by Slieker et al. (2021), identified a total of 64 prediction models.

It is evident that a sufficient number of risk prediction models was already developed. This highlights that one current research need is not to develop a new model, but rather to evaluate an existing model in a new population and appraise the model's transportability to different settings. Furthermore, the systematic reviews have not identified any studies that updated existing risk prediction by taking advantage of more recent analytical technologies (e.g., omics) and the integration of novel biomarkers to assess the added value of new markers to existing models. Considering these advancements may enhance the ability to capture disease risk beyond conventional clinical measures and lead to improvement in risk prediction, updating risk prediction models is unmet research need.

Slieker et al. (2021), reviewed currently available risk prediction models for CKD onset and progression for people who are also diagnosed with T2DM. The review found that most models predicted the risk within 5 years, and the most frequently used predictors were age, eGFR, and systolic blood pressure, all measured at baseline. The best apparent performance (based on results in the original model development publication) for prediction of new-onset CKD in people with diabetes was the model developed by Nelson et al. (2019), and it was also the second best in the external validation study carried out as part of the review (Slieker et al., 2021). Considering that there are no risk prediction models to predict incident CKD that have been recommended and implemented in a clinical setting, the model with the best performance stood out as the most appropriate candidate model for further research such as external model validation and update with new biomarkers (as detailed in the following Chapter 2).

Nelson et al. (2019) developed a risk prediction model for incident of CKD, defined as new-onset of eGFR below 60 ml min^{-1} (1.73 $m)^{-2}$ during the follow-up and produced an equation for 5-year absolute risk (referred to as the Nelson equation, hereafter). Nelson equation included age, black ethnicity, sex, baseline eGFR, history of CVD, HbA1c, diabetes mellitus medication use, hypertension status, smoking status, BMI and baseline albuminuria as predictors. The development of Nelson equation involved individual-level analysis of 34 multinational cohorts (any study within the consortium

with follow-up data, a minimum of 1000 participants and 50 events), which included 5,222,711 individuals from 28 countries. Eligible participants were adults (>18 years old) with eGFR of more than 60 ml min^{-1} $(1.73\ m)^{-2}$ at baseline. People with diabetes were included from 15 cohorts, comprising 15% (n= 781,627) of all participants included in this study. Additionally, the predictive performance of the Nelson equation was compared to the performance of two simpler models developed previously, referred to as O'Seaghdha equation (O'Seaghdha et al., 2012) and Chien equation (Chien et al., 2010). Nelson et al. (2019) conducted external validation using three models in an independent population to compare their predictive performance. For consistency these models were also used to meet the aims of this thesis.

Nelson equation was chosen as the primary candidate model to use in this thesis for the following reasons:

i) Developed specifically in people diagnosed with diabetes;

ii) Used a large number of participants for model development phase;

iii) Externally validated within the same consortium and outside;

iv) The model included readily available clinical variables, which enables defining the predictors and outcomes according to the definitions in terms of measures and units as provided in original publication that described the risk prediction model development;

v) The outcome of CKD onset was suitable to be studied in ET2DS (there were enough events during follow-up);

vi) The publication described the full linear predictor, where beta coefficients for predictors and estimate of baseline hazard (for at least the prediction time of interest, i.e., five years) were included. The predictor information can be used to calculate the linear predictor for each individual (LP_i, beta coefficients * predictor variable) and to generate predicted probabilities by adding baseline survival. This information is essential for external model validation as it allows a comprehensive evaluation of performance by the model using the same estimates for risk factors in a new set of individuals;

vii) Nelson equation showed top-performance when evaluated in individuals with T2DM and compared to other similar equations (Slieker et al., 2021);

viii) The model only used clinical and demographic variables, which provides an opportunity to update the model with more novel biomarkers discovered via new technologies, such as metabolomics.

1.6 Summary, project rationale and thesis structure

The increased prevalence of T2DM, which has now reached epidemic proportions (Saeedi et al., 2019), coupled with a demographic shift towards an aging population, is resulting in an increased incidence of medical complications associated with both conditions (Zheng et al., 2018). Characterized by insulin resistance, resulting in chronic hyperglycaemia, T2DM is a progressive metabolic disorder. Due to the complex interplay of genetic and environmental factors, most individuals with T2DM develop at least one common comorbidity, such as CVD, hypertension, dyslipidaemia, or CKD (Iglay et al., 2016). Detecting these conditions early allows for the modification of risk factors, which is pivotal in preventing further complications and mortality in people diagnosed with T2DM. However, the efficacy and cost-effectiveness of preventative strategies relies on how well we can identify patients at high risk and provide targeted intervention.

Considering that the mechanisms underlying the decline in kidney function are heterogeneous and complex, a single causal biomarker can only be linked to one pathway of the disease, such as fibrosis or tubular damage, capturing only a small component of the complex pathophysiological mechanisms involved in CKD development (Colhoun and Marcovecchio, 2018, Bidin et al., 2019). Consequently, the prognostic value of a single biomarker in predicting CKD onset or progression is limited. In contrast, investigating multiple biomarkers simultaneously, potentially involved in different disease pathways, enables the development of biomarker panels that better represent the intricate nature of the disease (Looker et al., 2015, Colhoun and Marcovecchio, 2018).

Some studies have used various selection techniques to develop molecular biomarker panels for successfully predicting CKD onset or renal decline people with diabetes (Looker et al., 2015, Pena et al., 2015, Mayer et al., 2017, Colombo et al., 2019a) as well as in the general population (Good et al., 2010, Argiles et al., 2013, Owens et al.,

2020), others have developed risk prediction models for CKD onset using clinical and demographic variables, as reviewed previously (Fraccaro et al., 2016, Slieker et al., 2021). A recent study attempted to improve an established risk prediction model with a metabolomic score (Zacharias et al., 2019). Here, the metabolomic score was added to the well-established four-variable Kidney Failure Risk Equation (KFRE) (Tangri et al., 2011) for the prediction of CKD progression to ESRD in the German Chronic Kidney Disease (GCKD) study (Zacharias et al., 2019). The KFRE, which includes age, sex, eGFR, and the natural log-transformed value of uACR as predictors, already yielded a notable c-statistic of 0.86 (95% CI, 0.83- 0.90) in the 4,640 CKD patients, of which 185 (3.99%) progressed to ESRD (Zacharias et al., 2019). Upon inclusion of 24 metabolomic features, the c-statistic modestly improved to 0.88 (95% CI, 0.85- 0.91), indicating that the metabolomic score achieved a 1.27% improvement in discrimination. To the best of my knowledge, no studies have integrated these two approaches, where a pre-selected metabolomic panel was incorporated into an established risk prediction model to improve the prediction of incident CKD in T2DM population.

After setting out the aims and objectives of this thesis, in the following chapter (Chapter 2), I provide a thorough description and results of a systematic overview of studies that explored the metabolomic profiles of incident CKD and new-onset kidney function decline. The remaining chapters are concerned with the results of my own risk prediction modelling of declining kidney function in people with T2DM. In this major part of thesis, I test my overarching hypothesis that metabolomics can be used to provide additional prognostic information beyond standard clinical risk factors in predicting new onset CKD in patients with T2DM.

Chapter 2. Aims and objectives

The key aims of this thesis were 1) to investigate the metabolomic profile of CKD in a cohort of older T2DM subjects, with the goal of identifying new biomarkers which could potentially be located within the underlying pathophysiological pathways leading to the development of CKD; 2) to identify a sparse set of metabolites associated with CKD in people with T2DM, in order to enhance the accuracy of predicting the risk of CKD onset in this high-risk population.

Specific objectives were to:

1) To conduct a scoping review of literature related to metabolomic profiling of people with declining kidney function in order to describe common statistical analysis methods and identify predictive/prognostic metabolites associated with incident kidney decline.

2) In an explorative cross-sectional study, to investigate the associations between individual metabolites and baseline kidney function (in terms of eGFR, uACR and CKD status) and to determine which metabolites were associated with kidney function independent of known CKD risk factors. Further, to strengthen the evidence for a possible biological and/or predictive association, to explore individual associations between the same metabolites and prospective outcomes (incident CKD and rapid decliner status), in order to identify associations which were consistent with findings from the cross-sectional study and to determine if altered metabolite levels might precede CKD onset.

3) To evaluate an existing kidney disease risk prediction model- the reference risk prediction model (Nelson et al., 2019) in the ET2DS dataset and assess if modifying this established model with a multi-marker score based on selected metabolites (the updated risk prediciton model) results in an improvement in the risk prediction of CKD onset in people with T2DM.

Chapter 3. Metabolomic signatures for the development, progression and prognosis of chronic kidney dysfunction: scoping review

In order to inform my own data analysis aimed at identifying metabolomic signatures for CKD in people with type 2 diabetes, I undertook a comprehensive search of published literature to identify previous studies using high-throughput analytical platforms to measure multiple metabolites and associate these with measures of kidney dysfunction, irrespective of study population characteristics/co-morbidities. The findings of this search, together with an assessment of the quality of the identified studies and a narrative summary of their key findings are presented in this chapter. I used this process to identify statistical approaches used previously for data analyses which were similar to my own planned analysis in the ET2DS.

3.1 Background

Previous efforts to summarize findings on metabolomic profiling for kidney disease have included a number of literature reviews, including a systematic review of metabolomic markers associated with diabetic nephropathy (Zhang et al., 2015, Saucedo et al., 2018, Kalantari and Nafar, 2019). However, these reviews are now somewhat out of date (in a fast-moving area), and, in the case of the review focused on identifying diagnostic markers of diabetic kidney disease, predominantly focused on publications with a case-control study design (Zhang et al., 2015). Considering that kidney function decline is a slow and asymptomatic, yet irreversible process, it may be of interest to gain a better understanding of biomarkers that better characterise the CKD risk early. In this context, the importance of the biomarker lies within its predictive ability, i.e., the strength of association with prospective outcomes, over and above clinical risk factors. In addition, since this is an emerging field, it is also of interest to determine if there are common methods to achieve the overarching objective of discovering predictive metabolites.

To the best of my knowledge, there are no recent scoping reviews (according to the PROSPERO register) that have comprehensively identified and summarised studies investigating metabolomic markers for incident chronic kidney disease and/or kidney function decline. To address this research need, this scoping review aims to answer the question:

i) What is the existing body of literature on high-throughput metabolomics and the association of metabolomic biomarkers with new-onset kidney decline or incident CKD?

A subsidiary question, aimed at informing my own approach to analysis in the ET2DS was:

ii) What statistical analysis methods were used to ascertain the relationships between metabolomic markers and new-onset kidney decline?

3.2 Methods

3.2.1 Data sources and searches

A comprehensive search was conducted in two databases, MEDLINE (via PubMed) and EMBASE (via Ovid). Search strategies are presented in Table 3-1. Further manual searches were performed of the references cited by studies identified in either of these two databases. My initial search was run in May 2020, prior to starting my own data analysis, and an updated search was run in February, 2024. The results of the updated search follow the original results and are presented in section 3.4 in this chapter.

Table 3-1 Key words used for search strategy on OVID Medline (1946 to May 2, 2020).

Database	Search items
EMBASE	(‘metabolome’/exp OR ‘metabolomics’/exp OR metabolom* OR ‘metabolic signature*’ OR ‘metabolic profile*’ OR lipidom*) AND (‘kidney function’/exp OR ‘renal function’/exp OR ‘chronic kidney disease’/exp OR ‘chronic kidney disease progression’/exp OR ‘kidney function decline’/exp OR ‘renal function decline’/exp OR 'Glomerular Filtration Rate'/exp OR 'end stage kidney disease'/exp OR 'end stage renal disease'/exp) LIMIT humans, English language
MEDLINE	(("metabolome"[MeSH Terms] OR "metabolomics"[MeSH Terms] OR "metabolo*"[All Fields] OR "metabolic signature*"[All Fields] OR "metabolic profile*"[All Fields] OR "lipidom*"[All Fields]) AND ((("chronic kidney disease"[All Fields] OR "renal function"[All Fields] OR "kidney function"[All Fields]) AND "Glomerular Filtration Rate"[MeSH Terms]) OR "kidney function decline"[All Fields] OR "renal function decline"[All Fields] OR "ESRD"[All Fields] OR "ESKD"[All Fields] OR "CKD"[All Fields] OR "chronic kidney disease progression"[All Fields]))

3.2.2 Eligibility Criteria

Types of Studies

Inclusion criteria (studies had to meet the criteria for all three points below):

1. Human observational studies (cohort studies or clinical trials) with a prospective cohort design, which reported association between metabolomic signatures and new-onset reduced kidney function, in either adjusted or unadjusted terms.
2. Studies that used high-throughput analytical platforms such as nuclear magnetic resonance or mass spectrometry, to identify metabolites in blood (plasma or serum) or urine samples.
3. Kidney function was estimated using an equation for glomerular filtration rate (GFR) or measured using urinary protein levels (e.g., urinary albumin to creatinine levels).

Exclusion criteria:

1. Study only analysed a single biomarker or specific selection of biomarkers such as those from a single chemical class/specific biological pathway.
2. Studies of biomarkers irrelevant to metabolomics i.e., non-metabolomic technologies or evaluated compounds were not metabolites.
3. Studies where metabolomic signatures were not analysed in relation to declining kidney function e.g., metabolomic signatures of other renal parameters (histopathological findings, kidney volume).
4. Animal studies and in vitro studies.
5. Non original studies (case reports, reviews, editorials, guidelines, conferences, abstract only).
6. Duplicate publications of the same data.
7. Not in English.

Participants/ population

Adult subjects (>18 years old), with no restriction to sex, race, or co-morbidity status.

Main outcomes (any of the following):

1. Incident CKD.
2. New-onset kidney function decline e.g., annual rate of change in kidney function absolute change from baseline to follow-up, end-stage renal disease (ESRD).
3. Incident albuminuria.

3.2.3 Study selection and coding process

The studies retrieved and imported into EndNote (version X9.3.1) were screened based on eligibility criteria. If deciding on the eligibility of an article is not possible by screening the title and abstract, the full-text article was considered. Studies were only selected if all inclusion criteria were met. Although an ideal review requires two reviewers working independently, it was not practical for me to involve a second reviewer as the review was not the main analysis of my PhD project. Nor was it realistic for me to translate studies published in languages other than English, so foreign language studies were excluded, unless there was at least an abstract published in English.

3.1.1 Data Items

Extracted data was narratively synthesised to describe both the statistical methods used and to summarise metabolites that showed statistically significant associations with prospective outcomes related to kidney function decline. The following data was extracted into an Excel spreadsheet developed specifically for the review:

i) Study characteristics: reference (first author, publication year, journal name, study name and location), study design, follow-up time, number and major characteristics of participants;
ii) Definition of main outcome;
iii) Metabolomics analysis methods (analytical platform, sample type, number of metabolites detected/measured);
iv) Statistical analysis approach (dimension reduction, multiple testing correction, modelling, adjustment for covariates);
v) Statistically significant metabolites.

3.2.4 Risk of bias

I evaluated methodological quality of studies using an eight point scale based on the relevant components of United States National Institutes of Health Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies (National Heart and Institute, 2014) (Table 3-2). Full marks for each item were awarded a score of 1, if the details were incomplete then 0.5 score was awarded and zero for not completed criteria. Studies scoring 0-3 were regarded as low quality, while scores 4-8, as high quality.

Table 3-2 Criteria for methodological quality assessment in studies that met inclusion criteria for scoping review.

Introduction: Was the research question or objective in this paper clearly stated?
Methods- population: Was the study population clearly specified and defined?
Methods- participation: Was the participation rate of eligible persons at least 50%?
Methods-outcome: Were outcome measures (dependent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?
Methods-exposure: Were the exposure measures (independent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?
Were confounding variables measured and adjusted statistically for their impact on the relationship between exposure(s) and outcome(s)?
Methods- missing data: Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.
Methods-Statistical analysis: Specify type of model, all model-building procedures (including any predictor selection) and method for internal validation.
Methodological Quality Score: 1= full marks, 0.5= incomplete, 0= not completed

3.3 Results

3.3.1 Study selection

There were 1,365 and 320 articles retrieved from EMBASE and MEDLINE, respectively (total of 1,581). All records were imported in a single library on EndNote for further management. Figure 3-1 shows the process of the literature search and study selection. After removal of duplication (n=219), 1,362 records were left for screening. These articles were screened by titles at first and 1,173 were removed because they were conference abstracts or reviews, mainly focused on animal research or drug/treatment effects in people with kidney disease and focused on other outcomes (particularly cardiovascular ones) in people with kidney disease. This left the number of articles to be further assessed at 189, and these articles were reviewed by screening the abstracts. This resulted in 154 studies being excluded (106 did not include longitudinal analysis, 26 did not apply high-throughput techniques to profile metabolites or focused on a pre-selected panel of metabolites, 22 were literature reviews or editorial letters and two focused on children). After the full texts of the remaining 35 articles were examined, 26 articles remained (five excluded because the study was designed using a hypothesis driven approach in regards to metabolites selected/analysed, one did not include prospective analysis, one used multi-omics

platform (an approach which integrated datasets obtained from multiple omics / high-throughput technologies) and one was a methodological study of alternative sample medium). Hand-searching the references did not reveal any additional literature.

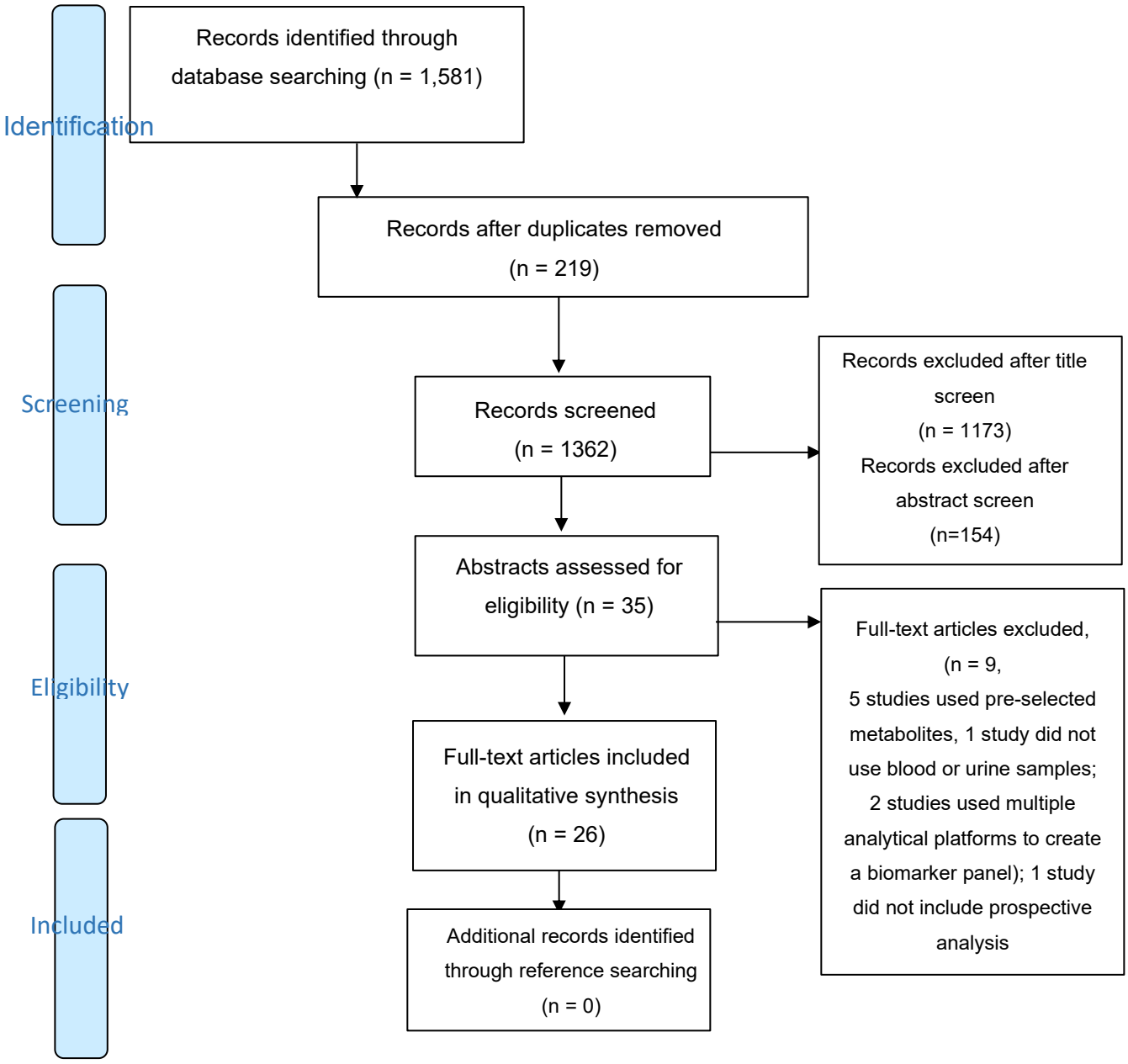


Figure 3-1 Flow Diagram of search results.

3.3.2 Characteristics of included studies

Detailed characteristics of the 26 identified papers were tabulated (Table 3-3) and the quality assessment of included studies is shown in Table 3-4. None of the studies were marked as low quality according to the eight-point scale. The lowest quality paper scored 5 points, because it did not provide information regarding participant attrition in cohorts studied, failed to provide details on how the missing data was handled and did not sufficiently report on the longitudinal analysis.

The earliest publication was released in 2013 (Goek et al., 2013). There was an average of three relevant articles released per year, but generally the number of publications increased over time and there were six publications in 2018 and 2019. The populations studied were from 14 countries (including the replication cohorts), with the predominant countries being the United States, which included studies of Native American populations (Rhee et al., 2013, Niewczas et al., 2014, Yu et al., 2014, Afshinnia et al., 2016, Rhee et al., 2016, McMahon et al., 2017b, Niewczas et al., 2017, Xu et al., 2018, Afshinnia et al., 2019, Afshinnia et al., 2020), Germany (Goek et al., 2013, Sekula et al., 2016, Zacharias et al., 2019) and Denmark (Pena et al., 2014, Tofte et al., 2019a). The remaining studies were mainly European cohorts, including those from UK (Sekula et al., 2016, Colombo et al., 2019b), Italy (Solini et al., 2016, Tavares et al., 2018), France (Nkuipou-Kenfack et al., 2014) and Belgium (Gil et al., 2018). Interestingly, one study used a meta-analysis approach, where multiple cohorts from four European countries (UK, Spain, Germany and Finland) were analysed together (Barrios et al., 2018).

Some publications were based on the same cohort population, but either applied different metabolomic analysis platform or analysed the metabolomic markers in terms of a different outcome. Three of the selected publications used Chronic Renal Insufficiency Cohort (CRIC) (Afshinnia et al., 2016, Rhee et al., 2016, Afshinnia et al., 2020), three used Cooperative Health Research in the Region of Augsburg (KORA) cohort (Goek et al., 2013, Sekula et al., 2016, Barrios et al., 2018), two used Twins UK cohort (Sekula et al., 2016, Barrios et al., 2018), Arthrosclerosis Sisk in Community Study (ARIC) (Yu et al., 2014, McMahon et al., 2017b) and Joslin Kidney study cohort (Niewczas et al., 2014, Niewczas et al., 2017).

In terms of cohort size, the number of participants ranged from 49 to 4,640 (Nkuipou-Kenfack et al., 2014, Zacharias et al., 2019). More than half of the selected publications included between 100 and 1000 participants (Afshinnia et al., 2016, Kimura et al., 2016a, Kimura et al., 2016b, Rhee et al., 2016, Sekula et al., 2016, Solini et al., 2016, Niewczas et al., 2017, Gil et al., 2018, Xu et al., 2018, Titan et al., 2019, Tofte et al., 2019a, Afshinnia et al., 2020, Colombo et al., 2019b). There were seven studies that included more than a thousand participants (Goek et al., 2013, Yu et al., 2014, Rhee et al., 2016, McMahon et al., 2017b, Barrios et al., 2018, Wang et al., 2018, Zacharias et al., 2019). Most studies included approximately even proportions of both sexes (males ~50%) and involved participants with a non-specific mixture of comorbidities such as hypertension, cardiovascular disease and diabetes and one study was conducted in sub-populations of both T2DM and non-T2DM population (Barrios et al., 2018). Furthermore, four publications focused specifically on the type 1 diabetes group of patients (Pena et al., 2014, Colombo et al., 2019b, Tofte et al., 2019a), and five publications focused on people with T2DM (Niewczas et al., 2014, Pena et al., 2014, Solini et al., 2016, Niewczas et al., 2017, Tavares et al., 2018). Lastly, one study specifically studied the risk of kidney function decline in a group of individuals diagnosed with sickle cell disease (a known risk factor for ESRD) (Xu et al., 2018).

Since the outcome of interest was the new-onset of kidney function decline, by the nature of the study selection criteria all of the studies included were prospective cohorts, in which the biomarker levels were measured before the onset of the outcome. Notably, some studies carried out analysis to compare a group of people who developed the outcome with an equivalently-sized age and sex matched group of people who did not develop the outcome, from the same cohort (Niewczas et al., 2014, Pena et al., 2014, Afshinnia et al., 2016, Rhee et al., 2016, McMahon et al., 2017a, Gil et al., 2018). The follow-up period ranged from 2.4 years to 19.6 years (Yu et al., 2014, Gil et al., 2018). The most common renal endpoint studied was the CKD progression to ESRD or requirement of dialysis, which was the focus of 11 articles included in the review (Niewczas et al., 2014, Afshinnia et al., 2016, Kimura et al., 2016a, Kimura et al., 2016b, Niewczas et al., 2017, Tavares et al., 2018, Titan et al., 2019, Tofte et al., 2019a, Zacharias et al., 2019, Afshinnia et al., 2020). The second most common outcome was incident CKD in terms of eGFR drooping below 60 ml

min^{-1} (1.73 m)$^{-2}$, which was investigated in seven studies (Goek et al., 2013, Rhee et al., 2013, Yu et al., 2014, Sekula et al., 2016, Solini et al., 2016, McMahon et al., 2017b, Wang et al., 2018). In addition, six studies investigated the metabolomic marker associations with eGFR decline which was defined either in terms of eGFR slope (commonly having a binary threshold for eGFR decline of >3 mL/min/1.73 m^2) or an absolute change in eGFR from baseline to follow-up (Goek et al., 2013, Nkuipou-Kenfack et al., 2014, Rhee et al., 2016, Barrios et al., 2018, Xu et al., 2018, Gil et al., 2018, Afshinnia et al., 2019, Colombo et al., 2019b). Finally, only one study investigated the progression of albuminuria (Pena et al., 2014) in a study of a T2DM cohort, were a progression from normal albumin levels to mildly increased level and a separate comparison in those with mild albuminuria to modestly increased albuminuria.

Table 3-3 Characteristics of the cohort studies associated with incident kidney function decline or new-onset kidney function decline.

First author, year, journal	Study name and country	Study design	Follow up (years)	Number of participants	Outcomes (definition and proportion of cases in cohort)	Baseline characteristics (mean age (years), proportions in %)
Goek et al. (2013), Nephrology Dialysis Transplant	KORA Survey 4/Follow-up4, Germany	Prospective, general population	7.1	1104	Incident CKD (eGFR <60 mL/min/1.73 m^2) n=106 (10%); eGFR decline >3 mL/min/1.73 m^2	Age 63.3 years Male 50.9% DM 6.2%; HTN 34%
Rhee et al. (2013), Journal of the American Society of Nephrology	Framingham Offspring Study, US	Prospective, general population	8	1434	Incident CKD (eGFR <60 mL/min/1.73 m^2 at FU) n=123 (9%)	Age 61 years Male 51% HTN 62%; DM 17%
Yu et al. (2014), Clinical Journal of the American Society of Nephrology	ARIC, US	Prospective, general population	19.6	1921	Incident CKD (eGFR <60 mL/min/1.73 m^2) n=204 (11%)	Age 54.4 Male 86% CHD 13%; DM 77%
Niewczas et al. (2014), Kidney International	Joslin Study, US	Prospective, select group comparison	8-12	80	CKD progression to ESRD n= 40 (50% cases)	Male 55% T2DM 100%
Nkuipou-Kenfack et al. (2014), PLoS One	Hospital patients, France	Prospective, select group comparison	2.86	49; Train: n=20; Test n=29	GFR slope (as continuous variable)	Age 66 years Male 70%
Pena et al (2014), Diabetic medicine	PREVEND, Netherlands; Steno Diabetes Center Copenhagen, Denmark	Prospective, T2DM cohort, select group comparison	2.9	48 and 42	Normo-albuminuria to micro-albuminuria (n=24); micro-albuminuria to macro-albuminuria (n=21	Age 65 years Male 17% T2DM 100%

Afshinnia et al. (2016), Kidney International	CRIC study, US	Prospective, select group comparison	6	200	CKD progression to ESRD or 25% decline in eGFR, n=79 or 39.65 % progressors , n=121 or 60.5% non-progressors)	Age 59 years Male 68% DM 47%; HTN 85%
Kimura et al. (2016a), Scientific Reports	Patients from Rinku General Medical Centre, Japan	Prospective	4.3	108	ESRD defined as kidney replacement therapy (n=58, 53.5%) or death (15, 13.9%)	Age 65 years Male 75% DM 30.6%
Kimura et al. (2016b), Scientific Reports	Patients from Rinku General Medical Centre, Japan	Prospective	4.1	112	ESRD defined as kidney replacement therapy (n=61, 54.5%) or death (17, 15.2%)	Age 65 years M%75 DM 31%
Rhee et al. (2016), American Journal of Nephrology	CRIC study, US	Prospective, select group comparison	2.84	400	eGFR decline < −3 ml/min/1.73 m2/yr (n=200, 50%)	Age 60 years Male 57% HTN 95%; DM 55%; CVD 47%
Sekula et al. (2016), Journal of the American Society of Nephrology	KORA Survey4/F4, Germany; Replication: TwinsUK, UK; AASK, US	Prospective	7.1	KORA n=991; TwinsUK n=1164; AASK n=188	Incident CKD KORA n=95, 10%; TwinsUK n=36; AASK n=188	KORA: Age 63 years Male 71% DM 0.6%
Solini et al. (2016), The Journal of Clinical Endocrinology & Metabolism	Academic T2DM outpatient clinics, Italy	Prospective	3	286	Incident CKD, eGFR < 60 mL/min/1.73 m^2, (n=45, 16%)	Age 68 years Male 56% T2DM 100%
McMahon et al. (2017b), Kidney International	Framingham Heart Study, United States; ARIC, US	Prospective, select group comparison	9.7	Framingham Heart Study n=386; ARIC n= 998 Tot. n= 1384	Incident CKD (eGFR <60 ml/min per 1.73 m^2) (n=193, 14%)	Age 63 years Male 51% HTN 47%; DM 10%

Niewczas et al. 2017, Diabetes Care	Joslin Kidney Study, US	Prospective, select T1DM patients with CKD at BL	11	158	eGFR slope (as continuous variable); ESRD (n=99, 63%)	Age 45 years Male 25%
Barrios et al. 2018, Scientific reports	GenodiabMar (n = 655), Spain; TwinsUK, UK (n = 1279, 111 with T2D); KORA, Denmark (n = 1784, 160 with T2D); Young Finns, Finland (n = 2046)	Prospective, meta-analysis		Subset of 3644 included in longitudinal analysis	eGFR slope (as continuous variable, average no stated)	Age 41-70 years Male 4.3- 60.9% T2DM 25%
Gil et al. (2018), Nephrology Dialysis Transplant	Nephrology Outpatient Clinic, Belgium	Prospective, select group comparison	2.4	114	eGFR change of −8% to −15% per year (n=57, 50%)	Age 62 years Male 60% SAH 96%
Tavares et al (2018), Metabolomics	Nephrology Outpatients with T2DM, Italy	Prspective, randomized control trial	2.5	56	ESRD defined as doubling of serum creatinine or death (n=17, 30.3%)	Age 58 years Male 33% T2DM 100%
Wang et al. (2018), Journal of the American Society of Nephrology	Community, China	Prospective	6	1765	Incident CKD (eGFR<60 ml/min per 1.73 m^2) (n=274, 16%)	Male 42% CVD 9%; HTN 50%; T2DM 10%
Xu et al. (2018), Am Journal of Hematology & Oncology	Outpatients with sickle cell disease, US	Prospective	5	193	eGFR decline of ≥3 mL/min/year), n=71 (36.8%)	Age 43 years Male 47%
Afshinnia et al (2019), JCI Insight.	Renoprotection in early diabetic nephropathy in Pima Indians trial, US	Prospective, randomized control trial	9.6	92	eGFR decline of 40% (n=32, 34.8%)	Age 43 years Male 16% HTN 40.6%

Colombo et al (2019), Diabetologia	SDRNT1BIO, UK; Replication- FinnDIane, Finland	Prospective	5.2 and 8.8	859 and 315	eGFR decline ≥3 mL/min/year (22.6% and 40.3%)	Age 55 and 46 years Male 44% and 66 T1DM 100%
Titan et al (2019), PLoS One	Progredir Cohort Study, Brazil	Prospective	3	454	ESRD or death, n=129 (28.4%)	Age 67 years Male 63% HTN 90%; DM 57%
Tofte et al (2019a), Scientific Reports	Steno Diabetes Center, Denmark	Prospective	5.8; 5.3; 6.2	669	normo- to microalbuminuria, or micro- to macroalbuminuria (n=37); eGFR ≥30% decrease (n=93); new-onset ESRD (n=21); death (n=58) Total: combined renal endpoint (any of the above) n=125	Age 55 years Male 55% T1DM 100%
Tofte et al. (2019a), Frontiers in Endocrinology	Steno Diabetes Center, Denmark	Prospective	5.2; 5.2; 6.2	669	eGFR decline from baseline (≥30%), n=91, progression to end-stage renal disease n=21 and all-cause mortality n=58; Total: combined renal endpoint (any of the above) n=123	T1DM 100%
Zacharias et al. (2019), Journal of Proteome	GCKD study, Germany	Prospective	3.7	4640	Incident ESRD, n=185 (3.99%)	Age 60 years Male 60%
Afshinnia et al. (2020), Nephrology Dialysis Transplant	CRIC, United States	Prospective	10	200	Incident ESRD, n=123 (61.5%)	Age 58 years Male 70 % DM 54%; HTN 94%

Abbreviations: AASK, African American Study of Kidney Disease and Hypertension; ARIC, atherosclerosis risk in communities; BL, baseline; CKD, chronic kidney disease; CRIC, Chronic Renal Insufficiency Cohort; CVD, cardiovascular disease; eGFR, estimated glomerular filtration rate; FinnDIane, Finnish Diabetic Nephropathy study; GCKD, German Chronic Kidney Disease; HTN, hypertension; KORA, Cooperative Health Research in the Region of Augsburg; M%, Percentage of males; PREVEND, Prevention of REnal and Vascular End-stage Disease; SDRNT1BIO, Scottish Diabetes Research Network Type 1 Bioresource; T1DM, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus.

Table 3-4 Quality assessment of included studies.

Study reference	Aims	Population	Attrition	Outcome	Exposure	Covariates	Missing data	Statistical methods	Total
Goek, et al. (2013)	1	1	0	1	1	1	1	1	7
Rhee, et al. (2013)	1	1	1	1	1	1	1	1	8
Yu, et al. (2014)	1	1	1	1	1	1	1	1	8
Nkuipou, et al. (2014)	1	1	0	1	1	0	1	1	6
Niewczas, et al. (2014)	1	1	1	1	1	1	1	1	8
Pena et al. (2014)	1	1	0	1	1	1	1	1	7
Afshinnia et al. (2016)	1	1	1	0.5	1	1	1	1	7.5
Kimura et al. (2016)	1	1	1	1	1	1	0	1	7
Kimura et al. (2016)	1	1	1	1	1	1	0	1	7
Rhee et al. (2016)	0.5	1	1	0.5	1	1	0	1	6
Solini,et al. (2016).	1	1	0	1	0	1	1	1	6
Sekula, et al. (2016).	1	1	1	1	1	1	1	1	8
McMahon, et al. (2017).	1	1	1	1	1	1	1	1	8
Niewczas et al. (2017)	1	1	1	1	1	1	0	1	7
Barios et al. (2018)	1	1	0	1	1	1	0	0	5
Gil, et al. (2018).	1	1	0	1	1	1	1	1	7
Tavares et al (2018)	1	1	1	1	1	0	0	1	6
Wang, F. J., et al. (2018).	1	1	1	1	1	1	1	1	8
Xu et al. (2018)	1	1	1	1	1	0	1	1	7
Afshinnia et al. (2019)	1	1	1	1	0.5	1	0	1	6.5
Colombo et al. (2019)	1	1	1	1	1	1	1	1	8
Titan et al. (2019)	1	1	1	1	1	1	1	1	8
Tofte, N., et al. (2019)	1	1	1	1	1	1	1	1	8
Tofte, et al. (2019)	1	1	1	1	1	1	1	1	8
Zacharias et al. (2019),	1	1	1	1	1	1	1	1	8
Afshinnia et al. (2020)	1	1	1	1	1	1	0	1	7

3.3.3 Methods used for metabolomic measurements

Metabolomic profiling analytical methods used in all included studies were summarised (Table 3-5). Metabolomic profiles were investigated using either blood serum or plasma in most studies, but some studies analysed urine samples (Nkuipou-Kenfack et al., 2014, Pena et al., 2014, Solini et al., 2016, McMahon et al., 2017a, Gil et al., 2018). The majority of studies (22 of 26) used mass spectrometry for analytics, while only three studies used nuclear magnetic resonance approach and one study used high-performance liquid chromatography to measure chiral amino acids (Kimura et al., 2016a). In terms of metabolomic profiling, half of the studies used a targeted method to measure and quantify know metabolites. The number of metabolites measured ranged from 16 amino acids to 406 lipids (Kimura et al., 2016a, Afshinnia et al., 2019). Targeted metabolites mainly included lipoprotein subclass, amino acids, and uremic solutes. The most commonly used kit for measurement and quantification of metabolites in targeted analysis was Absolute IDQTM kit p180 (Goek et al., 2013, Nkuipou-Kenfack et al., 2014). An untargeted profiling approach was used in the remaining half of included studies and identified up to 10,940 metabolite signals (Titan et al., 2019) and up to 510 known metabolite features analysed (Afshinnia et al., 2016). To identify known metabolites from the signals measured, a few studies used a Metabolon platform (Niewczas et al., 2014, Yu et al., 2014, Niewczas et al., 2017).

Table 3-5 Metabolomic techniques used in included studies.

Article reference	Assay	Sample type	Metabolomic profiling
Goek et al. (2013)	MS	S	Targeted (Absolute IDQTM kit p180) measurement of 140 metabolites and their 19 460 ratios; metabolite classes were mainly amino acids, lipids, energy related, vitamins, peptides.
Rhee et al. (2013)	MS	P	Targeted- 217 metabolites measured, 54 positively charged polar analytes, 59 negatively charged polar analytes (organic acids, sugars, bile acids) and 104 lipids.
Yu et al. (2014)	MS	S	Non-targeted (Metabolon), 204 serum metabolites quantified. 187 metabolites had ≥50% above the detection limit, and included amino acids, lipids and xenobiotics.
Niewczas et al. (2014)	MS	P	Non-targeted (Metabolon) detection of mass spectra of 123 metabolites global profiling including uremic solutes, amino acids and their derivatives
Nkuipou-Kenfack et al. (2014)	MS	UP	Targeted (Absolute IDQTM kit p180) measured and quantified amino acids, acylcarnitines, sphingomyelins, phosphatidylcholines, hexose (glucose), and biogenic amines
Pena et al. (2014)	MS	UP	Non-targeted, 265 plasma and 231 urine metabolites were quantified. Acylcarnitines, glycerophospholipids, sphingomyelins, amino acids, hexoses and biogenic amines were analysed in plasma and urine, whereas bile acids, eicosanoids and energy metabolism were measured in plasma.
Afshinnia et al. (2016)	MS	S	Non-targeted, 510 known lipids measured.
Kimura et al. (2016a)	HPLC	P	Targeted. 16 chiral amino acids (D-Amino acids, the enantiomers of L-amino acids)
Kimura et al. (2016b)	MS	P	218 metabolites, glycolysis, amino acids, amino sugar and glucoronate, biotin, glutathione, and taurine.
Rhee et al. (2016)	MS	P	160 metabolites measured. Positive probe- Amino acids, amino acid metabolites, acylcarnitines, dipeptides, nucleotides, and other cationic polar metabolites. Negative- Sugars, sugar phosphates, organic acids, bile acids, nucleotides and other anionic polar metabolites.
Sekula et al. (2016)	MS	S	Non-targeted; 321 identified known metabolites including 81 amino acids and related compounds, 14 carbohydrates, 15 cofactors and vitamins, 6 metabolites related to energy metabolism, 130 lipid metabolism derivatives, 14 purine and pyrimidine bases, 27 peptides, and 34 xenobiotics.
Solini et al. (2016)	MS	US	Non-targeted quantification of amino acids, lipids, carbohydrates, xenobiotics and energy related metabolites

McMahon et al. (2017b)	MS	U	Targeted, analysis of metabolites related to genetic variants associated with incident CKD, mainly amino acids.
Niewczas et al. (2017)	MS	S	Non-targeted (Metabolon) detected 110 amino acids and purine and pyrimidine metabolites
Barrios et al. (2018)	NMR	P	Targeted, 227 metabolites and ratios belonging to lipids, amino acids, ketone bodies, inflammation, fluid balance and energy-related metabolites.
Gil et al. (2018)	NMR	U	Non-targeted metabolites from the tricarboxylic acid cycle, amino acids (e.g., threonine), lipid metabolism (e.g., ethanolamine), gut microbiome–derived uraemic toxins (e.g., indoxyl sulphate and p-cresol sulphate), uracil and glycolic acid
Tavares et al. (2018)	MS	S	Non-targeted analysis detected and quantified 186 known metabolites including energy-related metabolites and amino acids
Wang et al. (2018)	MS	P	Targeted: Quantified 22 amino acids and 34 acylcarnitines
Xu et al. (2018)	MS	P	Targeted and non-targeted- 179 known measured metabolites related to genetic variant in APOL1 gene.
Afshinnia et al (2019)	MS	S	Targeted. Measured 406 lipids from 18 classes, including triacylglycerols, diacylglycerols, monoacylglycerols, phosphatidylcholines, cholesteryl-esters, cardiolipins, phosphatidic acids, phosphatidylinositols, phosphatidylglycerols, phosphatidylserines, sphingomyelins.
Colombo et al (2019)	MS	S	Targeted. 297 circulating biomarkers (30 proteins, 121 metabolites, 146 tryptic peptides)
Titan et al (2019)	MS	S	Nontargeted. 10940 metabolites detected; 293 metabolites were analysed including amino acids and energy related metabolites
Tofte et al. (2019a)	MS	S	Non-targeted. 106 known lipids from 5 classes (diacyl-phosphatidylcholines, alkyl-acyl-phosphatidylcholines, lyso-phosphatidylcholines, triacylglycerols, and sphingomyelins were identified and analysed
Tofte et al. (2019a)	MS	S	Non-targeted. 75 metabolites were identified and included in data analyses, which consisted of amino acids, free fatty acids, compounds from the energy metabolism pathways and polyols
Zacharias et al. (2019)	NMR	P	Non-targeted analysis identified 127 known metabolites including amino acids and energy-related metabolites.
Afshinnia et al. (2020)	MS	S	Targeted. 85 metabolites were analysed which consisted of products of arachidonic acid metabolites.

HPLC, high-performance liquid chromatograph; MS, mass-spectrometry; NMR, nuclear magnetic resonance

3.3.4 Statistical approaches for metabolomic association analysis

Statistical methods used in the included studies are described in Table 3-6. As the initial step, many studies opted to use traditional statistical methods in order to assess the associations between each metabolite measured and kidney function at baseline. There were nine studies that used linear regression models to investigate the associations between individual metabolites and markers of kidney function (Goek et al., 2013, Niewczas et al., 2014, Yu et al., 2014, Sekula et al., 2016, Barrios et al., 2018, Wang et al., 2018, Xu et al., 2018, Colombo et al., 2019b, Tofte et al., 2019a), while one study considered reduced kidney function status as a dichotomous variable and used a logistic regression model accordingly (Afshinnia et al., 2016). There was also three studies that used correlation analysis to evaluate the associations between each metabolite and kidney function (Kimura et al., 2016a, Kimura et al., 2016b, Niewczas et al., 2017) and three used a t-test (Rhee et al., 2013, Pena et al., 2014, Tavares et al., 2018).

Alternative approach using dimensional-reduction methods were used in eight studies (Nkuipou-Kenfack et al., 2014, Afshinnia et al., 2016, Solini et al., 2016, McMahon et al., 2017b, Gil et al., 2018, Afshinnia et al., 2019, Tofte et al., 2019a, Zacharias et al., 2019). Here, principal component analysis, partial least analysis or the least absolute shrinkage and selection operator (LASSO) methods were used to select a small group of informative metabolites. In addition, three of these studies combined the significant metabolites into a score and investigated the predictive performance either alone or in addition to clinical risk factors (Nkuipou-Kenfack et al., 2014, Solini et al., 2016, Zacharias et al., 2019).

As the final stage of analysis, majority of studies opted to use traditional modelling techniques where the statistically significant metabolites were analysed using either a logistic or cox regression analysis for their associations with prospective outcomes. Logistic regression was a slightly more popular approach with 11 studies using it as the final step of their analysis (Goek et al., 2013, Rhee et al., 2013, Niewczas et al., 2014, Pena et al., 2014, Afshinnia et al., 2016, Rhee et al., 2016, Solini et al., 2016, Sekula et al., 2016, Niewczas et al., 2017, Wang et al., 2018, Afshinnia et al., 2020). The second most popular analysis approach was using Cox proportional hazards method, with eight studies opting to use it (Yu et al., 2014, Kimura et al., 2016a, Kimura et al., 2016b, Tavares et al., 2018, Afshinnia et al., 2019, Titan et al., 2019, Tofte et

al., 2019a, Tofte et al., 2019b, Zacharias et al., 2019). In these studies, metabolites that remained significant after adjustment for covariates were considered as potentially associated with development or progression of kidney function decline.

All studies adjusted for traditional risk factors in regression models for incident outcomes (including age, sex, systolic blood pressure, smoking, body mass index, etc.). Among studies using traditional statistical methods to select most relevant metabolites, nine studies applied false discovery rates (FDR) procedures to take account of the issue of multiple comparisons (Niewczas et al., 2014, Pena et al., 2014, Kimura et al., 2016a, Kimura et al., 2016b, Niewczas et al., 2017, Barrios et al., 2018, Titan et al., 2019, Tofte et al., 2019a, Afshinnia et al., 2020). There were also seven studies that used a more stringent Bonferroni correction method to address the same issue (Goek et al., 2013, Rhee et al., 2013, Yu et al., 2014, Rhee et al., 2016, Sekula et al., 2016, Wang et al., 2018, Colombo et al., 2019b).

Table 3-6 Statistical analysis and results for metabolites associated with incident kidney function decline and progression.

Reference	Dimension reduction	Multiple comparison	Statistical analysis	Covariates	Statistically significant metabolites and adjusted effect sizes
Goek et al. (2012a)	Metabolites measured 186 → quantified 140 and 19460 ratios → discovery screen- LM→LR (incident CKD; eGFR decline)	Bonferroni i)p < 0.00036 ii)p<2.6×10^{-6}	LR	age, SBP, current smoking, serum glucose, eGFR antihypertensives,	Spermidine (OR 1.15 95% CI: 0.91–1.44) Kynurenine/Tryptophan (1.36 95% CI: 1.11–1.66), Phosphatidylcholine diacyl /acylalkyl (0.72 95% CI: 0.56–0.93)
Rhee et al. (2013)	measured 217 metabolites → two-tailed t tests → LR (incident CKD)	Bonferroni, P< 0.00023	LR	eGFR, age, sex, HTN, DM, BL proteinuria	Kynurenic acid OR 1.53 (95%CI 1.25, 1.88); Kynurenine 1.49 (1.22, 1.83); Citrulline1.48 (1.19, 1.83); Choline 1.46 (1.17, 1.82); Xanthosine 1.46 (1.21, 1.76); b-aminoisobutyric acid 1.41 (1.15, 1.72); Aconitate 1.32 (1.07, 1.62); Isocitrate 1.28 (1.05,1.58); 5-hydroxyindoleacetic acid, 0.62 (0.51, 0.76)
Yu et al. (2014)	quantified n=602→ 118 assigned ID→linear model-associations with eGFR at baseline → Cox (incident CKD)	Bonferroni P<0.001	Cox	age, sex, SBP, HDL, LDL, antihypertensive use, DM, CHD, smoking	n=40 named and n=34 unnamed metabolites significant in cross-sectional analysis. 5-oxoproline HR 0.70; (0.60 to 0.82) 1,5-anhydroglucitol HR0.68; (0.58 to 0.80)
Niewczas et al. (2014)	Measured, 262 metabolites→ 119 were stable → general linear model-analysis of fold difference →LR (ESRD)	FDR Benjamini-Hochberg, q-value <0.05	LR	UAER, eGFR, HbA1c	P-cresol OR 1.7 (95% CI 1.0, 2.8); Phenylacetylglutamine 1.7 (1.0, 2.9); pseudouridine 2.8 (1.5, 5.5); Indoleacetate 2.4 (1.3, 4.3); Urate 2.5 (1.3, 4.8)
Nkuipou-Kenfack et al. (2014)	Wilcoxon rank sum → MosaCluster used to build a classifier- vector machine → significant metabolites markers combined into a score	FDR Benjamini-Hochberg, q-value <0.05	SVM	-	Multi-marker panel- 17 plasma metabolites r= 20.6009, p = 0.0019 13 urinary metabolites r = 20.6574, p = 0.0005

Pena et al. (2014)	Metabolites below the detection limit in >70% of participants excluded → 265 plasma and 231 urine metabolites analysed with pairwise t-test→ LR to predict albuminuria status	FDR Benjamini-Hochberg, q-value <0.05	LR	UAER, and eGFR	plasma histidine (fold change=0.87,P=0.02) butenoylcarnitine (1.17,P=0.007); urine, hexose (0.2,P<0.001), glutamine (0.32,P<0.001), tyrosine (0.51,P=0.006)
Afshinnia et al. (2016)	compound-by-compound t-test→ FDR → the top lipids with nominal significance were analysed with three methods and compared: PLS-DA, RF and LR to predict ESRD	FDR Benjamini-Hochberg, q-value <0.05	LR	eGFR, UACR, age, sex, race, diabetes, HTN, CHD	diacylglycerols (OR 0.71, 95% CI: 0.43 – 0.85, P < 0.001) and 0.66 (95% CI: 0.33 to 0.83, P = 0.002),monoacylglycerol OR 5.45 (95% CI: 2.51 - 11.86, P < 0.001)
Kimura et al. (2016a)	Independent associations between metabolites and eGFR- Spearman rank regression analysis → Cox to predict ESRD	FDR Holm	Cox	eGFR, UAER, DM, age, sex, hemoglobin level, mean blood pressure, historical CVD, antihypertensive use.	D-Asn 3.07(1.30–7.26), P<0.05
Kimura et al. (2016b)	Independent associations between metabolites and eGFR- Spearman rank regression analysis → Cox to predict ESRD	FDR Holm	Cox	eGFR, UCPCR, DM, age, sex, calcium phosphate, blood pressure, historical CVD, haemoglobin	Isethionate 2.92 (1.76–4.84); Saccharate 2.60 (1.65–4.10); Trimethylamine N-oxide 2.30 (1.54–3.43); Gluconate 2.50 (1.41–4.44)
Rhee et al. (2016)	Metabolites associations (prospective and cross-sectional outcome, LR)→ significant in both, sum of coefficients (composite score) → LR (CKD decliner)	Bonferroni, P< 0.0003	LR	age, sex, race/ethnicity, HTN, SBP, DBP, diabetes, eGFR and eGFR, and proteinuria	None significantly related to prospective outcomes
Solini et al. (2016)	Random forest analysis selected top 30 metabolites associated with a baseline CKD	-	LR	female sex, age, fasting glucose, and baseline eGFR	MetIndex: C-glycosyl tryptophan, pseudouridine, and N-acetylthreonine combined: OR 5.48 (95% CI, 2.23–14.47)

	→ Multivariate LR (incident CKD) → multi-marker panel				
Sekula et al. (2016)	Quantified n=493, named n= 321→ 422 metabolites and 87,762 ratios; correlating with eGFR (r>0.5) → multivariate linear regression → LR to predict CKD onset	Bonferroni P<0.0001; for ratios P<0.0000005	LR	age, SBP, antihypertensive use, smoking, and HDL, eGFR	C-mannosyltryptophan, pseudouridine, and O-sulfo-L-tyrosine
McMahon et al. (2017a)	Measured 154 metabolites → quantified 151→ PCA (24 clusters) → LR between key metabolite in each cluster and incident CKD	Bonferroni P<0.0021	LR	BL eGFR, DM HTN, dipstick proteinuria	Glycine OR 0.59 (95% CI 0.43–0.80) Histidine OR 0.65, 95% CI 0.5–0.85)
Niewczas et al. (2017)	Metabolites detected in >80% participants→ spearman's rank correlation with eGFR slope→ Cox regression for ESRD	FDR Benjamini-Hochberg, <0.05	Cox	blood pressure, BMI, smoking status, HbA1c, ACR, eGFR, uric acid levels, treatment with renin-angiotensin system inhibitors, other antihypertensive treatment, and statins.	C-glycosyltryptophan, pseudouridine, O-sulfotyrosine, N-acetylthreonine, N-acetylserine, N6-carbamoylthreonyladenosine, N6-acetyllysine
Barrios et al. (2018)	Independent associations between metabolites and eGFR slope using LM	FDR Benjamini-Hochberg, <0.05	LM	Sex and baseline eGFR, age, and BMI	None
Gil et al. (2018)	Orthogonal PLS→ assigned metabolite ID→ quantified→ LM (baseline eGFR or eGFR slope)	None	LM	Age, sex, BMI and CKD disease group	Betaine beta -0.153 (SE 0.057 P=0.01) and myo-inositol beta -0.091 (SE 0.042 p=0.035)

Tavares et al. (2018)	Differences in metabolite peaks (Mann–Whitney)→ univariate and multivariate Cox (progression to ESRD)	None	Cox	eGFR, UACR, HbA1c	5-anhydroglucitol HR 0.1 (95%CI 0.01, 0.63), norvaline HR 0.004 (0.0001, 0.14) and L-aspartic acid 0.14 (0.03, 0.78)
Wang et al. (2018)	Metabolites measured n=56→ i) cross-sectional - multivariable LM→ Longitudinal associations with annual eGFR change n=12	Bonferroni Cross-sectional *P*=8.9E-04 Prospective P=3.8E−03	LR	Age, sex, region, education, smoking, drinking, physical activity, BMI, lipid-lowering meds, HDL, LDL, CVD, HTN, T2DM, BL eGFR	cysteine, long-chain acylcarnitines (C14:1OH, C18, C18:2, and C20:4), and other acylcarnitines (C3DC and C10), relative risks ranged from 1.16 to 1.25 per SD increment of metabolites; P<3.8E-03
Xu et al. (2018)	LM for association between baseline eGFR and each known metabolite and LR for eGFR decliner status→ Permutation resampling 100,000 times	None; P values <0.05		eGFR	asymmetric dimethylarginine and quinolinic acid
Afshinnia et al. (2019)	PCA to generate secondary variables representative of lipids→LM between outcome and PC components→ Cox to predict 40% eGFR decline	None	Cox	eGFR, UACR, fasting plasma glucose, HbA1c	Unsaturated free fatty acids, phosphatidylethanolamines, interaction terms of C16–C20 ACs and short-low-double-bond triacylglycerols
Colombo et al. (2019)	LM for cross-sectional analysis (eGFR) logistic regression models with adjustment for covariates to predict eGFR decliner status	Bonferroni, P<0.00026	LR	age, sex, DM duration, eGFR, follow-up time, BMI, SBP, DBP, HbA1c, HDL- cholesterol, smoking, UACR and historical eGFR	30 significant metabolites. Largest contributors in predicting eGFR decline: CD27 antigen (β -0.31 (95% CI-0.36 -0.26); kidney injury molecule 1(-0.26 (-0.31, -0.21)

Titan et al. (2019)	Independent Cox regression models→ significant metabolites re-analysed with adjustment to predict ESRD	FDR Benjamini-Hochberg, p <0.05	Cox	age, sex, DM, SBP, and eGFR	Lactose (HR 1.49, 95%CI 1.04–2.12), 2-O-glycerol-α-D-galactopyranoside (HR 1.76, 95%CI 1.06–2.92), tyrosine (HR 0.52, 95%CI 0.31–0.88)
Tofte et al. (2019b)	Metabolites measured n=75 → LASSO and extended Bayesian information criterion → i) cross-sectional linear regression (14 metabolites) ii) longitudinal - Cox to predict ESRD	FDR Benjamini-Hochberg, q-value <0.05	Cox	age, sex, HbA1c, SBP, smoking, statin treatment, BMI, TG, total cholesterol, eGFR, and UAER	Ribonic acid , myo-inositol HR 2.2–2.7, 95% CI [1.3–4.3], p < 0.001
Tofte et al. (2019)	LM: cross-sectional associations with eGFR →significant metabolites analysed with LR→ significant ones analysed in Cox to predict ESRD	FDR Benjamini-Hochberg, q-value <0.05	Cox	age, BMI, HbA1c, p-triglycerides, sex, smoking, statin treatment, SBP, total cholesterol, eGFR and logUACR	lipids, phosphatidylcholine (PC) and sphingomyelin (SM) species: PC(O-34:2), PC(O-34:3), SM(d18:1/24:0), SM(d40:1) and SM(d41:1)
Zacharias et al. (2019)	LASSO Cox to create sparse models of metabolites and combine with established clinical model to predict ESRD	None	Cox	age, sex, eGFR, and UACR	Of the 24 metabolites, top ones were creatinine, high-density lipoprotein, valine, acetyl groups of glycoproteins.
Afshinnia et al. (2020)	Multiple LR to predict ESRD	FDR Benjamini-Hochberg, q-value <0.05	LR	SBP, history of HTN, use of calcium channel blockers, baseline eGFR and urine UACR, sex, total number of antihypertensives, total number of medications.	20-hydroxyeicosatetraenoic OR 1.45 (95% CI 1.07–1.95; P = 0.017).

Abbreviations: BMI, body mass index; CKD, chronic kidney disease; CHD, congestive heart failure CVD, cardiovascular disease; DBP, diastolic blood pressure; DM, diabetes mellitus; eGFR, estimated glomerular filtration rate; ESRD, end stage renal disease; FDR, false discovery rate; LM, linear regression model; HbA1c, glycated haemoglobin; HDL, high-density lipoprotein; HTN, hypertension; LASSO, least absolute shrinkage and selection operator; LDL, low-density lipoprotein; LM, linear regression model; LR, logistic regression model; PCA, principal component analysis; PLS, partial least squares; SBP, systolic blood pressure; SVM, support vector machine; TG, triglycerides; UACR, urinary albumin to creatinine ratio; UAER, urinary albumin excretion rate.

3.3.5 Metabolites associated with kidney function decline

The included studies revealed over 153 metabolites in total that were associated with the risk of new-onset of kidney function decline (Table 3-6). In general, metabolites belonged to heterogenous sub-classes such as amino acids, lipids, sugars and acyl-carnitines. Given the large number of metabolites measured in the identified studies, together with diverse outcomes of kidney function decline considered in the included studies, my subsequent narrative synthesis of the results focuses only on the metabolites that were reported more than once.

Four metabolites were reported in multiple cohorts: i) Pseudouridine, a uremic solute, was the strongest predictor of increased risk of progression to ESRD in Joslin Study participants with T2DM (Niewczas et al., 2014) and T1DM (Niewczas et al., 2017). Pseudouridine was also a significant predictor of increased risk of incident CKD (Solini et al., 2016) and CKD progression in a study of T2DM outpatients and cohort of general population (Goek et al., 2013, Sekula et al., 2016, Solini et al., 2016). ii) Myo-inositol, which is also a uremic solute, was inversely associated with kidney function decline and significantly increased risk of progressive decline in study of patients from Nephrology Outpatient Clinic (Gil et al., 2018), Steno Type 1 Diabetes study (Tofte et al., 2019a) as well as ESRD incidence in Joslin study (Niewczas et al., 2014). iii) Tyrosine, a non-essential amino acid was associated with higher risk of incident CKD in KORA cohort and higher risk of more rapid kidney function decline in Joslin Kidney Study, while lower tyrosine was associated with progression of albuminuria in people with T2DM in PREVEND cohort. Apparently opposing effect, where inverse relationship between tyrosine and ESRD was reported in Progredir Cohort of people living in Brazil. iv) Finally, citrulline, a ubiquitous amino acid that plays a role in urea cycle, was a significant predictor of increased risk of incident CKD in Framingham Offspring Study (Rhee et al., 2013) and Korean Genome Epidemiology Study (Lee et al., 2020) as well as significantly higher risk of progression of eGFR decline in CKD patients (Nkuipou-Kenfack et al., 2014).

3.4 Update: studies published since completion of the original review

The original systematic scoping review was conducting in the first year of my PhD in order to inform my own analysis. The searches for the literature were completed in May, 2020, so I acknowledge that at the time of writing this thesis (between the end of 2023 and beginning of 2024), the main results of this chapter were outdated. While it

may be ideal to update the search of the published articles and completely re-write the results of this chapter, it was not possible complete this with the time constraints of my PhD. To address this, I have re-run the search and identified articles published between March 2020 and February 2024, based on the same selection criteria.

3.4.1 Results of articles published since 2020 and key findings

In brief, the search of publications released between March, 2020 and February 2024, retrieved 932 and 333 articles from EMBASE and MEDLINE, respectively (total of 1,265). All records were imported in a single library on EndNote for further management. After removal of duplication (n=165), 1,100 records were left for screening. These articles were screened by titles and abstracts, which resulted in removal of 1,070 articles, because they were conference abstracts or reviews, focused on animal research or did not include relevant prospective outcomes related to measures of kidney function decline. After the full texts of the remaining 30 articles were examined (two were excluded as they did not consider prospective outcomes related to kidney function, five studies focused on specific metabolites and one was conference abstract). Twenty-two relevant publications remained.

The summary of key information about the studies published in this period are presented in table 3-7. The selected articles were mainly prospective studies of general population, which commonly involved people with diabetes, prevalent CVD and hypertension. A few studies focused specifically on people diagnosed with T2DM, including a meta-analysis of sub-populations from Hoorn Diabetes Care Study and Maastrich Study (Tofte et al., 2020), hospital outpatients from China (Zhu et al., 2022) and Italy (Trischitta et al., 2023). Similarly, a few studies focused on T1DM patients (Afshinnia et al., 2021, Al-Sari et al., 2022, Mutter et al., 2022), while other included participants with any diabetes diagnosis (Kwan et al., 2020, Hirakawa et al., 2022, Kwon et al., 2023).

Some studies continued exploring metabolomic profiles in the same cohorts as reported in my earlier results, including CRIC (Kwan et al., 2020, Wen et al., 2022, Zhang et al., 2022), ARIC (Bernard et al., 2022, Su et al., 2023), Steno (Afshinnia et al., 2021, Al-Sari et al., 2022), GCKD (Steinbrenner et al., 2021) and KORA (Huang et al., 2020). The number of participants studied ranged from 30 T2DM patients with CKD (Zhu et al., 2022) to >90,000 participants from UK Biobank study (Geng et al.,

2024). Overall, more than half of studies (12 of 22) included >1000 participants (Huang et al., 2020, Kwan et al., 2020, Lee et al., 2020, Tofte et al., 2020, Steinbrenner et al., 2021, Bernard et al., 2022, Mutter et al., 2022, Wen et al., 2022, Zeng et al., 2022, Su et al., 2023, Geng et al., 2024, van der Burgh et al., 2024), which shows that more recent studies included larger populations compared to those identified in the original review. In contrast, a single study of T2DM patients used <100 participants (Zhu et al., 2022), which indicates that less studies analysed small populations compared to those identified in the original searches. In terms of outcomes, most studies (8 of 22) focused on progression to ESRD (Kwan et al., 2020, Steinbrenner et al., 2021, Bernard et al., 2022, Mutter et al., 2022, Wen et al., 2022, Zhang et al., 2022, Kwon et al., 2023). Incident CKD was also studied in the second largest proportion of studies (7 of 22) (Huang et al., 2020, Lee et al., 2020, Zeng et al., 2022, Zhu et al., 2022, Su et al., 2023, Geng et al., 2024, van der Burgh et al., 2024). This trend is similar to the results reported in the original review, where final end-point of kidney decline was the most popular outcome studied.

The most notable study, was the metabolomic analysis of UK Biobank participants, which is the largest study that investigated metabolomic associations with incident CKD to date (Geng et al., 2024). The study used targeted nuclear magnetic resonance metabolomics platform developed by Nightingale Health Ltd (Wurtz et al., 2015, Wurtz et al., 2017). The statistical analysis involved LASSO to reduce the dimension and select spare set of metabolites, which were then added along with covariates into a Cox model to predict the risk of incident CKD, which significantly improved the risk prediction in terms of discrimination (Geng et al., 2024). In addition, Rotterdam Study (van der Burgh et al., 2024) and the meta-analysis of Dutch cohorts (Tofte et al., 2020) also used Nightingale metabolomics platform, but neither reported any statistically significant findings in relation to new-onset kidney decline.

Table 3-7 Characteristics of the cohort studies selected in the updated search (2020-2024) associated with incident kidney function decline or new-onset kidney function decline.

First author, year, journal	Study	Follow up, years	Included participants	Outcomes	Baseline characteristics, mean age; proportions of participants (%)	Statistics	Main findings
Huang et al. (2020), Diabetes	KORA survey 4/F4, Germany	6.5	1838	Incident CKD (eGFR <60 mL/min/1.73 m2) or UACR ≥30 mg/g; n=200	Age 67 years Male 55%	multivariate LR, FDR adjusted P-value, only significant metabolites retained → LASSO→ LR with backward selection for metabolites and clinical variables	SM C18:1 and PC aa C38:0, and clinical variables improved prediction of incident CKD
Kwan et al. (2020), Am J Kidney Diseases	CRIC, US	8	1001	Annual eGFR slope and time to ESRD (kidney failure), n=359	Age 60 years Male 56% HTN 93% DM 100%	Stepwise selection and LASSO for eGFR slope analysis→ metabolites from most accurate model analysed in Cox for time to ESRD risk.	3-HIBA and aconitic acid levels were associated with higher and lower risk for ESRD
Lee et al. (2020), Biomedicines	Korean Genome Epidemiology Study, Korea	8	1741	Incident CKD, n=235	Age 55 years Male 47% HTN 30% DM 20%	Pearson's correlation: eGFR and metabolites, Bonferroni correction ($p < 0.00037$)→ LR for significant metabolites with adjustment for clinical covariates	citrulline [odds ratio (OR): 2.41, 95% confidence interval (CI): 1.26–4.59], kynurenine (OR: 1.98, 95% CI: 1.05–3.73), and phenylalanine (OR: 2.68, 95% CI: 1.00–7.16)
Tofte et al. (2020), The Journal of Clinical Endocrinology & Metabolism	Hoorn Study; Maastricht Study, Denmark	7	1100	eGFR slope	Age 64 years Male 59% CVD 19% T2DM 100	LM for associations between metabolites and eGFR slopes, with adjustment for covariates and FDR	Nightingale platform: None

Afshinnia et al. (2021), Diabetes care	Steno Diabetes Center study, EDC, CACTI	4	817	annual decline in eGFR ≥3 mL/min/1.73 $m^{2,}$ (cases n=123 vs non-cases n=284 and replication cases n=127, non-cases n=283)	Age 47 years Male 54% HTN 31% T1DM 100%	Independent t-test, FDR→ PCA to build secondary variables related to subclasses of metabolites→ LR models 1) top differentially regulated lipids, and 2) principal components as predictors adjusted for covariates	free fatty acid 20:2, phosphatidylcholine 16:0/22:6. A principal components of unsaturated free fatty acid and saturated phosphatidylcholinespredicted rapid eGFR decline
Lunyera et al. (2021), Metabolomics	STOP-DKD Trial	1.8	132	eGFR slope	Age 64 years Male 50%	PCA-derived principal component scores of TCA cycle organic anions; or PCA-derived principal component scores of untargeted urine metabolites)→ linear mixed models for eGFR slope	Methylmalonate, ethylmalonate and citrate/isocitrate loaded negatively on principal component (PC3). Higher PC3 was associated with lower eGFR.
Steinbrenner et al. (2021), Am J Kidney Diseases	GCKD	4	5087 (discovery, n=3391; replication	ESRD (kidney failure), n=241	Mean eGFR 44.9; Age 60; Male 59%; DM 100%; CVD 30%	Cox regression models were fitted for each metabolite separately, Bonferroni correction	C-glycosyltryptophan (hazard ratio of 1.43 [95% CI, 1.27-1.61])
Al-Sari et al. (2022), eBioMedicine	Steno Diabetes Center, Denmark	5	383	≥ 30% decline in eGFR, n=79	Age 55 years Male 55% CVD 67% T1DM 100%	Random Forest models were applied to predict future risk of outcome	Clinical variables and 3,4-dihydroxybutanoic acid, 2,4-dihydroxybutanoic acid, ribitol, ribonic acid, myo-inositol, and one unidentified metabolite (ketones and sugars)

Bernard et al. (2022) , Kidney Medicine	ARIC, US	23	3799 (split into training and validation set)	ESRD, n=160	Age 53-54 years Male 36-45% CVD 3.8- 5.9 %	Netboost dimension reduction technique to cluster metabolites→ Cox for associations between clusters and outcome	4 of 43 clusters were associated with ESRD
Hirakawa et al. (2022), Scientific reports	UT-DKD cohort, Japan	2.5	135	change rate of eGFR below − 10% of baseline eGFR	Age 70 years Male 79% DM 100%	Deep learning, logistic regression, random forest, and support vector machine (50 features selected)→ Piecewise linear model and handcrafted linear regression model to compare the selected features	None
Lee et al. (2022) , Metabolites	Korean Biobank Array, KoreaN Cohort Study	3.9 1.8	115 69	eGFR slope change −5 mL/min/1.73 m^2 per year	Age 57- 53 Male 56- 71%	LM to identify the metabolomic mechanisms underlying the genetic variants with FDR correction	pimelylcarnitine (beta = 0.030, SE = 0.007, FDR = 0.01) and octadecenoylcarnitine (beta = 0.167, SE = 0.049, FDR = 0.08)
Mutter et al. (2022) Diabetologia	Finnish Diabetic Nephropathy study, Finland	9	2670	ESRD, n=355	Age 36 years Male 50% T1DM 100%	Cox proportional hazard regression models for each urine metabolite and ESRD	Leucine (HR 1.47 [95% CI 1.30, 1.66], valine (1.38 [1.22, 1.56]), isoleucine (1.33 [1.18, 1.50]), pseudouridine (1.25 [1.11, 1.42]), threonine (1.27 [1.11, 1.46]) and citrate (0.84 [0.75, 0.93]). 2-Hydroxyisobutyrate (1.30 [1.16, 1.45])

Wen et al. (2022), American society for clinical investigation	CRIC, US	6.8	1773	ESRD or eGFR halving from baseline, n=675	Age 59 year Male 56% CVD 36% DM 49%	Cox proportional hazards models for each metabolite and outcome; FDR correction	Pseudouridine (HR 2.23 (95%CI 1.62, 2.38), methylimidazoleacetate (1.16 (1.07, 1.26), homocitrulline (1.28 (1.12, 1.44)
Zeng et al. (2022), Kidney International	Strong Heart Family, US	5	1910	Incident CKD, eGFR <60 ml/min per 1.73 m^2, n=228	Age 40 year Male 38% HTN 29% DM 18%	Mixed-effect LR with adjustment for covariates, FDR correction	29 baseline lipids associated with incident CKD
Zhang et al. (2022), American Journal of Nephrology	CRIC, US	10	995	eGFR slope and ESRD, n=360	Age 60 years Male 56% DM 50%	univariate and multivariate models for the eGFR slope using LASSO and random forest models for eGFR slope→ selected metabolites entered into Cox model for ESRD	Six eGFR slope models selected 9 to 30 metabolites. In the adjusted ESRD model, valine (or betaine) and 3-(4-methyl-3-pentenyl)thiophene were associated ($p < 0.05$) with 44% and 65% higher hazard of ESRD
Zhu et al. (2022), Nutirents	Patients from First Affiliated Hospital, China	5.7	30	annual eGFR decline ≥3 mL/min/1.73 m2, n=14 or incident CKD eGFR <60 mL/min/1.73 m2, n=12	Age 53 years Male 50% T2DM 100%	Pearson's correlation to assess the association between metabolites and kidney function→ amino acids significant → LR for amino acids and rapid eGFR decline, multivariate Cox analysis to determine independent risk factors	: L-valine (HR = 2.583, 95% CI = 1.006–6.629 and isoleucine (HR = 1.670, 95% CI = 1.206–2.312, $p = 0.002$)
Kim et al. (2023), Kidney360	Biopsy patients, China	5	936 (147 healthy controls, 789 biopsy patients	Composite: 30% decline in eGFR, doubling of	Age 50 years Male 61% DM 17% HTN 18%	Metabolite concentrations and the CKD stages were compared using the Jonckheere–Terpstra test	betaine, choline, glucose, fumarate, and citrate showed significant associations with the

			with eGFR>30)	serum creatinine levels, or ESRD n=309		at baseline→ eight metabolites that differed between CKD groups at baseline →Cox model for prospective outcomes	composite outcome after adjustment for covariates
Kwon et al. (2023), Kidney Research and Clinical Practice	Seoul Medical Center patients, China	4.5	234 (26 healthy controls, 208 CKD)	ESRD, n=103	Age 59 years Male 63% HTN 64% DM 100%	Pearson correlation between metabolites and eGFR or UPCR: 19 metabolites showed significant trends → Multivariate Cox regression with backward stepwise model selection	Choline, myo-inositol, and citrate were associated with ESRD progression with adjustment for covariates
Su et al. (2023), Clinical journal of American Society of Nephrology	ARIC, US	23	3751	incident CKD (eGFR <60 ml/min per 1.73 m2, accompanied by 25% decline of eGFR from baseline), n=1468	Age 54 years Male 40% HTN 44% DM 13%	LM to investigate the cross-sectional association between ultra-processed food consumption as the exposure and each metabolite→ Spearman rank for significantly associated metabolites (12)→ Cox models to assess association between related metabolites and incident CKD, FDR	N2-dimethylguanosine HR, 1.16, 95% CI, 1.09 to 1.24; glucose HR, 1.42, 95% CI, 1.19 to 1.69; mannose HR, 1.39, 95% CI, 1.22 to 1.58
Trischitta et al. (2023), BMJ Open Diabetes Research and Care	Endocrinology Unit patients, Italy	1.5	575	Annual eGFR change	Age 60 years Male 52% HTN 65% T2DM 100%	LR for serum metabolites levels and low eGFR, adjusted for covariates and Bonferroni→validation of significant findings→ LM for annual change in eGFR and significant metabolites with adjustment for	tiglylcarnitine, decadienylcarnitine, total dimethylarginine, decenoylcarnitine and kynurenine) (β range −0.11 to −0.19, p values range 4.8×10−2 to 3.0×10−3

						covariates→ each metabolite added to reference model	
Geng et al. (2024) American Journal of Kidney Diseases	UK Biobank study, UK	13.1	91532	Incident CKD identified using ICD codes, n= 2,269	Age 55 years Male 43% HTN 49% DM 2% CVD 3%	Cox regression for each metabolite and outcome with adjustment for covariates and FDR→LASSO to select metabolites which were added to a final model with covariates	Nightingale platform; LASSO selected: docosahexaenoic acid, saturated fatty acids to total fatty acids %, linoleic acid to total fatty acids %, histidine, isoleucine, valine, albumin, glycoprotein acetyls, and cholesteryl esters in small HDL
van der Burgh et al. (2024), Clinical Kidney Journal	Rotterdam Study, Netherlands	5.4 /5.2	3337 and 1540	2 eGFR's <60; single eGFR <45; kidney failure	Age 69- 67 years Male 42- 43% DM 14- 13% CVD 13- 9%	LM for associations between circulating metabolites and baseline kidney function, FDR adjusted→significant metabolites analysed with prospective outcome, adjusted for FDR and covariates	Nightingale and Metabolon platform, 16 metabolites were significant, strongest association found with C-glycosyltryptophan (HR 1.50, 95%CI 1.31;1.71).

Atherosclerosis Risk in Communities study, ARIC; CACTI, Coronary Artery Calcification in Type 1 Diabetes study; CKD, chronic kidney disease; CVD, cardiovascular disease; Chronic Renal Insufficiency, CRIC; EDC, Epidemiology of Diabetes Complications study; eGFR, estimated glomerular filtration rate; ESRD, end stage renal disease; DM, diabetes mellitus; FDR, false discovery rate; GCKD, German Chronic Kidney Disease Study; HTN, hypertension; LASSO, least absolute shrinkage and selection operator; LM, linear regression model, LR, logistic regression model; SM, sphingomyelins; M%, percentage of males; STOP-DKD, Simultaneous Risk Factor Control Using Telehealth to SlOw Progression of Diabetic Kidney Disease; PCA, principal component analysis; PC, phosphatidylcholine diacyl; T1DM, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus.

3.5 Discussion

3.5.1 Summary of findings

This scoping review of human observational prospective studies identified 26 publications (released up to March 2020) that evaluated the relationship between metabolomic fingerprints and the new onset of kidney function decline, either in terms of incident CKD, rate of kidney function decline, or incident ESRD. Updating the search to February, 2024, has almost doubled the number of relevant articles published, raising the total number of studies that investigated metabolomic profiles associated with new-onset kidney decline to 48. This suggests that the research body in this field continued to grow over the past 4 years, expanding our understanding of individual metabolites and related pathophysiological pathways associated with the risk of kidney function decline.

In general, most studies included in this review, analysed samples from the general population with common comorbidities, such as hypertension and diabetes mellitus, suggesting that conclusions might be generalized to a wider population. Both targeted and untargeted metabolomic profiling methods have been used equally to measure metabolites. The studies used dimension reduction or data selection methods to deal with high-dimensional data, followed by regression analysis of a sparse set of metabolites with adjustment for traditional risk factors. Statistically significant metabolites belonged to diverse categories, including amino acids and uremic solutes. While extensive coverage of the human metabolome enabled via high-throughput analytics is a great advantage in the search for biomarkers, it also makes it very time-consuming to undertake a systematic review for each potentially important metabolite. Therefore, the focus of my review was on describing the nature and extent of the studies already undertaken and identifying statistical approaches used to derive the list of prognostic metabolites in published studies.

3.5.2 Analytical techniques and metabolomic approaches

The nuclear magnetic resonance technique identifies compounds by analysing their resonance frequencies of energy released within an external magnetic field, while mass spectrometry achieves this by recording the mass-to-charge ratio of ionized analytes (McGarrah et al., 2018). Despite its advantages, such as being non-invasive, non-destructive, requiring minimal sample preparation, easy data interpretation, and

not needing chemical derivatization, which enhances reproducibility, nuclear magnetic resonance may capture fewer metabolites and exhibit lower sensitivity compared to mass spectrometry (Nicholson and Lindon, 2008). Majority of studies used the more sensitive mass spectrometry approach, which suggests that research in this area is focused on quantity of metabolites captured. This suggests that the overarching goal of many studies was to cover as much of human metabolome as possible to expand our current understanding of metabolomic profiles and kidney function.

Both nuclear magnetic resonance and mass spectrometry can profile metabolites with either targeted or untargeted approaches. Exactly half of the studies in the original review (13 out of 26) involved non-targeted high-throughput metabolomic profiling approaches. The major limitation of this method is the availability of recorded metabolomics spectra repositories, which restricts investigation of detected features. Most of these studies found associations with unrecognizable features, which prevented further interpretation. Nonetheless, the remaining half of the studies, chose targeted analysis, which allowed precise quantification of a limited set of known metabolites. The advantage here is that data produced consist of absolute metabolite concentrations detected in the sample, allowing more straightforward comparisons and hypothesis testing of chemically characterized and biochemically annotated metabolic compounds. However, the key limitation of the targeted method is that metabolites require clear definition, chemical characterization, and biochemical annotation before they are added to the analytical panel, and currently, a relatively small proportion meets this criterion (Wishart et al., 2022). Overall, the even division of profiling methods used suggests that there is high interest in both novel metabolite discovery related to kidney function, but also in looking for quantifiable metabolites that may be associated with kidney disease.

3.5.3 Most common statistical analysis methods

Metabolomic data presents challenges in statistical analysis due to its high dimensionality, collinearity, and non-normality. To mitigate the risk of false positives multiple testing correction methods are recommended (Chadeau-Hyam et al., 2013). Bonferroni correction may be too stringent for studies with small sample sizes, so it is not surprising that the majority of studies opted to use the less stringent FDR method to control Type I errors, which regulates the expected proportion of false discoveries, but allows for a higher number of Type I errors (Benjamini and Hochberg, 1995). This

suggested that studies of metabolomic profiles in this field were less concerned about a higher chance of false positives.

To handle collinearity among metabolites, the most popular method was variable selection technique using LASSO. Steno Diabetes Center and GCKD study (Tofte et al., 2019a, Zacharias et al., 2019) incorporated the LASSO method to prevent overfitting and to select a small panel of metabolites by constraining the sum of absolute regression coefficients (Chadeau-Hyam et al., 2013). The updated review also showed that many studies opted to use LASSO for variable selection (Kwan et al., 2020, Afshinnia et al., 2021, Zhang et al., 2022, Huang et al., 2023, Geng et al., 2024). The results of this scoping review suggested that LASSO is a commonly used method for analysing metabolomic profiles related to kidney function decline. This is in line with the notion that LASSO is a popular statistical method in 'omics' studies for risk prediction purposes as it is adapted for handling highly colinear datasets and effectively reduces the number of predictors.

3.5.4 Strengths and limitations

To the best of my knowledge, at the time of the inception of this scoping review, it is the first attempt to summarize methodologies and findings in studies exploring the association between metabolomic markers and new-onset kidney function decline. However, several limitations should be acknowledged. Firstly, the review only included studies of prospective design that evaluated associations with new-onset kidney function decline and ignored literature that focused on cross-sectional or case-control studies. Although a cross-sectional study has many advantages, including lower cost and suitability for large-scale epidemiological studies, it is not ideal because the metabolomic data and kidney function would be assessed at the same time, leaving the causal relationship unclear. Furthermore, the inclusion of two distinct outcomes would expand the list of studies to unmanageable levels, making it impractical within a PhD timeline. Secondly, the review ignored studies that pre-selected specific metabolites based on a knowledge-driven hypothesis. As a result, some metabolites that predict renal outcomes were missed, especially those with more evidence supporting their association with kidney function decline. However, evaluating a broader metabolomic profile enables novel biomarker discovery and allows for accounting for high collinearity using specialized statistical methods. Overall, the analysis of data produced by high-throughput analytics provided deeper insight into

the complicated relationship among metabolites. Thirdly, the review was limited to two major sample types - blood and urine, as these samples are easily obtainable and accessible in both research and clinical settings. However, this limited the number of publications identified, as metabolomic analysis may be conducted using a range of biological samples, including tissue, stool, and even exhaled breath.

Future studies should evaluate the effect of metabolite ratios, as these relationships were only considered in the KORA cohort, and some were identified as significant predictors (Goek et al., 2013). In addition, there was a relatively small number of T2DM cohorts. Even a smaller number of studies specifically focused on incident CKD in T2DM population. Considering that kidney function decline is irreversible, characterising risk of early CKD may enable more timely preventative interventions of CKD onset. Since clinical guidelines recommend kidney function screening for high-risk groups (such as people diagnosed with diabetes), the focus should be on trying to improve the risk prediction and stratification specifically designed for this population. The metabolites found in specific populations such as T2DM may reveal specific metabolic features related not only to the aetiology of the disease but also provide insight into novel therapeutic targets and improve risk prediction.

3.5.5 Conclusions of the scoping review

This scoping review of metabolomic studies summarised previous findings regarding metabolomics in people with incident CKD and its progression. Overall, there is enough evidence in the literature to suggest that analytics using high-throughput metabolomic tools have the potential to provide new insights into the molecular phenotypes associated with kidney function decline. Since the only studies included were those with a prospective cohort design, the associations provided evidence that metabolic markers may be useful in risk prediction for both individuals with higher levels and those at the lower extreme of kidney function. While, this is still an emerging field, with the first prospective study published only in 2013, it gathered a lot of research interest as there were over 40 prospective studies published over the last decade. Considering the huge complexity of the human metabolome, more high-quality studies with a sufficient number of participants and incident cases are necessary, especially to study early stages of decline such as incident CKD as this outcome was considerably less commonly investigated compared to the more extreme outcome.

Chapter 4. Data sources and statistical methods

This chapter provides a comprehensive description of the methodological foundation of the Edinburgh Type 2 Diabetes Study (ET2DS), essential for subsequent findings presented in this thesis. It begins by detailing the study design, measurement procedures, and variable derivation. The second part of the chapter outlines the statistical methods employed to achieve the key thesis objectives, divided into the following:

Association analysis: This section explains the analytical methods used to identify individual metabolites associated with kidney-related outcomes, independent of established risk factors.

Risk Prediction Analysis: This section focuses on statistical modelling to generate the best possible estimate for the risk of incident chronic kidney disease (CKD). It covers the validation of a reference risk prediction model, the selection process of predictive metabolites, the development of a metabolite-based risk score (MetS), and an evaluation of the added predictive value of metabolites for predicting the risk of 5-year incident CKD in the ET2DS.

4.1 Edinburgh type 2 diabetes study (ET2DS)

The ET2DS is a prospective cohort comprising 1,066 individuals aged 60 to 75 years, all diagnosed with type 2 diabetes mellitus (T2DM) and residing in the Lothian region of Scotland. The primary intent behind establishing this cohort was to explore potential risk factors associated with the complications of T2DM, particularly those linked to vascular diseases. The recruitment goal for the ET2DS was to enrol 1,000 subjects, a number chosen to provide a robust statistical foundation with 90% power at a two-sided 5% significance level. This statistical power would enable the detection of a Pearson correlation coefficient of ≥0.10 between two continuous variables. This sample size was therefore deemed sufficient for identifying risk factors explaining 1% or more of the variance in the outcome, both at baseline and during follow-up, as detailed in the comprehensive study protocol by Price et al. (2008). Within this section, I present a concise summary of aspects of the study design that hold particular relevance to the analyses I conducted for this thesis.

4.1.1 Study population

The ET2DS study recruited its participants from the Lothian Diabetes Register (LDR), which was established in 2001 to encompass individuals diagnosed with diabetes living in Lothian, Scotland, whether managed in primary or secondary care. Approximately 5,400 individuals were selected randomly from the LDR, between 2006 and 2007, ensuring representation across both sexes and equal recruitment across 5-year age bands within the target age range. Of those initially approached, 1,252 individuals responded positively to the invitation letters, and ultimately, 1,077 individuals agreed to attend the baseline clinic. To ensure comprehensive data collection for all relevant participant characteristics, including cognitive ability, essential exclusion criteria were applied. These criteria encompassed individuals who were not proficient in English, those with severe vision impairment, individuals unwilling to provide informed consent, and those unable to complete both physical and cognitive examinations.

Diabetes status was initially determined by inclusion on the LDR as type 2 diabetes, which is based on clinical diagnosis using World Health Organisation criteria (Puavilai et al., 1999). Clinical diabetologists in the research team reviewed the records of potential participants to confirm the accuracy of the T2DM diagnosis before their inclusion in the study. The confirmation of T2DM diagnosis was based on individuals either taking oral anti-diabetic medication and/or insulin, or solely managing diabetes through dietary modifications, coupled with a baseline glycosylated haemoglobin (HbA1c) level exceeding 6.5%. In cases where individuals controlled their diabetes only through dietary modifications and displayed an HbA1c measurement of 6.5% or lower at baseline, diabetologists scrutinised clinical records to validate the diagnosis. Additionally, diabetologists examined the clinical records of individuals treated with insulin within one year of diabetes diagnosis, to ensure no Type 1 Diabetes cases were included.

After the application of these exclusion criteria, a total of 1,066 out of the initial 1,077 participants who attended the baseline clinic were included in the ET2DS study. The resultant study population was shown to be largely representative

of all people in this age group with T2DM living in Lothian in terms of age, HbA1c, diabetes duration and insulin treatment (Marioni et al., 2010). These encompassed individuals receiving care across both primary and secondary healthcare settings and included a broad spectrum of disease severity based on treatment regimens.

4.1.2 Ethical approval for data collection

A written informed consent was given to each participant before data collection for clinical examinations and review of medical records. Ethical approval for the ET2DS was granted by the Lothian Medical Research Ethics Committee and complied with the Declaration of Helsinki (World Medical, 2001).

4.1.3 Data collection procedures

Baseline

Baseline data were gathered at specialised research clinics located within the Wellcome Trust Clinical Research Facility, situated at the Western General Hospital in Edinburgh, United Kingdom. To ensure accurate data acquisition, participants underwent specific procedures after an overnight fast.

i) Sample Collection: Each participant provided a urine specimen and venous blood samples (collected through venepuncture).
ii) Physical Examination: Specially trained research personnel conducted a physical examination for each participant, adhering to pre-established standard operating procedures.
iii) Questionnaires: Self-completion questionnaires were distributed to participants to collect data on demographic characteristics, detailed diabetes history and treatments, medical conditions, prescribed medications, and lifestyle-related information.

At baseline, a data linkage process was established with the Scottish Morbidity Records scheme via Information Services Division of NHS Scotland, which provided access to acute hospital discharge records over the 20 years preceding recruitment (between 1981 and 2007) for all participants. This baseline data linkage procedure enriched the dataset with comprehensive medical information, allowing for the validation of self-reported medical histories

and the inclusion of additional historical health metrics. Historical data recorded in the LDR were also selectively accessed. These historical records included key kidney-related measures for my analyses, such as urinary albumin to creatinine ratio (uACR), micro-albumin levels and serum creatinine. I used the kidney function measurements within two years of baseline visits to create a serial dataset for uACR and for estimated glomerular filtration rate (eGFR). I reviewed this data manually and with consultation from a nephrologist (Dr Bryan Conway), I excluded any results indicative of an episode of acute kidney injury and remaining results were used to derive CKD status as detailed below (section 4.1.5.7). The AKI was defined as eGFR measurement or a series of measurements recorded in closely together, which showed an extreme dip in eGFR that is inconsistent to the eGFR results before and after the potential acute kidney injury period.

Follow-up

The Lothian laboratory and LDR databases were interrogated and all out-patient or primary care serum creatinine results between 2005 to May 2014 were identified. Serum creatinine-based eGFR was calculated, episodes of acute kidney injury excluded and sorted into serial data by a previous student (Sara Jenks). The follow-up serum creatinine results were isotope dilution mass spectrometry (IDMS) traceable or were IDMS-corrected according to the equation provided below (section 4.1.5). Sara Jenks manually investigated if the correction is relevant based on the laboratory and the time of eGFR measurement to decide if the creatinine was already IDMS-traceable or not (NHS Lothian laboratories underwent standardisation around 2008, so results before standardisation needed IDMS-correction to be applied). I used this dataset to derive incident CKD and annual rate of kidney decline variables, as detailed below (section 4.1.5).

4.1.4 Data management and cleaning

Upon the completion of each clinic visit, the questionnaires, laboratory results, and clinical data collected at baseline were entered into a central database. This database operation was executed by members of the research team using

Microsoft Access 2003/2010 software. All data management procedures described below were completed prior the start of my PhD.

To ensure data integrity, an exacting process was implemented for paper records at baseline. Specifically, a dual-entry approach was adopted, with the data being entered twice into the database. This procedure revealed a low error rate of 0.02%, which was rectified by cross-referencing with the original paper records. Additionally, all data underwent a scrutiny for outliers through descriptive analyses. Any data points deemed potentially inaccurate, underwent further examination, which involved consulting the original paper files and subsequently making corrections where necessary.

To safeguard the confidentiality and security of paper records, they were securely stored in locked cabinets within a controlled access room at the University of Edinburgh. Meanwhile, electronic data was stored on a dedicated university server. Access to this server required electronic authorisation and password authentication, bolstering data protection and confidentiality measures.

4.1.5 Variable measurement and definitions

This section outlines the measurement procedures and definitions of the ET2DS variables used in this thesis. Additional data collected as part of the ET2DS can be found in Price et al. (2008).

Physical examination

During physical examination, systolic and diastolic blood pressure was measured in the right arm to the nearest 2 mmHg, with the participant in the supine position. To obtain the body mass index (BMI), height (in metres) was measured without shoes using a wall mounted ruler and weight (in kilograms) was measured without outdoor clothing using electronic scales. Then, the BMI was calculated as weight / height^2 (kg/m^2). This data was recorded in the ET2DS database prior my PhD.

Demographic information

Self-reported questionnaires were employed to gather demographic details, encompassing date of birth, sex, and ethnicity. The date of birth served as the

basis for calculating the age at the baseline visit (prior to my PhD). Furthermore, I used the date of birth for determining the age at every serum creatinine measurement before the baseline assessment for eGFR calculation.

Smoking

A smoking questionnaire was used to collect information about smoking status. Participants who had quitted smoking in the last 6 months before the baseline were regarded as current smokers in this study. Finally, smoking status was categorised into mutually exclusive categories: non-smoker, ex-smoker, current-smoker. The smoking status was derived in line with Nelson et al. (2019) publication in order to create comparable predictors in ET2DS.

Biochemical analysis

All blood samples were separated and frozen at -80^{o}C within 1 hour of venepuncture except for serum samples for creatinine and plasma for HbA1c and glucose, which were analysed on fresh samples. The uACR was measured on a fresh urine sample, which was subsequently frozen. Baseline creatinine was measured using a dry slide assay on a Vitros 5, 1 FS analyser (Ortho Clinical Diagnostics, High Wycombe, United Kingdom). Internal quality control data indicated that the between batch coefficient of variation (CV) were 1.9% at a creatinine concentration of 83.9 µmol/L and 1.3% at 804.6 µmol/L. HbA1c was measured using high performance liquid chromatography using a Biorad variant (II) DCCT aligned analyser (Bio-Rad Laboratories Ltd, Hemel Hempstead, United Kingdom). The between batch CVs were both 2.5% at HbA1c concentrations of 6.05% and 10.4%. The uACR was measured using a dry slide immunoturbidimetric assay on a Vitros 5 1 FS analyser (Ortho Clinical Diagnostics, High Wycombe, United Kingdom). The between batch CVs were 3% at an uACR of 42.2mg/mmol and 3.9% at an uACR of 84.4 mg/mmol.

Measurement of metabolites

Metabolite concentrations were measured using previously frozen (at -80 ^{o}C) blood serum samples. The metabolomic panel contained a total of 228 metabolites (n=149 concentrations of individual metabolites and n=79 derived ratios) that were quantified using automated high-throughput proton Nuclear

Magnetic Resonance spectroscopy platform developed by Nightingale Health Ltd. The metabolomic panel has been described previously and used in several large-scale epidemiological studies (Wurtz et al., 2017).

The metabolites quantified included lipids labelled according to size (very large to very small), density (high to low) and trait (e.g., triglyceride, phospholipid, cholesterol ester) as well as other low-molecular weight metabolic biomarkers. The whole panel can be categorised into 12 distinct subgroups:

(1) Components (including total lipids, phospholipids, free cholesterol, etc.) in very-low density lipoprotein (VLDL);
(2) Components in intermediate-density lipoprotein (IDL);
(3) Components in low density lipoprotein (LDL);
(4) Components in high density lipoprotein (HDL);
(5) Lipoprotein particles (e.g., total cholesterol in HDL);
(6) Apolipoproteins, namely apolipoprotein A-1 (ApoA-1), apolipoprotein-B;
(7) Fatty acids, including saturated fatty acids and unsaturated fatty acids;
(8) Glycolysis related metabolites, including lactate, glycerol, glucose;
(9) Amino acids, including branched-chain and aromatic amino acids;
(10) Ketone bodies, including acetate and acetoacetate;
(11) Fluid balance molecules, namely creatinine and albumin;
(12) Inflammation marker, namely glycoprotein acetyls (GlycA).

Categorisation of medical conditions and medications

Information on medical history and medication use was obtained from the self-completed questionnaires, record linkage results and findings on clinical examination. Research personnel systematically coded medical conditions in accordance with the International Classification of Diseases (ICD) - 10 codes (ICD-10, 1997), while medication data was coded using the British National Formulary (BNF) (Tallo, 2016). The categorisation of diabetes treatment type was determined from the baseline self-report questionnaire which was checked for accuracy by the research nurses by comparing against prescriptions provided by participants at the research clinic. Using the diabetes treatment information, I defined the diabetes treatment types into mutually exclusive categories, as follows:

Diet Controlled Only: Participants who managed their diabetes solely through dietary modifications.

Oral Anti-Diabetic Medication Only: This category included individuals using oral medications such as metformin, sulphonylureas, and thiazolidinediones as their primary form of treatment.

Insulin Use (Potentially with Oral Anti-Diabetic Medication): This category encompassed participants who relied on insulin therapy, either exclusively or in conjunction with oral anti-diabetic medications.

I defined the hypertension status as systolic blood pressure ≥140 and diastolic blood pressure ≥90 and/or use of blood pressure lowering medication. I defined the prevalent cardiovascular disease (CVD) as a composite of myocardial infarction, angina, stroke, transient ischemic attack and/or coronary intervention. Previous students verified the prevalent CVD at baseline through the use of multiple sources of data, including the self-reported questionnaires (questions on medical conditions and treatments related to CVD as well as the WHO chest pain questionnaire) and 12-lead ECG assessments from physical examination. For any events identified through these methods, further confirmation was sought by cross-referencing with hospital discharge records provided by the Information Services Division. In cases where data linkage records identified events not reported through self-reporting, notes from general practitioners and hospitalisation records were queried to establish the validity of these events.

4.1.6 Derivation of kidney function and renal disease variables

Kidney function

I used the equations detailed below to derive the eGFR for pre-baseline and baseline, while Sara Jenks derived the serial follow-up eGFR measurements. Participants who were missing baseline eGFR result, the most recent eGFR before baseline was entered as their baseline result.

The eGFR was estimated from serum IDMS-traceable creatinine measured using the CKD-EPI equation:

$eGFRcr = 142 \times \min(S_{cr}/\kappa, 1)^{\alpha} \times \max(S_{cr}/\kappa, 1)^{-1.200} \times 0.9938^{Age} \times 1.012$ [if female]

where:

S_{cr} = IDMS standardised serum creatinine in mg/dL
κ = 0.7 (females) or 0.9 (males)
α = -0.241 (female) or -0.302 (male)
$\min(S_{cr}/\kappa, 1)$ is the minimum of S_{cr}/κ or 1.0
$\max(S_{cr}/\kappa, 1)$ is the maximum of S_{cr}/κ or 1.0
Age (years)

Historical serum creatinine was not IDMS- traceable so I applied a correction using the following equation before calculating historical eGFR measurements:

S_{cr} = (Laboratory reported creatinine -7.71) / 0.988

S_{cr} = IDMS standardised serum creatinine in mg/dL.

Baseline uACR

The uACR levels from the laboratory were only recorded if urinary albumin was >6 mg/L, since only these levels are of clinical relevance indicating some kidney damage and potential albuminuria. The albumin < 6 mg/L was entered as 6 mg/L and I calculated the uACR levels manually by dividing urinary albumin by urinary creatinine.

Baseline albuminuria

I reviewed the uACR and microAlbumin measurements from baseline and up to two years before baseline to define albuminuria status. Participants with at least two consecutive uACR >2.5 mg/mmol for men, >3.5 mg/mmol for women or those with at least two consecutive microAlbumin results of > 2.5 mg/mmol, were classed as having albuminuria. Participants who had no historical uACR or micro-Albumin results, their baseline uACR was used to determine albuminuria status and similarly, those who did not have a baseline result, their status was based on historical microalbumin measurements.

Baseline chronic kidney disease (CKD)

Reduced kidney function (referred to as baseline CKD) was defined as eGFR <60 ml min^{-1} (1.73 $m)^{-2}$ in at least two measurements taken three months apart. The results were reviewed at baseline and up to two years before baseline. Participants who were missing baseline eGFR result, their CKD status was based on historical eGFR data and the most recent eGFR before baseline was entered as their baseline result. Also, for participants without historical serum creatinine results, the baseline CKD status was based on baseline eGFR alone and if the eGFR was below 60 ml min^{-1} (1.73 $m)^{-2}$ the participant was regarded as a baseline CKD case.

Incident CKD

New-onset of reduced kidney function (referred to as incident CKD) was investigated only in participants with eGFR > 60 ml min^{-1} (1.73 $m)^{-2}$ and without evidence of reduced kidney function at baseline either based on baseline CKD (as per definition above) alone or combined with albuminuria status.

Incident CKD was defined as two eGFRs <60 ml min^{-1} (1.73 $m)^{-2}$ at least three months apart during follow-up and at least 25% reduction in eGFR from baseline, as per the KDIGO guideline for the evaluation of CKD in research (Levin and Stevens, 2014).This stringent criteria for incident CKD ensured that participants classed as cases truly experienced a substantial reduction in their kidney function since baseline.

The annual rate of change in kidney function

The annual rate of eGFR change was determined from linear eGFR slope that was estimated within each participant using a linear regression of eGFR against follow-up time in participants with a minimum of one year follow-up and three eGFR measurements. Rapid decliner status was defined as annual rate of eGFR change of at least -5%. The percentage change was chosen over the absolute annual rate of change as the former is less dependent on baseline eGFR. The estimates from linear regression slopes were converted into percentage annual rate of eGFR change with the following equation:

Percentage slope = (linear coefficient / baseline eGFR) * 100.

A summary of the final kidney-related variables used in my cross-sectional and prospective analyses are described in table 4-1.

Table 4-1 Summary of the kidney-related outcomes used in association analyses.

Kidney related marker	Trait	Definition	Variable type	Analysis type
eGFR	eGFR	Kidney function calculated using serum creatinine-based CKD-EPI equation	Continuous	Cross-sectional
eGFR	CKD	At least 2 consecutive eGFR measurements <60 ml min^{-1} $(1.73\ m)^{-2}$ at baseline and up to 2 years before baseline	Binary	Cross-sectional
uACR	uACR	Ratio of urinary albumin and creatinine	Continuous	Cross-sectional
uACR	Albuminuria	2 of 3 consecutive uACR measurements: >2.5 mg/mmol for men, >3.5 mg/mmol for women measured at baseline and up to 2 years before baseline	Binary	Cross-sectional
eGFR	annual rate of eGFR change	Estimated coefficients from individual linear regression of follow-up eGFR over time	Continuous	Prospective
eGFR	Rapid decliner	Dichotomised version of eGFR slope -5% or lower	Binary	Prospective
eGFR	Incident CKD	At least 2 consecutive eGFR measurements <60 ml min^{-1} $(1.73\ m)^{-2}$ minimum 3 months apart during follow-up and 25% reduction in eGFR from baseline.	Binary	Prospective

Abbreviations: CKD, chronic kidney disease; eGFR, estimated glomerular filtration rate; uACR, urinary albumin to creatinine ratio.

4.1.7 Missing data

The association analysis was adjusted for 13 covariates. So, to avoid any loss of sample size caused by the complete-case strategy, single imputation was used to address missing values in covariate dataset. The imputation method carried out using the “mice” package (version 3.14.0) in R. This package handles missing data through imputation by employing predictive modelling techniques. It treats variables with missing values as dependent variables and uses the remaining variables as predictors to estimate the missing values. This process is iterated multiple times to create several imputed datasets, and these imputed datasets are eventually combined to form a single complete dataset.

4.1.8 Missing metabolomics data

Nightingale Health conducted metabolomic profiling on all frozen ET2DS blood serum samples, with the exception for eight samples, which did not have sufficient serum for metabolomic profiling. Further interrogation of the participants with completely missing metabolomics data did not reveal any particular pattern which would provide explanation for this, so I assumed the low sample volume was due to laboratory processing. An inspection of the dataset revealed 40 lipid ratios had missing values, ranging from 1 to 347, due to the denominator lipid falling below the detection limit (recorded as zero). These missing values were addressed through imputation where zero values were assigned a value equivalent to half of the lowest observed value across the participants. All affected metabolite ratios were recalculated using the updated values. However, even after the initial imputation, 30 metabolites (again, mainly lipids) still had missing values. Among these, 27 metabolites had only one missing value each, while a glycolysis-related metabolite called pyruvate had the highest number of missing values, totalling 21. Further imputation was not performed to address these remaining missing data points. This decision was made because these measurements were rejected during quality control checks, not because they were below the detection limit.

4.2 Statistical analysis

Continuous variables were reported as mean and standard deviation (SD) if normally distributed and as median with interquartile range, if skewed. Categorical variables were presented as total numbers with corresponding percentages. Histogram plots were employed to visually inspect the distribution of individual metabolites, ensuring the absence of any prominent outliers and to assess for normal distribution. All analyses were performed using R version 4.0.3 (R Foundation for Statistical Computing, Vienna, Austria). See table 4-2 for the summary of the main packages used in the analysis.

Table 4-2 List of main R packages and functions used for analysis.

R package	Function	Brief description of use	Relevant analysis
mice	mice	Single imputation	4.1.5.8 dealing with missing data of covariate variables
varhandle	to.dummy	Data manipulation for LASSO	4.2.1.3 complementary analysis for associations of kidney measures with metabolomic markers
glmnet	cv.glmnet	LASSO analysis	4.2.1.3 complementary analysis for associations of kidney measures with metabolomic markers; 4.2.2.2 Development of metabolites-based risk score
survminer	ggcoxzph	Schoenfeld residuals- testing proportionality of predictors and convariates in CHP models	4.2.2.1 data manipulation (risk prediction models)
survival	surv	CPH for incident CKD	4.2.2 CPH modelling analysis
rms	validate	Internal validation using 500 bootstrap repetitions	4.2.2.4.1 Reference Score modification and evaluation- to address overfitting and estimate optimism adjusted c-statistic
nricens	nricens	NRI metrics	4.2.2.4.2 Metrics assessing risk prediction performance
survIDINRI	IDI.INF	IDI metrics	4.2.2.4.2 Metrics assessing risk prediction performance
pheatmap	pheatmap	Correlational heatmap	Correlational heatmap figures
Ggplot2	ggplot	Figure creation	Figures.
dplyr	various	Data manipulation/management	Used throughout steps described in methods/analysis
arsenal	tableby	Data summaries	Results chapter 5 and 6- variable descriptions

Abbreviations: CKD, Chronic kidney disease; CPH, Cox proportional hazards; IDI, integrative discrimination index; NRI, net reclassification improvement.

4.2.1 Associations between metabolites and kidney function measures

In order to ensure comparability of association coefficients and to fulfil the necessary condition for LASSO analysis, the original metabolite levels were natural log-transformed and standardised (mean= 0 and standard deviation (SD)= 1). This was achieved by subtracting the mean and subsequently dividing by the SD for each individual metabolite. All of the metabolite ratio measurements were excluded in order to reduce the number of comparisons.

The distribution of uACR was very left-skewed, so the variable was log-transformed, which resulted in approximately normal distribution, which was then used as a covariate in regression analyses. Smoking status was re-defined from the ET2DS baseline to only consist of two categories: ever smoker (originally recorded as current-smoker or ex-smoker) and non-smoker.

Cross-sectional associations between continuous baseline measures (eGFR and uACR levels) and individual metabolite concentrations were analysed using linear regression. Associations with baseline CKD status and albuminuria status were analysed using binary logistic regression. All of the regression models investigated the independent associations between individual metabolites and each outcome. The initial regression models were adjusted for baseline age and sex (age and sex-adjusted model). To further explore whether the association between metabolites and CKD phenotypes was independent of traditional kidney disease risk factors potential confounders were selected in accordance to a published risk prediction model (Nelson et al., 2019) . The following baseline covariates were adjusted for in the full model: age, sex, diabetes control (diet only, tablets only or insulin), uACR (or eGFR for uACR and albuminuria analyses), hypertension status, CVD status, BMI, ever-smoker status and HbA1c. Considering that the metabolomic panel consisted of many lipoprotein and lipid measures, the results may be affected by use of lipid lowering medication. To test the robustness of the findings, I conducted sensitivity analysis of the associations for baseline kidney function (eGFR and uACR) with additional adjustment for medication use.

The prospective analyses, considered either incident CKD or rapid decliner status and each metabolite was modelled using logistic regression and annual change in eGFR as percentage was investigated using linear regression. The models were adjusted for baseline eGFR, age and sex (Model 1 plus eGFR), followed by further adjustment with established CKD risk factors (Full model: Model 1 plus eGFR, diabetes control, uACR, hypertension status, CVD status, BMI, ever-smoker status and HbA1c). To test the robustness of the prospective findings, I conducted sensitivity analysis of the associations for incident CKD in a sub-population of ET2DS which excluded participants with CKD in terms of

reduced eGFR and/or albuminuria at baseline as well as anyone with a single eGFR < 60 ml min^{-1} (1.73 m)$^{-2}$ during the baseline visit.

Regression estimates were reported with 95% confidence intervals. To account for large number of statistical tests, known to increase false positive findings, the false discovery rate was controlled using the Benjamini-Hochberg method with up to 5% threshold for false positive results (Benjamini and Hochberg, 1995). The result was considered to be significant if multiple comparison adjusted P_{FDR} <0.05.

To explore if the observed associations could lie on the pathophysiological pathways of traditional CKD risk factors, the relationships between key metabolites and important covariates were assessed using correlations. Pearson's coefficients were calculated for correlations between key metabolites and continuous risk factors.

In addition, to further explore the metabolomic profiles related to CKD phenotype and evaluate the stability of findings in traditional analyses I employed a modern statistical technique called least absolute shrinkage and selection operator (LASSO). Since metabolomics data is characterised by high dimensionality and multi-collinearity, LASSO was appropriate as it shrinks the coefficients of some variables into exactly zero by adding penalty, allowing selection of variables deemed most valuable for explaining variability in the outcome (Tibshirani, 1996). Unlike the traditional analysis, where each metabolite is analysed individually, LASSO accounts for complex information among the whole metabolomic panel. Nevertheless, it is important to note that LASSO was utilised as a supplementary tool in my analysis of metabolomic profiles related to CKD phenotypes. Its application was reserved for instances where significant findings emerged from traditional regression analyses, ensuring a focused and targeted use of LASSO.

In this analysis, I employed a five-fold cross-validation process to select the tuning parameter lambda (λ) for LASSO, and I applied the one standard error rule. The parameter λ plays a critical role in LASSO as a hyperparameter controlling the degree of regularisation in the model. It effectively manages model complexity, performs feature selection, and facilitates an optimal trade-

off between bias and variance. To choose the best λ, I utilised cross-validation, opting not to directly select the λ with the lowest error but instead applying the "one standard error rule." This rule favoured a slightly larger λ, resulting in more regularization, which helps guard against overfitting by promoting a simpler model with slightly increased bias, but reduced variance, in turn promoting a balance between model complexity and performance. This selection method was executed using the "cv.glmnet" function within the "glmnet" package (version 4.0–2) in R. Specifically, I employed linear LASSO and logistic LASSO for continuous and binary outcomes, respectively. To ensure specific covariates were included (i.e., the established CKD risk factors used in traditional analysis), I set the penalty for their coefficients to zero.

4.2.2 Risk prediction for CKD using metabolites in Type 2 Diabetes

This sub-section outlines the statistical analysis procedure I carried out in order to modify the risk prediction model for incident CKD, originally developed by Nelson et al. (2019) (here referred to as the reference risk prediction model). The modification was performed using metabolomics data, resulting in what is referred to as the updated risk prediction model. Initially, 100 iterations of Cox LASSO analysis were conducted to identify a stable set of metabolites associated with incident CKD. The metabolites with the highest frequency of occurrence were considered as potential components for constructing a metabolite-based risk score (MetS). Subsequently, the reference risk prediction model (Nelson et al., 2019) was validated, then re-fitted in the ET2DS cohort and finally used as the foundation for developing the updated risk prediction model. Finally, I assessed the additional value provided by the MetS in predicting the risk of incident CKD within a 5-year timeframe, beyond the predictive capacity of traditional risk factors in the reference risk prediction model.

4.2.2.1 Data manipulation

Before constructing the metabolomics-based risk score (MetS), creatinine was omitted from the list of metabolites. Serum creatinine is a major component in eGFR calculation, which is already included in the predictor list of the reference risk prediction model. Since the aim of this analysis is to identify a sparse panel

of novel metabolites, including creatinine would defeat this purpose. The measurements in the metabolomic panel underwent a two-step transformation process: first, a logarithmic transformation, and then standardisation by subtracting the mean and dividing by the SD (mean= 0, SD= 1). This transformation was performed to align the data more effectively with the underlying assumptions of the Cox regression model.

Table 4-3 describes any data manipulation that ensured the variable definitions in the original publications for reference risk prediction model align with those available in the ET2DS for risk prediction model validation and modification.

Table 4-3 Summary of variable definitions in the reference model (Nelson et al., 2019a) and ET2DS data manipulation.

Predictors in reference model	Units	Description of variable	Variable in ET2DS	Units/ Description	Handling to match the reference model
Age	Per 5 years	Age at BL of CKD-PC study.	Age_BL	Years, calculated from DOB.	Rounded up to 0 decimals and divided by 5
Sex	Male/ Female	Records as per CKD-PC guideline. Categorical.	Sex_BL	Male/ Female. Self-reported in BL Quest 1 personal history.	None
Race/ ethnicity	White, Asian, Black, Hispanic	Adjustment only for black race in the model.	Ethnic_BL	Self-reported in BL Quest 1 personal.	One subject was of black ethnicity who had albuminuria, so was excluded from further analysis. Predictor omitted.
eGFR	mL/min/1.73m^2	Cr based CKD- EPI equation. Linear splines with a knot at eGFR 60-90, per –5 mL, eGFR ≥90, per –5	eGFR at BL	mL/min/1.73 m^2 , (BL_biochemistry-serum creatinine ne and Quest 1 for self-reported age, sex)	CKD-EPI equation based on baseline serum creatinine as described in previous Methods section 4.1.5). Linear splines: Category 60-90 per 5mL: (eGFR)/-5 Category 90 or more per 5mL: (eGFR)/-5
Hypertension	BP - mmHg or medication use	SBP 140 and DBP 90 mmHg; use of antihypertensive medications	SBP, DBP; anti-HTN medication use	mmHG, measured (BL_data_Physiological); Meds- BP lowering: Y/N (BP_lowerBL from summary medication, derived from medication	SBP => 140 AND DBP=>90; OR Yes to medication use == Hypertension

				questionnaire, self-reported-MedBP_BL (Y/N/DN))	
Smoking	Ever smoking vs never smoking	Status reported by studies individually	Smoker- 1. No 2. Ever	BL_ quest 1 Smoking Questionnaire, self-reported	ever smoker= smoker present or ex-smoker.
History of CVD	myocardial infarction, coronary revasculari sation, angina, heart failure, or stroke	Status reported by studies individually	Prevalent events 2006/200 7 variables: myocardi al infarction, coronary interventi on, Stroke, TIA, angina	Derived from BL_quest Medical hist,. Self-reported questionnaire).	CVD if yes to any of the listed prevalent variables.
BMI	Kg/m^2 per 5 points	calculated as weight in kilograms divided by height in meters squared	BMI_BL	kg/m^2 , calculated from weight and height measured at BL (BL_data collection Physiological)	Rounded to 0 decimals, and divided by 5.
Urinary ACR	mg/g per 10-fold increase		albACR_ BL	mg/mmol, Urinary measurement (BL_ biochemistry)	Mg/mmol divided by conversion factor 0.113 = mg/g. Natural $\log_{10}$ of ACR
HbA1c	per 1%		HbA1c	% total; measured (BL_ biochemistry	Rounded to 0 decimal places.
Insulin vs tablets diabetes medication	at 7% HbA1c		Insulin use Oral diabetes meds.	summary medication, derived from medication questionnaire, self-reported: diabetes tablets, diabetes injections, dietary control	A categorical variable with 3 mutually exclusive categories: 0=diet only; 1=tablets only; 2= insulin injections (with or without tablet us/diet control)
None vs tablets diabetes medication					

Abbreviations: ACR, albumin to creatinine ratio; BL, baseline; BMI, body mass index; BP, blood pressure; CVD, cardiovascular disease; DOB, date of birth; eGFR, estimated glomerular filtration rate; HbA1c, glycated haemoglobin, TIA, Transient ischemic attack.

Moreover, the proportional hazards assumption in Cox models suggests that the effect of a predictor on the hazard of an event is consistent over the entire follow-up period. If this assumption is violated, it suggests that the effect of predictor varies over time, which directly impacts the interpretation and reliability of the model. To ensure the assumptions of Cox regression proportional hazard model were met, a check was conducted using scaled

Schoenfeld residuals (In and Lee, 2023). These residuals indicate the disparities between observed event times and the expected event times for each individual at various time points. If the assumption holds, then the Schoenfeld residuals are scattered randomly around zero when plotted against time (i.e., no visible trends). If it was found that some predictors violated these assumptions, interactions between age and those specific predictors would be introduced into the Cox model to address this issue. The interactions with age were chosen because it would allow the effect of predictor to vary as a function of age and capture any time-dependent changes.

4.2.2.2 Development of metabolites-based risk score (MetS)

Throughout my PhD project, I intended to primarily apply LASSO as a fundamental technique for analysing metabolomics data. LASSO offers an advantage over traditional stepwise regression models by addressing prediction optimism. Consequently, I have chosen to employ it as a filtering mechanism to identify a sparse and consistent set of metabolites that exhibit predictive capabilities for incident CKD risk. Once these metabolites have been identified, the next step involved amalgamating them into a unified score using an unpenalized and unadjusted Cox model. This methodology is derived from a study conducted by (Ganz et al., 2016) and was also utilised by my colleague who studied CVD risk in ET2DS, which we published previously (Huang et al., 2023).

To establish a robust set of components for the metabolite-based score, MetS, I conducted 100 iterations of Cox LASSO using Bootstrap with replacement (figure 4-1). During these iterations, I varied the cross-validation fold and random seeds, employing the "cv.glmnet" function within the 'glmnet' package (version 4.0–2) in R. In order to create MetS, which is a representative selection of metabolites that one would expect when employing LASSO with regularisation, guided by the 1-standard error rule (see section 4.2.1 for more detailed explanation), I decided to only consider metabolites that were selected most frequently in the LASSO iterations. This decision was guided by a ‘frequency threshold’ approach, where the threshold for frequency was chosen where a natural break in distribution of frequencies for each metabolite occurs.

Metabolites that emerged within the top frequencies were deemed eligible components for the MetS.

It is essential to note that LASSO was exclusively utilised for variable selection in this context. Following the determination of the metabolites, they were integrated into a single score, which I called MetS using a semi-parametric survival model, specifically, an unadjusted and unpenalized Cox model. The linear predictor within this Cox model, calculated as the sum of regression coefficients multiplied by the corresponding metabolite concentrations, served as the weighted MetS for each participant.

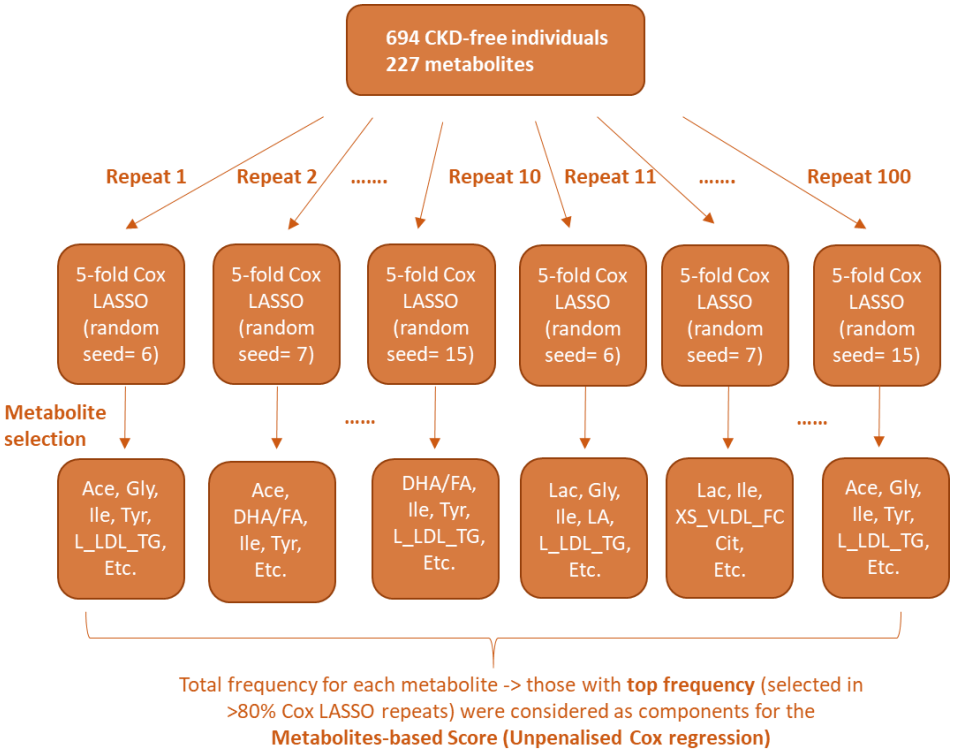


Figure 4-1 Selection procedure of candidate metabolites in 100 repeats of Cox LASSO.

4.2.2.3 Reference risk prediction model validation

As introduced in the background chapter, a recent systematic review (Slieker et al., 2021) suggested that the reference risk prediction model (Nelson et al., 2019) showed the best performance in terms of apparent discrimination among people with type 2 diabetes. Nelson equation predicted confirmed eGFR <60 ml min^{-1} $(1.73m)^{-2}$, which was defined as at least 2 consecutive eGFRs < 60

ml min^{-1} $(1.73m)^{-2}$ at least 3 months apart (but not longer than over 2 years in between measurements).

To ensure that the reference risk prediction model developed by (Nelson et al., 2019), performs as well as suggested by previous literature, it was examined in ET2DS using a process known as external validation (i.e., assessment of predictive performance in data not used for mode development). As per recommendation (Bleeker et al., 2003, Reilly and Evans, 2006, Toll et al., 2008, Royston and Altman, 2013), to obtain accurate and precise estimation of predictive performance by the reference risk prediction model in terms of calibration and discrimination, the definitions of predictors were applied exactly as described in the original publication by Nelson et al. (2019). In order to achieve this, I calculated the risk for the confirmed incident CKD event within 5 years of follow-up in CKD-free (at baseline) ET2DS participants using the published risk equation shown in Box 4-1 (Nelson et al., 2019).

Box 4-1. The reference risk prediction model equation for risk calculation:

Published prediction equation applied to individual patients for the 5-year absolute risk of incident eGFR 60 ml min^{-1} $(1.73m)^{-2}$

exp [-5^0.9212477 × exp [-3.070735 + 0.1351572 × (age/5 – 11) + 0.1381975 × (if female) + 0.0920208 × (if black*) + 0.3546697 × (15 – min(eGFR, 90)/5)– 0.1525133 × max(0, eGFR-90)/5 + 0.1870637 × (if has history of CVD) + 0.0619679 × (HbA1c –7) + 0.1078296 × (if insulin use) – 0.150944 × (if no DM medication use) + 0.023959 × (HbA1c –7) × (if insulin use) + 0.0398424 × (HbA1c –7) × (if no DM medication use) –0.00084 × (if ever smoking) + 0.3653268 × (if hypertensive) + 0.050306 × (BMI/5-5.4) + 0.3737905 × (log_{10}ACR – 1)]]

In addition, to test that the reference model is superior in terms of predicted risk in ET2DS, as suggested by previous studies (Slieker et al., 2021), I also compared it with risks calculated using two simpler equations (herein referred to as the Chien equation (Chien et al., 2010) and the O'Seaghdha equation

(O'Seaghdha et al., 2012) for predicting a 5-year risk of incident CKD. The equations were applied to ET2DS dataset as shown in Box 4-2 and Box 4-3 for Chien equation and O'Seaghdha equation, respectively. I chose these equations as comparators, as both were also used for comparison in the development study of the reference model, the models consisted of variables available in ET2DS, predicted a relevant outcome and had replicable equations published (Nelson et al., 2019).

Box 4-2. Chien equation:

Published prediction equation applied to individual patients for the 4-year absolute risk of incident eGFR 60 ml min^{-1} $(1.73m)^{-2}$*

1 - 0.9632 ^ exp (-6.8 + 0.077 × age (in years) + 0.366 × (if diabetic) + 1.24 × (if history of stroke) + 0.059 × Body mass index (kg/m^2) + 0.018 × diastolic blood pressure (in mmHg))

*Equation taken from Nelson model publication (supplementary material).

Box 4-3. O'Seaghdha equation:

Published prediction equation applied to individual patients for the 5-year absolute risk of incident eGFR <60 ml/min/1.73m2 *

1 – (1 – 0.092) ^ 0.5 / ((1 – 0.092) ^ 0.5 + (1 – (1 – 0.092) ^ 0.5) × exp(-6.235 + 0.095 × age (in years) + 0.476 × (if diabetic) + 0.761 × (if hypertensive) + 0.779 × (if 75 ≤ eGFR < 90) + 1.558 × (if 60 ≤ eGFR <75) + 0.300 × (if ACR ≥ 30 or dipstick ≥ trace)))

In brief, the Chien equation (Chien et al., 2010) originated from a study involving 5,168 Chinese individuals who underwent initial health assessments at the National Taiwan University Hospital, followed by yearly check-ups that included measurements of serum creatinine levels to evaluate the changes in eGFR. Over a median follow-up period of 2.2 years, 190 individuals in this group developed CKD. Chien et al. (2010) used the best subset multivariate Cox proportional hazards model to establish a parsimonious model according to backward selection strategy. I used the clinical equation that predicted incident

CKD defined as an eGFR of less than 60 ml min^{-1} $(1.73m)^{-2}$ and consisted of age, BMI, diastolic blood pressure, and history of T2DM and stroke as predictors. The variables in Chien equation, did not require any manipulation as they were recorded for ET2DS exactly as describe in the development study.

The O'Seaghdha equation (O'Seaghdha et al., 2012) was developed using data primarily from a white population in Framingham, Massachusetts. This model relied on baseline serum creatinine levels and a follow-up measurement taken a decade later. Within a group of 2,490 individuals aged 45 to 64 years in this particular study, 229 developed an eGFR of less than 60 mL/min/1.73m² after 10 years. O'Seaghdha et al. (2012), used a stepwise logistic regression to identify CKD prognostic factors and selected these to construct a risk score predicting 10-year chronic kidney disease risk. The O'Seaghdha equation included age, hypertension, diabetes, eGFR category, and albuminuria values as predictors. In terms of variables used in the O'Seaghdha equation, compared to those available in ET2DS, only hypertension status variable had to be created, which was defined the same as for the reference risk prediction model (systolic blood pressure >140 mmHg and diastolic blood pressure >90 mmHg and/or use of antihypertensive medications). Also, eGFR was categorised as defined in the equation (Box 4-3).

Brier scores (Brier, 1950), the mean squared difference between the predicted risk vs observed binary outcomes (see equation below), were used to evaluate which risk equation showed the best calibration within each cohort. In this equation, y_i is the observed outcome (y=0 for no CKD and y=1 for CKD) and p_i is the risk score for the i^{th} person:

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - p_i)^2$$

4.2.2.4 Reference Score modification and evaluation

Once I established that the risk prediction model developed by Nelson et al. (2019) was adequate to use as the reference model (Nelson et al., 2019), I modified it with MetS by combining the two. In order to achieve this, I retrained the reference model (Nelson et al., 2019) by re-estimating the refence model coefficients when applying it to the ET2DS data. Cox proportional hazards models were used to estimate the association between levels of individual variables and risk of incident CKD. Finally, this re-fitted reference model was modified by adding a new variable, named metabolites-based risk score (MetS) to develop the updated risk prediction model. To address the concern of overfitting, I conducted internal validation using 500 bootstrap repetitions, which was carried out using the "validate" function within the 'rms' R package. This process was employed to reduce the optimism of the predictive performance of the refitted reference score before and after the modification with MetS. The evaluation of predictive performance involved the re-fitted reference model, MetS in isolation, and their joint application, each assessed independently through discrimination and calibration measures (described in the following section 4.2.2.5). Furthermore, I assessed the additional predictive contribution of the MetS beyond the traditional risk factors CKD prediction. This assessment was based on continuous net reclassification improvement (NRI) and integrated discrimination improvement (IDI) analyses (described in the following section 4.2.2.5).

4.2.2.5 Metrics assessing risk prediction performance

Discrimination statistic evaluates the ability of the risk prediction model to distinguish between individuals at high and low risk (Alba et al., 2017). The primary metric used for assessing discrimination in general linear regression models is the concordance (c) statistic (Steyerberg, 2019). In binary prediction scenarios, this is equivalent to the area under the curve (AUC). It represents the probability that, among randomly selected "event" and "non-event" individuals, the "event" individual will have a higher predicted risk. A higher c-statistic indicates better discrimination, where 1 suggests perfect discrimination. However, traditional c-statistics are not directly applicable to time-to-event data. Therefore, Harrell's c-statistic was developed to evaluate

discrimination in survival analysis, measuring the proportion of usable patient pairs with concordant predictions and outcomes (Harrell et al., 1996).

While the c-statistic serves as a valuable tool to assess model performance, it has limitations when evaluating the addition of new biomarkers to models already containing strong predictors (Price et al., 2017, Moons et al., 2012b). To address this, metrics like the net reclassification improvement (NRI) were developed. The NRI assesses how the introduction of novel predictors reclassifies individuals into risk categories. Subjects correctly re-categorised as higher risk (+1) or lower risk (-1) in the new model are assigned specific values, while those remaining in the same category receive a value of 0. The sum of these values informs the NRI calculation (Pencina et al., 2008, Pencina et al., 2011). The NRI metrics were calculated with nricens function (nricens R package).

The integrated discrimination improvement (IDI) is another valuable tool for evaluating reclassification. The IDI is calculated as the disparity in discrimination slopes between the two models under comparison, with discrimination slopes indicating the slope of a linear regression relating the predicted probability of the 'event' group to that of the 'non-event' group (Pencina et al., 2017). It quantifies the difference in sensitivity and specificity integrals between models with and without novel biomarkers. IDI was calculated using the IDI.INF function in the survIDINRI R package. A larger IDI suggests improved risk prediction and it is calculated using the following equation (Kerr et al., 2011):

$$IDI = (IS_{new} – IS_{old}) – (IP_{new} – IP_{old})$$

Where:

IS: the integral sensitivity over all possible cut-off values

IP: the corresponding integral of ‘1-specificty’,

Subscript of ‘new’: the model with novel biomarkers

Subscript of ‘old’ : the reference model without the markers

Calibration, measures the agreement between observed outcomes and predicted risks, which is a vital property of a risk prediction model (Steyerberg, 2019). Calibration plots provide a visual representation of calibration, with the x-axis representing predicted risks and the y-axis indicating observed outcomes. A perfect prediction aligns with the 45° line (Steyerberg, 2019). When a model exhibits good calibration, the actual occurrence rate of the event within various groups should closely align with the predicted event probability generated by the model. I used an adapted calibration plot that compares predicted risks and observed event frequencies for enhanced visualisation (Ganz et al., 2016, Hippisley-Cox et al., 2017), and this adapted plot was used to present the results in my thesis.

In summary, I developed a multi-marker score called MetS using a multivariable selection method LASSO. The sum of selected metabolites coefficients was used as a new predictor, which was used to modify the reference model, in turn developing an updated risk prediction model. Then, I evaluated discriminative ability of the reference and updated models, the reclassification performance, and calibration to comprehensively assess its predictive accuracy in the context of risk prediction for 5-year risk of incident CKD.

Chapter 5. Results I: metabolomic profiles of declining kidney function in T2DM

This chapter describes baseline characteristics of the Edinburgh Type 2 Diabetes Study (ET2DS) participants, summaries of kidney function measures available, as well as the description of measured metabolite levels in this cohort of patients with type 2 diabetes mellitus (T2DM). The metabolomic profiles according to estimated glomerular filtration rate (eGFR), chronic kidney disease (CKD) status and albuminuria at baseline are presented. Longitudinal changes in eGFR and incident CKD are also described along with their associated metabolomic profiles.

5.1 Characteristics of the studied population

5.1.1 Baseline demographics

The baseline demographic and clinical characteristics of the ET2DS cohort (N=1,066) are summarised in Table 5-1. The average age of the entire population was 67.9 (standard deviation (SD)= 4.2) years, with 547 (51.3%) male participants. The average body mass index (BMI) for the population was 31.4 (SD= 5.7) kg/m^2. While the average systolic blood pressure was 133.3 (SD= 16.4) mmHg (below hypertensive level of 140 mmHg), the majority of individuals were considered to have evidence of hypertension (82.5%) based on the reported blood-pressure lowering medication use. The prevalence of cardiovascular disease (CVD) in the ET2DS at baseline was 35.3% (376 individuals having a history of one or more cardiovascular events prior to baseline). A minority of individuals (14.4%) were current smokers, but the majority of those who are not currently smoking had quit smoking more than 6 months prior to baseline visit (46.7%). The mean plasma HbA1c level was 7.4 (SD= 4.1) %, and the median duration of diabetes was 6 years (interquartile range (IQR) 2, 10). More than half of subjects (62.9%) received oral medication for diabetes control. Overall, the baseline characteristics of interest had low levels of missing data, with the maximum levels of 2.8% missing results seen in diabetes medications. The data were complete for age, sex, smoking status and prevalent CVD variables.

Table 5-1 Baseline characteristics of the ET2DS (N=1,066)

	Before Imputation			After imputation
Variable	N	Missing	Mean, SD/ IQR/ N (%)	Mean, SD/ IQR/ N (%)
Age (years)	1,066	0 (0.0)	67.9, SD= 4.2	67.9 SD= 4.2
Sex: male	1,066	0 (0.0)	547 (51.3)	547 (51.3)
Sex: female	1,066	0 (0.0)	519 (48.7)	519 (48.7)
BMI (kg/m^2)	1,065	1 (0.1)	31.4, SD= 5.7	31.4, SD= 5.7
Diabetes duration (years)	1,053	13 (1.2)	6 (2, 10)	NA
HbA1c (% Total)	1,057	9 (0.8)	7.4 SD= 1.1	7.4 SD= 1.1
Plasma glucose (mmol/l)	1,049	17 (1.6)	7.6 SD= 2.1	NA
Diabetes control				
Oral medication only (%)	1,036	30 (2.8)	652 (62.9)	678 (63.6)
Insulin injections (%)	1,036	30 (2.8)	186 (18.0)	190 (17.8)
Diet only (%)	1,036	30 (2.8)	198 (19.1)	198 (18.6)
Cardiovascular-related				
SBP (mmHg)	1,064	2 (0.2)	133.3, SD=16.4	133.3, SD= 16.4
DBP (mmHg)	1,064	2 (0.2)	69.1, SD= 9.0	69.1, SD= 9.0
Blood-pressure lowering meds (%)	1,059	7 (0.7)	873 (82.4)	880 (82.6)
Hypertension	1,059	7 (0.7)	874 (82.5)	881 (82.6)
Lipid-lowering meds (%)	1,064	2 (0.2)	912 (85.7)	914 (85.7)
Prevalent CVD	1,066	0 (0.0)	376 (35.3)	376 (35.3)
Smoking status				
Current smoker (%)	1,066	0 (0.0)	154 (14.4)	154 (14.4)
Ex-smoker (%)	1,066	0 (0.0)	498 (46.7)	498 (46.7)
Never smoked (%)	1,066	0 (0.0)	414 (38.8)	414 (38.8)

Data are presented as mean and standard deviation, median (lower, upper interquartile range (IQR)) or n (%). Abbreviations: BMI, Body mass index; CVD, cardiovascular disease; DBP, Diastolic blood pressure; HbA1c, glycosylated haemoglobin; SBP, Systolic blood pressure; NA, not applicable.

5.1.2 Serum metabolites

The metabolomic panel measured 228 different low-molecular weight metabolites (149 concentrations of individual metabolites and 79 derived ratios)

in a cohort of 1,058 participants from the ET2DS. After imputation of metabolites which had concentration of zero (as detailed in section 4.1.8), the average number of missing metabolite measurements was very low (0.03%, [range 0% – 1.99%]) and should not impact the results of subsequent metabolomic profile analyses. Appendix B Table S5-1 provides details on the summary statistics for each individual metabolite. Appendix B Figure S5-1 displays histograms depicting the distribution of these individual metabolites, which indicated that the majority followed a normal distribution pattern. Notably, there were substantial interrelationships observed among individual metabolites, especially within the category of lipid-related metabolites, with pairwise correlation coefficients between lipids ranging from -0.99 to +0.99 (Figure 5-1).

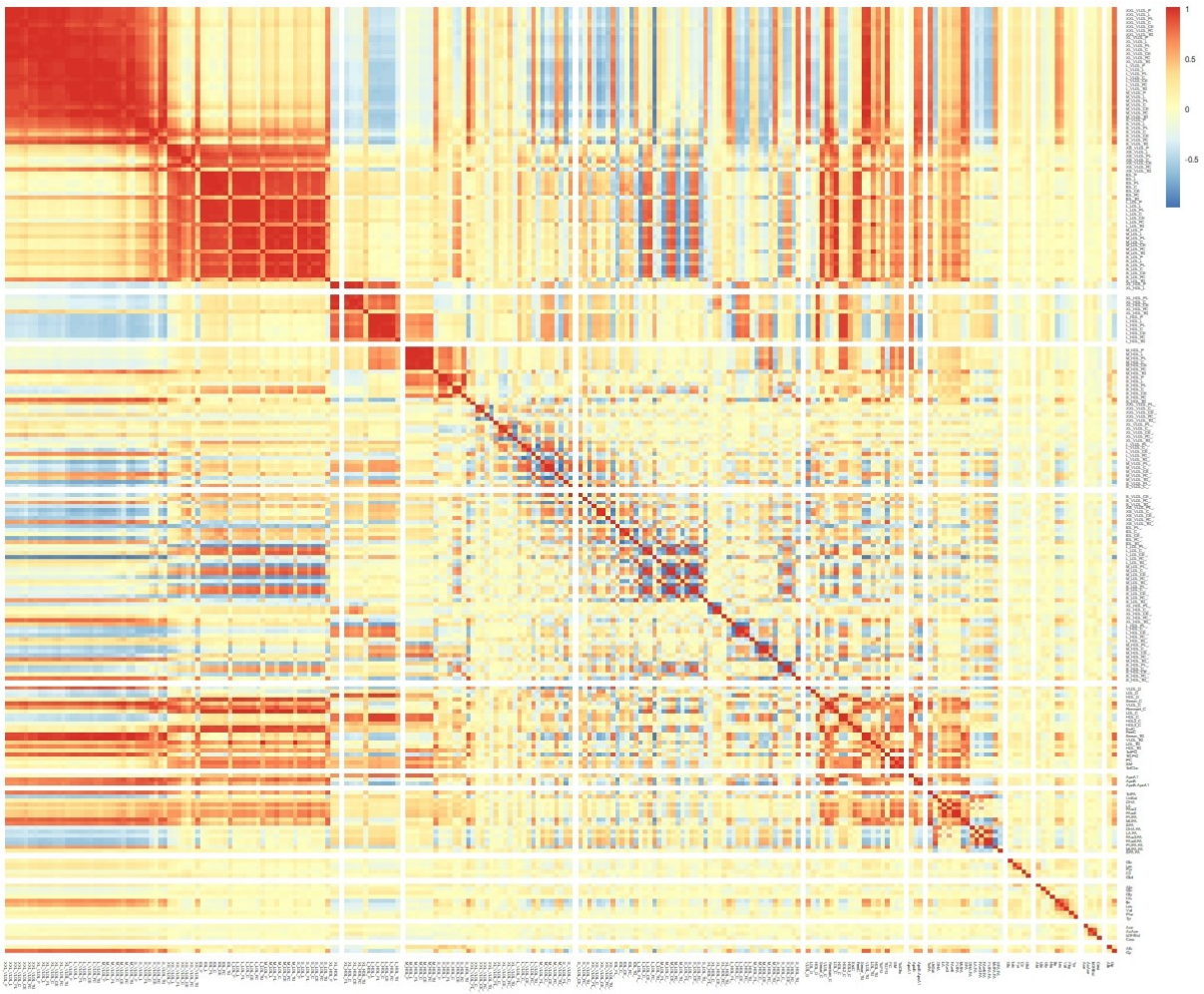


Figure 5-1 Correlation matrix of serum metabolites in the ET2DS (N=1,058).

5.2 Cross-sectional analyses of metabolomic profiles in ET2DS

This section explores the metabolomic profiles of kidney function in terms of associations between serum metabolites and relevant kidney function measures at baseline. After describing the individual outcomes studied, the focal point of the section summarises the results of the cross-sectional analysis, which involved traditional regression models as well as complementary multivariate analysis using least absolute shrinkage and selection operator (LASSO) models.

5.2.1 Description of kidney function measures and outcomes

5.2.1.1 Missing data and included participants

The total number of participants included in each cross-sectional analysis and by outcome, is presented in table 5-2, along with reasons for excluding subjects from each analysis. Historical serum creatinine was missing for 14 participants, so their CKD status was based on baseline eGFR result alone. At baseline, eGFR was missing in eight participants, so their CKD status was determined using historical eGFR results. All eight had a measurement done within 6 months of baseline visit, which was used to substitute their baseline eGFR. Overall, all 1,066 ET2DS participants had at least a historical or baseline eGFR to define their CKD status. The baseline uACR measurements were missing for 20 participants (15 missing urinary albumin and 11 missing urinary creatinine measurements), so their albuminuria status was based on historical albuminuria levels alone. There were 309 participants with no historical albuminuria results (n=236 with no historical data; n=73 with two or less consecutive measurements within two years before baseline). For 308 of these participants albuminuria status was determined using baseline uACR alone, and one who had missing urinary creatinine level had a urinary albumin of >100 mg/mmol, which suggested that albuminuria is highly likely. Overall, all 1,066 participants in ET2DS had either historical or baseline measurements to determine the albuminuria status at baseline. There were eight participants who did not have any metabolomic profiling data and thus were not included in any of the subsequent analyses, leaving 1,058 participants for cross-sectional analysis of metabolomic profiles related to CKD.

Table 5-2 Summary of number of participants included in the cross-sectional analysis of the metabolomic profiles of baseline kidney function in the ET2DS.

Trait	Included subjects	Missing	Reason missing data/ exclusion from analysis
eGFR, ml min^{-1} $(1.73\ m)^{-2}$	1058	8 (0.8)	Eight missing metabolomics data due to insufficient serum for metabolic profiling
CKD	1058	8 (0.8)	Eight missing metabolomics data
uACR, mg/mmol	1038	28 (2.6)	20 missing uACR (15 missing urinary albumin and 11 missing urinary creatinine measurements); eight missing metabolomics data
Albuminuria	1058	8 (0.8)	Eight missing metabolomics data

Data are presented as n (%). Abbreviations: CKD, chronic kidney disease; eGFR, estimated glomerular filtration rate; uACR, urinary albumin to creatinine ratio.

5.2.1.2 Description of kidney function measures at baseline

The mean (SD) baseline levels of eGFR in the included ET2DS participants (n=1,058) was 77.3 (SD= 18.9) ml min^{-1} $(1.73\ m)^{-2}$ and the median uACR for 1,038 available participants was 1.17 (IQR 0.77, 1.99). Both kidney function measures indicated that an average participant in ET2DS does not have evidence of CKD. The distribution of continuous kidney function measures (eGFR and uACR at baseline) is presented in figure 5-2. The eGFR at baseline (figure 5-2A), showed slightly left-skewed distribution, which appeared to be a result of participants with CKD at baseline. Similarly, uACR at baseline (figure 5-2B.), showed slightly right- skewed distribution even after log-transformation, but the skew appeared to be caused by the presence of a small number of patients with very high albuminuria in the ET2DS cohort. The overlap between cases and no cases (group 1 and 0, respectively) in the histograms of baseline eGFR reflects that CKD case status was defined as two of three eGFR measurments below 60, while the histograms show the single measure of baseline eGFR (Figure 5-2A). Likewise for albuminuria cases and uACR histogram Figure 5-2B).

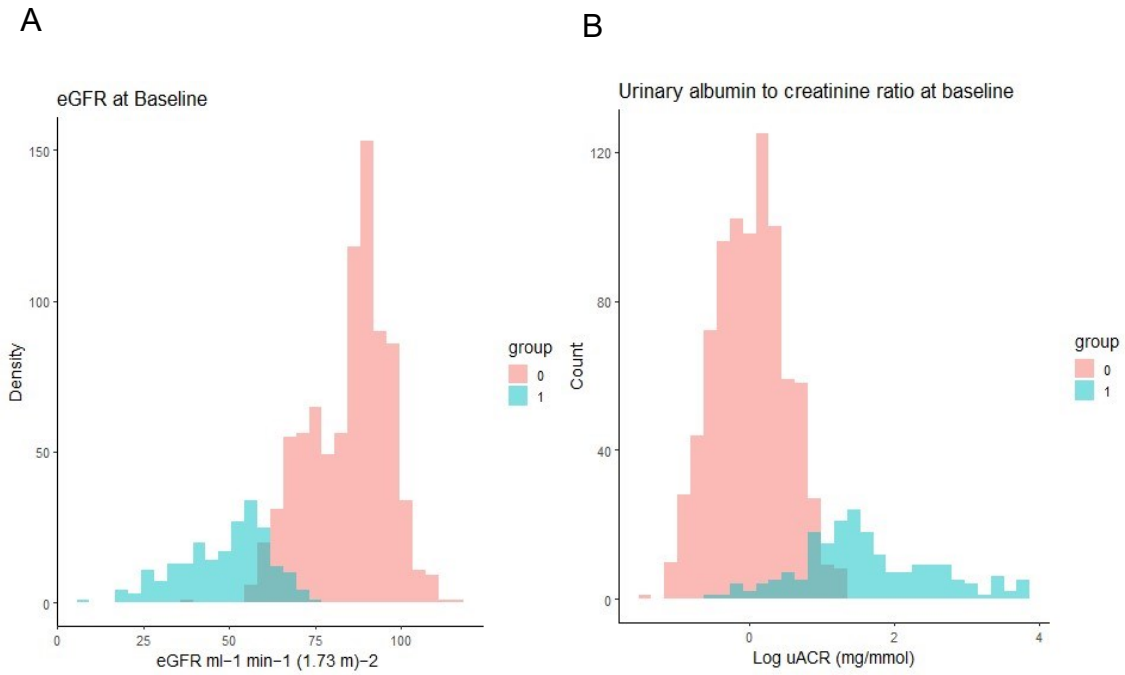


Figure 5-2 Distribution of baseline eGFR and log of uACR.

Left panel, Group 0 = no CKD, Group 1 = CKD at baseline; right panel, Group 0 = no albuminuria, Group 1 = albuminuria. eGFR, estimated glomerular filtration rate; uACR, urinary albumin to creatinine ratio.

Although creatinine measured via the metabolomics platform was excluded from the main analysis, its measurement by two independent methods (metabolomics and clinical blood serum biochemistry analyser) made it worthwhile to investigate the agreement between the two. I performed a Pearson correlation analysis, presented in Figure 5.3, which revealed a strong positive correlation (coefficient= 0.96, close to perfect correlation of 1) between the creatinine levels obtained from both methods. This, in turn, supports the validity of the case-control definitions for CKD.

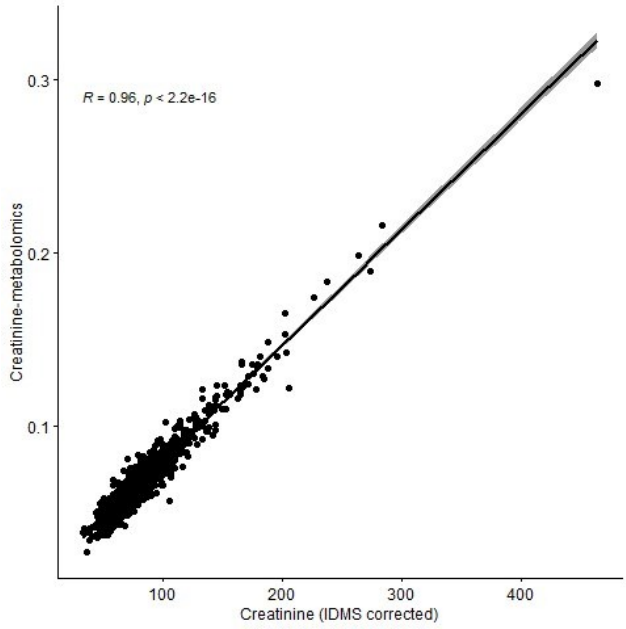


Figure 5-3 Correlation between creatinine measured using two independent methods: metabolomics and clinical chemistry analyser.

The number of cases and non-cases for binary outcomes (reduced eGFR and albuminuria) are summarised in table 5-3. Prevalence of persistently reduced eGFR (indicative of CKD) among included ET2DS participants at baseline was 20.4%. Prevalence of albuminuria among the included participants (n=1,058) was 19.9%. Based on persistent albuminuria and reduced eGFR status, 33.3% of individuals had evidence of CKD at baseline.

Table 5-3 Contingency table for number of study participants who achieved binary kidney function outcomes in the ET2DS.

	Reduced eGFR present	Normal eGFR	Total
Albuminuria present	75	136	211
Albuminuria absent	141	706	847
Total	216	842	1058

Abbreviations: CKD, chronic kidney disease.

Data presented as numbers of individuals in the study; Reduced eGFR defined as persistently reduced eGFR below 60 (indicative of CKD) at baseline based on consecutive baseline and historical eGFR measurements; normal eGFR, eGFR above 60 based on consecutive baseline and historical eGFR measurements; Albuminuria, evidence of increased urinary albumin to creatinine ratio based on consecutive baseline and historical measurements.

5.2.2 Cross-sectional analysis of metabolomic markers associated with baseline eGFR and CKD

The creatinine was included in the metabolomics panel (with 149 metabolites measured in total), however since it is the key component of the calculation for eGFR, it was not considered in the analyses concerned with baseline eGFR. Linear regression analysis revealed that 103 metabolites were significantly associated with baseline eGFR at $P_{FDR} < 0.05$ in age and sex-adjusted model. Further adjustment for CKD risk factors (full model included adjustment for age, sex, diabetes control, uACR, hypertension status, CVD status, BMI, ever-smoker status and HbA1c) decreased the number of significant associations ($P_{FDR} < 0.05$) to 88 metabolites (Figure 5-4; see appendix Table S5-2 for detailed P values and β coefficients).

The majority of metabolites were inversely associated with eGFR. The most prominent association was with very-low-density lipoprotein (VLDL) subclasses, specifically triglycerides in very small VLDL particles (β = -3.95 (95% Confidence interval (CI) -4.97, -2.92)), $P_{FDR}=7.04*10^{-12}$) and cholesterol in small VLDL (β = -3.92 (CI -4.94, -2.89), $P_{FDR}=7.04*10^{-12}$). Triglycerides in intermediate density lipoprotein (IDL) and low-density lipoprotein (LDL), glycolysis-related glycerol and citrate, inflammation-related glycoprotein acetyls as well as amino acids phenylalanine, glycine and isoleucine, were all inversely associated with eGFR. Conversely, glucose, apolipoprotein A1 (ApoA1) and high-density lipoprotein (HDL) subclass were positively associated with eGFR, with the strongest effect demonstrated in total HDL cholesterol (β = 3.53 (CI 2.44, 2.62), $P_{FDR}=4.36*10^{-9}$). Most of the significant associations ($P_{FDR}<0.05$) remained the same in terms of significance and the effect direction after additional adjustment for lipid lowering medication use (see appendix B, Table S5-3). The only exception was identified for free cholesterol in very large HDL, which only showed significant association with baseline eGFR before additional adjustment for lipid lowering medication.

Baseline CKD status (based on persistently reduced eGFR at baseline and up to 2 years before baseline) was significantly associated with 101 metabolites (excluding creatinine) of 148 included metabolite concentrations in age and

sex-adjusted model, which again slightly decreased to 87 significant associations ($P_{FDR} < 0.05$) in the fully adjusted model (Figure 5-4B, appendix B Table S5-4). Similarly, to baseline eGFR results, triglycerides in very small VLDL demonstrated the strongest associations (odds ratio (OR) 1.80 (95% CI 1.50, 2.19), $P_{FDR} = 7.14*10^{-8}$). Most of the associations reflected those described for eGFR, with exception for lipid traits in very-large HDL and total cholines, which were only associated with CKD. Conversely, free cholesterol and phospholipids in small HDL as well as total fatty acids, degree of unsaturation and glucose were only significantly associated with baseline eGFR.

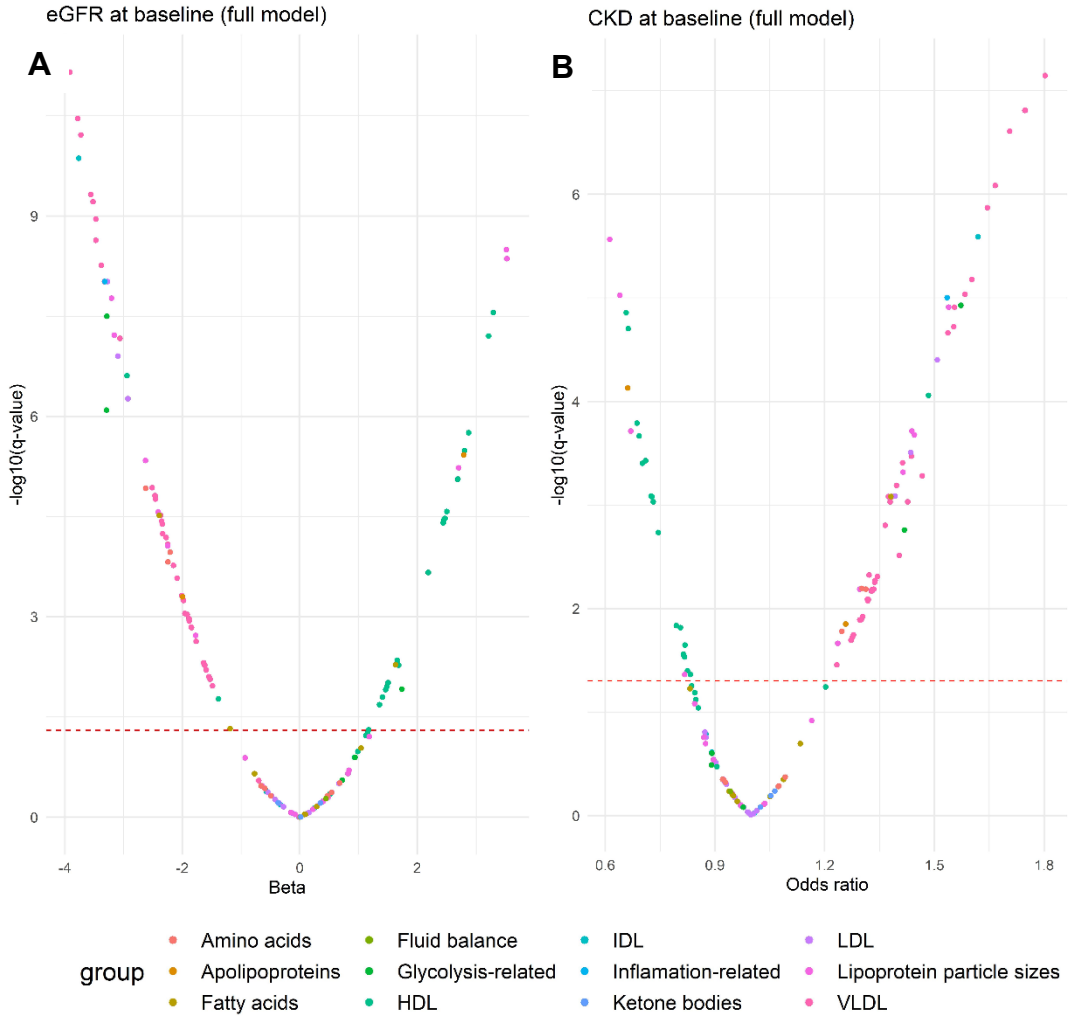


Figure 5-4 Association between 148 individual metabolites (except creatinine) and baseline eGFR and CKD in the ET2DS (n= 1,058).

Red dashed line represents significance level, metabolites above it were significant at $P_{FDR}<0.05$. **A panel:** Baseline eGFR (Full model adjusted for age, sex, diabetes control, uACR, hypertension status, CVD status, BMI, ever-smoker status and HbA1c); **B panel:** Baseline CKD (Full model adjusted for same covariates as listed above).

5.2.3 LASSO regression models of kidney function at baseline

In the complementary multivariate analysis using LASSO regression with age and sex forced into the models, nine metabolites were selected in association with baseline eGFR and seven metabolites with baseline CKD status (Table 5-4). As expected, creatinine was selected in addition to a sparse panel of uncorrelated metabolites predicting eGFR or CKD status at baseline (Figure 5-4A and B, respectively). Comparison between metabolites selected in LASSO and the statistically significant associations in traditional regression analyses with adjustment for age and sex showed a considerable overlap, with stable direction of effect in majority of the overlapping metabolites (see Venn diagrams for baseline eGFR and CKD status in Figures 5-5 and 5-6, respectively). This indicates that the associations found in the analysis that considered individual metabolites are robust in terms of reliability and consistency. The findings reported in the cross-sectional analysis are independent of the specific statistical analysis method I used for analysis.

Table 5-4 Selected metabolites in LASSO regressions for baseline eGFR and CKD status.

Outcome	**eGFR**		**CKD**
Metabolite	**Beta coefficient***	**Metabolite**	**Odds ratio***
Creatinine	-3.41	Creatinine	2.74
Albumin	0.47	Albumin	0.86
Cholesterol in very small VLDL	-0.11	Cholesterol in very small VLDL	1.01
Phospholipids in small LDL	0.32	Phospholipids in small LDL	0.90
VLDL diameter	0.65		
Omega-3	0.01		
Glucose	0.18		
Isoleucine	0.02		
Phenylalanine	-0.05		
		Free cholesterol in very small VLDL	1.01
		Cholesterol esters in medium HDL	0.99
		Citrate	1.03

*Models were adjusted for age and sex. Abbreviations: CKD, chronic kidney disease; GFR, estimated glomerular filtration rate; HDL, high density lipoproteins; LDL, low density lipoproteins; VLDL very-low density lipoproteins.

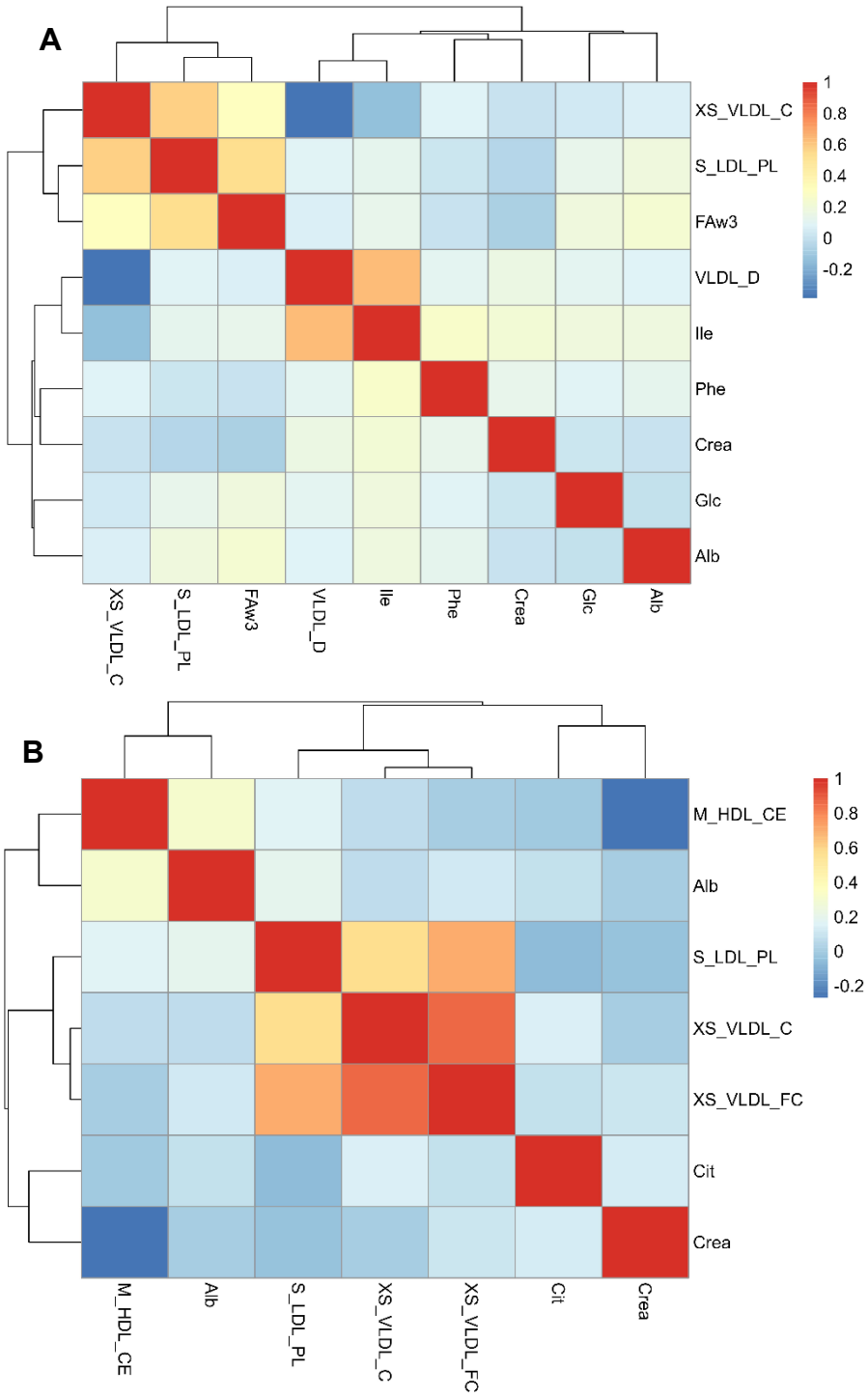


Figure 5-5 Correlational heatmap between the metabolites selected by the LASSO regression model in the ET2DS (n= 1,058).

A: Metabolites selected for eGFR at baseline as the outcome. **B:** Metabolites selected for CKD status at baseline as the outcome. Red and blue colours indicate positive and negative direction of correlation, respectively and colour intensity is relative to the strength of correlation coefficient. Abbreviations: Alb, albumin; Cit, citrate; Crea, creatinine; FAw3, Omega-3 fatty acids; Glc, glucose; Ile, isoleucine; M_HDL_CE, Cholesterol esters in medium HDL; Phe, phenylalanine; S_LDL_PL, Phospholipids in small LDL; VLDL_D, Mean diameter for VLDL particles; XS_VLDL_C, Total cholesterol in very small VLDL; XS_VLDL_FC, Free cholesterol in very small VLDL.

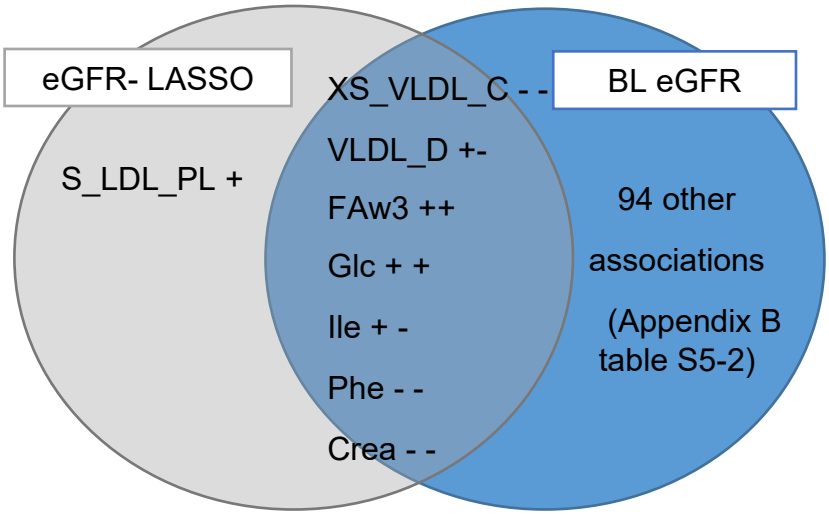


Figure 5-6 Venn diagram showing the overlap between metabolites associated with baseline eGFR and those selected by the LASSO linear regression model in the ET2DS (n= 1,058).

Abbreviations: . **+** positive direction of association; **-** negative direction of association**;** Alb, albumin; Crea, creatinine; eGFR, estimated glomerular filtration rate; FAw3, Omega-3 fatty acids; Glc, glucose; Ile, isoleucine; Phe, phenylalanine; S_LDL_PL, Phospholipids in small low-density lipoprotein; VLDL_D, Mean diameter for very low-density lipoprotein particles; XS_VLDL_C, Total cholesterol in very small very low-density lipoprotein.

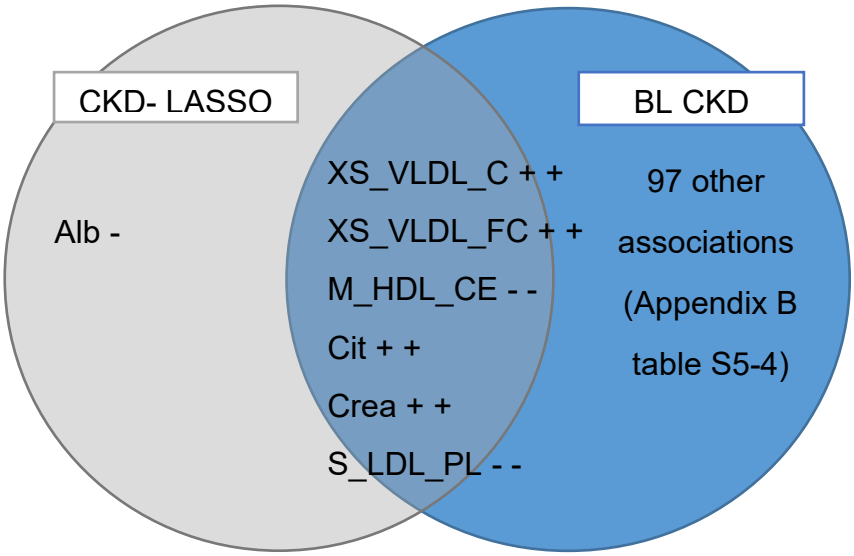


Figure 5-7 Venn diagram showing the overlap between metabolites associated with baseline CKD and those selected by the LASSO logistic regression model in the ET2DS (n= 1,058).

Abbreviations: . **+** positive direction of association; **-** negative direction of association**;** Alb, albumin; Cit, citrate; CKD, chronic kidney disease; Crea, creatinine; M_HDL_CE, Cholesterol esters in medium high-density lipoprotein ; S_LDL_PL, Phospholipids in small low-density lipoprotein; XS_VLDL_C, Total cholesterol in very small very low-density lipoprotein; XS_VLDL_FC, Free cholesterol in very small very low-density lipoprotein.

5.2.4 Metabolomic markers associated with baseline albuminuria

Baseline uACR

The uACR levels at baseline were significantly associated with 46 metabolites out of 149 metabolite concentrations measured in the ET2DS in age and sex-adjusted model with multiple testing correction (Figure 5-7A, appendix B Table S5-5). While the strongest positive association was demonstrated with inflammation-related glycoprotein acetyls (mainly a1-acid glycoprotein) (β=0.15 (95% CI 0.10, 0.21), P_{FDR} = 1.92*10^{-6}), most of metabolites belonged to HDL and VLDL subclass. The metabolites in HDL group were inversely associated (except for the triglycerides) with uACR with the effect size ranging from β -0.13 (95% CI -0.18, -0.08) to β -0.07 (95% CI -0.21, -0.02), all P_{FDR} <0.05). The metabolites in VLDL group were positively associated with uACR with the effect size ranging from β 0.07 (95% CI 0.02, 0.13) to β 0.07 (95% CI 0.01, 0.12), all P_{FDR} <0.05).

After adjustment for all of the CKD risk factors, the uACR levels were significantly associated with only two metabolites (P_{FDR} <0.05) (Figure 5-7B appendix B Table S5-5). The glycoprotein acetyls remained positively associated with uACR, although the effect size was slightly attenuated (β=0.10 (95% CI 0.05, 0.16), P_{FDR} = 0.02). The positive direction of effect, suggested that increased levels of glycoprotein acetyls were associated with higher uACR which is indicative of worse kidney damage. Glycoprotein acetyls were also associated with baseline eGFR after adjustment for all CKD risk factors and also increased with worsening kidney function. Aromatic amino acid tyrosine was also significantly in both models and showed a negative association with baseline uACR (β= -0.11 (95% CI -0.18, -0.07), P_{FDR} = 0.001), which indicated that tyrosine is potentially associated with lower levels of kidney damage. Tyrosine was not significantly associated with baseline eGFR.

In sensitivity analysis, additional adjustment for lipid-lowering medication use in the full model, both tyrosine and glycoprotein acetyls remained significantly associated with similar effect size [tyrosine: β= -0.11 (95% CI -0.16, -0.06), P_{FDR} = 0.002), β=0.09 (95% CI 0.04, 0.14), P_{FDR} = 0.04).

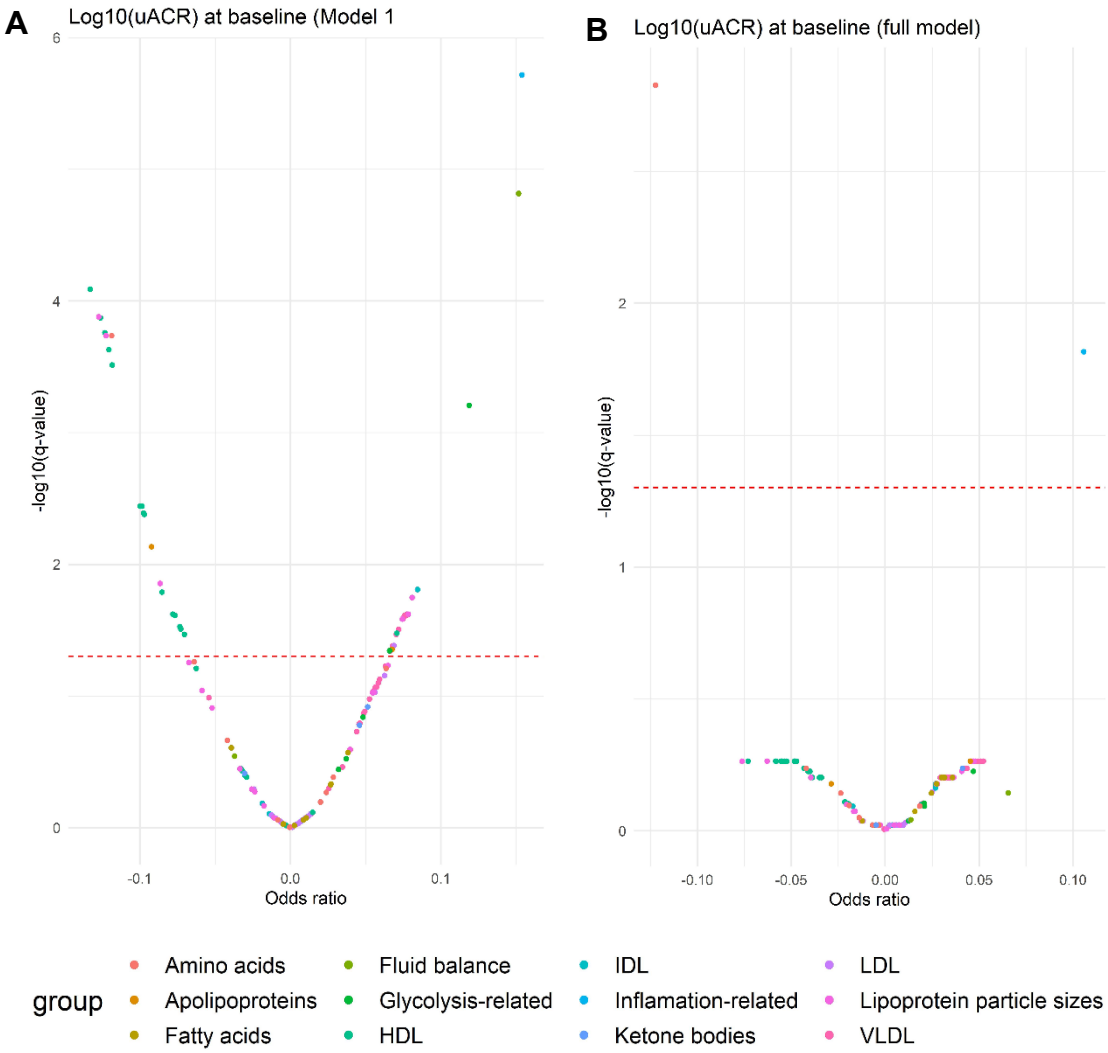


Figure 5-8 Association between 149 individual metabolites and natural log transformed baseline uACR in the ET2DS (n= 1,058).

Red dashed line represents multiple testing corrected significant level, metabolites above it were significant at P_{FDR}<0.05. **A:** Model 1 (age and sex-adjusted model); **B:** Full model (adjusted for age, sex, diabetes control, baseline eGFR, hypertension status, CVD status, BMI, ever-smoker status and HbA1c).

Baseline albuminuria status

There were 32 metabolites that showed significant associations with albuminuria status at baseline in age and sex adjusted model, (P_{FDR} <0.05) (Figure 5-8A, appendix B Table S5-6). The majority of metabolites belonged to HDL subclass and all had OR below 1, suggesting HDL was associated with lower risk of albuminuria (OR range 0.67 (95% CI 0.57, 0.78), P_{FDR} =1.57*10^{-5} to OR 0.82 (95% CI 0.71, 0.95) P_{FDR}= 0.03). Amino acid tyrosine (OR 0.77, (95% CI 0.65, 0.89), P_{FDR}=0.007) and glycoprotein acetyls (OR 1.38, (95% CI

1.19, 1.61), P_{FDR}=0.0003) demonstrated significant associations in the same direction as for baseline uACR levels. However, after adjustment for a full list of CKD risk factors, none of the metabolites remained significant (Figure 5-8B). Although tyrosine remained nominally significant with similar effect size (OR 0.75, (95% CI 0.62, 0.89), P=0.001) along some HDL particles that showed comparable effect size and direction as in model 1 (OR range 0.80 (95% CI 0.66, 0.96), P= 0.02 to 0.83 (95% CI 0.70, 0.99) P= 0.05).

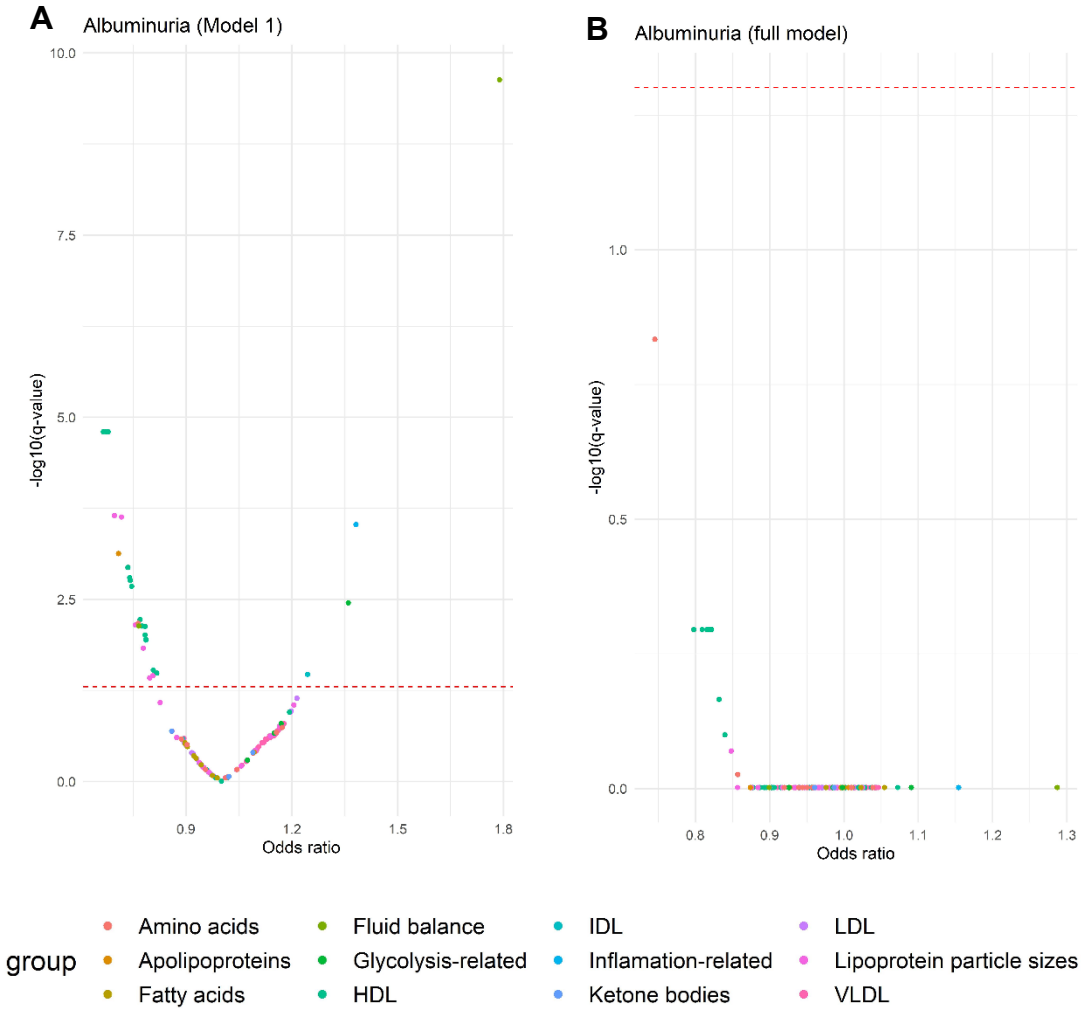


Figure 5-9 Association between 149 individual metabolites and baseline albuminuria status in the ET2DS (n= 1,058).

Red dashed line represents significance level, metabolites above it were significant at P_{FDR}<0.05. **A:** Model 1 (age and sex-adjusted model); **B:** Full model (adjusted for age, sex, diabetes control, eGFR, hypertension status, CVD status, BMI, ever-smoker status and HbA1c).

5.2.5 LASSO analysis of urinary kidney function markers at baseline

In the complementary LASSO analysis, five metabolites were selected in association with baseline uACR on top of age and sex (Table 5-5). As expected, the model selected a sparse panel of uncorrelated metabolites (Figure 5-9). Amino acid tyrosine and inflammation-related glycoprotein acetyls showed a stable association in both, LASSO and traditional regression analysis (Figure 5-10).

Table 5-5 Metabolites selected in LASSO regression for baseline uACR.

Metabolite	Beta coefficient
Medium HDL free cholesterol	-0.07
Glycerol	0.06
Tyrosine	-0.18
Creatinine	0.19
Glycoprotein acetyls	0.33

Note: Model was adjusted for age and sex. Abbreviations: HDL, high density lipoprotein

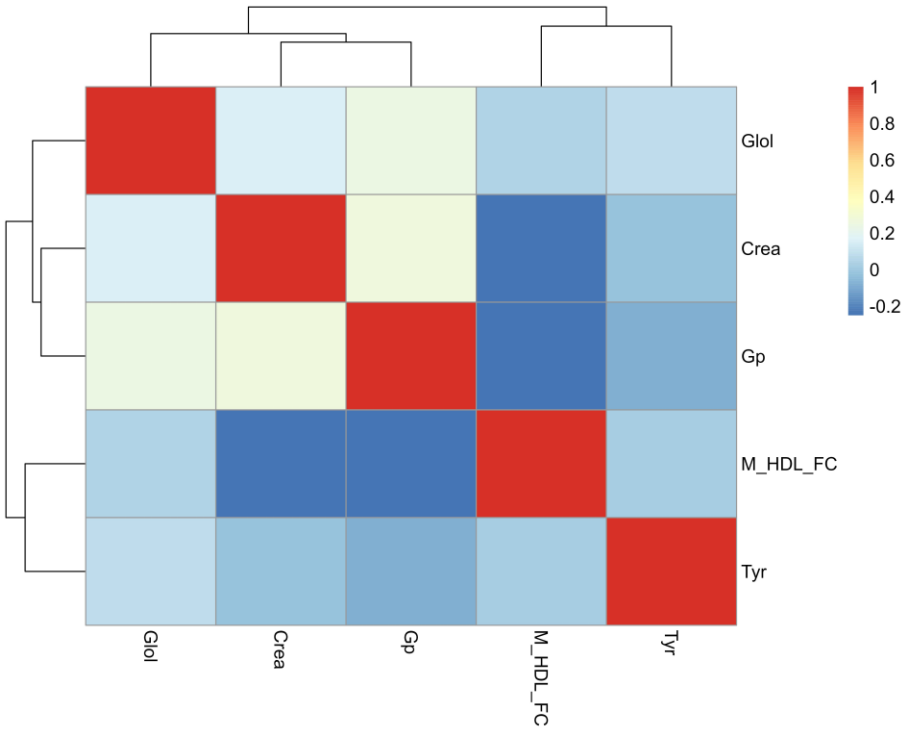


Figure 5-10 Correlational heatmap between the metabolites selected by the LASSO regression model for baseline uACR in the ET2DS (n= 1,058).

Red and blue colours indicate positive and negative direction of correlation, respectively and colour intensity is relative to the strength of coefficient. Abbreviations: Tyr, tyrosine; Free cholesterol in medium HDL; Gp, glycoprotein acetyls; Crea, creatinine; Glol, glycerol.

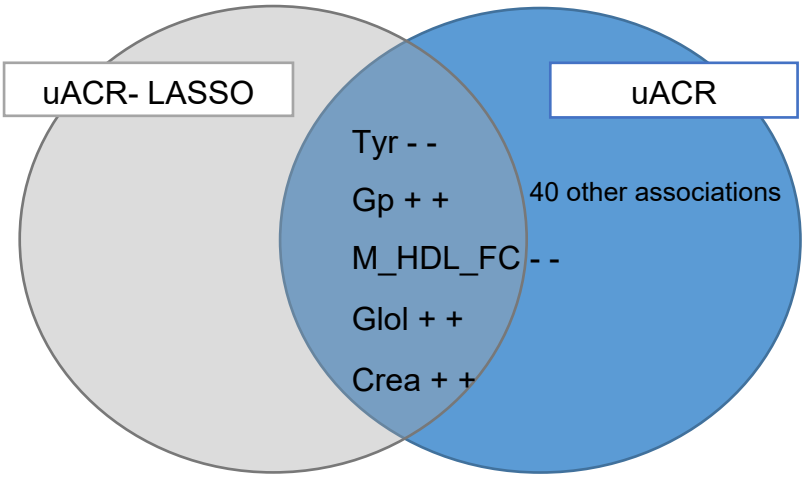


Figure 5-11 Venn diagram showing the overlap between metabolites associated with baseline uACR and those selected by the LASSO linear regression model in the ET2DS (n= 1,058).

Abbreviations: **+** positive direction of association; **-** negative direction of association uACR, urinary albumin to creatinine ratio; Tyr, tyrosine; Free cholesterol in medium HDL; Gp, glycoprotein acetyls; Crea, creatinine; Glol, glycerol

5.3 Prospective analyses of incident CKD metabolomic profiles in ET2DS

5.3.1 Missing data and included participants

The total number of participants included in each analysis and by outcome, is presented in table 5-6, along with reasons for excluding subjects from each analysis. There were 19 participants who had less than two follow-up eGFR measurements (deemed as insufficient) and/or were missing metabolomics data, thus got excluded from the prospective analysis concerned with incident CKD. In addition, 216 participants were excluded from this analysis as they had evidence of CKD at baseline based on persistently reduced eGFR at baseline and historically. In total, 831 participants were included in the prospective analysis of metabolomic markers of incident CKD. For the sensitivity analysis, in which those with albuminuria and/or CKD at baseline (n=374) as well as those with baseline eGFR <60 ml min^{-1} (1.73 $m)^{-2}$ (n=5) were excluded, a total of 694 participants remained for the sub-population analysis. For the annual rate of change in eGFR as percentage based on the estimate from eGFR slope, 30 participants were excluded due to not having any metabolomics data and/or having less than three eGFR measurements during the follow-up time of at least

one year. In total, leaving 1,036 participants with sufficient follow-up data for this prospective analysis.

Table 5-6 Summary of number of participants included in the analysis for different kidney function measures in the ET2DS.

Trait	Included subjects	Missing N (%)	Reason missing data/ exclusion from analysis
Annual rate of change in eGFR %	1036	30 (2.8)	30 insufficient number of FU eGFR records and/or missing metabolomics data
Rapid decliner *1	1036	30 (2.8)	30 insufficient number of FU eGFR records and/or missing metabolomics data
Incident CKD*2	831	19 (1.8)	216 excluded due to evidence of CKD at BL; 18 insufficient FU data; one missing metabolomics data.
Incident CKD*3	694	20 (1.9)	347 excluded due to evidence of CKD and/or albuminuria; five excluded due to having eGFR below 60 ml min^{-1} (1.73 $m)^{-2}$; 18 insufficient FU data; two missing metabolomics data

Notes: *1 Rapid decliner- participants with annual decline of 5% in eGFR; *2 Incident CKD- participants with no evidence of CKD at baseline based on historical and baseline eGFR; *3 Incident CKD- participants with no evidence of CKD and/or albuminuria at baseline and excluding anyone with baseline eGFR < 60 ml min^{-1} (1.73 $m)^{-2}$. Abbreviations: BL, baseline, CKD, chronic kidney disease; eGFR, estimated glomerular filtration rate; ET2DS, Edinburgh Type 2 Diabetes Study; FU, follow-up; uACR, urinary albumin to creatinine ratio.

5.3.2 Description of longitudinal kidney function measures

The distribution of longitudinal kidney function measures is summarised in table 5-8. Of 831 CKD-free participants with longitudinal eGFR data and metabolomic profiles, 155 (18.7%) participants developed incident CKD during the median follow-up of 6.7 years [IQR 6.4, 7.1]. When participants who had albuminuria as well as baseline CKD (N=347) or a single measure of eGFR below 60 ml min^{-1} (1.73 $m)^{-2}$ at baseline (n=5) were excluded, 16.9 % (N=117) of the remaining individuals (n=694) experienced incident CKD during the follow-up. The annual rate of change in eGFR as a percentage showed approximately normal distribution with an average of -1.2 % (SD 4.4) in 1,036 participants (Figure 5-11, Table 5-7). When the eGFR slope was dichotomised, 157 (15%) were classed as rapid decliners, having an annual rate of eGFR change of at least -5% (Table 5-7).

Table 5-7 Summary of longitudinal kidney function outcomes in ET2DS.

Trait	Included participants	N (%) / Mean, SD
Incident CKD [*1]	831	155 (18.7)
Incident CKD [*2]	694	117 (16.9)
Rapid decliner [*3]	1,036	157 (14.8)
Annual rate of change in eGFR %	1,036	-1.25, SD= 4.44

Notes: [*1] CKD – participants with no evidence of CKD at baseline based on baseline and historical eGFR measurements, n=831; [*2] CKD- participants with no evidence of CKD at baseline based on baseline and historical eGFR measurements and/or albuminuria status, n=694. [*3] Rapid decliner- participants with annual decline in eGFR of 5% or more, n=1036. Abbreviations: CKD, chronic kidney disease; eGFR, estimated glomerular filtration rate; N, number.

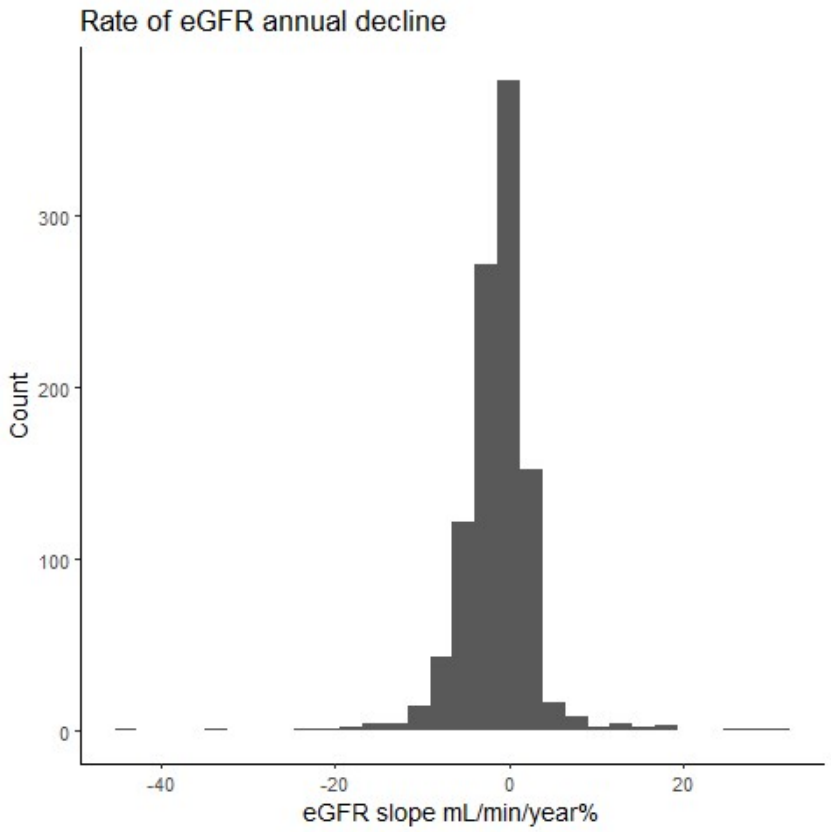


Figure 5-12 Distribution of annual change in eGFR as percentage.

5.3.3 Metabolomic profiles of longitudinal kidney function measures

After adjustment for multiple testing correction none of the metabolites were significantly associated with incident CKD. There were 16 metabolites associated with incident CKD at nominal significance level in baseline eGFR, age and sex-adjusted model (P<0.05, Figure 5-12A, appendix B Table S5-7). Of these, isoleucine, glucose, total cholines and ApoA1 also demonstrated

significant associations with baseline kidney function (CKD and/or eGFR) with comparable magnitude and direction of effect. Conversely, in the full model the metabolites that showed nominal association (Figure 5-12B, appendix B Table S5-7) were not significantly associated with baseline kidney function, except for creatinine, which was expected and amino acid phenylalanine, where there was an apparent opposing direction of effect.

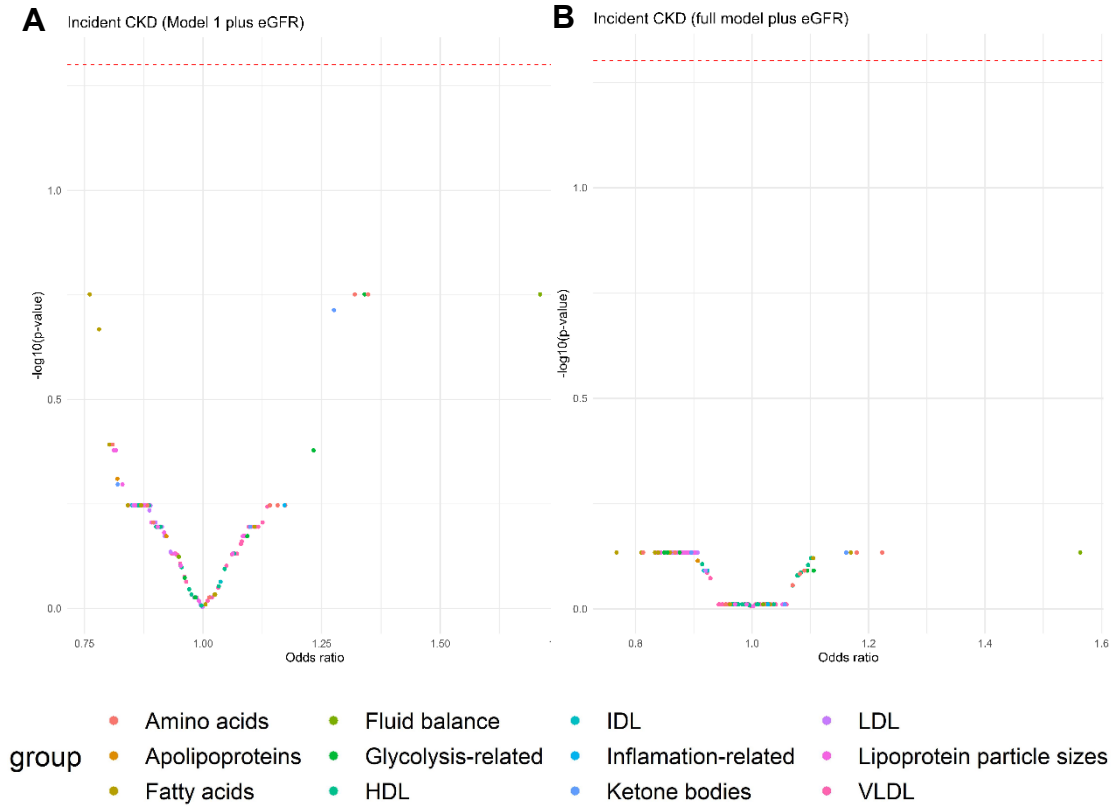


Figure 5-13 Association between 149 individual metabolites and incident CKD in the ET2DS (n= 831).

Metabolites above threshold (red dashed line) showed nominally significant association $P_{FDR}<0.05$. **A:** Model 1 (adjusted for age and sex and eGFR).**B:** Full model (adjusted for age, sex, diabetes control, baseline eGFR, hypertension status, CVD status, BMI, ever-smoker status and HbA1c).

Sensitivity analysis of sub-population which only included participants without any evidence of persistently reduced eGFR and/or albuminuria revealed 21 nominally significant metabolites in association with incident CKD after adjustment for all CKD risk factors (appendix B table S5-8). The majority of metabolites belonged to intermediate density lipoproteins (IDL) and low-density lipoproteins (LDL), all of which were associated with lower risk of incident CKD. While, isoleucine (OR 1.31 (95% CI 1.03, 1.68) P= 0.03) was the only metabolite that showed positive direction of effect. This was in agreement with

findings from baseline CKD analysis where isoleucine showed OR 1.31 (95% CI 1.10, 1.57, P_{FDR}= 0.0006).

Some lipid traits in medium HDL metabolites, glucose and total phosphoglycerides showed marginally non-significant positive association with annual rate of change in eGFR as a percentage (P_{FDR} = 0.055) in the eGFR, age and sex-adjusted model (appendix B table S5-9), which suggested potential association with slower rate of decline i.e., more stable kidney function. Amino acid valine, showed marginally non-significant (P_{FDR} = 0.055) inverse association with this outcome, which indicated potential relationship with faster rate of kidney function decline. Nonetheless, similarly as with incident CKD, after adjustment for CKD risk factors, metabolites were only associated at nominal significance level (Figure 5-13, appendix B Table S5-9). A third of these nominal associations (four of 12) were found with HDL subclass, which were associated with slower rate of decline.

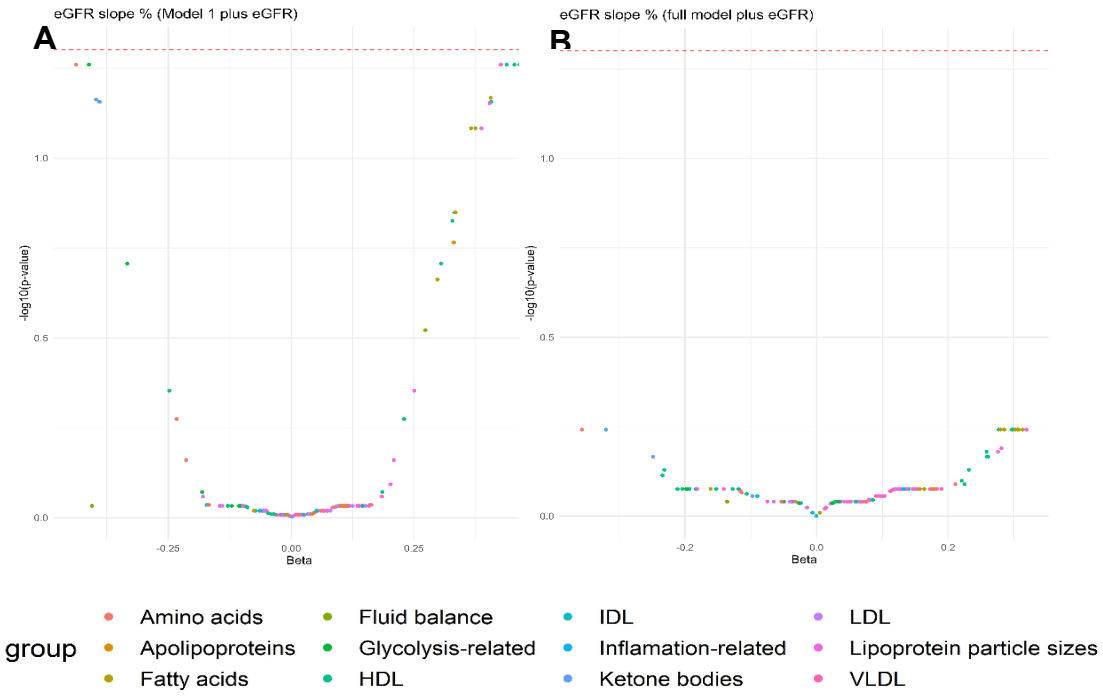


Figure 5-14Association between 149 individual metabolites and annual rate of change in eGFR % in the ET2DS (n= 1,036).

Metabolites above threshold (red dashed line) showed nominally significant association P_{FDR}<0.05. A: Model 1, adjusted for age and sex and eGFR. B: Full model, adjusted for age, sex, diabetes control, baseline eGFR, hypertension status, CVD status, BMI, ever-smoker status and HbA1c.

Rapid decliners demonstrated a significant association with glucose after adjustment for eGFR age and sex (OR 1.37, (95% CI 1.16, 1.62), P_{FDR} = 0.03), which indicated that higher serum glucose levels increased the risk of having rapid decline in kidney function. Additionally, the results also demonstrated marginally non-significant associations with some lipid traits in medium HDL subclass (OR 0.76 (95% CI 0.64, 0.90), P_{FDR} = 0.057) (appendix B Table S5-10), which suggested that HDL may have some protective effects against the risk of having rapid kidney function decline. At nominal significance level (P< 0.05) there were 20 associations with rapid decliner status after adjusting for eGFR, age and sex (Figure 5-13A). These associations reflected those found in the analysis of annual rate of change in eGFR as well as the findings in terms of metabolites significantly associated with baseline kidney function. After adjustment for CKD risk factors, the associations with metabolites from medium HDL were attenuated and the association with glucose was only at nominal significance level (Figure 5-14, appendix B Table S5-10). Nonetheless, the associations remained comparable in terms of effect direction and magnitude.

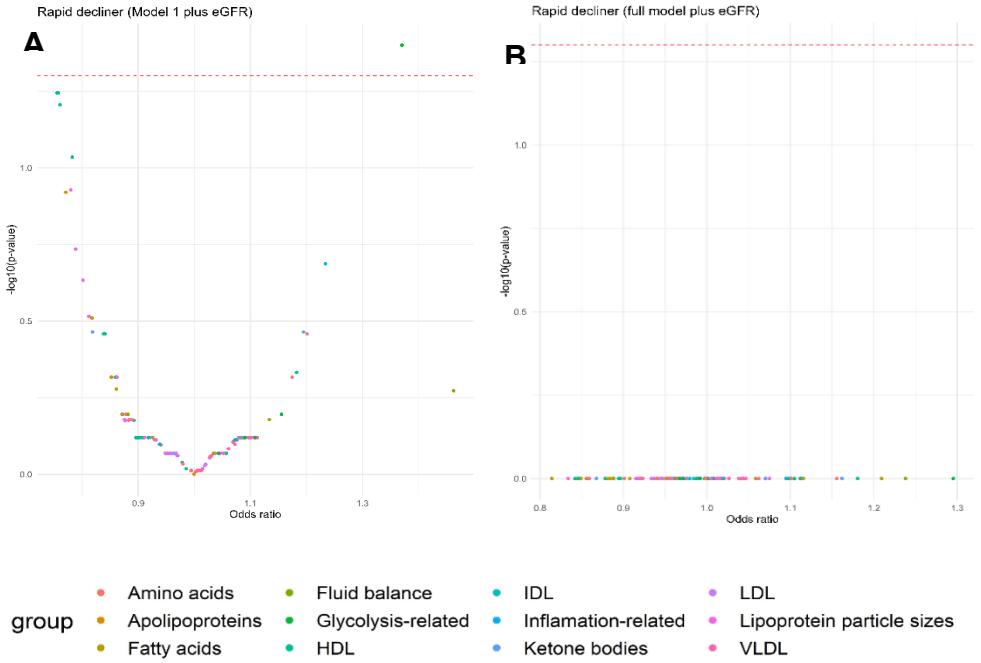


Figure 5-15 Association between 149 individual metabolites and decliner status (eGFR slope < -5%) in the ET2DS (n= 1,036).

Metabolites above threshold (red dashed line) showed nominally significant association P_{FDR}<0.05. **A:** Model 1, adjusted for age and sex and eGFR). **B:** Full model, adjusted for age, sex, diabetes control, baseline eGFR, hypertension status, CVD status, BMI, ever-smoker status and HbA1c.

5.4 Metabolites consistently associated with cross-sectional and prospective outcome

The metabolites that showed some consistency in their association with both cross-sectional and prospective outcomes, were isoleucine, total choline, glucose and apolipoprotein ApoA1 (Figure 5-15). The associations were significantly associated with baseline CKD and/or baseline eGFR (P_{FDR} <0.05) even after adjustment for all CKD risk factors and multiple testing correction. Moreover, isoleucine was selected in the LASSO regression for baseline eGFR and in sensitivity analysis of CKD-free and albuminuria-free subpopulation, it showed a nominally significant association even after adjustment for all CKD risk factors. While the associations with prospective outcomes were only at nominal significance level (P <0.05) and were attenuated after adjustment for the full list of covariates (Full model), the direction and magnitude of associations remained stable. Interestingly, higher glucose levels were associated with higher baseline renal function and also with faster subsequent fall in eGFR.

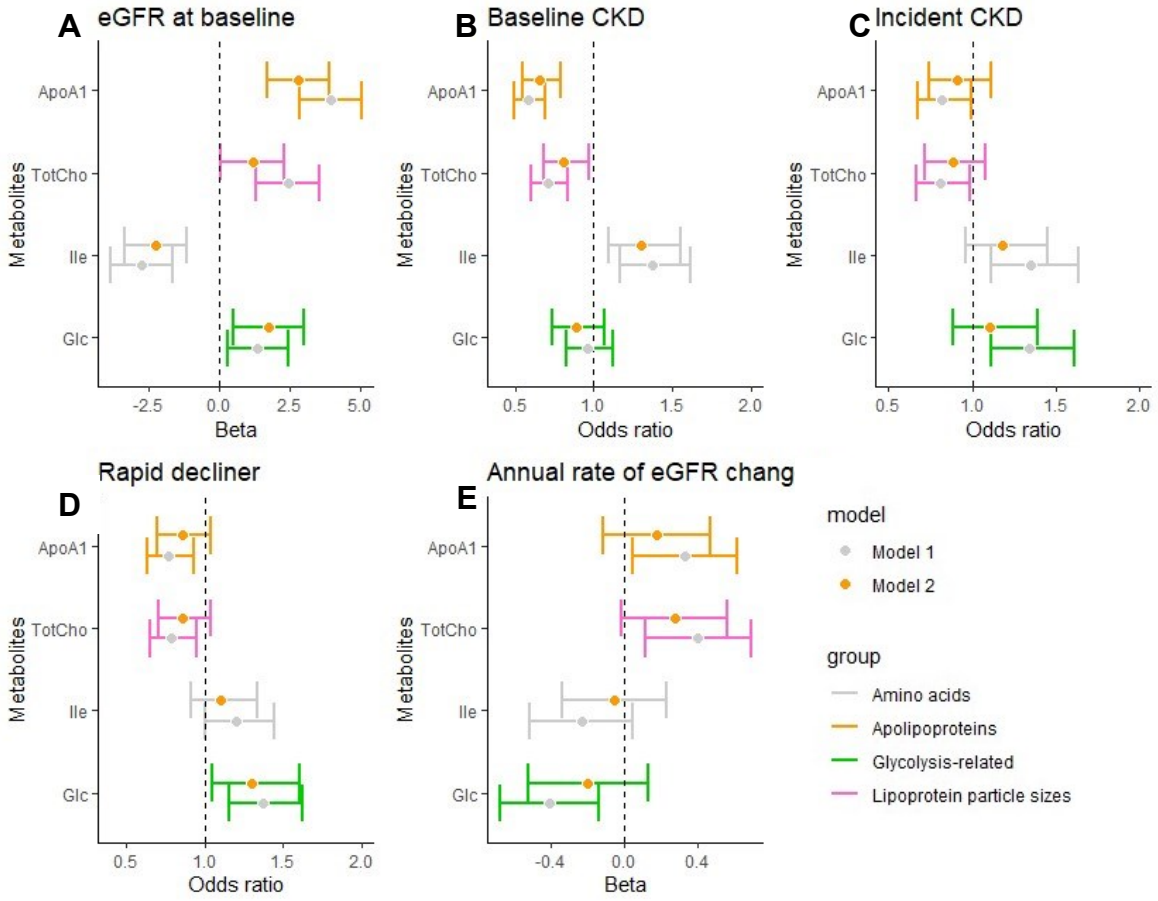


Figure 5-16 The key metabolites that showed stable associations with both cross-sectional and prospective outcomes.

A: associations with baseline eGFR (P_{FDR}<0.05, N=1058); **B:** associations with baseline CKD (P<0.05, N=1,058, events N=216); **C:** associations with incident CKD (P<0.05, N=831, events N=156); **D:** associations with rapid decliner (P_{FDR}<0.05, N=1,036, CKD cases=157); **E:** associations with annual rate of eGFR change % (P_{FDR}<0.05, N=1,036). Model 1: adjusted for age, sex (and eGFR for longitudinal outcomes); Model 2: adjusted for Model 1 + diabetes control, hypertension status, CVD status, BMI, ever-smoker status and HbA1c. Abbreviations: ApoA1, apolipoprotein A1; CKD, chronic kidney disease, eGFR, estimated glomerular filtration rate; Glc, glucose, Ile, Isoleucine, TotCho, total cholines.

5.5 Summary of findings for metabolomic association analyses

In this chapter, I explored the associations between a wide range of serum metabolites and kidney function-related outcomes in older Scottish individuals with type 2 diabetes (T2DM). Baseline kidney function, measured by eGFR, revealed numerous associations with circulating metabolites, extending beyond traditional panels. Of the 88 significant findings (P_{FDR}<0.05), the majority were linked to subgroups of VLDL and HDL, encompassing various sizes and lipid compounds, such as cholesterol, triglycerides, and phospholipids. Notably, three metabolites- isoleucine, ApoA1, and total cholines, were associated with

both baseline chronic kidney disease (CKD) and longitudinal outcomes. However, after adjusting for multiple testing and known CKD risk factors, none of these associations remained statistically significant for prospective outcomes. This suggests that the metabolites in this panel do not provide insight into potential causal pathways for CKD development. These findings are consistent with previous studies that reported similar associations with baseline eGFR (>100), but no significant associations with future outcomes (Tofte et al., 2020, Barrios et al., 2018, Aguilar-Ramirez et al., 2021).

In contrast to baseline eGFR, only two metabolites, namely tyrosine and inflammation-related glycoprotein acetyls, showed significant associations with baseline uACR ($P_{FDR}<0.05$). Since uACR is a urinary biomarker, fewer correlations with circulating serum metabolites were anticipated compared to eGFR, a blood serum-based biomarker. This aligns with previous studies on metabolomic associations with uACR, which also found fewer significant results (Tofte et al., 2019b, Tofte et al., 2020), but this may also be a results of smaller sample sizes available for albuminuria data.

Chapter 6. Results II: Incremental values of combining metabolites and traditional risk factors for predicting CKD in Type 2 Diabetes

This chapter presents the results of analyses exploring the value of adding metabolites to traditional risk factors for predicting the onset of chronic kidney disease (CKD) in people diagnosed with type 2 diabetes mellitus (T2DM). Firstly, to justify my subsequent decision on which existing risk prediction model to use as the reference model for my thesis, I present the predictive performance of the "best-performing model" (Nelson et al., 2019) in the T2DM population, according to a recent review (Slieker et al., 2021). In addition, I present the performance of two simpler risk prediction equations used for comparison. Then, the principal results of this chapter consist of the assessment of the reference risk prediction model re-fitted in the CKD-free ET2DS sub-population, followed by the development of a metabolite-based risk score (MetS). A thorough evaluation of the added value of combining the MetS with the traditional risk factors for predicting the 5-year risk of CKD in the ET2DS is also presented.

6.1 Risk prediction model validation in ET2DS

6.1.1 Choosing a suitable risk prediction model for validation

A recent systematic review and validation study suggested that the risk prediction model developed by Nelson et al. (2019) showed the best performance in predicting the risk of incident CKD in T2DM subjects (Slieker et al., 2021). Furthermore, this model (Nelson et al., 2019) consists of variables available in the ET2DS, and the outcome of CKD onset was applicable to study in this cohort, as there was a sufficient number of incident outcomes. The publication included a replicable equation that comprises the full linear predictor (i.e., beta coefficients for predictors and baseline hazard for predicting incidence within 5 years), allowing for model validation rather than re-fitting (Moons et al., 2012a). In light of this, I selected the risk prediction model developed by Nelson et al. (2019), as the reference risk prediction model, for evaluation in the ET2DS as the primary choice (described in Chapter 1, section 1.5.3).

6.1.2 Description of included participants

The baseline characteristics for all included participants and for those who did and did not experience incident CKD separately are summarised in Table 6-1. A total of 696 participants were included based on the criteria that they did not have baseline CKD in terms of eGFR and/or albuminuria and were followed for at least one year (N=347 were excluded as per definition of CKD detailed in Chapter 4 section 4.1.6; additional five participants were excluded as they had eGFR <60 mL min^{-1} $(1.73m)^{-2}$ based on a single measurement at baseline and N=18 were excluded due to insufficient follow-up duration). Among these participants, 118 subjects experienced incident CKD over a median follow-up time of 7.8 years (IQR 6.4, 7.1) and 106 subjects experienced incident CKD within 5-years. Overall, participants with incident CKD status were older, had lower eGFR at baseline, higher HbA1c levels, and a greater prevalence of CVD. Additionally, a larger proportion used medication for diabetes control.

Table 6-1 Summary of CKD risk factors in all ET2DS participants without CKD at baseline, sub-divided according to whether they developed CKD during follow-up.

	Mean, SD / Median (IQR) / N (%)			**P-value**
Renal Function Decline	No (578)	Yes (118)	All (696)	
eGFR (ml/min/1.73m^2)	86.80, SD=10.72	78.77, SD= 9.99	85.44, SD=11.02	< 0.001
Age (years)	67.09, SD= 4.06	68.92, SD= 3.85	67.40, SD=4.08	< 0.001
uACR (mg/g)	8.58 (2.24-36.33)	9.95 (2.75- 271.76)	8.82 (2.24- 271.76)	0.004
HbA1c (%)	7.27, SD= 1.05	7.54, SD=1.03	7.31, SD= 1.05	0.011
BMI (kg/m^2)	30.85, SD= 5.50	31.47, SD= 5.46	30.96, SD= 5.49	0.268
Male	297 (51.4%)	52 (44.1%)	349 (50.1%)	0.147
Female	281 (48.6%)	66 (55.9%)	347 (49.9%)	
Smoker (ever)	338 (58.5%)	68 (57.6%)	406 (58.3%)	0.864
Prevalent CVD	152 (26.3%)	46 (39.0%)	198 (28.4%)	0.005
Hypertension	446 (77.2%)	99 (83.9%)	545 (78.3%)	0.106
Diet only	148 (25.6%)	11 (9.3%)	159 (22.8%)	< 0.001
Tablets only	357 (61.8%)	90 (76.3%)	447 (64.2%)	
Insulin	73 (12.6%)	17 (14.4%)	90 (12.9%)	

Abbreviations: BMI, body mass index; CVD, cardiovascular disease; eGFR, estimated glomerular filtration rate; uACR, urinary albumin to creatinine ratio.

6.1.3 Predictive performance of the reference model

The reference risk prediction model (Nelson et al., 2019) included age, sex, baseline eGFR, history of CVD, HbA1c, hypertension status, diabetes mellitus medication use, smoking history, BMI and uACR as predictors (see section 4.2.2.3 Box 4-1 for detailed equation). The 5-year absolute risk of incident CKD was calculated using the published equation for the reference risk prediction model (Nelson et al., 2019). The discrimination, based on Harrell's C-statistic was 0.78 (95% CI 0.74, 0.82), which indicates relatively good ability in distinguishing those who experienced incident CKD from those who did not.

The calibration slope was 1.23 (the ideal value is 1, indicating perfect calibration). A slope greater than 1 suggests that the reference model is under-fitted, meaning the distribution of the predicted probabilities is somewhat too narrow and does not vary enough. The calibration plot (Figure 6-1.) compares the observed survival probability to the expected survival probability based on the reference risk prediction model equation, with the risk groups divided into groups by decile of risk (10 groups, 9 cut points). The plot shows that all of the points lie above the reference line, indicating poor calibration, with the reference risk prediction model consistently underestimating the event-free probability (i.e., being CKD-free at 5 years). The predictions cover a fairly wide range of probabilities, spanning approximately from 0.3 to 0.9. The majority of probabilities were clustered in the higher probability of being event-free, which reflects the fact that a large proportion of included participants did not experience CKD onset during the 5 years of follow-up. For a more intuitive graphical representation, Figure 6-2 presents the predicted versus observed risk divided by tertile, providing further evidence that the reference model over-predicted risk across all categories.

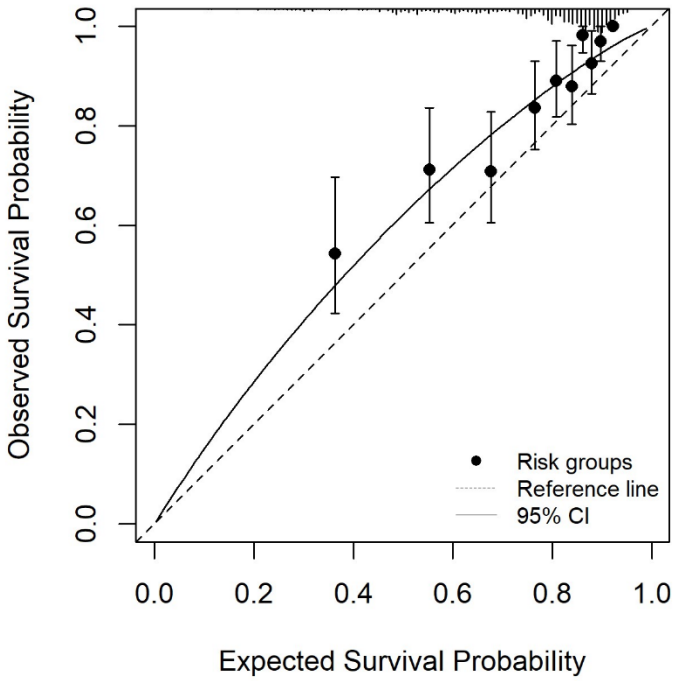


Figure 6-1 Calibration plot summarising the performance of the reference model for survival probability over 5 years, based on the risk prediction equation developed by Nelson et al. (2019). The risk groups divided into groups by decile of risk. The histogram on top of the graph shows the distribution of predicted values, which shows the spread of risk in the dataset. Solid line is a smoothed non-linear curve generated using predicted probabilities.

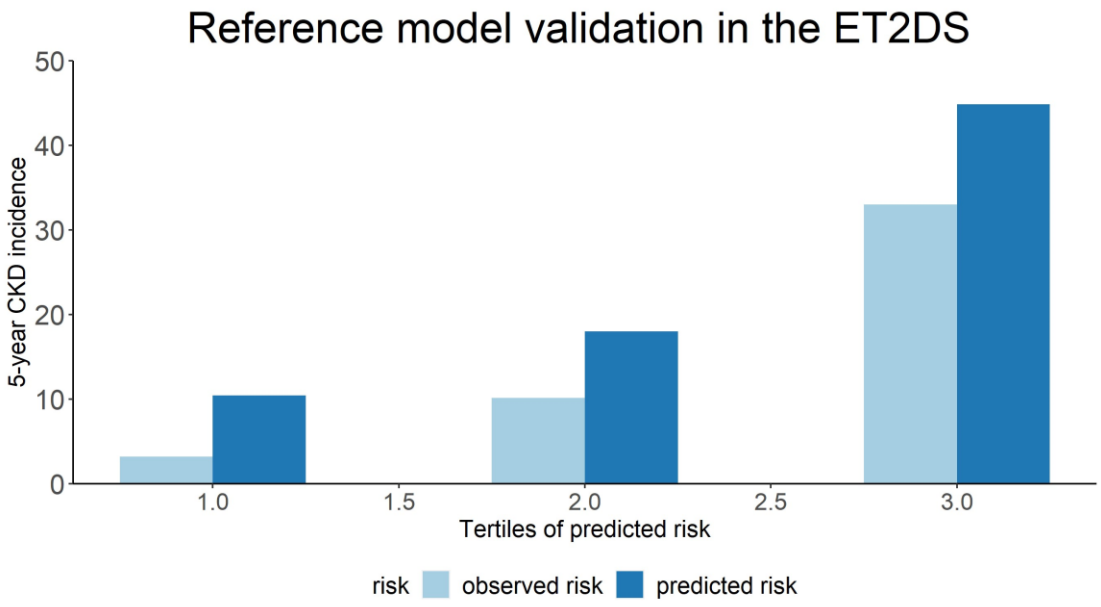


Figure 6-2 Predicted and observed 5-year CKD risk by tertile of risk estimated by the reference risk prediction model.

6.1.4 Reference model versus other risk prediction models

A simpler risk prediction model developed by Chien et al. (2010) (referred to as the Chien equation) was used for comparison with the reference model. The Chien equation included age, diabetes status, history of stroke, BMI and diastolic blood pressure as predictors (for full equation refer to Box 4-2 in Chapter 4, section 4.2.2.3). Harrell's C-statistic for Chien equation was 0.62 (95% CI 0.55, 0.68), suggesting that the discrimination ability for incident cases and non-cases was poor, as it was relatively close to 0.5 (which indicates that the model is no better than chance at making a prediction). The calibration plot (Figure 6-3) showed a narrow range of predicted survival (mainly between 0.6-0.8) and poor fit, with the Chien model consistently underestimating survival probability. This was further confirmed in the predicted versus observed risk tertile in Figure 6-4, which showed that Chien equation over-predicted the risk of an event across all groups.

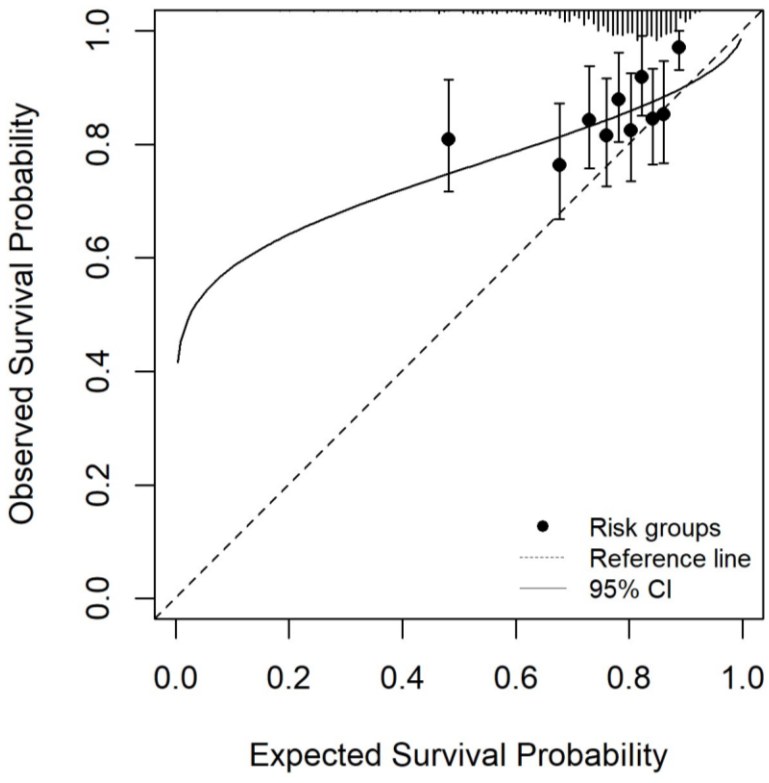


Figure 6-3 Calibration plot summarising the performance of the Chien equation for survival probability during the 4 years (Chien et al. 2010). The risk groups divided into groups by decile of risk. The histogram (top) shows the distribution of predicted values, which indicates the spread of risk in the dataset. Solid line is a smoothed non-linear curve generated using predicted probabilities.

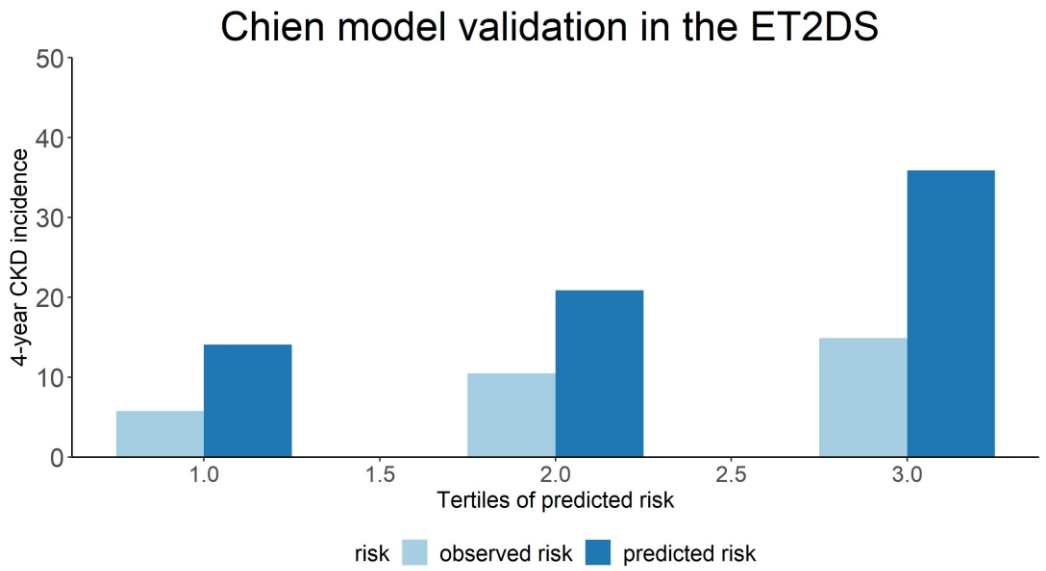


Figure 6-4 Predicted and observed 4-year CKD risk by tertile estimated by the reference risk prediction model.

In addition, Chien at al. (2010) developed a point system for simplified risk prediction: age ≥ 65 (8 points), BMI ≥ 26 (2 points), diastolic blood pressure ≥ 80 mmHg (2 points), history of T2DM (1 point) and history of stroke (4 points). In the ET2DS, on average, patients scored 11 ± 2 points. The receiver operating curve (ROC) was 0.57 (95% CI, 0.52-0.62), which was similar to the discrimination power achieved using coefficient estimates. The best threshold on the ET2DS dataset, calculated using the Youden's index method (Youden, 1950), was 10.5, which is much higher than the threshold of 7 proposed in the model development study (Chien et al., 2010). This provides further evidence of poor calibration of the Chien model and poor risk prediction performance in the ET2DS.

The second comparison was made with the O'Seaghdha equation (O'Seaghdha et al., 2012). This O'Seaghdha equation included age, eGFR, uACR, diabetes status and hypertension status as predictors (for full equation refer to Box 4-3 in Chapter 4, section 4.2.2.3). The c-statistic, based on the AUC, was 0.71 (95% CI 0.66, 0.76), which suggested that the performance of the model in terms of discrimination ability was acceptable and better compared to Chien equation. The observed over expected risk was 0.62, indicating that the observed probability of incident CKD was 38% lower than the expected proportion. The O'Seaghdha model underestimated number of events

compared to what was observed in the CKD-free sub-population of ET2DS, which is also illustrated in the calibration plot (Figure 6-5).

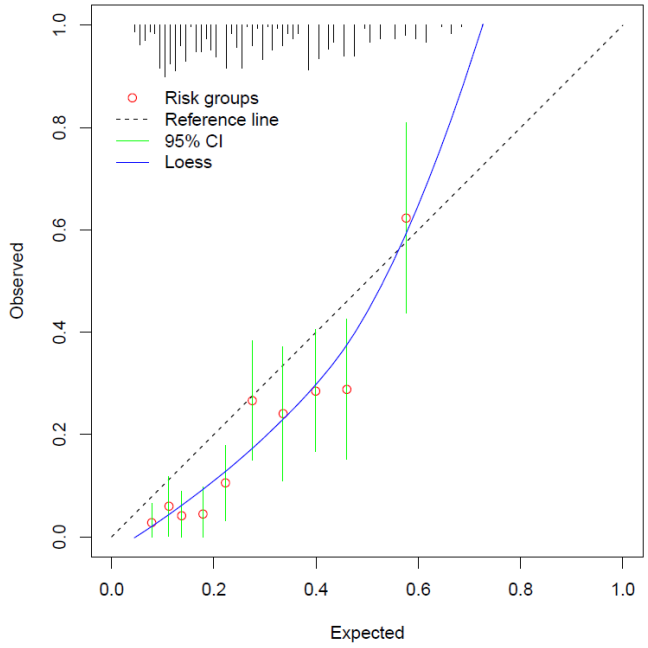


Figure 6-5 Calibration plot of observed risk probabilities and expected risk probabilities calculated in ET2DS based on O'Seaghdha equation. The risk groups divided into groups by decile of risk. The histogram on top of the graph shows the distribution of predicted values, which shows the spread of risk in the dataset. Solid blue line is a smoothed non-linear curve generated using LOESS smoother of predicted probabilities.

6.1.5 Comparison of the reference model vs other risk prediction models

In terms of discrimination, the equation from the reference model demonstrated the best performance with a c-statistic of 0.78, outperforming the other two equations. Calibration for each of the three equations was assessed using the Brier score (Brier, 1950), where the lowest score suggests best calibration. The Brier scores were 0.11, 0.12 and 0.16 for the reference risk prediction model, Chien model and O'Seaghdha model, respectively, which indicated that the Nelson equation had the best calibration. Overall, the discrimination and calibration statistics suggests that the reference risk prediction model (Nelson et al., 2019) outperformed the simpler equations in predicting the risk of incident CKD in ET2DS, which provides sufficient evidence in support of choosing this as the reference model.

6.2 Metabolites-based risk score development

This section describes the metabolites selected as the components of the metabolites-based risk score (MetS), followed by its addition to the reference model (i.e., the refitted risk prediction model developed by Nelson et al. (2019)).

6.2.1 Missing data and incidence of CKD

There were 676 ET2DS participants included in the prospective analysis concerning the development of the metabolites-based risk score (MetS) and its subsequent addition to the reference risk prediction model of established CKD risk factors and the development of the updated risk prediction model. Originally, there were 696 participants with baseline eGFR >60 ml min^{-1} $(1.73m)^{-2}$ who were CKD-free and had sufficient data, as detailed in the previous section (section 6.1.2). However, 18 participants were excluded due to missing data in metabolite variables (as described in section 4.1.8). Among the included participants, 115 experienced incident CKD during the median follow-up of 6.8 (IQR, 0.9, 8.1) years, of which 102 experienced incident CKD within 5 years of follow-up.

6.2.2 Development of metabolites-based risk score

In the Cox LASSO analysis undertaken to select an optimal small group of metabolites for CKD risk prediction, each single repetition identified a group of metabolites with numbers ranging from 6 to 25. The large range in variation reflects some instability in the modelling. Overall, 27 metabolites (mainly subclasses of lipoproteins, amino acids, and ketone bodies) were identified after 100 repetitions as being associated with 5-year risk of incident CKD in at least one repetition of the analysis. The frequency of inclusion of each of these 27 selected metabolites in the analysis is shown in Table 6-2. A distinct ‘frequency threshold’ emerged around the frequency of 72, after which a sudden drop in number of times particular metabolites was selected (Methods chapter 4, section 4.2.2.2). As a result, 18 metabolites with frequencies exceeding 72 in the 100 repetitions were considered as the final components of MetS. Notably, the selection predominantly consisted of lipoprotein subclasses (HDL, LDL and VLDL), ketone bodies, amino acids, fatty acids, and metabolites related to glycolysis and inflammation.

Table 6-2 Frequency of the 27 selected metabolites in 100 repeats of Cox LASSO (n=676).

Metabolite	Subclass	Frequency
Acetate	Ketone bodies	95
3-hydroxybutyrate	Ketone bodies	95
Glycine	Amino acids	95
Isoleucine	Amino acids	95
Phospholipids to total lipids ratio in large VLDL	VLDL	95
Citrate	Glycolysis	94
Lactate	Glycolysis	94
Triglycerides to total lipids ratio in large HDL	HDL	93
18-2, Linoleic acid	Fatty acids	93
Free cholesterol to total lipids ratio in very small VLDL	VLDL	93
Cholesterol esters to total lipids ratio in large LDL	LDL	88
Glutamine	Amino acids	87
Triglycerides to total lipids ratio in medium LDL	LDL	87
Triglycerides in very large HDL	HDL	87
Glycoprotein acetyls	Inflammation	72
Cholesterol esters to total lipids ratio in small LDL	LDL	72
Free cholesterol to total lipids ratio in very large VLDL	VLDL	72
Cholesterol esters in very small VLDL	VLDL	72
Ratio of 22:6 docosahexaenoic acid to total fatty acids	Fatty acids	40
Albumin	Fluid balance	26
LDL particles	Mean diameter	26
Cholesterol esters to total lipids ratio in very large HDL	HDL	26
Total cholesterol to total lipids ratio in very small VLDL	VLDL	26
Triglycerides to total lipids ratio in large LDL	LDL	23
Omega-3 fatty acids	Fatty acids	1
Glucose	Glycolysis	1
Leucine	Amino acids	1

Abbreviations: HDL, high density lipoprotein; IDL, intermediate density lipoprotein; LDL low density lipoprotein; VLDL, very-low density lipoprotein.

6.2.3 Sensitivity analysis

In order to maintain consistency with the approach I used for handling missing metabolite measurements throughout my thesis, I refrained from applying imputation to metabolite measurements before constructing the MetS in the principal analysis of this chapter. Consequently, 18 participants with missing metabolite data were excluded. Although theoretically, excluding only a small number of individuals (N=18) before constructing the MetS is unlikely to have a major impact on the final results, I conducted a sensitivity analysis. In this analysis, missing measurements for those metabolites underwent imputation using a single imputation strategy that was employed for handling missing data

in variables related to CKD risk factors (as described in the Methods chapter, section 4.1.7). Following the same selection criteria applied in the principal analysis (i.e., frequency >72%), 13 metabolites met this threshold and all were selected as the MetS components in the main analysis.

Overall, the sensitivity analysis showed that imputation had some impact on the selection of metabolite components for MetS as the sensitivity analysis left out five metabolites that were present in the main analysis. Considering that the difference between sensitivity and the main analysis is additional 18 participants (only three of which were incident CKD cases) it was not expected to have any effect on the selection of MetS components. Nonetheless, despite five metabolites not getting selected for MetS in the sensitivity analysis, four of these still appeared to be selected by some LASSO iterations in the sensitivity analysis. However, they have not met the frequency threshold, which included glutamine (selected in 48 iterations), glycoprotein acetyls (selected in 48 iterations), triglycerides to total lipids ratio in medium LDL (selected in 19 iterations) and cholesterol esters in very small VLDL (selected in 19 iterations).

6.2.4 Performance of the metabolite-based risk score

Hazard ratios (HRs) for all 18 selected metabolites in the unpenalized and unadjusted Cox model are displayed in Figure 6-6. The HRs for seven of the metabolites were below one, suggesting that these metabolites may have an inverse association with incident CKD. Specifically, glycolysis-related metabolite lactate (HR [95% CI] 0.66 [0.51, 0.84]), fatty acid called linoleic acid (0.72 [0.56, 0.94]), and the amino acid glutamine (HR 0.76 [0.60, 0.97]) demonstrated a significant association with incident CKD ($P<0.05$). The remaining 11 metabolites showed HRs above one, suggesting a positive association with incident CKD. Among these, four metabolites exhibited a statistically significant association with incident CKD, including amino acids isoleucine (HR 1.53, [95% CI 1.13, 2.07]) and glycine (1.28 [1.04, 1.57]), lipoprotein subclass free cholesterol to total lipids ratio in very small VLDL (1.74 [1.28, 2.36]), and ketone body 3-hydroxybutyrate (1.32 [1.06, 1.64]); all $P<0.05$). The distribution of MetS, which combines the linear part of the predictor from the Cox LASSO (as described in Chapter 4, section 4.2.2.2) for

the 18 selected metabolites and metabolite ratios is shown in Figure 6-7. The mean of the MetS was 0.0 (SD= 0.94), with the minimum and maximum values of -3.44 and +2.43, respectively. The distribution of the MetS appeared to be approximately normal.

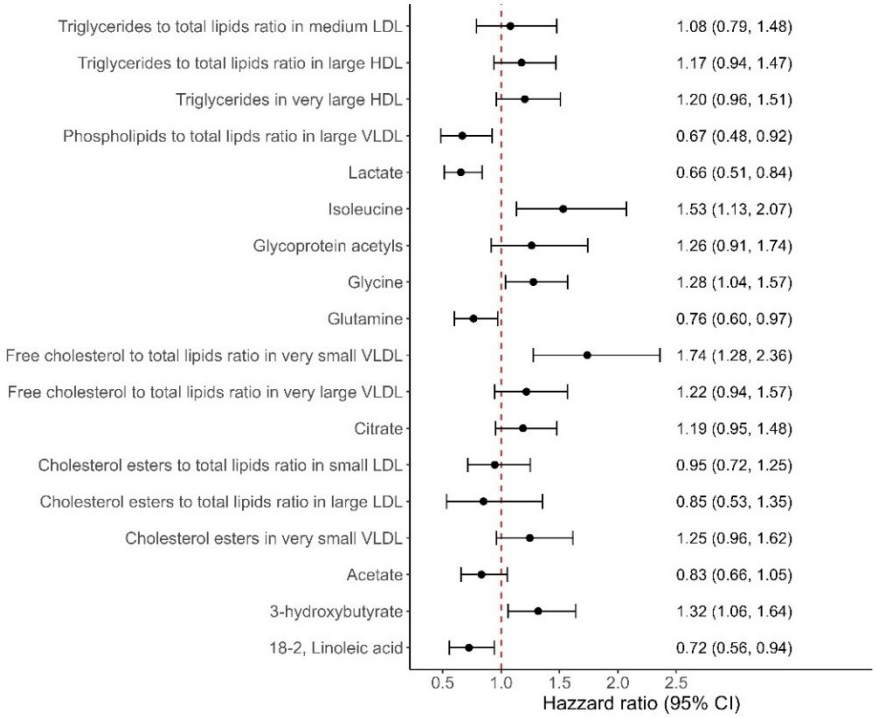


Figure 6-6 Association between 18 selected individual metabolites and 5-year risk of incident CKD in the unpenalized and un-adjusted Cox model (n=676).

Abbreviations: CI, confidence interval; HDL, High density lipoprotein; LDL, Low density lipoprotein; VLDL, very low-density lipoprotein.

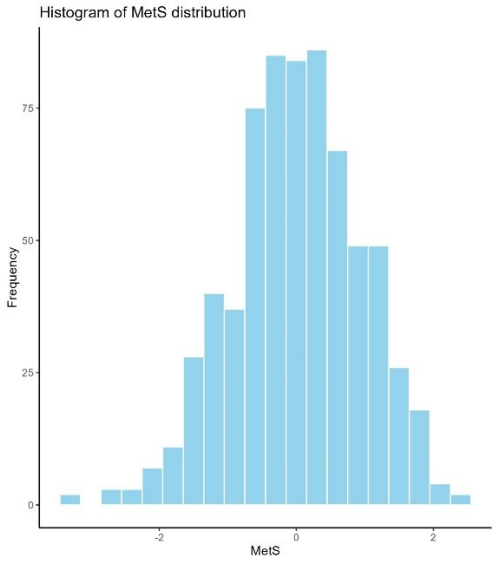


Figure 6-7 Distribution of the metabolites-based risk score (MetS) in the ET2DS. (N=676).

The metabolites-based risk score, when adjusted for baseline eGFR (with linear splines at 90 mL min^{-1} 1.73 m^{-2}, per -5 mL, for details see table 4-3 in Chapter 4, section 4.2.2.1), exhibited a strong positive association with the risk of incident CKD (HR 2.52, [95% CI 1.98, 3.19], P=3.25*10^{-14}). Even after being added to the reference model, the association with the risk of the outcome remained strong (HR 2.21, [1.71, 2.84], P=8.62*10^{-10}). The associations between individual components of the reference model and incident CKD, both before and after the addition of MetS, are shown in Table 6-3. Older age, declining eGFR, increasing uACR, and the use of diabetes tablets alone and/or with insulin were significant predictors in both the reference model and the updated model (reference model plus MetS). After adding MetS as a predictor, all the predictors remained significant with similar HR values. The only exception was for uACR, where the addition of MetS resulted in reduced effect size, rendering uACR as a non-significant predictor. Correlations of MetS and its constituent metabolites with traditional CKD risk factors used as predictors are displayed in Figure 6-8. The correlations ranged from weak to moderate, with values between -0.27 to +0.21. The strongest correlation was found between the amino acid glutamine and HbA1c.

Table 6-3 Association of individual kidney decline risk factors and the MetS with incident CKD in the ET2DS (N=676).

	Reference model				**Updated model**			
	HR	95% CI		P	HR	95% CI		P
Age (per 5 years)	1.44	1.11	1.86	0.006*	1.39	1.07	1.80	0.013*
Female sex	1.20	0.77	1.89	0.418	1.01	0.65	1.59	0.952
eGFR (60-90 per -5 mL)	1.31	1.16	1.48	1.69e-5*	1.27	1.12	1.44	0.0002*
eGFR (≥90 per -5 mL)	1.06	1.02	1.09	0.003*	1.05	1.02	1.09	0.005*
HbA1c (per 1%)	1.10	0.84	1.44	0.48	1.08	0.82	1.43	0.58
BMI (per 5 units, kg m^{-2})	1.12	0.94	1.35	0.21	1.10	0.92	1.33	0.30
Smoker- ever	0.40	0.03	5.42	0.49	0.41	0.03	5.95	0.51
Hypertension status	0.90	0.53	1.55	0.71	0.88	0.51	1.52	0.65
Prevalent CVD	1.42	0.93	2.18	0.10	1.40	0.92	2.14	0.12
uACR (log_{10})	1.54	1.03	2.28	0.03*	1.38	0.95	2.01	0.09
Diabetes tablets vs diet	3.68	1.76	7.73	0.001*	2.68	1.27	5.65	0.01*
Insulin vs diet	5.23	2.10	13.06	0.001*	3.02	1.18	7.74	0.02*
Smoker: Age	1.14	0.81	1.60	0.47	1.14	0.80	1.62	0.48
MetS					2.21	1.71	2.84	8.61E-10*

*p-value < 0.05; HR, Hazzard ration, MetS- metabolites based risk score, BMI, body mass index, uACR, urinary albumin to creatinine ratio, CVD, cardiovascular disease, up.CI- upper confidence interval, p, p-value; HbA1c- glycated haemoglobin.

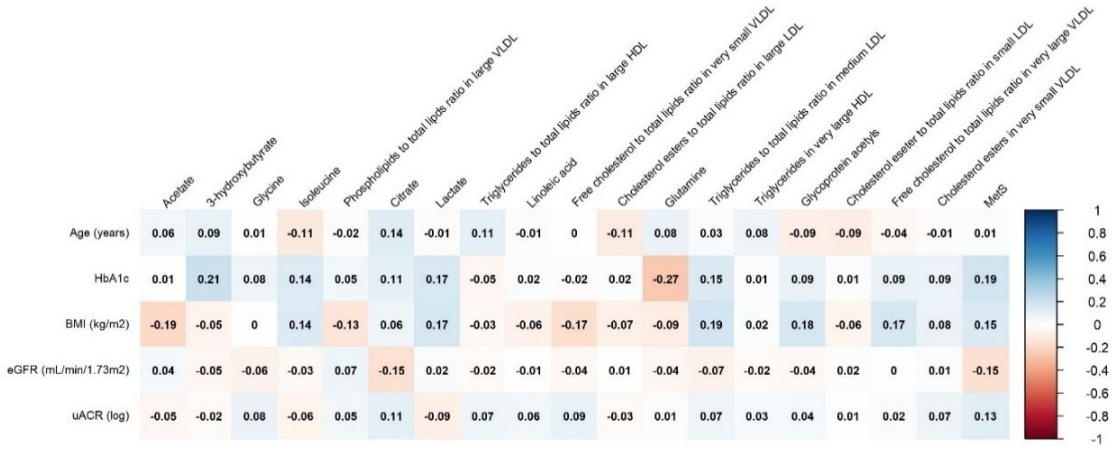


Figure 6-8 Correlations of the MetS and its 18 constitutive metabolites with individual CKD risk factors in the reference model (N=676).

Notes: The darker the colour, the stronger the correlations. Abbreviations: BMI, body mass index; HbA1c, glycosylated haemoglobin; HDL, High density lipoprotein; eGFR, estimated glomerular filtration rate; LDL, Low density lipoprotein; MetS, metabolites-based risk score; uACR (log), log transformed urinary albumin to creatinine ratio; VLDL, very low-density lipoprotein.

6.3 Assessment for the updated risk prediction model: combining the MetS and the reference model

6.3.1 Discrimination

When the MetS was added to the reference model (referred to as updated risk prediction model), the discriminative ability as indicated by concordance statistic (c-statistic) marginally increased from 0.80 [95% CI (0.76, 0.84)] to 0.84 [0.80, 0.87] (Table 6-4.). The internal validation method using 500-time bootstrapping was used to adjust for optimism in predictions, which resulted in slightly reduced c-statistics. The optimism-adjusted c-statistic for the updated model was 0.82 [0.78, 0.85], which remained higher compared to the reference model, which showed an optimism-adjusted c-statistic of 0.78 [0.740, 0.82]. Furthermore, when MetS was added to baseline eGFR in a standalone model, the discrimination performance was slightly better than the reference model, with a modest improvement in optimism-adjusted c-statistic from 0.78 [0.740, 0.82] in the reference model to 0.81 [0.78, 0.85] in the MetS plus eGFR model. Notably, there is a lot of overlap between the 95% CI, so the improvement is not substantial, but positive nonetheless.

Table 6-4 Metrics of predictive performance of the reference model and its combination with the MetS for 5-year incident CKD risk (N=676).

Metrics	Reference model	Reference model plus MetS	eGFR plus MetS
C-statistic	0.799 [0.760, 0.838]	0.837 [0.804, 0.871]	0.813 [0.777, 0.850]
Adjusted c-statistic *	0.777 [0.739, 0.816]	0.818 [0.784, 0.852[	0.812 [0.775, 0.848]
NRI	Ref.	0.612 [0.302, 0.806]	
NRI+	Ref.	0.307 [0.133, 0.450]	
NRI-	Ref.	0.305 [0.158, 0.416]	
IDI	Ref.	0.064 [0.022, 0.108], p-value <0.01	

Notes: Data presented as estimate (95% confidence interval) and significance value (P-value) as appropriate. C-statistic was adjusted for optimism using bootstrap method (described in section 4.2.2.4). Abbreviations: c-statistic, concordance statistic; IDI, integrative discriminatory index; MetS, metabolites-based risk score; NRI, net reclassification improvement; NRI+, net reclassification improvement for events, NRI-, net reclassification improvement for non-events.

As illustrated below in Figure 6-9 (top panel reference model, middle panel-updated model), the discrimination between the 1st and 2nd tertile was poor, as indicated by the small gap between the risk groups in both models. On the other hand, the discrimination between risk groups 2 and 3 was very good in both the reference and updated model. The addition of MetS to the reference model improved the separation between higher risk groups very slightly. The standalone model (MetS plus baseline eGFR, figure 6-9, bottom panel) showed similar gaps between the risk groups as in the reference model indicating similar discrimination performance (Figure 6-9).

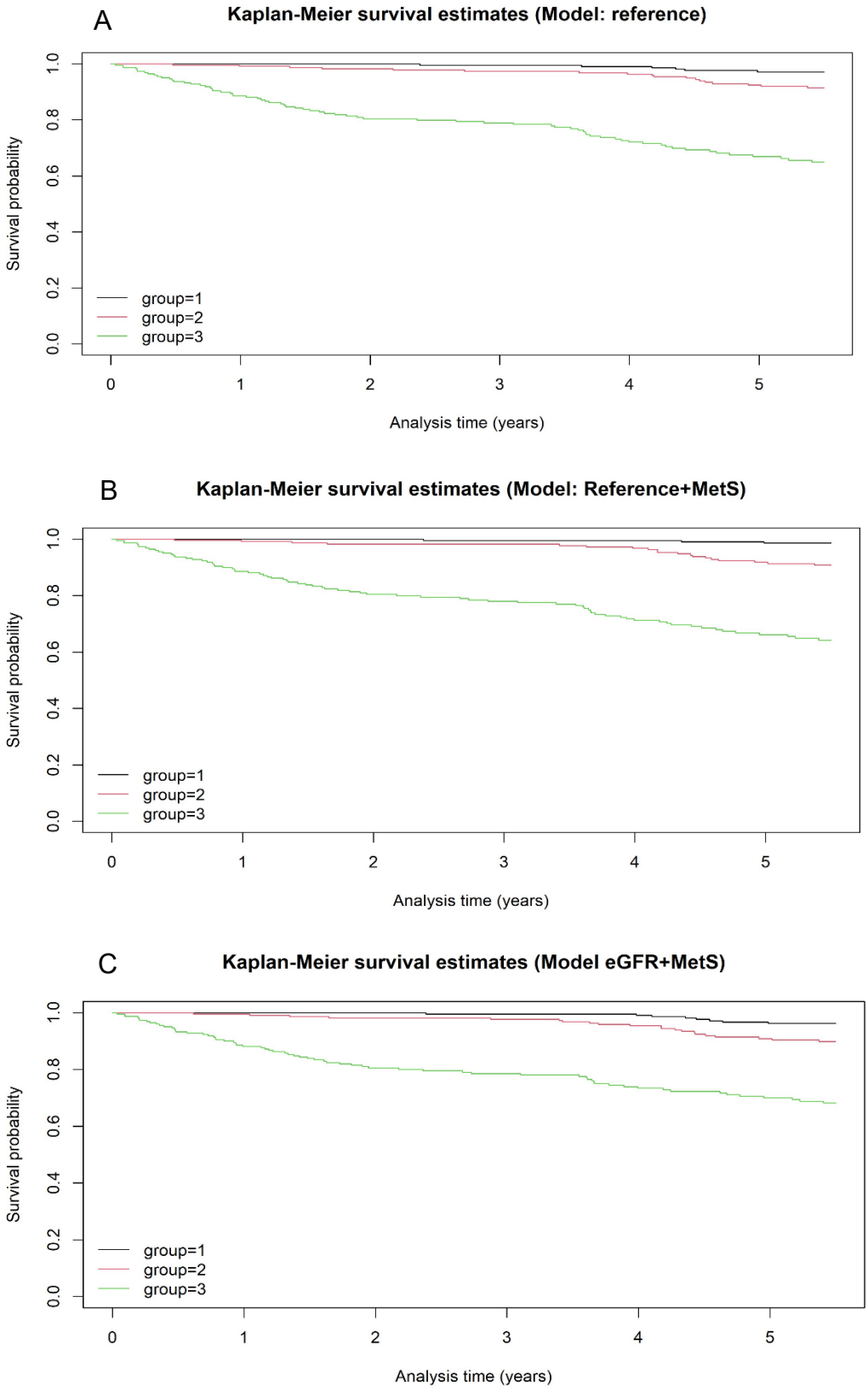


Figure 6-9 Kaplan-Meir survival estimates for analysis time in years (N=676). Panel A: CKD-free survival stratified by tertile in the reference model. Panel B: updated model (reference model plus the MetS). Panel C MetS plus baseline eGFR.

6.3.2 Reclassification

As shown in Table 6-4, the addition of MetS to the reference model resulted in a continuous net reclassification improvement (NRI) of 0.61 [95% CI 0.302, 0.806]. The NRI is the primary metric and represents the overall improvement in risk classification after adding the new predictors to the model (the updated model). It ranges from -2 to 2. A positive NRI indicates an improvement in risk prediction, while a negative NRI suggests a worsening of prediction. We demonstrated that after the addition of MetS, the estimated NRI had 95% CI entirely within the positive region, indicating a significant improvement in risk classification. The NRI+ represents the improvement in correctly reclassifying events to higher risk categories. It ranges from 0 to 1, where a higher value indicates better reclassification. Here, the estimated NRI+ was 0.306 [95% CI 0.133, 0.450], indicating an improvement in correctly reclassifying events to higher risk categories. The NRI- represents the improvement in correctly reclassifying non-events to lower risk categories and also ranges from 0 to 1. In this case, the estimated NRI- is 0.305 [95% CI 0.158, 0.415], indicating an enhancement in correctly reclassifying non-events to lower risk categories. The integrative discriminatory index (IDI) of 0.064 (95% CI 0.022, 0.108, p-value <0.01), which indicated that the gap in risk between those with and without events is 6.4% wider for the updated risk prediction model over the reference risk prediction model.

6.3.3 Calibration

Calibration of the reference risk prediction model in the ET2DS is displayed in Figure 6-10 (top panel), indicating very good agreement between observed and predicted risk, with only slight differences in the 2nd tertile. Calibration of the updated risk prediction model in the ET2DS is shown in Figure 6-10 (middle panel). This was similar to the calibration of the reference model, with almost perfect agreement between observed and predicted risk across all categories, but some difference in the highest risk tertile . The calibration for the stand-alone model (MetS plus eGFR) also showed almost perfect calibration (Figure 6-10, bottom panel).

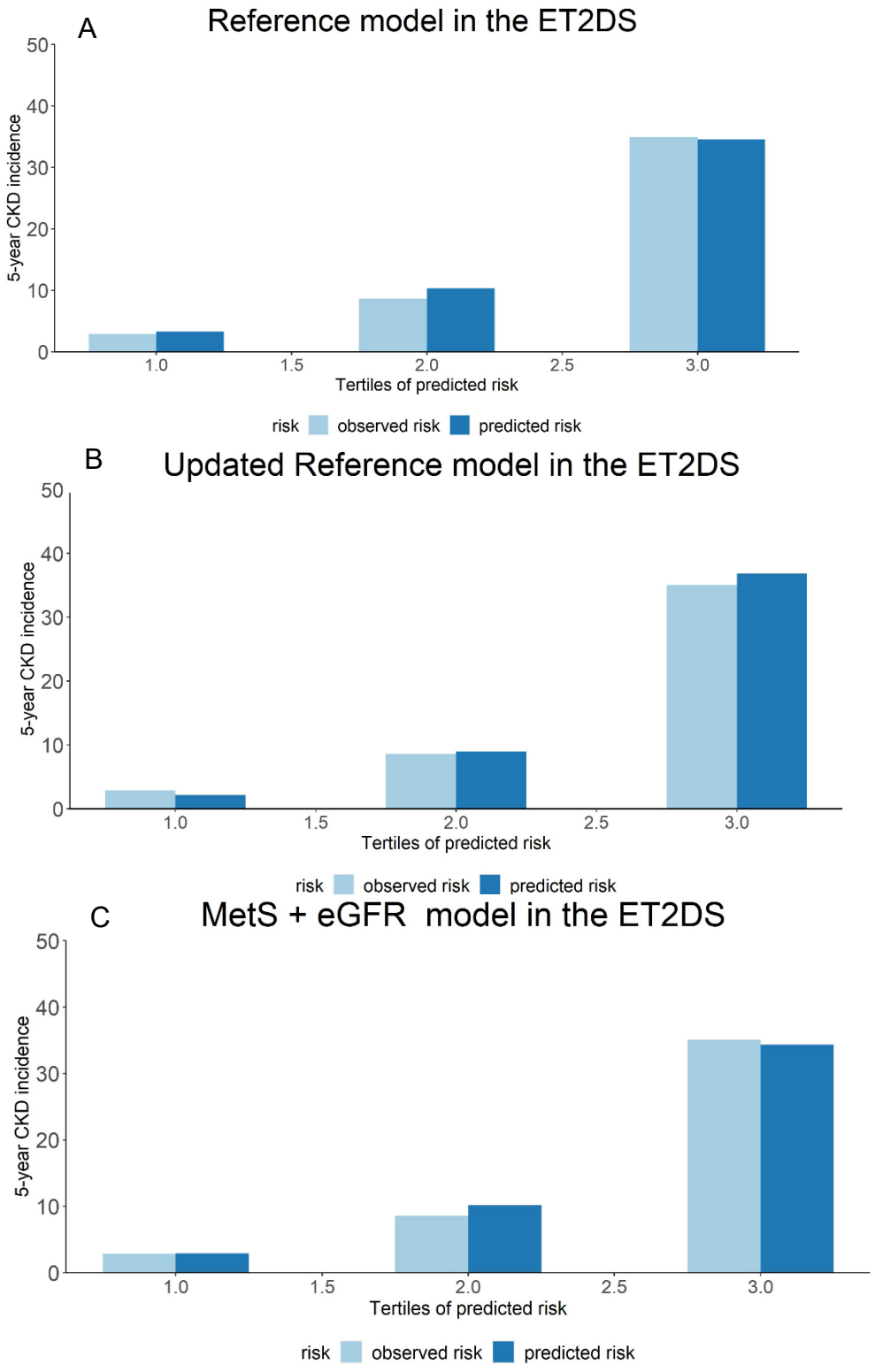


Figure 6-10 Predicted and observed 5-year CKD risk by tertile estimated in the reference model (A), reference model plus MetS (B), and MetS plus baseline eGFR (C).

6.4 Summary of key findings from risk prediction study

Several previous studies have developed metabolite-based scores to predict the risk of incident CKD (Rhee et al., 2013), rapid CKD decline (Nkuipou-Kenfack et al., 2014; Rhee et al., 2016), and progression to ESRD (Zacharias et al., 2019) in the general population, as well as CKD progression in people with type 2 diabetes (T2DM) (Solini et al., 2016). These studies demonstrated improved risk prediction when combining clinical factors with metabolomic-based scores. However, to date, no studies have created a multi-metabolite marker score specifically to predict incident CKD in individuals with T2DM.

The reference risk prediction model, in its original form (using the published equation by Nelson et al., 2019), showed good predictive performance, excelling in calibration and discrimination, and outperforming simpler models. As expected, when the model was refitted using the same components but tailored to the ET2DS population, its performance improved. To further enhance prediction, eighteen metabolites related to inflammation, glycolysis, ketone bodies, amino acids, and lipoproteins were selected to create a metabolite score, referred to as MetS. This score was positively associated with CKD risk after adjusting for traditional CKD risk factors. Incorporating the MetS into the reference model led to moderate improvements in predicting 5-year CKD risk in the ET2DS cohort, particularly in terms of discrimination and reclassification, while maintaining similar calibration between the models. The positive findings here, not only warrant external validation in larger independent cohorts, but also provide evidence that combining metabolomics with more traditional clinical variables may provide valuable insights into kidney decline.

Chapter 7. Discussion

This final discussion chapter of my thesis includes a brief summary of the main findings, a comparison of results with previous studies and a discussion of possible mechanistic explanations for the results. Strengths and limitations of the research are discussed, including those of the study population, the pros and cons of the statistical approaches and the methods I used to estimate kidney function from the data available. The concluding part of the chapter outlines recommendations for the future direction of research on the topic of improving risk prediction for new-onset kidney function decline using metabolomic markers.

7.1 Key findings of the research

The metabolomic profiles of kidney function phenotypes were investigated in the Edinburgh Type 2 Diabetes Study (ET2DS), a cohort of men and women aged 60-75 years with T2DM, in which the prevalence of CKD at baseline was found to be 33%. Based on both kidney measures, eGFR and uACR, the incidence of CKD in those people free of disease at baseline was 17% (based on eGFR alone) over a median follow-up of 7.8 years.

My results on this type 2 diabetes population, revealed cross-sectional associations between baseline kidney function measured using eGFR and a large number of circulating metabolites. Among 88 statistically significant findings independent of CKD risk factors ($P_{FDR}<0.05$), very low-density lipoprotein (VLDL) and high-density lipoprotein (HDL) subgroups, encompassing various lipid compounds, represented the majority of such associations. In contrast to findings for baseline eGFR, only two metabolites, namely amino acid tyrosine and inflammation-related glycoprotein acetyls, were significantly associated with baseline uACR. In addition, the metabolites selected in the complementary Least Absolute Shrinkage and Selection Operator (LASSO) analysis overlapped with those that showed statistically significant association with cross-sectional measures of kidney function. Consistent results obtained from two different analytical approaches support the robustness of my findings in relation to baseline kidney measures.

Notably, four metabolites (isoleucine, apolipoprotein-A1, glucose, and total choline) which showed modest, but statistically significant associations with baseline kidney function in terms of eGFR, also showed nominally significant association with both

incident CKD and the rate of kidney function decline during follow-up. Isoleucine showed an inverse association, while the remaining three metabolites showed positive associations with baseline eGFR. The direction of effect remained consistent with the prospective outcomes for all metabolites, except for glucose. Higher levels of ApoA1 and total choline were nominally associated with lower risk of incident CKD, while higher levels of isoleucine and glucose levels were nominally associated with higher risk of incident CKD. However, these associations became weaker and lost statistical significance for prospective outcomes after adjustment for CKD risk factors and multiple testing correction.

Considering that none of the metabolites demonstrated a consistent and statistically significant association independent of risk factors with both cross-sectional and prospective outcomes, I did not focus on a particular metabolite for the risk prediction analysis. Consequently, a deliberate decision was made to collectively consider the entire panel of additional metabolomic markers, rather than concentrating on one or two potential metabolomic markers with the strongest associations revealed in the initial analysis. This approach mitigated the risk of prematurely fixating on any singular marker, without sufficient evidence in support of the decision. Indeed, lactate was one of the most frequently selected metabolites in LASSO analyses for incident CKD and showed relatively strong effect size in the multi-marker metabolomics-based risk score (MetS), despite its weaker, non-significant association with incident CKD. Furthermore, when considering the metabolites that were analysed in both, the association analysis and the risk prediction analysis, only four metabolites were statistically significantly associated with baseline CKD and also selected for MetS. This included, glycine, isoleucine, citrate and glycoprotein acetyls all of which showed consistency in terms of direction of effect, but only for baseline and not prospective outcomes. Although the underlying biological mechanisms behind this phenomenon are not understood, the result underscores the importance of this unbiased approach to investigating high-dimensional metabolomic datasets.

A reference risk prediction model, based on components of the Nelson risk equation (Nelson et al., 2019) demonstrated a good predictive performance in terms of discrimination for five-year risk of incident CKD in the ET2DS. However, it also indicated that there is room for improvement. The MetS, incorporated 18 metabolites related to ketone bodies, amino acids, fatty acids, glycolysis, and lipoproteins.

Although there were no statistically significant associations between individual metabolites and prospective outcomes, four of the 18 metabolites selected for MetS (acetate, 3-hydroxybutyrate, isoleucine, and 18-2 linoleic acid) showed nominally significant associations with incident CKD independent of age and sex in the prior analysis. Combining MetS with the reference model resulted in a modest yet statistically significant improvement in the prediction of five-year incident CKD. This improvement was observed in terms of discrimination, calibration, and reclassification.

7.2 Strengths and limitations of the research

7.2.1 Completeness and accuracy of data available for analysis

A key advantage of the ET2DS in this research context is that the participants were well-phenotyped, which provided an extensive array of risk factors and biomarkers for in-depth analyses. The comprehensive data collected on ET2DS participants enabled me to align the predictors with those of the reference model. Matching the predictors is crucial to ensure consistency, generalizability, validity of comparisons, applicability, and accurate performance assessment when evaluating risk prediction models in external populations. Thus, the availability of relevant data in ET2DS reduced bias and allowed me to thoroughly assess the performance of the reference risk prediction model in ET2DS. Furthermore, the accuracy and reliability of the data collected in the ET2DS were ensured through well-trained researchers, adherence to standard operating procedures, and stringent quality control measures for physical examinations and laboratory work. A wide range of clinical and demographic variables were meticulously collected with minimal missing data, which enabled the exploration of the metabolomic associations independent of traditional CKD risk factors.

7.2.2 Use of Nightingale metabolomics dataset

The Nightingale metabolomic platform used in the ET2DS incorporated stringent quality control measures to ensure the reliability of metabolite measurements. For instance, each sample plate included two control samples, an imitation serum and a blend of two low-molecular-weight metabolites to monitor quantification consistency and assess the performance of the automated liquid handler and spectrometer. The widespread use of the Nightingale platform in large epidemiological studies enhances the potential for future validation of findings from the ET2DS in diverse study

populations (Soininen et al., 2015, Wurtz et al., 2017, Tofte et al., 2020, Geng et al., 2024, van der Burgh et al., 2024).

Instead of focusing on a specific subclass of metabolites, I employed global metabolomic profile analysis, allowing for a comprehensive depiction of metabolic status. However, while plasma metabolomics measurements encompass diverse metabolic processes taking place in various tissues, it is important to note that the kidneys are the central metabolite filtration unit. Thus, caution is advised when interpreting changes in plasma metabolomic profiles as they may be indicative of changes in systemic metabolism or a consequence of altered renal clearance. Validation of these findings requires in vivo experimental studies to confirm the underlying mechanisms.

7.2.3 Representativeness of the study population

The establishment of the ET2DS, as outlined in Price et al. (2008) was motivated by the goal of exploring the associations between potential risk factors and complications related to T2DM, including cognitive decline, CVD and CKD. The ET2DS cohort comprises a representative sample according to a previous study (Marioni et al., 2010), which was systematically drawn from the entire population of individuals with T2DM residing in Lothian, encompassing the city of Edinburgh as well as surrounding urban and rural areas. Diverging from the recruitment methods employed in prior studies such as those by Tofte et al. (2020) and Barrios et al. (2018), where subgroup of individuals with T2DM were sourced from a mixture of smaller cohort studies, the ET2DS methodology incorporates individuals with T2DM spanning a diverse spectrum of disease severity. This spectrum is characterized by various treatment modalities ranging from dietary management to insulin therapy. Thus, the ET2DS not only mirrors the characteristics of the local T2DM population, but also holds the potential for generalizability to the broader target population of individuals with T2DM residing throughout the UK.

Although ET2DS represents a cohort with diverse disease severity, it is noteworthy that the study intentionally enrolled participants who were older adults, specifically aged between 60 and 75 years at the baseline assessment. Furthermore, the study participants are predominantly white. These two characteristics of the study population limit the generalisability of findings to other ethnicities and younger individuals with

diabetes. In addition, the impact of healthy survivor bias (i.e., the possibility that people who have healthy kidney function at the age of 60 years or above may be biologically different to those who do not) is unknown.

7.2.4 Kidney measures and follow-up

Longitudinal dataset on kidney function markers with a relatively long period of follow-up enabled me to study both biomarkers at baseline and their changes over time, which was vital for valid analysis of the association between biomarkers and CKD risk. The availability of repeat kidney function measurements during the two years before baseline and during a period of over seven years of follow-up enabled me to determine CKD status according to the globally approved recommendations from KDIGO, which relies on kidney function measures repeated at least twice on three-monthly intervals (Levin et al., 2013). Furthermore, taking full advantage of this longitudinal data allowed the exclusion of participants with CKD at baseline and permitted the prospective study of incident CKD in individuals free of CKD at baseline. Compared to previous studies, which have mainly had a cross-sectional design, the prospective nature of the ET2DS allowed for exploration of the temporal relationship between metabolites and incident CKD. Furthermore, the longitudinal records of kidney function for up to seven years of follow-up enabled the identification of time of CKD onset and facilitated time-to-event analysis for the risk of incident CKD.

Applying the definition of CKD in line with KDIGO guidelines (Levin et al., 2013) allowed me to determine clinically-relevant cases, which is preferable to dichotomising a single-point kidney function measurement to define CKD phenotype, as it may lead to misclassification caused by temporal variability of eGFR (Preiss et al., 2007). This is because serum creatinine concentrations may exhibit regular natural fluctuations influenced by factors like the dietary consumption of cooked red meat or other creatinine precursors, as well as diurnal rhythms. Indeed, it has been suggested that up to 25–30% of individuals initially diagnosed as being in stage 3 CKD will eventually improve and move out of this category when eGFR measurement is repeated later (Glassock and Winearls, 2008). So, taking full advantage of the available eGFR data from clinical records in ET2DS enabled me to classify CKD status at baseline and follow-up with minimum misclassification.

On the other hand, the nature of kidney function measurements available in the ET2DS database is a potential limitation. Despite the prospective design of the study, the reliability and accuracy of the kidney function measurements has restricted my ability to explore the etiological roles of metabolites in early CKD development. In this thesis, CKD was defined using the eGFR calculated using a creatinine-based CKD-EPI calculation, widely employed for CKD diagnosis and monitoring in a clinical setting (Levey and Stevens, 2010). While the eGFR calculation incorporates patient sex, ethnicity, and age to eliminate certain sources of variation present when using serum creatinine alone, it is crucial to note that since serum creatinine is the sole measured parameter in the eGFR equation, inaccuracies in creatinine levels will directly result in errors in the calculated eGFR.

The most important factor to consider is the accuracy of biomarker assay calibration. Variability in reported serum creatinine levels can occur due to differences in the assays used for analysis, even when examining the same sample from a single patient. While discrepancies in measured serum creatinine may seem minor, it is important to note that these assay readouts are integral to eGFR equations, which apply an exponent to the serum creatinine value. A previous study used two different Jaffe assays to measure serum creatinine levels in non-diabetic patients from the National Health and Nutrition Examination Survey III (NHANES III) and observed a difference of 20.3 µmol/l (0.23 mg/dl) between measurements (Coresh et al., 2002). As a result, the eGFR calculations were significantly affected, which led to variations in the prevalence of specific eGFR categories. For example, prevalence of participants classed as having CKD stage 3 (eGFR between 30 and 59 ml/min/1.73 m²) varied from 3.2% to 12.5% solely due to the differences in serum creatinine measurement method (Coresh et al., 2002). Furthermore, the exponential relationship between serum creatinine levels and eGFR exacerbates the impact of inaccuracies in creatinine measurements, especially in higher GFR ranges. The results from NHANES III cohort showed that prevalence of eGFR >80 (ml/min/1.73 m²) ranged from 41.8% to 82.1% in the same study, depending on the assay used (Coresh et al., 2002).

Coresh et al. (2002) suggested that the fluctuations in eGFR calculations caused by deficiencies in assay calibration could be minimized by using each eGFR equation exclusively with the particular assay it was originally developed in conjunction with. However, this is impractical, as new assays are developed over time, replacing the

older ones. Therefore, to address this potential source of variation, I ensured that serum creatinine measurements were isotope dilution mass spectrometry (IDMS) traceable (Pieroni et al., 2011). One significant benefit of employing eGFR calibrated to IDMS method is that an eGFR equation created using an IDMS traceable assay is applicable to any other IDMS traceable creatinine measurement (Stevens et al., 2007, Levey et al., 2009c, Schaeffner et al., 2012, Bjork et al., 2015, Pottel et al., 2016). Consequently, using IDMS traceable creatinine measurements ensured that CKD-EPI equation was used appropriately to estimate kidney function in this study and minimised the estimation error for the eGFR results.

Since the initial kidney decline progresses silently, some studies suggested that serum creatinine-based estimation of glomerular filtration rate may not be sensitive enough to detect early changes (Macisaac et al., 2006, Lin et al., 2016). In particular, patients with T2DM and comorbidities with early stages of CKD present specific diagnostic challenges, which were explored extensively in previous research (Oh et al., 2012, Schottker et al., 2012, Lamb et al., 2014, Maahs et al., 2014, Tsai et al., 2014, Avinash et al., 2015, Einhorn and Mende, 2015, MacIsaac et al., 2015, Lin et al., 2016, Mende and Katz, 2016, Tsuda et al., 2016). Studies suggested that some creatinine-based eGFR formulas may lack accuracy in detecting early CKD compared to cystatin-based formulas, especially in individuals diagnosed with diabetes (Borges et al., 2010, Iliadis et al., 2011, Oh et al., 2012, Schottker et al., 2012, Lamb et al., 2014, Tsai et al., 2014, Avinash et al., 2015, Einhorn and Mende, 2015, Mende and Katz, 2016). However, both serum creatinine and cystatin C equations for eGFR may be affected by non-GFR determinants, which directly influence whether an individual is classified as having CKD or not. Factors like muscle wasting or increased muscle mass can impact creatinine levels, while inflammation and obesity can influence cystatin C levels (Knight et al., 2004, Melsom et al., 2015, Schei et al., 2016). Studies that used inulin clearance, an established gold-standard method for measured glomerular filtration rate (mGFR), indicated significant inaccuracies in both cystatin-based and creatinine-based eGFR formulas for diagnosing and staging diabetes-related CKD (Tsuda et al., 2016). Consequently, the discrepancies in eGFR formulas led to the proposal of using mGFR for kidney function assessment, which may provide a more accurate evaluation independent of variables like age or sex (Maahs et al., 2014, Porrini et al., 2019). However, this technique is expensive and impractical for large-scale epidemiological

studies, as it requires intravenous continuous infusion and multiple urine collections over 60 minutes (White et al., 2019). Alternatively, the plasma clearance of iodinated contrast agents like iohexol and iothalamate can be employed for mGFR in smaller cohort studies and clinical trials without urine collections (Maahs et al., 2014). However, this approach also has its drawbacks, including the need for repetitive plasma samples over an extended period, making it unsuitable for larger studies of CKD due to cost and inconvenience. Considering that ET2DS was not primarily a CKD study, but rather a study exploring vascular disease, these intensive methods related specifically to CKD diagnosis would not be considered practical or appropriate, especially when the difference between methods is marginal, and the correlation is high.

7.2.5 Strengths and limitations of the analysis plan for study of metabolomic marker associations with kidney function

To the best of my knowledge, the ET2DS stands as the largest single-cohort study investigating associations between incident CKD and the nuclear magnetic resonance metabolomic panel, specifically in an older population with T2DM. Although previously three studies analysed a similar nuclear magnetic resonance panel (Ibarra-Gonzalez et al., 2018, Barrios et al., 2018, Tofte et al., 2019b), they did not focus on the older age group, which is most susceptible to CKD development. Notably, the studies by Tofte et al. (2019) and Barrios et al. (2018) conducted meta-analyses of smaller cohorts, likely selected based on nuclear magnetic resonance panel availability. This approach is suboptimal as it may introduce biases due to unaccounted heterogeneity between cohorts, such as systemic differences in participant recruitment and pre-analytical influences on metabolomic measurements.

While the primary analysis of the study relied on traditional multivariable regression models examining individual metabolites, LASSO was also employed to assess the robustness of the findings from the analysis which considered associations between individual metabolites and the cross-sectional outcomes. This dual approach allowed for a comprehensive exploration of the metabolomic profile of kidney function phenotypes, considering both clinical interpretability and intricate relationships within the metabolomic data. The consistent results obtained from two different analytical

approaches strengthened the confidence in the associations identified and contributed to the reliability of my study.

The most notable limitation of this study is the lack of comparison with the general population or patients with Type 1 Diabetes Mellitus (T1DM). Without exploring whether similar patterns exist in other groups, it was not possible to determine which metabolic traits were specifically associated with CKD in individuals with T2DM and hence we may have overlooked potential differences in metabolomic profiles between populations. An understanding of metabolite profile associations in isolation could limit the clinical relevance and utility of the findings as comparative studies have the potential to identify markers specific to T2DM, facilitating the development of targeted diagnostic or therapeutic strategies. Additionally, I may have missed the opportunity for a comparative analysis between T1DM and T2DM populations, potentially hindering insights into shared and distinct pathways that potentially contribute to diabetes-related kidney function decline.

Although no participants were lost to follow-up for renal failure or ESRD in the ET2DS, those who died from non-CKD causes were treated as non-cases in the regression models, potentially introducing misclassification error. This misclassification might affect the estimated associations between metabolites and CKD, with the likelihood of estimates leaning toward the null hypothesis and exhibiting smaller effect sizes, reducing statistical power to identify true associations. The 'dilution' effect could suggest that currently identified metabolites might have larger coefficients if misclassification had been avoided.

7.2.6 Strengths and limitations of the analysis plan for improving risk prediction of incident CKD

In terms of analysing risk prediction for incident CKD, special attention was dedicated to formulating an analysis plan which would maximise the informativeness and clinical relevance of the results. The selection of a risk score for constructing a reference risk prediction model, to which metabolomic markers would be added, was guided by specific characteristics detailed in Chapter 1 (section 1.5.3). Emphasis was placed on models that incorporated only clinical variables and were subject to extensive external validation. The incorporation of metabolomic markers into the reference risk prediction

model serves to elucidate their additional value beyond the information available from routine clinical measures alone.

Ideally, the process of selecting a reference model should involve a rigorous evaluation, including conducting a systematic review to compile available risk scores and carefully comparing their predictive performance in the ET2DS to identify the optimal score. However, a recent study of this nature was already performed in the T2DM population (Slieker et al., 2021), so conducting this analysis specifically in ET2DS would not contribute substantially to current research knowledge. Furthermore, systematic reviewing was not a primary objective of this thesis, so it would be unrealistic to conduct a second comprehensive review of the literature (in addition to the one carried out for Chapter 3) within the time constraints of the PhD. Nonetheless, I have selected a reference model based on the results from the previous systematic review, and it demonstrated good performance in the ET2DS, enabling further investigation of the added benefit by metabolomic profiles.

Multiple repeats of LASSO were used to handle high-dimensional metabolomics data, effectively addressing issues of high correlation. By incorporating information from various instances of the LASSO procedure, where predictive metabolites selected across different runs are combined, I was able to identify a stable set of metabolomic predictors to construct a multi-marker risk score, MetS. The identification of predictors that consistently stand out reduced the likelihood of false positives and enhanced the reliability of selected variables. However, LASSO modelling exhibited some instability, as indicated by a wide variation in metabolites selected in association with the outcomes, ranging from six to twenty-five. This could be attributed to the limited statistical power as a result of a relatively small sample size (119 incident CKD outcomes) compared to the number of measured exposures (228 metabolites measured).

Another and most critical limitation, particularly relevant to covariate selection methods like LASSO, is the absence of external validation using a comparable and independent population, as this approach tends to overfit the data and provide optimistic results. While internal validation through bootstrapping have been cautiously employed to address overfitting concerns, these measures only represent the minimum requirement for ensuring the generalizability of the present findings according to

TRIPOD guidelines (Moons et al., 2015). A future replication study is therefore warranted to evaluate if the findings are generalizable to other independent populations of older people with T2DM as well as other populations to accurately assess the performance of the updated risk prediction model. In addition, external validation could offer further insights, especially in determining whether overfitting might contribute to the slightly improved predictive performance observed in the reference risk prediction model (the re-fitted Nelson model) compared to the original Nelson risk equation (evaluation of the model with direct application of the published equation) within the ET2DS. Furthermore, the clinical interpretation of certain predictive metrics, such as continuous NRI, remains unclear (Kerr et al., 2011), and hence the assessment of the clinical utility of the MetS was unfortunately unattainable.

7.3 Comparisons of findings on metabolomic profiles with previous studies and possible mechanistic explanations

The metabolites that showed some consistency in their direction of effect with both cross-sectional and prospective outcomes, were isoleucine, glucose, apolipoprotein-A1 and total choline. These findings were considered as most notable, so are discussed below in terms of comparisons with previous research and possible mechanistic explanations for the associations. However, it is important to note that these associations lost statistical significance for prospective outcomes after adjustment for CKD risk factors and multiple testing correction, suggesting that they are not independent risk factors, but rather a part of complex pathways involved in early stages of CKD development. In addition, the metabolites that were only significantly associated with baseline kidney measures are also described.

7.3.1 Isoleucine (amino acid)

Amino acids serve as fundamental building blocks for all life forms, and their absorption and transport occur in various tissues such as the small intestine, colon, liver, kidneys, and others (Neis et al., 2017, Broer and Fairweather, 2018). This widespread distribution enables amino acids to play a crucial role in influencing the growth and overall health of humans. Branched-chain amino acids, specifically valine, leucine, and isoleucine, function as signalling molecules that regulate the metabolism of proteins, glucose, and lipids, thereby playing essential roles in maintaining energy homeostasis. Changes in branched-chain amino acid levels were identified in people with diabetes several decades ago, which showed that the disruptions in branched-

chain amino acid metabolism occur prior to the onset of diabetes, thereby suggesting its involvement in diabetes pathogenesis (Wang et al., 2011, Nagata et al., 2013, Zheng et al., 2016, Lotta et al., 2016) and insulin resistance, which is a key contributor to T2DM onset (Lynch and Adams, 2014, Lee et al., 2016a). The increased risk of T2DM as a result of high circulating levels of branched-chain amino acids was verified in multiple cohorts (Floegel et al., 2013, Tillin et al., 2015). A previous study also demonstrated a connection between elevated isoleucine levels and an increased likelihood of having elevated HbA1c (>53 mmol/L), indicating a significant correlation with disrupted diabetes control (t Hart et al., 2018).

In the current study, I found the branched-chain amino acid isoleucine showed the strongest inverse association with baseline kidney function in terms of eGFR compared to other amino acids. Additionally, it showed a nominally significant association with prospective outcomes, including incident CKD, rapid decliner status and annual rate of eGFR change in percentage, albeit only after adjustment for baseline eGFR, age and sex, but not after adjustment for other CKD risk factors. Consistent with these findings, a previous study showed that isoleucine levels were increased in CKD patients (DeFronzo and Felig, 1980). More recently, a UK Biobank study also demonstrated that isoleucine significantly associated with increased risk of incident CKD in general population (Geng et al., 2024). Similarly, a small prospective study also demonstrated that elevated isoleucine levels were associated with an increased risk of experiencing a rapid decline in eGFR and incident CKD in individuals with T2DM, independent of CKD risk factors (Zhu et al., 2022). Together these findings suggest that isoleucine may serve as a signal not only for disruptions in diabetes and insulin resistance, but also for the development of CKD. While the mechanisms for these findings remain unclear, it is possible that a diminished renal role in protein synthesis as a result of kidney function decline could account for the increase in circulating branched-chain amino acids observed in association with CKD (Garibotto et al., 2010). It is also possible that observed inverse relationship (isoleucine levels increased as kidney function decreased) could be attributed to the fact that isoleucine may be metabolised by the healthy kidneys. Further exploration is warranted to delve into the pathophysiology that underlies these connections as it may provide valuable insights into potential targets for treatment.

7.3.2 Glucose (glycolysis-related metabolite)

The renal system maintains glucose homeostasis by engaging in gluconeogenesis, glucose filtration, glucose reabsorption, and glucose consumption processes (Mather and Pollock, 2011). In people with diabetes, the reabsorption of glucose from glomerular filtrate is increased to reduce the loss of glucose via urinary excretion. Elevated blood glucose levels were shown to cause increased renal tubule reabsorption through upregulated glucose transporters (Bakris et al., 2009, Gorboulev et al., 2012, Uehara-Watanabe et al., 2022). Some of these glucose transporters are also sodium transporters (e.g., sodium-glucose transporter 2), leading to increased sodium reabsorption and reduced sodium and chloride delivery to the macula densa (a specialized region in the kidney, specifically located in the distal tubule of the nephron, involved in regulating renal blood flow and GFR). This triggers a decrease in the signal for tubule-glomerular feedback, resulting in elevated single nephron GFR (Vallon et al., 1999, Thomson et al., 2001). Consequently, elevated blood glucose levels play a role in inducing glomerular hyperfiltration (Wiseman et al., 1987, Vallon and Komers, 2011), which may contribute to the onset or progression of CKD in people with diabetes, as suggested by several observational studies, which reported associations between glomerular hyperfiltration and rapid kidney function loss in patients with T1DM or T2DM (Magee et al., 2009, Ruggenenti et al., 2012, Bjornstad et al., 2015, Low et al., 2018). Although some studies reported conflicting results where no such associations were found (Ficociello et al., 2009, Thomas et al., 2012, Molitch et al., 2019), it is possibly influenced by the use of GFR-estimating equations, which did not provide an accurate estimation of kidney function, especially at higher levels of GFR function (Gaspari et al., 2013).

This evidence presents the hypothesis that hyperglycaemia or elevated blood glucose levels may cause glomerular hyperfiltration, which exacerbates the rate of decline in renal function in the future. In line with this hypothesis, the analyses of metabolomic associations in ET2DS revealed that circulating glucose levels were positively associated with baseline renal function, yet it was linked to a faster subsequent decline in eGFR. Previous studies also showed that raised glucose levels increased the risk of a rapid reduction in GFR, ultimately culminating in the faster development and progression of CKD (Magee et al., 2009, Ruggenenti et al., 2012, Bjornstad et al., 2015, Low et al., 2018). Additionally, circulating glucose levels were associated with

ESRD independent of other risk factors in people with T2DM (Colhoun et al., 2001) and the general population (Zacharias et al., 2019) as well as incident CKD in UK Biobank cohort study (Geng et al., 2024).

7.3.3 Apolipoprotein-A1 (apolipoprotein)

Apolipoproteins play essential roles in stabilizing the lipoprotein structure through their hydrophobic and hydrophilic regions (Rader et al., 1994, Bolanos-Garcia and Miguel, 2003, Sniderman and Faraj, 2007). They also serve as cofactors for enzymes in lipid metabolism and act as ligands for cell surface lipoprotein receptors, mediating the modification and absorption of lipids (Mahley et al., 1984). A type of Apolipoprotein-A (ApoA), namely ApoA1, is produced in the liver and predominantly found in plasma as well as the extravascular compartment. It is a primary protein component contained in the HDL particles, accounting for 70% of the total HDL protein content (Gursky, 2005). The ApoA1 is the main acceptor of cholesterol during the transport of cholesterol from tissues to the liver for excretion from the body. It is also involved in transporting cholesterol from the liver to cells (Talmud et al., 2002). In addition, ApoA1 demonstrated anti-atherogenic effects through prostacyclin stabilization (Yui et al., 1988, Flores et al., 2019, van der Vorst, 2020).

Consistent findings from previous studies affirm that ApoA1 serves as an indicator of CVD risk (Berg and Borresen, 1976, Ishikawa et al., 1978, Cremer et al., 1988, Walldius et al., 2001, Bodde et al., 2019). Reduced ApoA1 levels were shown as a negative prognostic factor for CVD outcomes (Florvall et al., 2006) while higher levels were associated with a lowered risk of CVD, implying a cardioprotective effects (Erqou et al., 2010). Considering that CKD and CVD share common risk factors and pathophysiologic mechanisms, such as oxidative stress, inflammation, hypertension, endothelial dysfunction, vascular calcification, and dyslipidaemia, (Matsushita et al., 2022, Zoccali et al., 2023), it is not surprising that previous studies also observed alterations in HDL composition and decreased levels of ApoA1 in individuals with CKD (Attman et al., 1987, Attman and Alaupovic, 1991, Shah et al., 1996, Vaziri et al., 1999, Muntner et al., 2004, Vaziri, 2006, Lacquaniti et al., 2010, Goek et al., 2012b, Zewinger et al., 2014) and in haemodialysis patients (Moradi et al., 2009, Moradi et al., 2010). Positive correlations between ApoA1 and baseline eGFR were shown in the general population as well as in people with diabetes, including individuals with T2DM (Goek et al., 2012b, Barrios et al., 2018, Tofte et al., 2020). In line with previous findings, the

study of metabolomic profiles in ET2DS showed that ApoA1 displayed a statistically significant inverse association with baseline CKD and statistically significant positive association with baseline eGFR, independent of CKD risk factors. Considering that in patients with CKD, the filtration, catabolism, salvage, and excretion of HDL and its components are impacted, it is not surprising that the concentration, composition, and functionality of HDL particles is also influenced by kidney dysfunction (Marsche et al., 2020). Suggesting that ApoA1 changes may be a consequence of kidney dysfunction rather than being predictive of it. Moreover, the biosynthesis of ApoA1 is reduced in people with CKD (Vaziri et al., 1999). The changes in HDL particles and reduced levels of ApoA1 in CKD patients results in lowered cholesterol efflux capacity along with reduced antioxidant function of HDL, affecting CVD risk in CKD patients (Holzer et al., 2011).

In terms of longitudinal findings in ET2DS cohort, the associations between ApoA1 levels and new-onset CKD or annual rate of eGFR change were only nominally significant. However, point estimates indicated trends were consistent with the results shown in cross-sectional analyses. Similar findings were observed in two large multi-ethnic population-based cohorts, namely the Atherosclerosis Risk in Communities study (ARIC) and NHANES III (Goek et al., 2012b). Here, the ApoA1 levels were not significantly associated with incident CKD, but showed consistent trends in line with cross-sectional analysis of CKD status and baseline eGFR. While some studies revealed a statistically significant positive association between ApoA1 and annual change in renal function in an non-diabetic population (Barrios et al., 2018), others have not (Samuelsson et al., 1997). Similarly, the CRIC study revealed that ApoA1 was not independently associated with the risk of CKD progression to ESRD in a large cohort of patients with mild or moderate CKD (Rahman et al., 2014). Taken together these findings suggest that there is no robust evidence to support that ApoA1 independently predicts CKD development or progression. Although the observational nature of the ET2DS study prevents me from addressing mechanisms of CKD development, there is a lack of evidence to suggest that ApoA1 has a direct role, indicating that CKD onset may be mediated through other pathways.

7.3.4 Total cholines

Total cholines measurement encompasses the total concentration of various choline-containing compounds circulating in the blood serum. Cholines are essential nutrients

and precursors for the synthesis of phospholipids, neurotransmitters, and other vital molecules. Previous research indicates that choline metabolism involves both the liver and kidneys (Bligh, 1953). It plays a key role in diverse physiological processes and stands as an independent predictor for several cardiovascular conditions, including hypertension, atherosclerosis, and myocardial infarction (Danne et al., 2003, Danne and Mockel, 2010, Song et al., 2021, Yang et al., 2023).

Higher levels of total cholines significantly reduced odds of baseline CKD independent of CKD risk factors ($P_{FDR}<0.05$) and demonstrated consistent direction of effect in terms of reduced incident CKD and rapid decliner status. Similarly, total cholines levels also showed non-significant inverse association with incident CKD in UK Biobank study (Geng et al., 2024). In contrast, previous studies showed an opposite relationship, where higher levels of cholines were associated with reduced eGFR in people with T2DM and general population (Aguilar-Ramirez et al., 2021). Moreover, cholines demonstrated strong inverse association with eGFR (Missailidis et al., 2016) and increased the risk of incident CKD during 8 years of follow-up (Rhee et al., 2013). Cholines are acquired from dietary sources, so total choline levels are directly influenced by consumption of meat, fish and eggs (Ufnal et al., 2015). Older people tend to consume less protein rich foods, thus, conflicting results may be due to fundamental demographic differences in the populations studied. This indicates that substantial differences may exist in metabolic profiles due to demographic variations such as age and lifestyle, which highlights the importance of considering variables such as diet when studying the metabolome.

7.3.5 Metabolites associated with kidney damage

In contrast to findings for baseline eGFR, only two metabolites, namely tyrosine and glycoprotein acetyls, were significantly associated with baseline uACR, which is a clinical indicator of damage to the glomerular filtration barrier. This may, at least in part, be explained by uACR being a urinary biomarker, for which fewer correlations with circulating serum metabolites may be expected compared to eGFR, a serum-based biomarker. Notably, previous studies that also considered metabolomic associations with uACR, also reported fewer significant findings in people with either T2DM (Tofte et al., 2020) or T1DM (Tofte et al., 2019b), possibly due to the smaller number of participants with available data on albuminuria.

Tyrosine (amino acid)

The amino acid tyrosine was found to be inversely associated with uACR independent of CKD risk factors in ET2DS. Tyrosine is considered to be a semi-essential amino acid, as it can only be synthesized by the hydroxylation of phenylalanine, an essential amino acid. Apart from the enzymatic process facilitated by phenylalanine hydroxylase, leading to the conversion of phenylalanine to tyrosine, the sole origin of these two amino acids is through dietary intake (Chazot et al., 1997). In individuals and animals experiencing CKD, there is often a decrease in plasma tyrosine concentrations, while, plasma phenylalanine levels tend to be normal to slightly increased, leading to a reduced ratio of tyrosine to phenylalanine (Gulyassy et al., 1970, Young and Parsons, 1973, Jones et al., 1978, Fukuda and Kopple, 1980, Furst et al., 1980, Kopple et al., 1982, Flugel-Link et al., 1983, Laidlaw et al., 1994). These observations suggest a potential impairment in the conversion of phenylalanine to tyrosine catalysed by phenylalanine hydroxylase.

Phenylalanine hydroxylase enzyme has been identified in the liver, kidney, and pancreatic cells of mice (Kaufman, 1993), so the loss of the enzyme activity in the kidneys may contribute to the diminished conversion of phenylalanine to tyrosine during kidney decline. Indeed, some functional studies in humans and animals demonstrated lower levels of circulating tyrosine in ESRD due to decreased rate of conversion of phenylalanine to tyrosine (Boirie et al., 2004, Kopple, 2007). Moreover, a number of observational studies showed that tyrosine was linked to different CKD stages (based on eGFR) in people with diabetes (Lee et al., 2016b) and the general population (Mika et al., 2018) as well as incident CKD (Sekula et al., 2016). Although, we did not find that tyrosine was significantly associated with eGFR, it was inversely associated with baseline uACR levels. So, ET2DS participants with lower levels of uACR i.e., those with less renal damage, had higher levels of tyrosine. This is in line with previous studies, which demonstrated that higher levels of circulating tyrosine were associated with lower odds of diabetic nephropathy (Barrios et al., 2018, Zhang et al., 2020a) and inversely associated with albuminuria (Tofte et al., 2020). Furthermore, the ADVANCE trial demonstrated an inverse association between tyrosine and microvascular diseases in ~3500 individuals with T2DM (Welsh et al., 2018).

The findings from ET2DS, consistent with earlier research (as discussed above), suggest that kidney damage as indicated by increased uACR, is linked to lower concentrations of plasma tyrosine, which may result from dysfunctional activity of phenylalanine hydroxylase (Kopple, 2007). Additionally, tetrahydrobiopterin, a crucial cofactor for phenylalanine hydroxylase, may also influence its activity and production of tyrosine (Werner et al., 2011). The oxidative stress present in diabetes patients could potentially reduce tetrahydrobiopterin availability, thereby impairing phenylalanine hydroxylase activity and contributing to reduced plasma tyrosine levels (Werner et al., 2011).

Tyrosine plays a role in gluconeogenesis and glucose transport (Li et al., 2019). Free tyrosine has the potential to combine with free radicals, forming 3-nitrotyrosine, which may harm pancreatic islet beta cells (Chi et al., 2005). The question of whether lower tyrosine levels solely indicate worsening nephropathy/ kidney damage or if tyrosine itself acts as a protective and/or aggravating factor is still a subject of controversy. Recent research, demonstrated heightened nitro-tyrosine staining in the kidneys of individuals with diabetic nephropathy (Thuraisingham et al., 2000). This warrants further study to investigate if tyrosine predicts albuminuria in people with diabetes and its use as a prognostic marker for renal decline.

Glycoprotein Acetyls (inflammation-related factor)

Recent studies described glycoprotein acetyls as an inflammatory biomarker related to glycosylation of proteins in blood and may be more reflective of systemic inflammation (Connelly et al., 2017). Low-grade inflammation has been implicated with various chronic diseases (Fuertes-Martin et al., 2020) including diabetes (Duncan et al., 2003, Duncan and Schmidt, 2006). Titan et al. (2017) demonstrated that glycoprotein acetyls was associated with albuminuria and kidney function independent of major risk factors in the general population and showed better discriminative ability compared to the traditional inflammatory biomarker high-sensitivity c-reactive protein. Moreover, Tofte et al. (2020), showed that glycoprotein acetyls were associated with baseline eGFR and albuminuria in people with T2DM, which was also revealed in the analysis of metabolomic profiles in ET2DS. However, glycoprotein acetyls were not identified as significant metabolite in association with prospective outcomes in either ET2DS or previous studies, so the associations with baseline CKD in people with

T2DM may be the result of low-systemic inflammation caused by metabolic disorder rather than CKD itself.

7.3.6 Other metabolites associated with kidney function

Lipoproteins and their traits

Patients with CKD often have increased cardiovascular risk as a result of dyslipidaemia, the hallmark of which is decreased HDL cholesterol and increased triglyceride levels measured by traditional methods (Mikolasevic et al., 2017). In agreement, I found nuclear magnetic resonance-measured total serum triglycerides and total cholesterol in HDL were associated with eGFR and baseline CKD, indicating that individuals experiencing kidney decline are likely to exhibit typical traits of dyslipidaemia.

Large cohort studies suggested that dyslipidaemia may also be a risk factor for development of kidney dysfunction (defined as new-onset of reduced eGFR and/or albuminuria) in people with T2DM (Sacks et al., 2014, Russo et al., 2016) and albuminuria progression in T1DM as suggested by a prospective FinnDiane Study Group (Makinen et al., 2013). Although I did not find that measures akin to the traditional total HDL levels were associated with the prospective outcomes, I did find phospholipid and cholesterol traits in medium HDL particles showed positive nominal association with eGFR slope and demonstrated protective effect against the decliner status, while triglycerides in very-large-HDL had the opposite effect. The associations were in agreement with cross-sectional results, suggesting that HDL particles may exert effect not only on kidney function at baseline, but also on the rate of kidney function decline. Similarly, Tofte et al. (2020) revealed positive associations with eGFR slope and HDL subclass, but only in very large HDL particles. Although, in both studies the associations became non-significant after correction for multiple testing, it is possible these associations are true, considering these HDL particles were significant in baseline eGFR analysis. A potential explanation for the lack of statistical significance may be related to the statistical power as indicated by relatively wide confidence intervals.

Overall, the cross-sectional and longitudinal findings discussed above provide some evidence that lipids, particularly cholesterols and phospholipids in HDL subclass play a role in development of CKD in people with diabetes. The HDL cholesterol or

phospholipids could serve as target for therapies in the future, but so far, the effects of increasing HDL did not show favourable outcomes. For example, torcetrapib caused an increase of morbidity and mortality leading to early termination of the ILLUMINATE trial (Barter et al., 2007). Another trial specifically in T2DM patients, used fenofibrate to raise HDL cholesterol levels and while the lipid profile was improved it had no effect on CKD or diabetes (Keech et al., 2005). It is important to note that while HDL cholesterol is commonly referred to as "good cholesterol," the functionality and composition of HDL particles can vary, as shown in the metabolomic profiles of ET2DS. Merely raising HDL cholesterol levels may not necessarily confer clinical benefits, as it is the functional properties of HDL particles that are thought to be more relevant in terms of metabolic health.

Besides, a more distinct dyslipidaemia pattern emerged via nuclear magnetic resonance -profiling, where a number of non-traditional lipoproteins were associated with reduced kidney function, in some cases irrespective of lipid content. All sizes of very low-density lipoprotein (VLDL) particles with various lipid components showed inverse association with continuous eGFR levels at baseline. The direction of the same lipoprotein profile was mirrored when a binary measure of CKD was explored. These relationships were consistent with previous studies, which included comparative associations with many lipoproteins for people with and without diabetes (Barrios et al., 2018, Tofte et al., 2020, Aguilar-Ramirez et al., 2021). The results suggest that alterations of the lipoprotein composition beyond traditional dyslipidaemia measures may contribute to the risk observed in CKD patients.

Interestingly, my study, similarly to previous studies (Barrios et al., 2018, Tofte et al., 2020), showed that the associations persisted even with adjustment for lipid lowering medication use. Lipid-lowering is recommended for reducing CVD and mortality risk in general population, but its effectiveness in the CKD population (with and without diabetes) is controversial as it may be dependent on severity of renal dysfunction (Wanner et al., 2005, Fellstrom et al., 2009, Baigent et al., 2011). Statin therapies (i.e. lipid-lowering medication) did not have an effect on increased triglycerides, decreased HDL or ApoA1 levels (Mesquita et al., 2010). Indeed, the results presented here showed substantial variation in lipid content in the different lipoprotein particles associated with eGFR independent of lipid-lowering treatment, which suggested that

more subtle risk patterns in people with CKD not detected by traditional lipid measures may affect the comorbidity risk and alter the effectiveness of lipid-lowering treatment.

Unmeasured metabolites associated with incident CKD in previous studies of people with Type 2 Diabetes

The metabolites analysed on the Nightingale platform represent only a fraction of the well-known metabolites, and the Human Metabolome Database continues to expand in scope (Wishart et al., 2022). While certain metabolites were not evaluated in the ET2DS, they could potentially have a significant impact on the development of kidney function decline in individuals with T2DM. A recent systematic review described seven blood plasma biomarkers that were most frequently associated with incident CKD as well as CKD progression and the onset of ESRD in a T2DM population (Liu et al., 2022). Plasma tumour necrosis factor receptor 1 (TNFR1), a receptor found on inflammatory cells, was the most studied plasma biomarker in relation to kidney dysfunction, which showed positive association with CKD outcomes. Statistically significant results for TNFR1 were reported in 13 cohort studies comprising of individuals only with T2DM, as well as three cohorts of T1DM, 11 cohorts of mixed diabetes mellitus populations and four cohorts of the general population. Another top-studied biomarker was fibroblast growth factor-23, a marker related to fibrosis, which also showed statistically significant positive association with CKD outcomes, including five cohorts of individuals with T2DM only. Notably, the performance in clinical practice was not yet evaluated even for the most studied biomarkers (Liu et al., 2022).

As discussed in Chapter 3 (section 3.6.4), some studies comprised people with various comorbidities, including individuals with diabetes mellitus and adjusted for diabetes status, but very few studies to date separately depict metabolomic profiles of kidney decline specifically in people with T2DM. Solini et al. (2016) identified significantly correlated metabolites with baseline eGFR, which included C-glycosyl tryptophan, pseudouridine, and N-acetylthreonine, the sum of which predicted the eGFR decline during the follow up. However, due to heterogeneity of the metabolomic platforms applied, these findings could not be replicated in the ET2DS.

7.4 Risk prediction- incremental values of combining metabolites and traditional risk factors for predicting five-year CKD in T2DM

One of the primary aims of my research was to improve the prediction of incident CKD. The risk prediction analysis involved an evaluation of an existing risk prediction model (the reference model) followed by incorporation of its components with metabolites-based risk score and evaluation of predictive performance before and after addition of the new score. In the following section, I discuss the current state of research in the area of risk prediction for incident CKD and compare the performance of the selected reference risk prediction model in ET2DS with previous studies. I also consider the value of metabolomics in this field and evaluate its use for developing risk scores.

7.4.1 Performance of existing CKD risk scores

The Nelson risk prediction model, chosen as the reference risk prediction model for this thesis, was developed in a diverse diabetes population encompassing multiple cohorts from various countries worldwide to predict the risk of incident CKD (Nelson et al., 2019). It underwent evaluation in several external populations, demonstrating very good discrimination and calibration performance (Nelson et al., 2019, Slieker et al., 2021). When evaluated in ET2DS, the Nelson risk equation also showed relatively similar performance in terms of the discrimination based on concordance statistic (c-statistic) compared to previous studies. The c-statistic for the five-year predicted probability of incident CKD in ET2DS was 0.78 (95% CI 0.74, 0.82), while Nelson et al. (2019) reported an average c-statistic of 0.81 (IQR 0.80-0.82) when it was evaluated in nine external diabetes cohorts. Similarly, Slieker et al. (2021) showed a c-statistic of 0.81 (95% CI 0.81, 0.82) when Nelson risk equation was evaluated in the Hoorn Diabetes Care System cohort, which consisted of people with T2DM living in the Hoorn region of the Netherlands. In terms of calibration, the ET2DS showed a slope of 1.23, which fell within the higher-end of the range of calibration values (0.80 to 1.25) that were shown when Nelson et al. (2019) evaluated the CKD onset risk prediction using the reference model in a number of diabetes mellitus cohorts. Overall, previous external validation studies as well as the evaluation carried out in ET2DS consistently report good performance for predicting the 5-year risk of incident CKD.

The similar performance of the reference risk prediction model in the ET2DS and previous cohorts may, at least in part, reflect a similar case-mix. For example, the mean age of the Hoorn cohort was 62 (SD= 11) years (Slieker et al., 2021), which

overlaps with the mean age of included participants from ET2DS (67, SD= 4 years). Also, both studies included approximately 50% males, and the majority used oral glucose-lowering drugs (approximately 60% in both studies) for diabetes control. Similarly, BMI was also high in both cohorts, with an average of 29.3 (SD= 5.4) kg/m^2 and 30.9 (SD= 5.5) kg/m^2 points in the Hoorn and ET2DS, respectively. Furthermore, the cohorts used for the development of the Nelson risk equation was also similar in terms of case-mix, with an average age of 62 (SD= 11) years and mean BMI of 32 (SD= 6) kg/m^2.

However, despite the use of readily available demographic, clinical, and laboratory variables, the Nelson equation has not yet been implemented into clinical practice, a common flaw shared with many previously developed models in nephrology research. In contrast to the cardiology field, where risk calculators have been integrated into the clinical decision-making process, such as QRISK3 for predicting CVD risk in UK (Samarasekera et al., 2023), nephrology has lagged behind. While accurate risk calculators for predicting coronary heart disease incidence were initially developed in the late 1980s (Wilson et al., 1987), it was not until 2003 that the first reasonable attempt to predict the progression of CKD was published (Dimitrov et al., 2003). Since then, the area of risk prediction research in nephrology has expanded considerably, with at least 41 studies reporting on 64 models that predicted outcomes including albuminuria, diabetic kidney disease, CKD, and ESRD (Slieker et al., 2021). Nonetheless, so far only Kidney Failure Risk equation (Tangri et al., 2011, Tangri et al., 2016), which predicts two- and five-year risk of ESRD in CKD patients has been endorsed by CKD guidelines (Levin et al., 2013, Farrington et al., 2016).

The implementation of risk prediction models for incident CKD into clinical practice faces several challenges and complexities, contributing to limited adoption. First and foremost, research on predictive models for renal outcomes is suboptimal, often failing to account for competing events and measurement errors. Common shortcomings in current studies include a lack of calibration analyses and external validation in different populations, contributing to prevalent fallacies in the existing literature (Slieker et al., 2021). Second, there is a lack of robust clinical trial evidence to demonstrate the effectiveness and improved patient outcomes resulting from the implementation of CKD risk prediction models. Without such evidence, healthcare providers may be hesitant to integrate these models into routine practice. Third, the heterogeneity in the

underlying CKD causes, such as diabetes, hypertension, or glomerulonephritis, means that developing a universal risk prediction model that applies across diverse populations is difficult. Consequently, until further evidence proving that risk prediction models for incident CKD are accurate in multiple disease-specific subpopulations, ethnicities and continents along with robust evidence for clinical impact (e.g., referral process, wait times, health outcome improvement and economic benefits), these models remain within the research setting.

7.4.2 Individual metabolites and improved CKD risk prediction

Prior metabolomics studies aimed at predicting incident CKD or CKD progression have predominantly focused on the general population (many of these studies were described in Chapter 3). Few of these studies are directly comparable to mine due to variations in case-mix and the methods employed to integrate metabolomics data. Unlike the approach of constructing a multi-marker risk score using a metabolomic panel, the majority of these studies tend to augment traditional risk factors by incorporating individual metabolites.

In general, prospective metabolomic profiling studies recruited broader populations, some of which included people with diabetes mellitus. For instance, in the Framingham Heart Study, 120 metabolites were measured using liquid chromatography - mass spectrometry method in urine specimens from 193 cases along with age- and sex-matched controls (McMahon et al., 2017a). The analysis revealed that glycine and histidine were significantly associated with incident CKD independent of clinical covariates (including baseline eGFR, presence of diabetes, hypertension, and dipstick proteinuria). Addition of these two metabolites to established CKD risk factors led to a modestly improved discrimination of individuals at risk of developing CKD. Compared to a reference risk prediction model of established risk factors, the c-statistic increased from 0.72 (95% CI 0.67–0.77) to 0.76 (95% CI 0.71–0.81), resulting in a 4% improvement in discrimination. Another study, that also used the same metabolomic analysis method in CRIC study, participants with mild renal impairment, found a number of alterations in lipid profile associated with progression to ESRD (Afshinnia et al., 2016). Combining the significant lipids with baseline eGFR and uACR resulted in improved c-statistic, rising from 0.83 (95% CI 0.76- 0.90) to 0.92 (95% CI 0.88-0.97). Goek et al. (2013) used a different metabolomic analysis method of flow injection and liquid chromatography coupled to tandem mass spectrometry to

determine the baseline serum concentrations of 140 metabolites in participants from the KORA study. The study revealed that kynurenine-to-tryptophan ratio, phosphatidylcholine diacyl C42:5-to-phosphatidylcholine acylalkyl C36:0 ratio and spermidine were associated with a significantly higher incidence of CKD at follow-up visits. Combining the individual metabolites with clinical variables led to some improvement in discrimination, with c-statistic rising from 0.81 (95% CI 0.77 – 0.86) to 0.86 (95% CI 0.83- 0.90). More recently, a largest metabolomics study in this area of research, investigated the associations between incident CKD and metabolomic profiling measurements obtained using Nightingale platform in UK Biobank cohort (Geng et al., 2024). Similarly, to smaller studies, the results showed an improvement in c-statistic from 0.82 (0.81-0.83) to 0.83 (0.82-0.84), when individual metabolites were combined with some clinical measurements (Geng et al., 2024). Of the nine metabolites added to the model developed by Geng et al (2024), only isoleucine and glycoprotein acetyls overlapped with those selected in the results I presented in this thesis. The differences in the metabolite selection, could at least in part be explained by the fundamental differences between populations studied (e.g., UK Biobank consists of general population with only a small proportion of those with T2DM (2%) while ET2DS was specifically based on older people with T2DM), as the metabolome is likely to be affected by factors such as age and disease status. In turn, this highlights that the score based on LASSO selected metabolites in my study may not be generalisable to a wider population and may be specific to the population studied, potentially suggesting that the results of analysis may have been impacted by overfitting.

Metabolomics studies focusing on improving CKD risk prediction in populations with T2DM were more scarce. The metabolite profiling of baseline plasma specimens using untargeted mass-spectrometry revealed several associations between uremic solutes (metabolites known to accumulate in blood when kidneys are not filtering effectively) and ESRD in Joslin Kidney Study cohort of T2DM patients (Niewczas et al., 2014). Individually, these metabolites were independent predictors of kidney decline, demonstrating similar effect size and statistical significance before and after adjustment for clinical covariates such as albumin excretion rate, eGFR, and glycated hemoglobin (HbA1c), and multiple comparisons. When all of these metabolites were added to established CKD risk factors, the discrimination of individuals at risk of

developing ESRD was very good with a c-statistic of 0.89 compared to c-statistic range of 0.74–0.75 in models that only included a single metabolite combined with the risk factors. However, the discrimination was not compared to a model without any metabolites.

In summary, these studies demonstrate that, depending on the cohort and the chosen metabolomics approach, different metabolites may be identified to potentially enhance risk prediction. In contrast to my metabolomic analysis, which examined global metabolomic profiles, many analyses from the exemplified studies evaluated the predictive performance of metabolites identified from a specific subclass of metabolites, such as uremic solutes, amino acids or lipids.

7.4.3 Predictive value of the MetS

A few previous studies developed metabolites-based scores to predict the risk of incident CKD (Rhee et al., 2013), rapid CKD decline (Nkuipou-Kenfack et al., 2014, Rhee et al., 2016) and CKD progression to ESRD (Zacharias et al., 2019) in the general population as well as risk of CKD progression in people with T2DM (Solini et al., 2016). All of these studies have shown some improvement in the risk prediction when clinical covariates were combined with a metabolomic-based score. To the best of my knowledge, no studies have combined multiple metabolites into a single multi-metabolomic marker score to predict incident CKD specifically in people with T2DM.

In contrast to the well-established genetic risk scores within genomics, constructing an equivalent score in metabolomics lacks a standardized approach and each study varied in terms of methods. For example, Rhee et al. (2013) identified individual metabolites significantly associated with incident CKD after adjustment for covariates and multiple testing correction. Then, a stepwise logistic model on candidate metabolites was applied, and those metabolites that remained significant in the multivariable model were used to construct a multi-marker score based on the regression coefficients. Alternative approach was adopted by Solini et al. (2016), who utilized a machine learning technique called random forest (Breiman, 2001) to identify the metabolites that are most correlated with the decline in kidney function, which were then summed to calculate a score. Other similar examples of multi-marker score development are available in the cardiovascular field. For instance, Wurtz et al. (2015) employed a similar strategy as Rhee et al. (2013), where traditional regression models

were employed to identify candidate metabolites, followed by a stepwise selection process to determine the optimal combination of candidate metabolites, which were combined into a risk score. In a different approach, Wang et al. (2019) employed LASSO to initially select a compact set of candidate metabolites, which were incorporated into traditional multivariable regression model to calculate the metabolite-based risk score. This score was computed as the summation of selected metabolite levels, taking into consideration the direction of their effects. Inspired by the concept of polygenic risk scores in genetics and the previous studies described above, the Metabolomics-based risk score (MetS) in ET2DS was formulated using a weighted approach, a methodology extensively applied in various omics studies for integrating high-dimensional data (Ganz et al., 2016, Hu et al., 2021).

Based on their weights, the top five contributors to the MetS in my study were isoleucine, free cholesterol to lipids ratio in very small very low-density lipoprotein, 3-hydroxybutyrate, lactate and 18-2, linoleic acid. Although some moderate correlations were observed between traditional CKD predictors and components of the MetS (e.g., acetate and BMI), the updated risk prediction model with the MetS still showed some improvement in predictive performance. This might suggest that some metabolites of the MetS could lie on the pathways of traditional risk factors, but residual risk, which could not be reflected by these factors, might be complemented by the MetS. It is possible that metabolomic profiles could serve as mediators of association between CKD and some risk factors unmeasured in the ET2DS, such as physical activities and diet.

7.4.4 Metrics of the predictive performance

The updated risk prediction model in my study, incorporating components of the reference risk prediction model combined with MetS, demonstrated a modest improvement in discriminative capability, as indicated by the c-statistic rising from 0.80 (95% CI 0.76, 0.84) for re-fitted reference model to 0.84 (95% CI 0.80, 0.87) for the updated model. The marginal improvement in my study could be attributed to the effective performance of the reference model, consisting of 13 predictors with strong effect sizes.

Interestingly, the prediction model containing only MetS combined with baseline eGFR showed c-statistic of 0.81 (0.78, 0.85), which is very similar to the discrimination

performance of the reference model. While this highlights the significant ability of a plasma metabolite panel to predict disease to a comparable extent as clinical parameters, even in the absence of additional patient information, it does not advocate for the routine use of metabolomic measurements for prediction. In a similar vein, observations for other markers identified from high-throughput screens suggest that they may not substantially enhance the prediction of common complex diseases, such as CKD (Paynter et al., 2010, O'Seaghdha et al., 2012).

Notably, c-statistics often exhibit insensitivity in reflecting improved model performance when new biomarkers are introduced into a model already incorporating robust predictors (Moons et al., 2012b, Price et al., 2017). As a result, there is a recommendation to present reclassification metrics alongside c-statistics to provide a comprehensive understanding of the additional predictive value offered by novel biomarkers (Moons et al., 2012b, Pencina et al., 2011). The updated risk prediction model incorporating the MetS, demonstrated an improved net reclassification in both, the events group (NRI_{event} = 0.31; 95% CI: 0.13, 0.45) and the non-events group ($NRI_{non\text{-}event}$ = 0.31; 95%CI: 0.16, 0.42). The NRI statistics presented here focus on the movement of absolute predicted values after metabolomics score, MetS, was added to the updated risk prediction model (Pencina et al., 2011). The positive statistics for both groups affirmed that the updated model had shown improvement in risk classification and suggested that the updated model is better at correctly assigning individuals to their appropriate risk categories. The reclassification improvement in the non-event group might help these people avoid unnecessary anxiety about potential diagnosis and medication treatments that yield limited or no benefit. In contrast, improvement in identification of individuals at higher risk of developing CKD permits more intensive preventative strategies, which is crucial for slowing down kidney function decline and developing further complications, such as increasing the risk of CVD. Overall, improvement in classification of high and low risk patients has a potential to optimise the use of healthcare resources. Nonetheless, considering that the improvement in risk prediction was fairly modest (NRI values far from ideal of 1), the cost of additional time and money for metabolomic profiling, MetS or similar scores may not yield a sufficient benefit to advocate its use in a clinical setting. Furthermore, while the statistically significant result indicated that adding MetS yielded improvement in risk stratification, without further understanding of what the value means in terms of

the proportion of individuals who would be reclassified to high or low risk it is difficult to understand how clinically meaningful benefit is. However, at the time of writing this thesis, there were no published guidelines on CKD risk categories, so I could not conduct analysis of categorical NRI, which would determine proportions of individuals that moved from one risk category to another. Hence limiting the clinical interpretation of this analysis.

7.5 Future research

7.5.1 Metabolomic profiles of kidney function phenotypes in T2DM

While certain metabolites demonstrated a consistent association in both cross-sectional and longitudinal analyses, the next natural step involves external validation of these findings in larger independent cohorts. These populations should encompass identical metabolites as measured in the ET2DS, and the cohorts should include individuals with T2DM spanning a broader range of age and ethnicities to comprehensively assess the generalizability of the findings in this thesis. Once identified metabolites are replicated in external populations, their etiological roles could be further examined using genetic approaches, such as Mendelian Randomization. For example, Mendelian Randomization found that blood triglycerides had a potential causal role for incident CKD (Zhang et al., 2020b), while another study revealed that genetically higher HDL cholesterol level may be causally associated with improved kidney function (Lanktree et al., 2018). Moreover, systems biology approaches, such as ingenuity pathways analysis, could also be conducted to integrate gene and metabolite networks, in order to highlight links between biological pathways (Komorowsky et al., 2012).

It would also be relevant to investigate whether the metabolomic profiles of CKD vary between T1DM and T2DM, considering that T1DM presents with varying clinical features, disease progression and risk of complications (Teng et al., 2014, Ricciardi and Gnudi, 2021). In T1DM, CKD is primarily attributed to diabetic nephropathy, where chronic hyperglycaemia acts as the main pathogenetic factor and albuminuria is a well-established risk factor (Ricciardi and Gnudi, 2021). In contrast, CKD in T2DM is a combination of diabetic nephropathy, vascular, and interstitial kidney disease (Teng et al., 2014). A previous study has identified some metabolites that differentiate between CKD progression in T1DM and T2DM (Manca et al., 2021), but there is a need for studies to also explore the metabolomic features related to new-onset CKD in these

groups of patients. On a similar note, it would also be interesting to investigate whether the metabolomic profiles of CKD vary among different subtypes of T2DM. Despite diabetes traditionally being classified as T1DM and T2DM, the latter exhibits considerable heterogeneity in clinical presentation and progression. Ahlqvist et al. (2018) using six clinical variables, proposed a refined classification of T2DM into five subgroups with distinct disease progression and complication risks. A previous study already identified some distinct features of metabolomic profiles in the four T2DM subtypes (Zaghlool et al., 2022). Future research could delve further into the metabolomic signatures of CKD within these diabetes subtypes, potentially shedding light on the pathophysiological mechanisms contributing to the heterogeneity of renal decline risk among individuals with T2DM. In addition, identifying metabolomic features that differentiate between albuminuric and non-albuminuric CKD in people with diabetes would also be of interest as there is a growing body of research, which revealed significant differences in the sets of risk factors, renal morphology, comorbidity, and outcomes for these phenotypes (Pugliese et al., 2020). However, the molecular mechanisms of albuminuric and non-albuminuric CKD remain to be explored and metabolomic profiling may help to elucidate the pathophysiology behind these different phenotypes.

The variation in metabolite concentrations over time can be influenced by factors like intra-individual fluctuations and physiological conditions (e.g., physical exercise) (Nayor et al., 2020). In the ET2DS and most existing metabolomic studies, metabolomic profiles are typically measured at a single time point. Implementing repeated measurements of metabolomic data enables the calculation of mean and deviation in metabolite concentrations, providing a more accurate depiction of metabolomic signatures compared to a single-time snapshot. Exploring dynamic changes (e.g., trajectory) in metabolites through repeated measurements at multiple time points could offer valuable insights into the association between these changes and CKD. This approach may contribute to a better understanding of how environmental factors (e.g., diet, exercise, drug use) impact the molecular-level development of CKD (Kofink et al., 2017).

7.5.2 Combining metabolites and traditional risk factors for predicting 5-year CKD in Type 2 Diabetes Mellitus

The ultimate goal for any risk stratification tool is its integration into clinical practice. The research presented in this thesis should be viewed as an initial step in exploring the potential application of metabolomics for developing a novel tool. Before progressing to an evaluation of clinical utility, it is imperative to replicate these findings in external populations with T2DM. Once the generalizability and consistency of the findings are established across studies, clinical trials treating the MetS as an 'intervention' will be essential to determine if these equations can improve care and CKD outcomes in real-world scenarios. For instance, a forthcoming study could explore whether directing resources toward individuals at the highest risk of developing CKD enhances the management of blood pressure and/or promotes weight loss. Subsequent research could also investigate whether administering medications to address albuminuria or manage diabetes might serve as a preventive measure against the onset of reduced eGFR in individuals at risk.

Additionally, economic modelling will be necessary to evaluate the cost-effectiveness of implementing the MetS in clinical practice. Given the constrained additional value of MetS compared to readily available demographic variables and clinical measures like age, medication use and BMI, one could argue that the modest improvement in predictive capability may not be significant enough to warrant the cost. Alternatively, the risk prediction score updated with MetS may be best employed in smaller settings, such as clinical trials where stratifying the randomisation of patients by high (or low) levels of risk may ensure that different groups in clinical trials have similar risk levels for outcomes. Alternatively, it could be used to recruit high-risk patients who are more likely to have clinical outcomes, therefore reducing the sample size required to achieve sufficient power to address the question,

7.6 Conclusions

In conclusion, this current study involving a cohort of elderly individuals with T2DM revealed extensive alterations in the metabolomic profile associated with CKD, particularly in lipoproteins and their lipid constituents. Elevated HDL components were observed in individuals with better kidney function, whereas VLDL, triglycerides, and specific amino acids exhibited increased levels in those with poorer kidney function.

Low molecular-weight molecules, namely ApoA1, total choline, glucose and isoleucine, had a consistent association with cross-sectional and prospective outcomes. Subsequent replication studies are imperative to validate the longitudinal observations and elucidate whether the metabolic signals identified at baseline can reliably predict outcomes of kidney function decline.

Combining metabolites related to fluid balance, ketone bodies, amino acids, fatty acids, glycolysis, and lipoproteins into a multi-marker risk score, MetS, demonstrated a modest, but positive improvement in predicting the risk of five-year incident CKD in individuals with T2DM, over and above traditional risk factors. The subsequent step involves external validation of these findings, and, contingent upon the outcomes, it would be relevant to explore the potential use of a metabolomic risk score in other research settings such as clinical trials.

Bibliography

AFKARIAN, M., SACHS, M. C., KESTENBAUM, B., HIRSCH, I. B., TUTTLE, K. R., HIMMELFARB, J. & DE BOER, I. H. 2013. Kidney disease and increased mortality risk in type 2 diabetes. *J Am Soc Nephrol,* 24**,** 302-8.

AFSHINNIA, F., NAIR, V., LIN, J., RAJENDIRAN, T. M., SONI, T., BYUN, J., SHARMA, K., FORT, P. E., GARDNER, T. W., LOOKER, H. C., NELSON, R. G., BROSIUS, F. C., FELDMAN, E. L., MICHAILIDIS, G., KRETZLER, M. & PENNATHUR, S. 2019. Increased lipogenesis and impaired B-oxidation predict type 2 diabetic kidney disease progression in American Indians. *JCI Insight,* 4(21) (no pagination).

AFSHINNIA, F., RAJENDIRAN, T. M., HE, C., BYUN, J., MONTEMAYOR, D., DARSHI, M., TUMOVA, J., KIM, J., LIMONTE, C. P., MILLER, R. G., COSTACOU, T., ORCHARD, T. J., AHLUWALIA, T. S., ROSSING, P., SNELL-BERGEON, J. K., DE BOER, I. H., NATARAJAN, L., MICHAILIDIS, G., SHARMA, K. & PENNATHUR, S. 2021. Circulating Free Fatty Acid and Phospholipid Signature Predicts Early Rapid Kidney Function Decline in Patients With Type 1 Diabetes. *Diabetes Care,* 44(9)**,** 2098-2106.

AFSHINNIA, F., RAJENDIRAN, T. M., KARNOVSKY, A., SONI, T., WANG, X., XIE, D., YANG, W., SHAFI, T., WEIR, M. R., HE, J., BRECKLIN, C. S., RHEE, E. P., SCHELLING, J. R., OJO, A., FELDMAN, H., MICHAILIDIS, G. & PENNATHUR, S. 2016. Lipidomic Signature of Progression of Chronic Kidney Disease in the Chronic Renal Insufficiency Cohort. *Kidney Int Rep,* 1**,** 256-268.

AFSHINNIA, F., ZENG, L., BYUN, J., WERNISCH, S., DEO, R., CHEN, J., HAMM, L., MILLER, E. R., RHEE, E. P., FISCHER, M. J., SHARMA, K., FELDMAN, H. I., MICHAILIDIS, G. & PENNATHUR, S. 2020. Elevated lipoxygenase and cytochrome P450 products predict progression of chronic kidney disease. *Nephrology Dialysis Transplantation,* 35(2)**,** 303-312.

AGUILAR-RAMIREZ, D., ALEGRE-DIAZ, J., HERRINGTON, W. G., STAPLIN, N., RAMIREZ-REYES, R., GNATIUC, L., HILL, M., ROMER, F., TORRES, J., TRICHIA, E., WADE, R., COLLINS, R., EMBERSON, J. R., KURI-MORALES, P. & TAPIA-CONYER, R. 2021. Association of Kidney Function With NMR-Quantified Lipids, Lipoproteins, and Metabolic Measures in Mexican Adults. *J Clin Endocrinol Metab,* 106**,** 2828-2839.

AHLQVIST, E., STORM, P., KARAJAMAKI, A., MARTINELL, M., DORKHAN, M., CARLSSON, A., VIKMAN, P., PRASAD, R. B., ALY, D. M., ALMGREN, P., WESSMAN, Y., SHAAT, N., SPEGEL, P., MULDER, H., LINDHOLM, E., MELANDER, O., HANSSON, O., MALMQVIST, U., LERNMARK, A., LAHTI, K., FORSEN, T., TUOMI, T., ROSENGREN, A. H. & GROOP, L. 2018. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol,* 6**,** 361-369.

AL-SARI, N., KUTUZOVA, S., SUVITAIVAL, T., HENRIKSEN, P., POCIOT, F., ROSSING, P., MCCLOSKEY, D. & LEGIDO-QUIGLEY, C. 2022. Precision diagnostic approach to predict 5-year risk for microvascular complications in type 1 diabetes. *eBioMedicine,* 80(no pagination).

ALBA, A. C., AGORITSAS, T., WALSH, M., HANNA, S., IORIO, A., DEVEREAUX, P. J., MCGINN, T. & GUYATT, G. 2017. Discrimination and Calibration of Clinical Prediction Models: Users' Guides to the Medical Literature. *JAMA,* 318**,** 1377-1384.

ANDERS, H. J., HUBER, T. B., ISERMANN, B. & SCHIFFER, M. 2018. CKD in diabetes: diabetic kidney disease versus nondiabetic kidney disease. *Nat Rev Nephrol,* 14**,** 361-377.

ANWARDEEN, N. R., DIBOUN, I., MOKRAB, Y., ALTHANI, A. A. & ELRAYESS, M. A. 2023. Statistical methods and resources for biomarker discovery using metabolomics. *BMC Bioinformatics,* 24**,** 250.

ARGILES, A., SIWY, J., DURANTON, F., GAYRARD, N., DAKNA, M., LUNDIN, U., OSABA, L., DELLES, C., MOURAD, G., WEINBERGER, K. M. & MISCHAK, H. 2013. CKD273, a new proteomics classifier assessing CKD and its prognosis. *PLoS One,* 8**,** e62837.

ATTMAN, P. O. & ALAUPOVIC, P. 1991. Lipid and apolipoprotein profiles of uremic dyslipoproteinemia--relation to renal function and dialysis. *Nephron,* 57**,** 401-10.

ATTMAN, P. O., ALAUPOVIC, P. & GUSTAFSON, A. 1987. Serum apolipoprotein profile of patients with chronic renal failure. *Kidney Int,* 32**,** 368-75.

AVINASH, S., SINGH, V. P., AGARWAL, A. K., CHATTERJEE, S. & ARAYA, V. 2015. Identification and Stratification of Diabetic Kidney Disease Using Serum Cystatin C and Serum Creatinine Based Estimating Equations in Type 2 Diabetes: A Comparative Analysis. *J Assoc Physicians India,* 63**,** 28-35.

BAIGENT, C., LANDRAY, M. J., REITH, C., EMBERSON, J., WHEELER, D. C., TOMSON, C., WANNER, C., KRANE, V., CASS, A., CRAIG, J., NEAL, B., JIANG, L., HOOI, L. S., LEVIN, A., AGODOA, L., GAZIANO, M., KASISKE, B., WALKER, R., MASSY, Z. A., FELDT-RASMUSSEN, B., KRAIRITTICHAI, U., OPHASCHAROENSUK, V., FELLSTROM, B., HOLDAAS, H., TESAR, V., WIECEK, A., GROBBEE, D., DE ZEEUW, D., GRONHAGEN-RISKA, C., DASGUPTA, T., LEWIS, D., HERRINGTON, W., MAFHAM, M., MAJONI, W., WALLENDSZUS, K., GRIMM, R., PEDERSEN, T., TOBERT, J., ARMITAGE, J., BAXTER, A., BRAY, C., CHEN, Y., CHEN, Z., HILL, M., KNOTT, C., PARISH, S., SIMPSON, D., SLEIGHT, P., YOUNG, A., COLLINS, R. & INVESTIGATORS, S. 2011. The effects of lowering LDL cholesterol with simvastatin plus ezetimibe in patients with chronic kidney disease (Study of Heart and Renal Protection): a randomised placebo-controlled trial. *Lancet,* 377**,** 2181-92.

BAKRIS, G., OSHIMA, M., MAHAFFEY, K. W., AGARWAL, R., CANNON, C. P., CAPUANO, G., CHARYTAN, D. M., DE ZEEUW, D., EDWARDS, R., GREENE, T., HEERSPINK, H. J. L., LEVIN, A., NEAL, B., OH, R., POLLOCK, C., ROSENTHAL, N., WHEELER, D. C., ZHANG, H., ZINMAN, B., JARDINE, M. J. & PERKOVIC, V. 2020. Effects of Canagliflozin in Patients with Baseline eGFR <30 ml/min per 1.73 m(2): Subgroup Analysis of the Randomized CREDENCE Trial. *Clin J Am Soc Nephrol,* 15**,** 1705-1714.

BAKRIS, G. L., FONSECA, V. A., SHARMA, K. & WRIGHT, E. M. 2009. Renal sodium-glucose transport: role in diabetes mellitus and potential clinical implications. *Kidney Int,* 75**,** 1272-1277.

BAKRIS, G. L., WEIR, M. R., SHANIFAR, S., ZHANG, Z., DOUGLAS, J., VAN DIJK, D. J., BRENNER, B. M. & GROUP, R. S. 2003. Effects of blood pressure level on progression of diabetic nephropathy: results from the RENAAL study. *Arch Intern Med,* 163**,** 1555-65.

BARNES, S., BENTON, H. P., CASAZZA, K., COOPER, S. J., CUI, X., DU, X., ENGLER, J., KABAROWSKI, J. H., LI, S., PATHMASIRI, W., PRASAIN, J. K., RENFROW, M. B. & TIWARI, H. K. 2016. Training in metabolomics research. II. Processing and statistical analysis of metabolomics data, metabolite identification, pathway analysis, applications of metabolomics and its future. *J Mass Spectrom,* 51**,** 535-548.

BARRIOS, C., ZIERER, J., WURTZ, P., HALLER, T., METSPALU, A., GIEGER, C., THORAND, B., MEISINGER, C., WALDENBERGER, M., RAITAKARI, O., LEHTIMAKI, T., OTERO, S., RODRIGUEZ, E., PEDRO-BOTET, J., KAHONEN, M., ALA-KORPELA, M., KASTENMULLER, G., SPECTOR, T. D., PASCUAL, J. & MENNI, C. 2018. Circulating metabolic biomarkers of renal function in diabetic and non-diabetic populations. *Sci Rep,* 8**,** 15249.

BARTER, P. J., CAULFIELD, M., ERIKSSON, M., GRUNDY, S. M., KASTELEIN, J. J., KOMAJDA, M., LOPEZ-SENDON, J., MOSCA, L., TARDIF, J. C., WATERS, D. D., SHEAR, C. L., REVKIN, J. H., BUHR, K. A., FISHER, M. R., TALL, A. R., BREWER, B. & INVESTIGATORS, I. 2007. Effects of torcetrapib in patients at high risk for coronary events. *N Engl J Med,* 357**,** 2109-22.

BELLO, A. K., ALRUKHAIMI, M., ASHUNTANTANG, G. E., BASNET, S., ROTTER, R. C., DOUTHAT, W. G., KAZANCIOGLU, R., KOTTGEN, A., NANGAKU, M., POWE, N. R., WHITE, S. L., WHEELER, D. C. & MOE, O. 2017. Complications of chronic kidney disease: current state, knowledge gaps, and strategy for action. *Kidney Int Suppl (2011),* 7**,** 122-129.

BENJAMINI, Y. & HOCHBERG, Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological),* 57**,** 289-300.

BERG, K. & BORRESEN, A. L. 1976. Serum-high-density-lipoprotein and atherosclerotic heart-disease. *Lancet,* 1**,** 499-501.

BERNARD, L., ZHOU, L., SURAPANENI, A., CHEN, J., REBHOLZ, C. M., CORESH, J., YU, B., BOERWINKLE, E., SCHLOSSER, P. & GRAMS, M. E. 2022. Serum Metabolites and Kidney Outcomes: The Atherosclerosis Risk in Communities Study. *Kidney Medicine,* 4(9) (no pagination).

BETZ, B. B., JENKS, S. J., CRONSHAW, A. D., LAMONT, D. J., CAIRNS, C., MANNING, J. R., GODDARD, J., WEBB, D. J., MULLINS, J. J., HUGHES, J., MCLACHLAN, S., STRACHAN, M. W. J., PRICE, J. F. & CONWAY, B. R. 2016. Urinary peptidomics in a rodent model of diabetic nephropathy highlights epidermal growth factor as a biomarker for renal deterioration in patients with type 2 diabetes. *Kidney Int,* 89**,** 1125-1135.

BICTASH, M., EBBELS, T. M., CHAN, Q., LOO, R. L., YAP, I. K., BROWN, I. J., DE IORIO, M., DAVIGLUS, M. L., HOLMES, E., STAMLER, J., NICHOLSON, J. K. & ELLIOTT, P. 2010. Opening up the "Black Box": metabolic phenotyping and metabolome-wide association studies in epidemiology. *J Clin Epidemiol,* 63**,** 970-9.

BIDANI, A. K. & GRIFFIN, K. A. 2004. Pathophysiology of hypertensive renal damage: implications for therapy. *Hypertension,* 44**,** 595-601.

BIDIN, M. Z., SHAH, A. M., STANSLAS, J. & SEONG, C. L. T. 2019. Blood and urine biomarkers in chronic kidney disease: An update. *Clin Chim Acta,* 495**,** 239-250.

BJORK, J., GRUBB, A., LARSSON, A., HANSSON, L. O., FLODIN, M., STERNER, G., LINDSTROM, V. & NYMAN, U. 2015. Accuracy of GFR estimating equations combining standardized cystatin C and creatinine assays: a cross-sectional study in Sweden. *Clin Chem Lab Med,* 53**,** 403-14.

BJORNSTAD, P., CHERNEY, D. Z., SNELL-BERGEON, J. K., PYLE, L., REWERS, M., JOHNSON, R. J. & MAAHS, D. M. 2015. Rapid GFR decline is associated with renal hyperfiltration and impaired GFR in adults with Type 1 diabetes. *Nephrol Dial Transplant,* 30**,** 1706-11.

BLAISE, B. J., SHINTU, L., ELENA, B., EMSLEY, L., DUMAS, M. E. & TOULHOAT, P. 2009. Statistical recoupling prior to significance testing in nuclear magnetic resonance based metabonomics. *Anal Chem,* 81**,** 6242-51.

BLAND, J. M. & ALTMAN, D. G. 1995. Multiple significance tests: the Bonferroni method. *BMJ,* 310**,** 170.

BLEEKER, S. E., MOLL, H. A., STEYERBERG, E. W., DONDERS, A. R., DERKSEN-LUBSEN, G., GROBBEE, D. E. & MOONS, K. G. 2003. External validation is necessary in prediction research: a clinical example. *J Clin Epidemiol,* 56**,** 826-32.

BLIGH, J. 1953. The role of the liver and the kidneys in the maintenance of the level of free choline in plasma. *J Physiol,* 120**,** 53-62.

BODDE, M. C., HERMANS, M. P. J., JUKEMA, J. W., SCHALIJ, M. J., LIJFERING, W. M., ROSENDAAL, F. R., ROMIJN, F., RUHAAK, L. R., VAN DER LAARSE, A. & COBBAERT, C. M. 2019. Apolipoproteins A1, B, and apoB/apoA1 ratio are associated with first ST-segment elevation myocardial infarction but not with recurrent events during long-term follow-up. *Clin Res Cardiol,* 108**,** 520-538.

BOIRIE, Y., ALBRIGHT, R., BIGELOW, M. & NAIR, K. S. 2004. Impairment of phenylalanine conversion to tyrosine in end-stage renal disease causing tyrosine deficiency. *Kidney Int,* 66**,** 591-6.

BOLANOS-GARCIA, V. M. & MIGUEL, R. N. 2003. On the structure and function of apolipoproteins: more than a family of lipid-binding proteins. *Prog Biophys Mol Biol,* 83**,** 47-68.

BORGES, R. L., HIROTA, A. H., QUINTO, B. M., RIBEIRO, A. B., ZANELLA, M. T. & BATISTA, M. C. 2010. Is cystatin C a useful marker in the detection of diabetic kidney disease? *Nephron Clin Pract,* 114**,** c127-34.

BOUSSAGEON, R., BEJAN-ANGOULVANT, T., SAADATIAN-ELAHI, M., LAFONT, S., BERGEONNEAU, C., KASSAI, B., ERPELDINGER, S., WRIGHT, J. M., GUEYFFIER, F. & CORNU, C. 2011. Effect of intensive glucose lowering treatment on all cause mortality, cardiovascular death, and microvascular events in type 2 diabetes: meta-analysis of randomised controlled trials. *BMJ,* 343**,** d4169.

BOYLE, J. P., THOMPSON, T. J., GREGG, E. W., BARKER, L. E. & WILLIAMSON, D. F. 2010. Projection of the year 2050 burden of diabetes in the US adult population: dynamic modeling of incidence, mortality, and prediabetes prevalence. *Popul Health Metr,* 8**,** 29.

BRANTSMA, A. H., BAKKER, S. J., DE ZEEUW, D., DE JONG, P. E. & GANSEVOORT, R. T. 2006. Urinary albumin excretion as a predictor of the development of hypertension in the general population. *J Am Soc Nephrol,* 17**,** 331-5.

BREIMAN, L. 2001. Random forests. *Machine learning,* 45**,** 5-32.

BRENNER, B. M., COOPER, M. E., DE ZEEUW, D., KEANE, W. F., MITCH, W. E., PARVING, H. H., REMUZZI, G., SNAPINN, S. M., ZHANG, Z., SHAHINFAR, S. & INVESTIGATORS, R. S. 2001. Effects of losartan on renal and cardiovascular outcomes in patients with type 2 diabetes and nephropathy. *N Engl J Med,* 345**,** 861-9.

BRIER, G. W. 1950. Verification of forecasts expressed in terms of probability. *Monthly weather review,* 78**,** 1-3.

BROER, S. & FAIRWEATHER, S. J. 2018. Amino Acid Transport Across the Mammalian Intestine. *Compr Physiol,* 9**,** 343-373.

CAREY, R. N., WOLD, S. & WESTGARD, J. O. 1975. Principal component analysis: an alternative to "referee" methods in method comparison studies. *Anal Chem,* 47**,** 1824-9.

CARLSSON, L. M., ROMEO, S., JACOBSON, P., BURZA, M. A., MAGLIO, C., SJOHOLM, K., SVENSSON, P. A., HARALDSSON, B., PELTONEN, M. & SJOSTROM, L. 2015. The incidence of albuminuria after bariatric surgery and usual care in Swedish Obese Subjects (SOS): a prospective controlled intervention trial. *Int J Obes (Lond),* 39**,** 169-75.

CARVILLE, S., WONDERLING, D., STEVENS, P. & GUIDELINE DEVELOPMENT, G. 2014. Early identification and management of chronic kidney disease in adults: summary of updated NICE guidance. *BMJ,* 349**,** g4507.

CENTERS FOR DISEASE CONTROL AND PREVENTION, C. 2021. Chronic kidney disease (CKD) surveillance system: 2021.

CHADEAU-HYAM, M., CAMPANELLA, G., JOMBART, T., BOTTOLO, L., PORTENGEN, L., VINEIS, P., LIQUET, B. & VERMEULEN, R. C. 2013. Deciphering the complex: methodological overview of statistical models to derive OMICS-based biomarkers. *Environ Mol Mutagen,* 54**,** 542-57.

CHARLTON, J. R., SPRINGSTEEN, C. H. & CARMODY, J. B. 2014. Nephron number and its determinants in early life: a primer. *Pediatr Nephrol,* 29**,** 2299-308.

CHAUFFARD, A. 1913. Address in medicine, ON MEDICAL PROGNOSIS: ITS METHODS, ITS EVOLUTION, ITS LIMITATIONS: delivered at the seventeenth international congress of medicine. *British Medical Journal,* 2**,** 286.

CHAZOT, C., SHAHMIR, E., MATIAS, B., LAIDLAW, S. & KOPPLE, J. D. 1997. Dialytic nutrition: provision of amino acids in dialysate during hemodialysis. *Kidney Int,* 52**,** 1663-70.

CHEN, Y., LI, E. M. & XU, L. Y. 2022. Guide to Metabolomics Analysis: A Bioinformatics Workflow. *Metabolites,* 12.

CHENG, H. T., XU, X., LIM, P. S. & HUNG, K. Y. 2021. Worldwide Epidemiology of Diabetes-Related End-Stage Renal Disease, 2000-2015. *Diabetes Care,* 44**,** 89-97.

CHERTOW, G. M., VART, P., JONGS, N., TOTO, R. D., GORRIZ, J. L., HOU, F. F., MCMURRAY, J. J. V., CORREA-ROTTER, R., ROSSING, P., SJOSTROM, C. D., STEFANSSON, B. V., LANGKILDE, A. M., WHEELER, D. C., HEERSPINK, H. J. L., COMMITTEES, D.-C. T. & INVESTIGATORS 2021. Effects of Dapagliflozin in Stage 4 Chronic Kidney Disease. *J Am Soc Nephrol,* 32**,** 2352-2361.

CHEUNG, A. K., RAHMAN, M., REBOUSSIN, D. M., CRAVEN, T. E., GREENE, T., KIMMEL, P. L., CUSHMAN, W. C., HAWFIELD, A. T., JOHNSON, K. C., LEWIS, C. E., OPARIL, S., ROCCO, M. V., SINK, K. M., WHELTON, P. K., WRIGHT, J. T., JR., BASILE, J., BEDDHU, S., BHATT, U., CHANG, T. I., CHERTOW, G. M., CHONCHOL, M., FREEDMAN, B. I., HALEY, W., IX, J. H., KATZ, L. A., KILLEEN, A. A., PAPADEMETRIOU, V., RICARDO, A. C., SERVILLA, K., WALL, B., WOLFGRAM, D., YEE, J. & GROUP, S. R. 2017. Effects of Intensive BP Control in CKD. *J Am Soc Nephrol,* 28**,** 2812-2823.

CHI, Q., WANG, T. & HUANG, K. 2005. Effect of insulin nitration by peroxynitrite on its biological activity. *Biochem Biophys Res Commun,* 330**,** 791-6.

CHIEN, K. L., LIN, H. J., LEE, B. C., HSU, H. C., LEE, Y. T. & CHEN, M. F. 2010. A prediction model for the risk of incident chronic kidney disease. *Am J Med,* 123**,** 836-846 e2.

CHU, L., FULLER, M., JERVIS, K., CIACCIA, A. & ABITBOL, A. 2021. Prevalence of Chronic Kidney Disease in Type 2 Diabetes: The Canadian REgistry of Chronic Kidney Disease in Diabetes Outcomes (CREDO) Study. *Clin Ther,* 43**,** 1558-1573.

COCA, S. G., NADKARNI, G. N., HUANG, Y., MOLEDINA, D. G., RAO, V., ZHANG, J., FERKET, B., CROWLEY, S. T., FRIED, L. F. & PARIKH, C. R. 2017. Plasma biomarkers and kidney function decline in early and established diabetic kidney disease. *Journal of the American Society of Nephrology: JASN,* 28**,** 2786.

COLHOUN, H. M., LEE, E. T., BENNETT, P. H., LU, M., KEEN, H., WANG, S. L., STEVENS, L. K. & FULLER, J. H. 2001. Risk factors for renal failure: the WHO Mulinational Study of Vascular Disease in Diabetes. *Diabetologia,* 44 Suppl 2**,** S46-53.

COLHOUN, H. M. & MARCOVECCHIO, M. L. 2018. Biomarkers of diabetic kidney disease. *Diabetologia,* 61**,** 996-1011.

COLLABORATION, G. B. D. C. K. D. 2020. Global, regional, and national burden of chronic kidney disease, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet,* 395**,** 709-733.

COLLABORATORS, G. B. O. D. D. A. H. 2018. Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet,* 392**,** 1859-1922.

COLLINS, G. S., OMAR, O., SHANYINDE, M. & YU, L. M. 2013. A systematic review finds prediction models for chronic kidney disease were poorly reported and often developed using inappropriate methods. *J Clin Epidemiol,* 66**,** 268-77.

COLOMBO, M., LOOKER, H. C., FARRAN, B., HESS, S., GROOP, L., PALMER, C. N. A., BROSNAN, M. J., DALTON, R. N., WONG, M., TURNER, C., AHLQVIST, E., DUNGER, D., AGAKOV, F., DURRINGTON, P., LIVINGSTONE, S., BETTERIDGE, J., MCKEIGUE, P. M., COLHOUN, H. M. & INVESTIGATORS, S. 2019a. Serum kidney injury molecule 1 and beta(2)-microglobulin perform as well as larger biomarker panels for prediction of rapid decline in renal function in type 2 diabetes. *Diabetologia,* 62**,** 156-168.

COLOMBO, M., VALO, E., MCGURNAGHAN, S. J., SANDHOLM, N., BLACKBOURN, L. A. K., DALTON, R. N., DUNGER, D., GROOP, P. H., MCKEIGUE, P. M., FORSBLOM, C. & COLHOUN, H. M. 2019b. Biomarker panels associated with progression of renal disease in type 1 diabetes. *Diabetologia,* 62(9)**,** 1616-1627.

CONNELLY, M. A., OTVOS, J. D., SHALAUROVA, I., PLAYFORD, M. P. & MEHTA, N. N. 2017. GlycA, a novel biomarker of systemic inflammation and cardiovascular disease risk. *J Transl Med,* 15**,** 219.

CORESH, J., EKNOYAN, G. & LEVEY, A. S. 2002. Estimating the prevalence of low glomerular filtration rate requires attention to the creatinine assay calibration. *J Am Soc Nephrol,* 13**,** 2811-2; author reply 2812-6.

CRAIG, A., CLOAREC, O., HOLMES, E., NICHOLSON, J. K. & LINDON, J. C. 2006. Scaling and normalization effects in NMR spectroscopic metabonomic data sets. *Anal Chem,* 78**,** 2262-7.

CREMER, P., ELSTER, H., LABROT, B., KRUSE, B., MUCHE, R. & SEIDEL, D. 1988. Incidence rates of fatal and nonfatal myocardial infarction in relation to the lipoprotein profile: first prospective results from the Gottingen Risk, Incidence, and Prevalence Study (GRIPS). *Klin Wochenschr,* 66 Suppl 11**,** 42-9.

DANNE, O. & MOCKEL, M. 2010. Choline in acute coronary syndrome: an emerging biomarker with implications for the integrated assessment of plaque vulnerability. *Expert Rev Mol Diagn,* 10**,** 159-71.

DANNE, O., MOCKEL, M., LUEDERS, C., MUGGE, C., ZSCHUNKE, G. A., LUFFT, H., MULLER, C. & FREI, U. 2003. Prognostic implications of elevated whole blood choline levels in acute coronary syndromes. *Am J Cardiol,* 91**,** 1060-7.

DARLINGTON, O., DICKERSON, C., EVANS, M., MCEWAN, P., SORSTADIUS, E., SUGRUE, D., VAN HAALEN, H. & GARCIA SANCHEZ, J. J. 2021. Costs and Healthcare Resource Use Associated with Risk of Cardiovascular Morbidity in Patients with Chronic Kidney Disease: Evidence from a Systematic Literature Review. *Adv Ther,* 38**,** 994-1010.

DE BOER, I. H., CARAMORI, M. L., CHAN, J. C., HEERSPINK, H. J., HURST, C., KHUNTI, K., LIEW, A., MICHOS, E. D., NAVANEETHAN, S. D. & OLOWU, W. A. 2020. Executive summary of the 2020 KDIGO Diabetes Management in CKD Guideline: evidence-based advances in monitoring and treatment. *Kidney international,* 98**,** 839-848.

DE BOER, I. H., RUE, T. C., HALL, Y. N., HEAGERTY, P. J., WEISS, N. S. & HIMMELFARB, J. 2011. Temporal trends in the prevalence of diabetic kidney disease in the United States. *JAMA,* 305**,** 2532-9.

DE LIVERA, A. M., OLSHANSKY, M. & SPEED, T. P. 2013. Statistical analysis of metabolomics data. *Methods Mol Biol,* 1055**,** 291-307.

DE VRIES, A. P., RUGGENENTI, P., RUAN, X. Z., PRAGA, M., CRUZADO, J. M., BAJEMA, I. M., D'AGATI, V. D., LAMB, H. J., PONGRAC BARLOVIC, D., HOJS, R., ABBATE, M., RODRIQUEZ, R., MOGENSEN, C. E., PORRINI, E. & DIABESITY, E.-E. W. G. 2014. Fatty kidney: emerging role of ectopic lipid in obesity-related renal disease. *Lancet Diabetes Endocrinol,* 2**,** 417-26.

DE ZEEUW, D., ANZALONE, D. A., CAIN, V. A., CRESSMAN, M. D., HEERSPINK, H. J., MOLITORIS, B. A., MONYAK, J. T., PARVING, H. H., REMUZZI, G., SOWERS, J. R. & VIDT, D. G. 2015. Renal effects of atorvastatin and rosuvastatin in patients with diabetes who have progressive renal disease (PLANET I): a randomised clinical trial. *Lancet Diabetes Endocrinol,* 3**,** 181-90.

DE ZEEUW, D., REMUZZI, G., PARVING, H. H., KEANE, W. F., ZHANG, Z., SHAHINFAR, S., SNAPINN, S., COOPER, M. E., MITCH, W. E. & BRENNER, B. M. 2004. Proteinuria, a target for renoprotection in patients with type 2 diabetic nephropathy: lessons from RENAAL. *Kidney Int,* 65**,** 2309-20.

DEFRONZO, R. A. & FELIG, P. 1980. Amino acid metabolism in uremia: insights gained from normal and diabetic man. *Am J Clin Nutr,* 33**,** 1378-86.

DERBY, L., WARRAM, J. H., LAFFEL, L. M. & KROLEWSKI, A. S. 1989. Elevated blood pressure predicts the development of persistent proteinuria in the presence of poor glycemic control, in patients with type I diabetes. *Diabete Metab,* 15**,** 320-6.

DIMITROV, B. D., RUGGENENTI, P., STEFANOV, R., PERNA, A. & REMUZZI, G. 2003. Chronic nephropathies: individual risk for progression to end-stage renal failure as predicted by an integrated probabilistic model. *Nephron Clin Pract,* 95**,** c47-59.

DISEASE, K. 2020. Improving global outcomes (KDIGO) Diabetes Work Group—KDIGO 2020 Clinical Practice Guideline for Diabetes Management in Chronic Kidney Disease. *Kidney Int,* 98**,** S1-S115.

DOSHI, S. M. & FRIEDMAN, A. N. 2017. Diagnosis and Management of Type 2 Diabetic Kidney Disease. *Clin J Am Soc Nephrol,* 12**,** 1366-1373.

DUNCAN, B. B. & SCHMIDT, M. I. 2006. The epidemiology of low-grade chronic systemic inflammation and type 2 diabetes. *Diabetes Technol Ther,* 8**,** 7-17.

DUNCAN, B. B., SCHMIDT, M. I., PANKOW, J. S., BALLANTYNE, C. M., COUPER, D., VIGO, A., HOOGEVEEN, R., FOLSOM, A. R., HEISS, G. & ATHEROSCLEROSIS RISK IN COMMUNITIES, S. 2003. Low-grade systemic inflammation and the development of type 2 diabetes: the atherosclerosis risk in communities study. *Diabetes,* 52**,** 1799-805.

ECHOUFFO-TCHEUGUI, J. B. & KENGNE, A. P. 2012. Risk models to predict chronic kidney disease and its progression: a systematic review. *PLoS Med,* 9**,** e1001344.

EINHORN, D. & MENDE, C. W. 2015. COMBINING CREATININE-BASED eGFR WITH CYSTATIN C-BASED EGFR TO BETTER ASSESS RENAL FUNCTION IN PATIENTS WITH DIABETES AND CHRONIC KIDNEY DISEASE 3A: IMPLICATIONS FOR DRUG SELECTION AND DOSAGE IN TYPE 2 DIABETES. *Endocr Pract,* 21**,** 1301-2.

EJERBLAD, E., FORED, C. M., LINDBLAD, P., FRYZEK, J., MCLAUGHLIN, J. K. & NYREN, O. 2006. Obesity and risk for chronic renal failure. *J Am Soc Nephrol,* 17**,** 1695-702.

EMDIN, C. A., RAHIMI, K., NEAL, B., CALLENDER, T., PERKOVIC, V. & PATEL, A. 2015. Blood pressure lowering in type 2 diabetes: a systematic review and meta-analysis. *JAMA,* 313**,** 603-15.

ERQOU, S., THOMPSON, A., DI ANGELANTONIO, E., SALEHEEN, D., KAPTOGE, S., MARCOVINA, S. & DANESH, J. 2010. Apolipoprotein(a) isoforms and the risk of vascular disease: systematic review of 40 studies involving 58,000 participants. *J Am Coll Cardiol,* 55**,** 2160-7.

ESMEIJER, K., DE VRIES, A. P., MOOK-KANAMORI, D. O., DE FIJTER, J. W., ROSENDAAL, F. R., RABELINK, T. J., SMIT, R. A. J., DE MUTSERT, R. & HOOGEVEEN, E. K. 2019. Low Birth Weight and Kidney Function in Middle-Aged Men and Women: The Netherlands Epidemiology of Obesity Study. *Am J Kidney Dis,* 74**,** 751-760.

ESMEIJER, K., GELEIJNSE, J. M., GILTAY, E. J., STIJNEN, T., DEKKER, F. W., DE FIJTER, J. W., KROMHOUT, D. & HOOGEVEEN, E. K. 2018. Body-fat indicators and kidney function decline in older post-myocardial infarction patients: The Alpha Omega Cohort Study. *Eur J Prev Cardiol,* 25**,** 90-99.

ETGEN, T., CHONCHOL, M., FORSTL, H. & SANDER, D. 2012. Chronic kidney disease and cognitive impairment: a systematic review and meta-analysis. *Am J Nephrol,* 35**,** 474-82.

FARRINGTON, K., COVIC, A., AUCELLA, F., CLYNE, N., DE VOS, L., FINDLAY, A., FOUQUE, D., GRODZICKI, T., IYASERE, O., JAGER, K. J., JOOSTEN, H., MACIAS, J. F., MOONEY, A., NITSCH, D., STRYCKERS, M., TAAL, M., TATTERSALL, J., VAN ASSELT, D., VAN DEN NOORTGATE, N., NISTOR, I., VAN BIESEN, W. & GROUP, E. G. D. 2016. Clinical Practice Guideline on management of older patients with chronic kidney disease stage 3b or higher (eGFR <45 mL/min/1.73 m2). *Nephrol Dial Transplant,* 31**,** ii1-ii66.

FELLSTROM, B. C., JARDINE, A. G., SCHMIEDER, R. E., HOLDAAS, H., BANNISTER, K., BEUTLER, J., CHAE, D. W., CHEVAILE, A., COBBE, S. M., GRONHAGEN-RISKA, C., DE LIMA, J. J., LINS, R., MAYER, G., MCMAHON, A. W., PARVING, H. H., REMUZZI, G., SAMUELSSON, O., SONKODI, S., SCI, D., SULEYMANLAR, G., TSAKIRIS, D., TESAR, V., TODOROV, V., WIECEK, A., WUTHRICH, R. P., GOTTLOW, M., JOHNSSON, E., ZANNAD, F. & GROUP, A. S. 2009. Rosuvastatin and cardiovascular events in patients undergoing hemodialysis. *N Engl J Med,* 360**,** 1395-407.

FICOCIELLO, L. H., PERKINS, B. A., ROSHAN, B., WEINBERG, J. M., ASCHENGRAU, A., WARRAM, J. H. & KROLEWSKI, A. S. 2009. Renal hyperfiltration and the development of microalbuminuria in type 1 diabetes. *Diabetes Care,* 32**,** 889-93.

FIORENTINO, M., BOLIGNANO, D., TESAR, V., PISANO, A., BIESEN, W. V., TRIPEPI, G., D'ARRIGO, G., GESUALDO, L. & GROUP, E.-E. I. W. 2017. Renal biopsy in patients with diabetes: a pooled meta-analysis of 48 studies. *Nephrol Dial Transplant,* 32**,** 97-110.

FLOEGEL, A., STEFAN, N., YU, Z., MUHLENBRUCH, K., DROGAN, D., JOOST, H. G., FRITSCHE, A., HARING, H. U., HRABE DE ANGELIS, M., PETERS, A., RODEN, M., PREHN, C., WANG-SATTLER, R., ILLIG, T., SCHULZE, M. B., ADAMSKI, J., BOEING, H. & PISCHON, T. 2013. Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. *Diabetes,* 62**,** 639-48.

FLORES, R., JIN, X., CHANG, J., ZHANG, C., COGAN, D. G., SCHAEFER, E. J. & KRUTH, H. S. 2019. LCAT, ApoD, and ApoA1 Expression and Review of Cholesterol Deposition in the Cornea. *Biomolecules,* 9.

FLORVALL, G., BASU, S. & LARSSON, A. 2006. Apolipoprotein A1 is a stronger prognostic marker than are HDL and LDL cholesterol for cardiovascular disease and mortality in elderly men. *J Gerontol A Biol Sci Med Sci,* 61**,** 1262-6.

FLUGEL-LINK, R. M., JONES, M. R. & KOPPLE, J. D. 1983. Red cell and plasma amino acid concentrations in renal failure. *JPEN J Parenter Enteral Nutr,* 7**,** 450-6.

FOLEY, R. N., MURRAY, A. M., LI, S., HERZOG, C. A., MCBEAN, A. M., EGGERS, P. W. & COLLINS, A. J. 2005. Chronic kidney disease and the risk for cardiovascular disease, renal replacement, and death in the United States Medicare population, 1998 to 1999. *J Am Soc Nephrol,* 16**,** 489-95.

FOREMAN, K. J., MARQUEZ, N., DOLGERT, A., FUKUTAKI, K., FULLMAN, N., MCGAUGHEY, M., PLETCHER, M. A., SMITH, A. E., TANG, K., YUAN, C. W., BROWN, J. C., FRIEDMAN, J., HE, J., HEUTON, K. R., HOLMBERG, M., PATEL, D. J., REIDY, P., CARTER, A., CERCY, K., CHAPIN, A.,

DOUWES-SCHULTZ, D., FRANK, T., GOETTSCH, F., LIU, P. Y., NANDAKUMAR, V., REITSMA, M. B., REUTER, V., SADAT, N., SORENSEN, R. J. D., SRINIVASAN, V., UPDIKE, R. L., YORK, H., LOPEZ, A. D., LOZANO, R., LIM, S. S., MOKDAD, A. H., VOLLSET, S. E. & MURRAY, C. J. L. 2018. Forecasting life expectancy, years of life lost, and all-cause and cause-specific mortality for 250 causes of death: reference and alternative scenarios for 2016-40 for 195 countries and territories. *Lancet,* 392**,** 2052-2090.

FOROUHI, N. G. & WAREHAM, N. J. 2014. Epidemiology of diabetes. *Medicine (Abingdon),* 42**,** 698-702.

FOSTER, M. C., HWANG, S. J., LARSON, M. G., LICHTMAN, J. H., PARIKH, N. I., VASAN, R. S., LEVY, D. & FOX, C. S. 2008. Overweight, obesity, and the development of stage 3 CKD: the Framingham Heart Study. *Am J Kidney Dis,* 52**,** 39-48.

FRACCARO, P., VAN DER VEER, S., BROWN, B., PROSPERI, M., O'DONOGHUE, D., COLLINS, G. S., BUCHAN, I. & PEEK, N. 2016. An external validation of models to predict the onset of chronic kidney disease using population-based electronic health records from Salford, UK. *BMC Med,* 14**,** 104.

FUERTES-MARTIN, R., CORREIG, X., VALLVE, J. C. & AMIGO, N. 2020. Title: Human Serum/Plasma Glycoprotein Analysis by (1)H-NMR, an Emerging Method of Inflammatory Assessment. *J Clin Med,* 9.

FUKUDA, S. & KOPPLE, J. D. 1980. Chronic uremia syndrome in dogs induced with uranyl nitrate. *Nephron,* 25**,** 139-43.

FURST, P., ALVESSTRAND, A. & BERGSTROM, J. 1980. Effects of nutrition and catabolic stress on intracellular amino acid pools in uremia. *Am J Clin Nutr,* 33**,** 1387-95.

GAJJALA, P. R., SANATI, M. & JANKOWSKI, J. 2015. Cellular and Molecular Mechanisms of Chronic Kidney Disease with Diabetes Mellitus and Cardiovascular Diseases as Its Comorbidities. *Front Immunol,* 6**,** 340.

GANSEVOORT, R. T., CORREA-ROTTER, R., HEMMELGARN, B. R., JAFAR, T. H., HEERSPINK, H. J., MANN, J. F., MATSUSHITA, K. & WEN, C. P. 2013. Chronic kidney disease and cardiovascular risk: epidemiology, mechanisms, and prevention. *Lancet,* 382**,** 339-52.

GANZ, P., HEIDECKER, B., HVEEM, K., JONASSON, C., KATO, S., SEGAL, M. R., STERLING, D. G. & WILLIAMS, S. A. 2016. Development and Validation of a Protein-Based Risk Score for Cardiovascular Outcomes Among Patients With Stable Coronary Heart Disease. *JAMA,* 315**,** 2532-41.

GARCIA SANCHEZ, J. J., POWER, A., ABDUL SULTAN, A., ÄRNLÖV, J., CABRERA, C., DE NICOLA, L., HALIMI, J. M., MENNINI, F. S., NAVARRO-GONZÁLEZ, J. F., NOLAN, S., RETAT, L., WEBBER, L. & XU, M. 2021. POS-323 INSIDE CKD: PROJECTING THE FUTURE BURDEN OF CHRONIC KIDNEY DISEASE IN EUROPE USING MICROSIMULATION MODELLING. *Kidney International Reports,* 6**,** S139-S140.

GARIBOTTO, G., SOFIA, A., SAFFIOTI, S., BONANNI, A., MANNUCCI, I. & VERZOLA, D. 2010. Amino acid and protein metabolism in the human kidney and in patients with chronic kidney disease. *Clin Nutr,* 29**,** 424-33.

GAROFALO, C., BORRELLI, S., MINUTOLO, R., CHIODINI, P., DE NICOLA, L. & CONTE, G. 2017. A systematic review and meta-analysis suggests obesity predicts onset of chronic kidney disease in the general population. *Kidney Int,* 91**,** 1224-1235.

GASPARI, F., RUGGENENTI, P., PORRINI, E., MOTTERLINI, N., CANNATA, A., CARRARA, F., JIMENEZ SOSA, A., CELLA, C., FERRARI, S., STUCCHI, N., PARVANOVA, A., ILIEV, I., TREVISAN, R., BOSSI, A., ZALETEL, J., REMUZZI, G. & INVESTIGATORS, G. F. R. S. 2013. The GFR and GFR decline cannot be accurately estimated in type 2 diabetics. *Kidney Int,* 84**,** 164-73.

GENG, T. T., CHEN, J. X., LU, Q., WANG, P. L., XIA, P. F., ZHU, K., LI, Y., GUO, K. Q., YANG, K., LIAO, Y. F., ZHOU, Y. F., LIU, G. & PAN, A. 2024. Nuclear Magnetic Resonance-Based Metabolomics and Risk of CKD. *Am J Kidney Dis,* 83**,** 9-17.

GIL, R. B., ORTIZ, A., SANCHEZ-NINO, M. D., MARKOSKA, K., SCHEPERS, E., VANHOLDER, R., GLORIEUX, G., SCHMITT-KOPPLIN, P. & HEINZMANN, S. S. 2018. Increased urinary osmolyte excretion indicates chronic kidney disease severity and progression rate. *Nephrology Dialysis Transplantation,* 33**,** 2156-2164.

GLASSOCK, R. J., WARNOCK, D. G. & DELANAYE, P. 2017. The global burden of chronic kidney disease: estimates, variability and pitfalls. *Nat Rev Nephrol,* 13**,** 104-114.

GLASSOCK, R. J. & WINEARLS, C. 2008. CKD--fiction not fact. *Nephrol Dial Transplant,* 23**,** 2695-6; author reply 2696-9.

GO, A. S., CHERTOW, G. M., FAN, D., MCCULLOCH, C. E. & HSU, C. Y. 2004. Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *N Engl J Med,* 351**,** 1296-305.

GOEK, O. N., DORING, A., GIEGER, C., HEIER, M., KOENIG, W., PREHN, C., ROMISCH-MARGL, W., WANG-SATTLER, R., ILLIG, T., SUHRE, K., SEKULA, P., ADAMSKI, J., KOTTGEN, A. & MEISINGER, C. 2012a. Serum metabolites associate with kidney function decline and incident chronic kidney disease in the general population. *Nephrology Dialysis Transplantation,* 27.

GOEK, O. N., KOTTGEN, A., HOOGEVEEN, R. C., BALLANTYNE, C. M., CORESH, J. & ASTOR, B. C. 2012b. Association of apolipoprotein A1 and B with kidney function and chronic kidney disease in two multiethnic population samples. *Nephrol Dial Transplant,* 27**,** 2839-47.

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. 2013. Metabolites associate with kidney function decline and incident chronic kidney disease in the general population. *Nephrol Dial Transplant,* 28**,** 2131-8.

GOOD, D. M., ZURBIG, P., ARGILES, A., BAUER, H. W., BEHRENS, G., COON, J. J., DAKNA, M., DECRAMER, S., DELLES, C., DOMINICZAK, A. F., EHRICH, J. H., EITNER, F., FLISER, D., FROMMBERGER, M., GANSER, A., GIROLAMI, M. A., GOLOVKO, I., GWINNER, W., HAUBITZ, M., HERGET-ROSENTHAL, S., JANKOWSKI, J., JAHN, H., JERUMS, G., JULIAN, B. A., KELLMANN, M., KLIEM, V., KOLCH, W., KROLEWSKI, A. S., LUPPI, M., MASSY, Z., MELTER, M., NEUSUSS, C., NOVAK, J., PETER, K., ROSSING, K., RUPPRECHT, H., SCHANSTRA, J. P., SCHIFFER, E., STOLZENBURG, J. U., TARNOW, L., THEODORESCU, D., THONGBOONKERD, V., VANHOLDER, R., WEISSINGER, E. M., MISCHAK, H. & SCHMITT-KOPPLIN, P. 2010. Naturally occurring human urinary peptides for use in diagnosis of chronic kidney disease. *Mol Cell Proteomics,* 9**,** 2424-37.

GORBOULEV, V., SCHURMANN, A., VALLON, V., KIPP, H., JASCHKE, A., KLESSEN, D., FRIEDRICH, A., SCHERNECK, S., RIEG, T., CUNARD, R., VEYHL-WICHMANN, M., SRINIVASAN, A., BALEN, D., BRELJAK, D., REXHEPAJ, R., PARKER, H. E., GRIBBLE, F. M., REIMANN, F., LANG, F., WIESE, S., SABOLIC, I., SENDTNER, M. & KOEPSELL, H. 2012. Na(+)-D-glucose cotransporter SGLT1 is pivotal for intestinal glucose absorption and glucose-dependent incretin secretion. *Diabetes,* 61**,** 187-96.

GRIFFIN, S. J., BORCH-JOHNSEN, K., DAVIES, M. J., KHUNTI, K., RUTTEN, G. E., SANDBAEK, A., SHARP, S. J., SIMMONS, R. K., VAN DEN DONK, M., WAREHAM, N. J. & LAURITZEN, T. 2011. Effect of early intensive multifactorial therapy on 5-year cardiovascular outcomes in individuals with type 2 diabetes detected by screening (ADDITION-Europe): a cluster-randomised trial. *Lancet,* 378**,** 156-67.

GROUP, I. G. O. K. D. W. 2022. KDIGO 2022 clinical practice guideline for diabetes management in chronic kidney disease. *Kidney international,* 102**,** S1-S127.

GULYASSY, P. F., AVIRAM, A. & PETERS, J. H. 1970. Evaluation of amino acid and protein requirements in chronic uremia. *Arch Intern Med,* 126**,** 855-9.

GUO, K., ZHANG, L., ZHAO, F., LU, J., PAN, P., YU, H., BAO, Y., CHEN, H. & JIA, W. 2016. Prevalence of chronic kidney disease and associated factors in Chinese individuals with type 2 diabetes: Cross-sectional study. *J Diabetes Complications,* 30**,** 803-10.

GURSKY, O. 2005. Apolipoprotein structure and dynamics. *Curr Opin Lipidol,* 16**,** 287-94.

HALAMA, A., HORSCH, M., KASTENMULLER, G., MOLLER, G., KUMAR, P., PREHN, C., LAUMEN, H., HAUNER, H., HRABE DE ANGELIS, M., BECKERS, J., SUHRE, K. & ADAMSKI, J. 2016. Metabolic switch during adipogenesis: From branched chain amino acid catabolism to lipid synthesis. *Arch Biochem Biophys,* 589**,** 93-107.

HALIMI, J.-M. 2012. The emerging concept of chronic kidney disease without clinical proteinuria in diabetic patients. *Diabetes & metabolism,* 38**,** 291-297.

HARRELL, F. E., JR., LEE, K. L. & MARK, D. B. 1996. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med,* 15**,** 361-87.

HAY, S. & COLLABORATORS, G. B. O. D. O. 2017. Health effects of overweight and obesity in 195 countries over 25 years. *New England journal of medicine,* 377.

HAYEK, S. S., QUYYUMI, A. A. & REISER, J. 2016. Soluble Urokinase Receptor and Chronic Kidney Disease. *N Engl J Med,* 374**,** 891.

HAYNES, R., LEWIS, D., EMBERSON, J., REITH, C., AGODOA, L., CASS, A., CRAIG, J. C., DE ZEEUW, D., FELDT-RASMUSSEN, B., FELLSTROM, B., LEVIN, A., WHEELER, D. C., WALKER, R., HERRINGTON, W. G., BAIGENT, C., LANDRAY, M. J., GROUP, S. C. & GROUP, S. C. 2014. Effects of lowering LDL cholesterol on progression of kidney disease. *J Am Soc Nephrol,* 25**,** 1825-33.

HEERSPINK, H. J. L., PERCO, P., MULDER, S., LEIERER, J., HANSEN, M. K., HEINZEL, A. & MAYER, G. 2019. Canagliflozin reduces inflammation and fibrosis biomarkers: a potential mechanism of action for beneficial effects of SGLT2 inhibitors in diabetic kidney disease. *Diabetologia,* 62**,** 1154-1166.

HEMINGWAY, H., CROFT, P., PEREL, P., HAYDEN, J. A., ABRAMS, K., TIMMIS, A., BRIGGS, A., UDUMYAN, R., MOONS, K. G., STEYERBERG, E. W., ROBERTS, I., SCHROTER, S., ALTMAN, D. G., RILEY, R. D. & GROUP, P. 2013. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ,* 346**,** e5595.

HILL, C. J., CARDWELL, C. R., PATTERSON, C. C., MAXWELL, A. P., MAGEE, G. M., YOUNG, R. J., MATTHEWS, B., O'DONOGHUE, D. J. & FOGARTY, D. G. 2014. Chronic kidney disease and diabetes in the national health service: a cross-sectional survey of the U.K. national diabetes audit. *Diabet Med,* 31**,** 448-54.

HILL, N. R., FATOBA, S. T., OKE, J. L., HIRST, J. A., O'CALLAGHAN, C. A., LASSERSON, D. S. & HOBBS, F. D. 2016. Global Prevalence of Chronic Kidney Disease - A Systematic Review and Meta-Analysis. *PLoS One,* 11**,** e0158765.

HINGORANI, A. D., WINDT, D. A., RILEY, R. D., ABRAMS, K., MOONS, K. G., STEYERBERG, E. W., SCHROTER, S., SAUERBREI, W., ALTMAN, D. G., HEMINGWAY, H. & GROUP, P. 2013. Prognosis research strategy (PROGRESS) 4: stratified medicine research. *BMJ,* 346**,** e5793.

HIPPISLEY-COX, J., COUPLAND, C. & BRINDLE, P. 2017. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ,* 357**,** j2099.

HIRAKAWA, Y., YOSHIOKA, K., KOJIMA, K., YAMASHITA, Y., SHIBAHARA, T., WADA, T., NANGAKU, M. & INAGI, R. 2022. Potential progression biomarkers of diabetic kidney disease determined using comprehensive machine learning analysis of non-targeted metabolomics. *Scientific reports,* 12(1)**,** 16287.

HOERGER, T. J., SEGEL, J. E., GREGG, E. W. & SAADDINE, J. B. 2008. Is glycemic control improving in U.S. adults? *Diabetes Care,* 31**,** 81-6.

HOLMAN, R. R., PAUL, S. K., BETHEL, M. A., MATTHEWS, D. R. & NEIL, H. A. 2008. 10-year follow-up of intensive glucose control in type 2 diabetes. *N Engl J Med,* 359**,** 1577-89.

HOLMES, E., LOO, R. L., STAMLER, J., BICTASH, M., YAP, I. K., CHAN, Q., EBBELS, T., DE IORIO, M., BROWN, I. J., VESELKOV, K. A., DAVIGLUS, M. L., KESTELOOT, H., UESHIMA, H., ZHAO, L., NICHOLSON, J. K. & ELLIOTT, P. 2008a. Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature,* 453**,** 396-400.

HOLMES, E., WILSON, I. D. & NICHOLSON, J. K. 2008b. Metabolic phenotyping in health and disease. *Cell,* 134**,** 714-7.

HOLZER, M., BIRNER-GRUENBERGER, R., STOJAKOVIC, T., EL-GAMAL, D., BINDER, V., WADSACK, C., HEINEMANN, A. & MARSCHE, G. 2011. Uremia alters HDL composition and function. *J Am Soc Nephrol,* 22**,** 1631-41.

HOOGEVEEN, E. K., ROTHMAN, K. J., VOSKAMP, P. W. M., DE MUTSERT, R., HALBESMA, N., DEKKER, F. W. & GROUP, P.-S. 2017. Obesity and risk of death or dialysis in younger and older patients on specialized pre-dialysis care. *PLoS One,* 12**,** e0184007.

HOVIND, P., TARNOW, L., ROSSING, K., ROSSING, P., EISING, S., LARSEN, N., BINDER, C. & PARVING, H. H. 2003. Decreasing incidence of severe diabetic microangiopathy in type 1 diabetes. *Diabetes Care,* 26**,** 1258-64.

HSU, C. Y., MCCULLOCH, C. E., IRIBARREN, C., DARBINIAN, J. & GO, A. S. 2006. Body mass index and risk for end-stage renal disease. *Ann Intern Med,* 144**,** 21-8.

HU, C. & DIGNAM, J. J. 2019. Biomarker-driven oncology clinical trials: Key design elements, types, features, and practical considerations. *JCO Precision Oncology,* 1**,** 1-12.

HU, T., CHITNIS, N., MONOS, D. & DINH, A. 2021. Next-generation sequencing technologies: An overview. *Hum Immunol,* 82**,** 801-811.

HUANG, J., HUTH, C., COVIC, M., TROLL, M., ADAM, J., ZUKUNFT, S., PREHN, C., WANG, L., NANO, J., SCHEERER, M. F., NESCHEN, S., KASTENMÜLLER, G., SUHRE, K., LAXY, M., SCHLIESS, F., GIEGER, C., ADAMSKI, J., HRABE DE ANGELIS, M., PETERS, A. & WANG-SATTLER, R. 2020. Machine Learning Approaches Reveal Metabolic Signatures of Incident Chronic Kidney Disease in Individuals With Prediabetes and Type 2 Diabetes. *Diabetes,* 69**,** 2756-2765.

HUANG, Z., KLARIC, L., KRASAUSKAITE, J., KHALID, W., STRACHAN, M. W. J., WILSON, J. F. & PRICE, J. F. 2023. Combining serum metabolomic profiles with traditional risk factors improves 10-year cardiovascular risk prediction in people with type 2 diabetes. *Eur J Prev Cardiol*.

HUTCHISON, S. R. 1928. *Principles of Diagnosis Prognosis and Treatment: A Trilogy*, John Wright and Sons Limited.

IBARRA-GONZALEZ, I., CRUZ-BAUTISTA, I., BELLO-CHAVOLLA, O. Y., VELA-AMIEVA, M., PALLARES-MENDEZ, R., NEVAREZ, D., SALAS-TAPIA, M. F., ROSAS-FLOTA, X., GONZALEZ-ACEVEDO, M., PALACIOS-PENALOZA, A., MORALES-ESPONDA, M., AGUILAR-SALINAS, C. A. & DEL BOSQUE-PLATA, L. 2018. Optimization of kidney dysfunction prediction in diabetic kidney disease using targeted metabolomics. *Acta Diabetologica,* 55**,** 1151-1161.

IBRAHIM, N. E., GAGGIN, H. K., KONSTAM, M. A. & JANUZZI JR, J. L. 2016. Established and emerging roles of biomarkers in heart failure clinical trials. *Circulation: heart failure,* 9**,** e002528.

ICD-10 1997. Implementation of the International Statistical Classification of Diseases and Related Health Problems, Tenth Revision (ICD-10). *Epidemiol Bull,* 18**,** 1-4.

IGLAY, K., HANNACHI, H., JOSEPH HOWIE, P., XU, J., LI, X., ENGEL, S. S., MOORE, L. M. & RAJPATHAK, S. 2016. Prevalence and co-prevalence of comorbidities among patients with type 2 diabetes mellitus. *Curr Med Res Opin,* 32**,** 1243-52.

ILIADIS, F., DIDANGELOS, T., NTEMKA, A., MAKEDOU, A., MORALIDIS, E., GOTZAMANI-PSARAKOU, A., KOULOUKOURGIOTOU, T. & GREKAS, D. 2011. Glomerular filtration rate estimation in patients with type 2 diabetes: creatinine- or cystatin C-based equations? *Diabetologia,* 54**,** 2987-94.

IMPERATORE, G., CADWELL, B. L., GEISS, L., SAADINNE, J. B., WILLIAMS, D. E., FORD, E. S., THOMPSON, T. J., NARAYAN, K. M. & GREGG, E. W. 2004. Thirty-year trends in cardiovascular risk factor levels among US adults with diabetes: National Health and Nutrition Examination Surveys, 1971-2000. *Am J Epidemiol,* 160**,** 531-9.

IN, J. & LEE, D. K. 2023. Survival analysis: Part II - applied clinical data analysis. *Korean J Anesthesiol,* 76**,** 84-85.

INTERNATIONAL HUMAN GENOME SEQUENCING, C. 2004. Finishing the euchromatic sequence of the human genome. *Nature,* 431**,** 931-45.

ISHIKAWA, T., FIDGE, N., THELLE, D. S., FORDE, O. H. & MILLER, N. E. 1978. The Tromso Heart Study: serum apolipoprotein AI concentration in relation to future coronary heart disease. *Eur J Clin Invest,* 8**,** 179-82.

JAGER, K. J., KOVESDY, C., LANGHAM, R., ROSENBERG, M., JHA, V. & ZOCCALI, C. 2019. A single number for advocacy and communication-worldwide more than 850 million individuals have kidney diseases. *Kidney Int,* 96**,** 1048-1050.

JANG, C., OH, S. F., WADA, S., ROWE, G. C., LIU, L., CHAN, M. C., RHEE, J., HOSHINO, A., KIM, B., IBRAHIM, A., BACA, L. G., KIM, E., GHOSH, C. C., PARIKH, S. M., JIANG, A., CHU, Q., FORMAN, D. E., LECKER, S. H., KRISHNAIAH, S., RABINOWITZ, J. D., WELJIE, A. M., BAUR, J. A., KASPER, D. L. & ARANY, Z. 2016. A branched-chain amino acid metabolite drives vascular fatty acid transport and causes insulin resistance. *Nat Med,* 22**,** 421-6.

JENKS, S. J., CONWAY, B. R., MCLACHLAN, S., TEOH, W. L., WILLIAMSON, R. M., WEBB, D. J., WELSH, P., SATTAR, N., STRACHAN, M. W. J. & PRICE, J. F. 2017. Cardiovascular disease biomarkers are associated with declining renal function in type 2 diabetes. *Diabetologia,* 60**,** 1400-1408.

JHA, V., GARCIA-GARCIA, G., ISEKI, K., LI, Z., NAICKER, S., PLATTNER, B., SARAN, R., WANG, A. Y. & YANG, C. W. 2013. Chronic kidney disease: global dimension and perspectives. *Lancet,* 382**,** 260-72.

JITRAKNATEE, J., RUENGORN, C. & NOCHAIWONG, S. 2020. Prevalence and Risk Factors of Chronic Kidney Disease among Type 2 Diabetes Patients: A Cross-Sectional Study in Primary Care Practice. *Sci Rep,* 10**,** 6205.

JONES, M. R., KOPPLE, J. D. & SWENDSEID, M. E. 1978. Phenylalanine metabolism in uremic and normal man. *Kidney Int,* 14**,** 169-79.

JU, W., NAIR, V., SMITH, S., ZHU, L., SHEDDEN, K., SONG, P. X. K., MARIANI, L. H., EICHINGER, F. H., BERTHIER, C. C., RANDOLPH, A., LAI, J. Y., ZHOU, Y., HAWKINS, J. J., BITZER, M., SAMPSON, M. G., THIER, M., SOLIER, C., DURAN-PACHECO, G. C., DUCHATEAU-NGUYEN, G., ESSIOUX, L., SCHOTT, B., FORMENTINI, I., MAGNONE, M. C., BOBADILLA, M., COHEN, C. D., BAGNASCO, S. M., BARISONI, L., LV, J., ZHANG, H., WANG, H. Y., BROSIUS, F. C., GADEGBEKU, C. A., KRETZLER, M., ERCB, C. P. N. & CONSORTIUM, P. K.-I. 2015. Tissue transcriptome-driven identification of epidermal growth factor as a chronic kidney disease biomarker. *Sci Transl Med,* 7**,** 316ra193.

JUN, M., ZHU, B., TONELLI, M., JARDINE, M. J., PATEL, A., NEAL, B., LIYANAGE, T., KEECH, A., CASS, A. & PERKOVIC, V. 2012. Effects of fibrates in kidney disease: a systematic review and meta-analysis. *J Am Coll Cardiol,* 60**,** 2061-71.

KALANTARI, S. & NAFAR, M. 2019. An update of urine and blood metabolomics in chronic kidney disease. *Biomark Med,* 13**,** 577-597.

KANEGAE, H., SUZUKI, K., FUKATANI, K., ITO, T., HARADA, N. & KARIO, K. 2020. Highly precise risk prediction model for new-onset hypertension using artificial intelligence techniques. *J Clin Hypertens (Greenwich),* 22**,** 445-450.

KAUFMAN, S. 1993. The phenylalanine hydroxylating system. *Adv Enzymol Relat Areas Mol Biol,* 67**,** 77-264.

KAYSEN, G. 2006. Dyslipidemia in chronic kidney disease: Causes and consequences. *Kidney International,* 70**,** S55-S58.

KAZANCIOGLU, R. 2013. Risk factors for chronic kidney disease: an update. *Kidney Int Suppl (2011),* 3**,** 368-371.

KEECH, A., SIMES, R. J., BARTER, P., BEST, J., SCOTT, R., TASKINEN, M. R., FORDER, P., PILLAI, A., DAVIS, T., GLASZIOU, P., DRURY, P., KESANIEMI, Y. A., SULLIVAN, D., HUNT, D., COLMAN, P., D'EMDEN, M., WHITING, M., EHNHOLM, C., LAAKSO, M. & INVESTIGATORS, F. S. 2005. Effects of long-term fenofibrate therapy on cardiovascular events in 9795 people with type 2 diabetes mellitus (the FIELD study): randomised controlled trial. *Lancet,* 366**,** 1849-61.

KEITH, D. S., NICHOLS, G. A., GULLION, C. M., BROWN, J. B. & SMITH, D. H. 2004. Longitudinal follow-up and outcomes among a population with chronic kidney disease in a large managed care organization. *Arch Intern Med,* 164**,** 659-63.

KENT, D. M. & HAYWARD, R. A. 2007. Limitations of applying summary results of clinical trials to individual patients: the need for risk stratification. *JAMA,* 298**,** 1209-12.

KERR, K. F., MCCLELLAND, R. L., BROWN, E. R. & LUMLEY, T. 2011. Evaluating the incremental value of new biomarkers with integrated discrimination improvement. *Am J Epidemiol,* 174**,** 364-74.

KERR, M., BRAY, B., MEDCALF, J., O'DONOGHUE, D. J. & MATTHEWS, B. 2012. Estimating the financial cost of chronic kidney disease to the NHS in England. *Nephrol Dial Transplant,* 27 Suppl 3**,** iii73-80.

KESTENBAUM, B., RUDSER, K. D., DE BOER, I. H., PERALTA, C. A., FRIED, L. F., SHLIPAK, M. G., PALMAS, W., STEHMAN-BREEN, C. & SISCOVICK, D. S. 2008. Differences in kidney function and incident hypertension: the multi-ethnic study of atherosclerosis. *Ann Intern Med,* 148**,** 501-8.

KHALSA, D. D., BEYDOUN, H. A. & CARMODY, J. B. 2016. Prevalence of chronic kidney disease risk factors among low birth weight adolescents. *Pediatr Nephrol,* 31**,** 1509-16.

KHAN, M. A. B., HASHIM, M. J., KING, J. K., GOVENDER, R. D., MUSTAFA, H. & AL KAABI, J. 2020. Epidemiology of Type 2 Diabetes - Global Burden of Disease and Forecasted Trends. *J Epidemiol Glob Health,* 10**,** 107-111.

KIANMEHR, H., ZHANG, P., LUO, J., GUO, J., PAVKOV, M. E., BULLARD, K. M., GREGG, E. W., OSPINA, N. S., FONSECA, V., SHI, L. & SHAO, H. 2022. Potential Gains in Life Expectancy Associated With Achieving Treatment Goals in US Adults With Type 2 Diabetes. *JAMA Netw Open,* 5**,** e227705.

KIM, Y., LEE, J., KANG, M. S., SONG, J., KIM, S. G., CHO, S., HUH, H., LEE, S., PARK, S., JO, H. A., YANG, S. H., PAEK, J. H., PARK, W. Y., HAN, S. S., LEE, H., LEE, J. P., JOO, K. W., LIM, C. S., HWANG, G. S. & KIM, D. K. 2023. Urinary Metabolite Profile Predicting the Progression of CKD. *Kidney360,* 4**,** 1048-1057.

KIMURA, T., HAMASE, K., MIYOSHI, Y., YAMAMOTO, R., YASUDA, K., MITA, M., RAKUGI, H., HAYASHI, T. & ISAKA, Y. 2016a. Chiral amino acid metabolomics for novel biomarker screening in the prognosis of chronic kidney disease. *Scientific reports,* 6**,** 26137.

KIMURA, T., YASUDA, K., YAMAMOTO, R., SOGA, T., RAKUGI, H., HAYASHI, T. & ISAKA, Y. 2016b. Identification of biomarkers for development of end-stage kidney disease in chronic kidney disease by metabolomic profiling. *Scientific reports,* 6**,** 26138.

KNIGHT, E. L., VERHAVE, J. C., SPIEGELMAN, D., HILLEGE, H. L., DE ZEEUW, D., CURHAN, G. C. & DE JONG, P. E. 2004. Factors influencing serum cystatin C levels other than renal function and the impact on renal function measurement. *Kidney Int,* 65**,** 1416-21.

KNOWLER, W. C., PETTITT, D. J., BENNETT, P. H. & WILLIAMS, R. C. 1983. Diabetes mellitus in the Pima Indians: genetic and evolutionary considerations. *Am J Phys Anthropol,* 62**,** 107-14.

KOFINK, D., EPPINGA, R. N., VAN GILST, W. H., BAKKER, S. J. L., DULLAART, R. P. F., VAN DER HARST, P. & ASSELBERGS, F. W. 2017. Statin Effects on Metabolic Profiles: Data From the PREVEND IT (Prevention of Renal and Vascular End-stage Disease Intervention Trial). *Circ Cardiovasc Genet,* 10.

KOMOROWSKY, C. V., BROSIUS, F. C., 3RD, PENNATHUR, S. & KRETZLER, M. 2012. Perspectives on systems biology applications in diabetic kidney disease. *J Cardiovasc Transl Res,* 5**,** 491-508.

KOPPLE, J. D. 2007. Phenylalanine and tyrosine metabolism in chronic kidney failure. *J Nutr,* 137**,** 1586S-1590S; discussion 1597S-1598S.

KOPPLE, J. D., BLUMENKRANTZ, M. J., JONES, M. R., MORAN, J. K. & COBURN, J. W. 1982. Plasma amino acid levels and amino acid losses during continuous ambulatory peritoneal dialysis. *Am J Clin Nutr,* 36**,** 395-402.

KOTTGEN, A., HWANG, S. J., LARSON, M. G., VAN EYK, J. E., FU, Q., BENJAMIN, E. J., DEHGHAN, A., GLAZER, N. L., KAO, W. H., HARRIS, T. B., GUDNASON, V., SHLIPAK, M. G., YANG, Q., CORESH, J., LEVY, D. & FOX, C. S. 2010. Uromodulin levels associate with a common UMOD variant and risk for incident CKD. *J Am Soc Nephrol,* 21**,** 337-44.

KOYE, D. N., MAGLIANO, D. J., NELSON, R. G. & PAVKOV, M. E. 2018. The Global Epidemiology of Diabetes and Kidney Disease. *Adv Chronic Kidney Dis,* 25**,** 121-132.

KRUGLYAK, L. 2008. The road to genome-wide association studies. *Nat Rev Genet,* 9**,** 314-8.

KWAN, B., FUHRER, T., ZHANG, J., DARSHI, M., VAN ESPEN, B., MONTEMAYOR, D., DE BOER, I. H., DOBRE, M., HSU, C. Y., KELLY, T. N., RAJ, D. S., RAO, P. S., SARAF, S. L., SCIALLA, J., WAIKAR, S. S., SHARMA, K., NATARAJAN, L. & INVESTIGATORS, C. S. 2020. Metabolomic Markers of Kidney Function Decline in Patients With Diabetes: Evidence From the Chronic Renal Insufficiency Cohort (CRIC) Study. *Am J Kidney Dis,* 76**,** 511-520.

KWON, S., HYEON, J. S., JUNG, Y., LI, L., AN, J. N., KIM, Y. C., YANG, S. H., KIM, T., KIM, D. K., LIM, C. S., HWANG, G. S. & LEE, J. P. 2023. Urine myo-inositol as a novel prognostic biomarker for diabetic kidney disease: a targeted metabolomics study using nuclear magnetic resonance. *Kidney Research and Clinical Practice,* 42(4)**,** 445-459.

LACQUANITI, A., BOLIGNANO, D., DONATO, V., BONO, C., FAZIO, M. R. & BUEMI, M. 2010. Alterations of lipid metabolism in chronic nephropathies: mechanisms, diagnosis and treatment. *Kidney Blood Press Res,* 33**,** 100-10.

LAIDLAW, S. A., BERG, R. L., KOPPLE, J. D., NAITO, H., WALKER, W. G. & WALSER, M. 1994. Patterns of fasting plasma amino acid levels in chronic renal insufficiency: results from the feasibility phase of the Modification of Diet in Renal Disease Study. *Am J Kidney Dis,* 23**,** 504-13.

LAMB, E. J., BRETTELL, E. A., COCKWELL, P., DALTON, N., DEEKS, J. J., HARRIS, K., HIGGINS, T., KALRA, P. A., KHUNTI, K., LOUD, F., OTTRIDGE, R. S., SHARPE, C. C., SITCH, A. J., STEVENS, P. E., SUTTON, A. J., TAAL, M. W. & E, G. F. R. C. S. G. 2014. The eGFR-C study: accuracy of glomerular filtration rate (GFR) estimation using creatinine and cystatin C and albuminuria for monitoring disease progression in patients with stage 3 chronic kidney disease--prospective longitudinal study in a multiethnic population. *BMC Nephrol,* 15**,** 13.

LANKTREE, M. B., THERIAULT, S., WALSH, M. & PARE, G. 2018. HDL Cholesterol, LDL Cholesterol, and Triglycerides as Risk Factors for CKD: A Mendelian Randomization Study. *Am J Kidney Dis,* 71**,** 166-172.

LARANJINHA, I., MATIAS, P., MATEUS, S., AGUIAR, F., PEREIRA, P., PERNETA SANTOS, M., COSTA, R., LOURENCO, A., GUIA, J., BARATA, J. D. & CAMPOS, L. 2016. Diabetic kidney disease: Is there a non-albuminuric phenotype in type 2 diabetic patients? *Nefrologia,* 36**,** 503-509.

LEE, C. C., WATKINS, S. M., LORENZO, C., WAGENKNECHT, L. E., IL'YASOVA, D., CHEN, Y. D., HAFFNER, S. M. & HANLEY, A. J. 2016a. Branched-Chain Amino Acids and Insulin Metabolism: The Insulin Resistance Atherosclerosis Study (IRAS). *Diabetes Care,* 39**,** 582-8.

LEE, H., JANG, H. B., YOO, M. G., PARK, S. I. & LEE, H. J. 2020. Amino Acid Metabolites Associated with Chronic Kidney Disease: An Eight-Year Follow-Up Korean Epidemiology Study. *Biomedicines,* 8.

LEE, J., CHOI, J.-Y., KWON, Y.-K., LEE, D., JUNG, H.-Y., RYU, H.-M., CHO, J.-H., RYU, D. H., KIM, Y.-L. & HWANG, G.-S. 2016b. Changes in serum metabolites with the stage of chronic kidney disease: Comparison of diabetes and non-diabetes. *Clinica chimica acta; international journal of clinical chemistry,* 459**,** 123-131.

LEE, S., HAN, M., MOON, S., KIM, K., AN, W. J., RYU, H., OH, K. H. & PARK, S. K. 2022. Identifying Genetic Variants and Metabolites Associated with Rapid Estimated Glomerular Filtration Rate Decline in Korea Based on Genome-Metabolomic Integrative Analysis. *Metabolites,* 12(11) (no pagination).

LEES, J. S., MCQUARRIE, E. P., MORDI, N., GEDDES, C. C., FOX, J. G. & MACKINNON, B. 2017. Risk factors for bleeding complications after nephrologist-performed native renal biopsy. *Clin Kidney J,* 10**,** 573-577.

LEIHERER, A., MUENDLEIN, A., SAELY, C. H., BRANDTNER, E. M., GEIGER, K., FRAUNBERGER, P. & DREXEL, H. 2018. The value of uromodulin as a new serum marker to predict decline in renal function. *J Hypertens,* 36**,** 110-118.

LEVEY, A. S., CORESH, J., BALK, E., KAUSZ, A. T., LEVIN, A., STEFFES, M. W., HOGG, R. J., PERRONE, R. D., LAU, J., EKNOYAN, G. & NATIONAL KIDNEY, F. 2003. National Kidney Foundation practice guidelines for chronic kidney disease: evaluation, classification, and stratification. *Ann Intern Med,* 139**,** 137-47.

LEVEY, A. S., CORESH, J., BOLTON, K., CULLETON, B., HARVEY, K. S., IKIZLER, T. A., JOHNSON, C. A., KAUSZ, A., KIMMEL, P. L. & KUSEK, J. 2002. K/DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification. *American Journal of Kidney Diseases,* 39**,** i-ii+ S1-S266.

LEVEY, A. S., ECKARDT, K. U., TSUKAMOTO, Y., LEVIN, A., CORESH, J., ROSSERT, J., DE ZEEUW, D., HOSTETTER, T. H., LAMEIRE, N. & EKNOYAN, G. 2005. Definition and classification of chronic kidney disease: a position statement from Kidney Disease: Improving Global Outcomes (KDIGO). *Kidney Int,* 67**,** 2089-100.

LEVEY, A. S., SCHOOLWERTH, A. C., BURROWS, N. R., WILLIAMS, D. E., STITH, K. R., MCCLELLAN, W., CENTERS FOR DISEASE, C. & PREVENTION EXPERT, P. 2009a. Comprehensive public health strategies for preventing the development, progression, and complications of CKD: report of an expert panel convened by the Centers for Disease Control and Prevention. *Am J Kidney Dis,* 53**,** 522-35.

LEVEY, A. S. & STEVENS, L. A. 2010. Estimating GFR using the CKD Epidemiology Collaboration (CKD-EPI) creatinine equation: more accurate GFR estimates, lower CKD prevalence estimates, and better risk predictions. *Am J Kidney Dis,* 55**,** 622-7.

LEVEY, A. S., STEVENS, L. A. & CORESH, J. 2009b. Conceptual model of CKD: applications and implications. *Am J Kidney Dis,* 53**,** S4-16.

LEVEY, A. S., STEVENS, L. A., SCHMID, C. H., ZHANG, Y. L., CASTRO, A. F., 3RD, FELDMAN, H. I., KUSEK, J. W., EGGERS, P., VAN LENTE, F., GREENE, T., CORESH, J. & CKD, E. P. I. 2009c. A new equation to estimate glomerular filtration rate. *Ann Intern Med,* 150**,** 604-12.

LEVIN, A. & STEVENS, P. E. 2014. Summary of KDIGO 2012 CKD Guideline: behind the scenes, need for guidance, and a framework for moving forward. *Kidney Int,* 85**,** 49-61.

LEVIN, A., STEVENS, P. E., BILOUS, R. W., CORESH, J., DE FRANCISCO, A. L., DE JONG, P. E., GRIFFITH, K. E., HEMMELGARN, B. R., ISEKI, K. & LAMB, E. J. 2013. Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney international supplements,* 3**,** 1-150.

LI, J., CAO, Y. F., SUN, X. Y., HAN, L., LI, S. N., GU, W. Q., SONG, M., JIANG, C. T., YANG, X. & FANG, Z. Z. 2019. Plasma tyrosine and its interaction with low high-density lipoprotein cholesterol and the risk of type 2 diabetes mellitus in Chinese. *J Diabetes Investig,* 10**,** 491-498.

LIN, C. H., CHANG, Y. C. & CHUANG, L. M. 2016. Early detection of diabetic kidney disease: Present limitations and future perspectives. *World J Diabetes,* 7**,** 290-301.

LIU, C., DEBNATH, N., MOSOYAN, G., CHAUHAN, K., VASQUEZ-RIOS, G., SOUDANT, C., MENEZ, S., PARIKH, C. R. & COCA, S. G. 2022. Systematic Review and Meta-Analysis of Plasma and Urine Biomarkers for CKD Outcomes. *J Am Soc Nephrol,* 33**,** 1657-1672.

LOOK, A. R. G. 2014. Effect of a long-term behavioural weight loss intervention on nephropathy in overweight or obese adults with type 2 diabetes: a secondary analysis of the Look AHEAD randomised clinical trial. *Lancet Diabetes Endocrinol,* 2**,** 801-9.

LOOKER, H. C., COLOMBO, M., HESS, S., BROSNAN, M. J., FARRAN, B., DALTON, R. N., WONG, M. C., TURNER, C., PALMER, C. N., NOGOCEKE, E., GROOP, L., SALOMAA, V., DUNGER, D. B., AGAKOV, F., MCKEIGUE, P. M., COLHOUN, H. M. & INVESTIGATORS, S. 2015. Biomarkers of rapid chronic kidney disease progression in type 2 diabetes. *Kidney Int,* 88**,** 888-96.

LOTTA, L. A., SCOTT, R. A., SHARP, S. J., BURGESS, S., LUAN, J., TILLIN, T., SCHMIDT, A. F., IMAMURA, F., STEWART, I. D., PERRY, J. R., MARNEY, L., KOULMAN, A., KAROLY, E. D., FOROUHI, N. G., SJOGREN, R. J., NASLUND, E., ZIERATH, J. R., KROOK, A., SAVAGE, D. B., GRIFFIN, J. L., CHATURVEDI, N., HINGORANI, A. D., KHAW, K. T., BARROSO, I., MCCARTHY, M. I., O'RAHILLY, S., WAREHAM, N. J. & LANGENBERG, C. 2016. Genetic Predisposition to an Impaired Metabolism of the Branched-Chain Amino Acids and Risk of Type 2 Diabetes: A Mendelian Randomisation Analysis. *PLoS Med,* 13**,** e1002179.

LOW, S., ZHANG, X., WANG, J., YEOH, L. Y., LIU, Y. L., ANG, K. K. L., TANG, W. E., KWAN, P. Y., TAVINTHARAN, S., SUM, C. F. & LIM, S. C. 2018. Long-term prospective observation suggests

that glomerular hyperfiltration is associated with rapid decline in renal filtration function: A multiethnic study. *Diab Vasc Dis Res,* 15**,** 417-423.

LU, J. L., MOLNAR, M. Z., NASEER, A., MIKKELSEN, M. K., KALANTAR-ZADEH, K. & KOVESDY, C. P. 2015. Association of age and BMI with kidney function and mortality: a cohort study. *Lancet Diabetes Endocrinol,* 3**,** 704-14.

LUNYERA, J., DIAMANTIDIS, C. J., BOSWORTH, H. B., PATEL, U. D., BAIN, J., MUEHLBAUER, M. J., ILKAYEVA, O., NGUYEN, M., SHARMA, B., MA, J. Z., SHAH, S. H. & SCIALLA, J. J. 2021. Urine tricarboxylic acid cycle signatures of early-stage diabetic kidney disease. *Metabolomics,* 18**,** 5.

LYNCH, C. J. & ADAMS, S. H. 2014. Branched-chain amino acids in metabolic signalling and insulin resistance. *Nat Rev Endocrinol,* 10**,** 723-36.

MAAHS, D. M., BUSHMAN, L., KERR, B., ELLIS, S. L., PYLE, L., MCFANN, K., BOUFFARD, A., BISHOP, F. K., NGUYEN, N. & ANDERSON, P. L. 2014. A practical method to measure GFR in people with type 1 diabetes. *J Diabetes Complications,* 28**,** 667-73.

MACISAAC, R. J., EKINCI, E. I., PREMARATNE, E., LU, Z. X., SEAH, J. M., LI, Y., BOSTON, R., WARD, G. M. & JERUMS, G. 2015. The Chronic Kidney Disease-Epidemiology Collaboration (CKD-EPI) equation does not improve the underestimation of Glomerular Filtration Rate (GFR) in people with diabetes and preserved renal function. *BMC Nephrol,* 16**,** 198.

MACISAAC, R. J., TSALAMANDRIS, C., THOMAS, M. C., PREMARATNE, E., PANAGIOTOPOULOS, S., SMITH, T. J., POON, A., JENKINS, M. A., RATNAIKE, S. I., POWER, D. A. & JERUMS, G. 2006. Estimating glomerular filtration rate in diabetes: a comparison of cystatin-C- and creatinine-based methods. *Diabetologia,* 49**,** 1686-9.

MAGEE, G. M., BILOUS, R. W., CARDWELL, C. R., HUNTER, S. J., KEE, F. & FOGARTY, D. G. 2009. Is hyperfiltration associated with the future risk of developing diabetic nephropathy? A meta-analysis. *Diabetologia,* 52**,** 691-7.

MAHLEY, R. W., INNERARITY, T. L., RALL, S. C., JR. & WEISGRABER, K. H. 1984. Plasma lipoproteins: apolipoprotein structure and function. *J Lipid Res,* 25**,** 1277-94.

MAKINEN, V. P., KANGAS, A. J., SOININEN, P., WURTZ, P., GROOP, P. H. & ALA-KORPELA, M. 2013. Metabolic phenotyping of diabetic nephropathy. *Clin Pharmacol Ther,* 94**,** 566-9.

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,* 36(10)**,** 1859-1866.

MARIONI, R. E., STRACHAN, M. W., REYNOLDS, R. M., LOWE, G. D., MITCHELL, R. J., FOWKES, F. G., FRIER, B. M., LEE, A. J., BUTCHER, I., RUMLEY, A., MURRAY, G. D., DEARY, I. J. & PRICE, J. F. 2010. Association between raised inflammatory markers and cognitive decline in elderly people with type 2 diabetes: the Edinburgh Type 2 Diabetes Study. *Diabetes,* 59**,** 710-3.

MARSCHE, G., HEINE, G. H., STADLER, J. T. & HOLZER, M. 2020. Current Understanding of the Relationship of HDL Composition, Structure and Function to Their Cardioprotective Properties in Chronic Kidney Disease. *Biomolecules,* 10.

MATHER, A. & POLLOCK, C. 2011. Glucose handling by the kidney. *Kidney Int Suppl***,** S1-6.

MATSUSHITA, K., BALLEW, S. H., WANG, A. Y., KALYESUBULA, R., SCHAEFFNER, E. & AGARWAL, R. 2022. Epidemiology and risk of cardiovascular disease in populations with chronic kidney disease. *Nat Rev Nephrol,* 18**,** 696-707.

MAYER, G., HEERSPINK, H. J., ASCHAUER, C., HEINZEL, A., HEINZE, G., KAINZ, A., SUNZENAUER, J., PERCO, P., DE ZEEUW, D., ROSSING, P., PENA, M., OBERBAUER, R. & CONSORTIUM, S. 2017. Systems Biology-Derived Biomarkers to Predict Progression of Renal Function Decline in Type 2 Diabetes. *Diabetes Care,* 40**,** 391-397.

MCCLELLAN, W. M. & FLANDERS, W. D. 2003. Risk factors for progressive chronic kidney disease. *J Am Soc Nephrol,* 14**,** S65-70.

MCCORMACK, S. E., SHAHAM, O., MCCARTHY, M. A., DEIK, A. A., WANG, T. J., GERSZTEN, R. E., CLISH, C. B., MOOTHA, V. K., GRINSPOON, S. K. & FLEISCHMAN, A. 2013. Circulating branched-chain

amino acid concentrations are associated with obesity and future insulin resistance in children and adolescents. *Pediatr Obes,* 8**,** 52-61.

MCCULLOUGH, K. P., MORGENSTERN, H., SARAN, R., HERMAN, W. H. & ROBINSON, B. M. 2019. Projecting ESRD incidence and prevalence in the United States through 2030. *Journal of the American Society of Nephrology: JASN,* 30**,** 127.

MCCULLOUGH, P. A., LI, S., JURKOVITZ, C. T., STEVENS, L., COLLINS, A. J., CHEN, S. C., NORRIS, K. C., MCFARLANE, S., JOHNSON, B., SHLIPAK, M. G., OBIALO, C. I., BROWN, W. W., VASSALOTTI, J., WHALEY-CONNELL, A. T., BRENNER, R. M., BAKRIS, G. L. & INVESTIGATORS, K. 2008. Chronic kidney disease, prevalence of premature cardiovascular disease, and relationship to short-term mortality. *Am Heart J,* 156**,** 277-83.

MCMAHON, G. M., HWANG, S.-J., CLISH, C. B., TIN, A., YANG, Q., LARSON, M. G., RHEE, E. P., LI, M., CONSORTIUM, C. K., LEVY, D., O'DONNELL, C. J., CORESH, J., YOUNG, J. H., GERSZTEN, R. E. & FOX, C. S. 2017a. Urinary metabolites along with common and rare genetic variations are associated with incident chronic kidney disease. *Kidney international,* 91**,** 1426-1435.

MCMAHON, G. M., HWANG, S. J., CLISH, C. B., TIN, A., YANG, Q., LARSON, M. G., RHEE, E. P., LI, M., LEVY, D., O'DONNELL, C. J., CORESH, J., YOUNG, J. H., GERSZTEN, R. E. & FOX, C. S. 2017b. Urinary metabolites along with common and rare genetic variations are associated with incident chronic kidney disease. *Kidney International,* 91(6)**,** 1426-1435.

MELSOM, T., FUSKEVAG, O. M., MATHISEN, U. D., STRAND, H., SCHEI, J., JENSSEN, T., SOLBU, M. & ERIKSEN, B. O. 2015. Estimated GFR is biased by non-traditional cardiovascular risk factors. *Am J Nephrol,* 41**,** 7-15.

MENDE, C. & KATZ, A. 2016. Cystatin C- and Creatinine-Based Estimates of Glomerular Filtration Rate in Dapagliflozin Phase 3 Clinical Trials. *Diabetes Ther,* 7**,** 139-51.

MENON, V., GUL, A. & SARNAK, M. J. 2005. Cardiovascular risk factors in chronic kidney disease. *Kidney Int,* 68**,** 1413-8.

MESQUITA, J., VARELA, A. & MEDINA, J. L. 2010. Dyslipidemia in renal disease: causes, consequences and treatment. *Endocrinol Nutr,* 57**,** 440-8.

MIKA, A., WOJTOWICZ, W., ZABEK, A., MLYNARZ, P., CHMIELEWSKI, M., SLEDZINSKI, T. & STEPNOWSKI, P. 2018. Application of nuclear magnetic resonance spectroscopy for the detection of metabolic disorders in patients with moderate kidney insufficiency. *Journal of Pharmaceutical and Biomedical Analysis,* 149**,** 1-8.

MIKOLASEVIC, I., ZUTELIJA, M., MAVRINAC, V. & ORLIC, L. 2017. Dyslipidemia in patients with chronic kidney disease: etiology and management. *Int J Nephrol Renovasc Dis,* 10**,** 35-45.

MISSAILIDIS, C., HALLQVIST, J., QURESHI, A. R., BARANY, P., HEIMBURGER, O., LINDHOLM, B., STENVINKEL, P. & BERGMAN, P. 2016. Serum Trimethylamine-N-Oxide Is Strongly Related to Renal Function and Predicts Outcome in Chronic Kidney Disease. *PLoS One,* 11**,** e0141738.

MOGENSEN, C. E., CHRISTENSEN, C. K. & VITTINGHUS, E. 1983. The stages in diabetic renal disease. With emphasis on the stage of incipient diabetic nephropathy. *Diabetes,* 32 Suppl 2**,** 64-78.

MOLITCH, M. E., GAO, X., BEBU, I., DE BOER, I. H., LACHIN, J., PATERSON, A., PERKINS, B., SAENGER, A. K., STEFFES, M., ZINMAN, B., DIABETES, C., COMPLICATIONS TRIAL/EPIDEMIOLOGY OF DIABETES, I. & COMPLICATIONS RESEARCH, G. 2019. Early Glomerular Hyperfiltration and Long-Term Kidney Outcomes in Type 1 Diabetes: The DCCT/EDIC Experience. *Clin J Am Soc Nephrol,* 14**,** 854-861.

MOONS, K. G., ALTMAN, D. G., REITSMA, J. B., IOANNIDIS, J. P., MACASKILL, P., STEYERBERG, E. W., VICKERS, A. J., RANSOHOFF, D. F. & COLLINS, G. S. 2015. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med,* 162**,** W1-73.

MOONS, K. G., KENGNE, A. P., GROBBEE, D. E., ROYSTON, P., VERGOUWE, Y., ALTMAN, D. G. & WOODWARD, M. 2012a. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart,* 98**,** 691-8.

MOONS, K. G., KENGNE, A. P., WOODWARD, M., ROYSTON, P., VERGOUWE, Y., ALTMAN, D. G. & GROBBEE, D. E. 2012b. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart,* 98**,** 683-90.

MORADI, H., GANJI, S., KAMANNA, V., PAHL, M. V. & VAZIRI, N. D. 2010. Increased monocyte adhesion-promoting capacity of plasma in end-stage renal disease - response to antioxidant therapy. *Clin Nephrol,* 74**,** 273-81.

MORADI, H., PAHL, M. V., ELAHIMEHR, R. & VAZIRI, N. D. 2009. Impaired antioxidant activity of high-density lipoprotein in chronic kidney disease. *Transl Res,* 153**,** 77-85.

MPONDO, B. C., NEILSON, E. & ERNEST, A. 2016. Prevalence of chronic kidney disease in diabetic adult out-patients in Tanzania. *BMC Nephrol,* 17**,** 71.

MUNTNER, P., HAMM, L. L., KUSEK, J. W., CHEN, J., WHELTON, P. K. & HE, J. 2004. The prevalence of nontraditional risk factors for coronary heart disease in patients with chronic kidney disease. *Ann Intern Med,* 140**,** 9-17.

MUTTER, S., VALO, E., AITTOMAKI, V., NYBO, K., RAIVONEN, L., THORN, L. M., FORSBLOM, C., SANDHOLM, N., WURTZ, P. & GROOP, P. H. 2022. Urinary metabolite profiling and risk of progression of diabetic nephropathy in 2670 individuals with type 1 diabetes. *Diabetologia,* 65(1)**,** 140-149.

NAGATA, C., NAKAMURA, K., WADA, K., TSUJI, M., TAMAI, Y. & KAWACHI, T. 2013. Branched-chain amino acid intake and the risk of diabetes in a Japanese community: the Takayama study. *Am J Epidemiol,* 178**,** 1226-32.

NARVA, A. 2018. Population Health for CKD and Diabetes: Lessons From the Indian Health Service. *Am J Kidney Dis,* 71**,** 407-411.

NATIONAL HEART, L. & INSTITUTE, B. 2014. Quality assessment tool for observational cohort and cross-sectional studies.

NATIONAL KIDNEY FOUNDATION, K. 2012. KDOQI Clinical Practice Guideline for Diabetes and CKD: 2012 Update. *Am J Kidney Dis,* 60**,** 850-86.

NATIONALGUIDELINECENTRE. 2017. *RRT and conservative management. Economic analysis report: Cost-effectiveness analysis of HDF versus high flux HD* [Online]. National Institute of Health and Care Excellence. Available: https://www.nice.org.uk/guidance/ng107/documents/supporting-documentation-2 [Accessed 16 October 2023].

NAYOR, M., SHAH, R. V., MILLER, P. E., BLODGETT, J. B., TANGUAY, M., PICO, A. R., MURTHY, V. L., MALHOTRA, R., HOUSTIS, N. E., DEIK, A., PIERCE, K. A., BULLOCK, K., DAILEY, L., VELAGALETI, R. S., MOORE, S. A., HO, J. E., BAGGISH, A. L., CLISH, C. B., LARSON, M. G., VASAN, R. S. & LEWIS, G. D. 2020. Metabolic Architecture of Acute Exercise Response in Middle-Aged Adults in the Community. *Circulation,* 142**,** 1905-1924.

NEIS, E. P., SABRKHANY, S., HUNDSCHEID, I., SCHELLEKENS, D., LENAERTS, K., OLDE DAMINK, S. W., BLAAK, E. E., DEJONG, C. H. & RENSEN, S. S. 2017. Human splanchnic amino-acid metabolism. *Amino Acids,* 49**,** 161-172.

NELSON, R. G., GRAMS, M. E., BALLEW, S. H., SANG, Y., AZIZI, F., CHADBAN, S. J., CHAKER, L., DUNNING, S. C., FOX, C., HIRAKAWA, Y., ISEKI, K., IX, J., JAFAR, T. H., KOTTGEN, A., NAIMARK, D. M. J., OHKUBO, T., PRESCOTT, G. J., REBHOLZ, C. M., SABANAYAGAM, C., SAIRENCHI, T., SCHOTTKER, B., SHIBAGAKI, Y., TONELLI, M., ZHANG, L., GANSEVOORT, R. T., MATSUSHITA, K., WOODWARD, M., CORESH, J., SHALEV, V. & CONSORTIUM, C. K. D. P. 2019. Development of Risk Prediction Equations for Incident Chronic Kidney Disease. *JAMA*.

NELSON, R. G., KNOWLER, W. C., MCCANCE, D. R., SIEVERS, M. L., PETTITT, D. J., CHARLES, M. A., HANSON, R. L., LIU, Q. Z. & BENNETT, P. H. 1993. Determinants of end-stage renal disease in Pima Indians with type 2 (non-insulin-dependent) diabetes mellitus and proteinuria. *Diabetologia,* 36**,** 1087-93.

NELSON, R. G., MORGENSTERN, H. & BENNETT, P. H. 1998. Birth weight and renal disease in Pima Indians with type 2 diabetes mellitus. *Am J Epidemiol,* 148**,** 650-6.

NELSON, R. G., PETTITT, D. J., CARRAHER, M. J., BAIRD, H. R. & KNOWLER, W. C. 1988. Effect of proteinuria on mortality in NIDDM. *Diabetes,* 37**,** 1499-504.

NHS 2021. NHS reference costs 2018/2019. *In:* HEALTH, D. O. (ed.).

NICHOLSON, J. K. & LINDON, J. C. 2008. Metabonomics. *Nature,* 455**,** 1054-1056.

NIEWCZAS, M. A., MATHEW, A. V., CROALL, S., BYUN, J., MAJOR, M., SABISETTI, V. S., SMILES, A., BONVENTRE, J. V., PENNATHUR, S. & KROLEWSKI, A. S. 2017. Circulating modified metabolites and a risk of ESRD in patients with type 1 diabetes and chronic kidney disease. *Diabetes Care,* 40(3)**,** 383-390.

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. 2014. Uremic solutes and risk of end-stage renal disease in type 2 diabetes: metabolomic study. *Kidney international,* 85**,** 1214-24.

NKUIPOU-KENFACK, E., DURANTON, F., GAYRARD, N., ARGILES, A., LUNDIN, U., WEINBERGER, K. M., DAKNA, M., DELLES, C., MULLEN, W., HUSI, H., KLEIN, J., KOECK, T., ZURBIG, P. & MISCHAK, H. 2014. Assessment of metabolomic and proteomic biomarkers in detection and prognosis of progression of renal function in chronic kidney disease. *PLoS ONE,* 9**,** e96955.

O'SEAGHDHA, C. M., LYASS, A., MASSARO, J. M., MEIGS, J. B., CORESH, J., D'AGOSTINO, R. B., SR., ASTOR, B. C. & FOX, C. S. 2012. A risk score for chronic kidney disease in the general population. *Am J Med,* 125**,** 270-7.

OH, S. J., LEE, J. I., HA, W. C., JEONG, S. H., YIM, H. W., SON, H. S. & SOHN, T. S. 2012. Comparison of cystatin C- and creatinine-based estimation of glomerular filtration rate according to glycaemic status in Type 2 diabetes. *Diabet Med,* 29**,** e121-5.

OLIVER, S. G., WINSON, M. K., KELL, D. B. & BAGANZ, F. 1998. Systematic functional analysis of the yeast genome. *Trends Biotechnol,* 16**,** 373-8.

OWENS, E., TAN, K. S., ELLIS, R., DEL VECCHIO, S., HUMPHRIES, T., LENNAN, E., VESEY, D., HEALY, H., HOY, W. & GOBE, G. 2020. Development of a Biomarker Panel to Distinguish Risk of Progressive Chronic Kidney Disease. *Biomedicines,* 8.

PACKER, M., ANKER, S. D., BUTLER, J., FILIPPATOS, G., POCOCK, S. J., CARSON, P., JANUZZI, J., VERMA, S., TSUTSUI, H., BRUECKMANN, M., JAMAL, W., KIMURA, K., SCHNEE, J., ZELLER, C., COTTON, D., BOCCHI, E., BOHM, M., CHOI, D. J., CHOPRA, V., CHUQUIURE, E., GIANNETTI, N., JANSSENS, S., ZHANG, J., GONZALEZ JUANATEY, J. R., KAUL, S., BRUNNER-LA ROCCA, H. P., MERKELY, B., NICHOLLS, S. J., PERRONE, S., PINA, I., PONIKOWSKI, P., SATTAR, N., SENNI, M., SERONDE, M. F., SPINAR, J., SQUIRE, I., TADDEI, S., WANNER, C., ZANNAD, F. & INVESTIGATORS, E. M.-R. T. 2020. Cardiovascular and Renal Outcomes with Empagliflozin in Heart Failure. *N Engl J Med,* 383**,** 1413-1424.

PARSA, A., KAO, W. H., XIE, D., ASTOR, B. C., LI, M., HSU, C. Y., FELDMAN, H. I., PAREKH, R. S., KUSEK, J. W., GREENE, T. H., FINK, J. C., ANDERSON, A. H., CHOI, M. J., WRIGHT, J. T., JR., LASH, J. P., FREEDMAN, B. I., OJO, A., WINKLER, C. A., RAJ, D. S., KOPP, J. B., HE, J., JENSVOLD, N. G., TAO, K., LIPKOWITZ, M. S., APPEL, L. J., INVESTIGATORS, A. S. & INVESTIGATORS, C. S. 2013. APOL1 risk variants, race, and progression of chronic kidney disease. *N Engl J Med,* 369**,** 2183-96.

PARVING, H. H., LEHNERT, H., BROCHNER-MORTENSEN, J., GOMIS, R., ANDERSEN, S., ARNER, P., IRBESARTAN IN PATIENTS WITH TYPE, D. & MICROALBUMINURIA STUDY, G. 2001. The effect of irbesartan on the development of diabetic nephropathy in patients with type 2 diabetes. *N Engl J Med,* 345**,** 870-8.

PARVING, H. H., LEWIS, J. B., RAVID, M., REMUZZI, G., HUNSICKER, L. G. & INVESTIGATORS, D. 2006. Prevalence and risk factors for microalbuminuria in a referred cohort of type II diabetic patients: a global perspective. *Kidney Int,* 69**,** 2057-63.

PAVKOV, M. E., KNOWLER, W. C., BENNETT, P. H., LOOKER, H. C., KRAKOFF, J. & NELSON, R. G. 2006. Increasing incidence of proteinuria and declining incidence of end-stage renal disease in diabetic Pima Indians. *Kidney Int,* 70**,** 1840-6.

PAVKOV, M. E., MASON, C. C., BENNETT, P. H., CURTIS, J. M., KNOWLER, W. C. & NELSON, R. G. 2009. Change in the distribution of albuminuria according to estimated glomerular filtration rate in Pima Indians with type 2 diabetes. *Diabetes Care,* 32**,** 1845-50.

PENA, M. J., HEINZEL, A., HEINZE, G., ALKHALAF, A., BAKKER, S. J., NGUYEN, T. Q., GOLDSCHMEDING, R., BILO, H. J., PERCO, P., MAYER, B., DE ZEEUW, D. & LAMBERS HEERSPINK, H. J. 2015. A panel of novel biomarkers representing different disease pathways improves prediction of renal function decline in type 2 diabetes. *PLoS One,* 10**,** e0120995.

PENA, M. J., LAMBERS HEERSPINK, H. J., HELLEMONS, M. E., FRIEDRICH, T., DALLMANN, G., LAJER, M., BAKKER, S. J. L., GANSEVOORT, R. T., ROSSING, P., DE ZEEUW, D. & ROSCIONI, S. S. 2014. Urine and plasma metabolites predict the development of diabetic nephropathy in individuals with Type 2 diabetes mellitus. *Diabetic Medicine,* 31(9)**,** 1138-1147.

PENCINA, M. J., D'AGOSTINO, R. B., SR., D'AGOSTINO, R. B., JR. & VASAN, R. S. 2008. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med,* 27**,** 157-72; discussion 207-12.

PENCINA, M. J., D'AGOSTINO, R. B., SR. & STEYERBERG, E. W. 2011. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med,* 30**,** 11-21.

PENCINA, M. J., FINE, J. P. & D'AGOSTINO, R. B., SR. 2017. Discrimination slope and integrated discrimination improvement - properties, relationships and impact of calibration. *Stat Med,* 36**,** 4482-4490.

PENNO, G., SOLINI, A., ZOPPINI, G., FONDELLI, C., TREVISAN, R., VEDOVATO, M., GRUDEN, G., LAMACCHIA, O., PONTIROLI, A. E., AROSIO, M., ORSI, E., PUGLIESE, G., RENAL, I. & CARDIOVASCULAR EVENTS STUDY, G. 2015. 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,* 10**,** e0125512.

PENNO, G., SOLINI, A., ZOPPINI, G., ORSI, E., ZERBINI, G., TREVISAN, R., GRUDEN, G., CAVALOT, F., LAVIOLA, L., MORANO, S., NICOLUCCI, A., PUGLIESE, G., RENAL, I. & CARDIOVASCULAR EVENTS STUDY, G. 2012. Rate and determinants of association between advanced retinopathy and chronic kidney disease in patients with type 2 diabetes: the Renal Insufficiency And Cardiovascular Events (RIACE) Italian multicenter study. *Diabetes Care,* 35**,** 2317-23.

PETTITT, D. J., SAAD, M. F., BENNETT, P. H., NELSON, R. G. & KNOWLER, W. C. 1990. Familial predisposition to renal disease in two generations of Pima Indians with type 2 (non-insulin-dependent) diabetes mellitus. *Diabetologia,* 33**,** 438-43.

PIERONI, L., DELANAYE, P., BOUTTEN, A., BARGNOUX, A. S., ROZET, E., DELATOUR, V., CARLIER, M. C., HANSER, A. M., CAVALIER, E., FROISSART, M., CRISTOL, J. P. & SOCIETE FRANCAISE DE BIOLOGIE, C. 2011. A multicentric evaluation of IDMS-traceable creatinine enzymatic assays. *Clin Chim Acta,* 412**,** 2070-5.

POLLOCK, C., JAMES, G., GARCIA SANCHEZ, J. J., CARRERO, J. J., ARNOLD, M., LAM, C. S. P., CHEN, H. T., NOLAN, S., PECOITS-FILHO, R. & WHEELER, D. C. 2022. Healthcare resource utilisation and related costs of patients with CKD from the UK: a report from the DISCOVER CKD retrospective cohort. *Clin Kidney J,* 15**,** 2124-2134.

PORRINI, E., RUGGENENTI, P., LUIS-LIMA, S., CARRARA, F., JIMENEZ, A., DE VRIES, A. P. J., TORRES, A., GASPARI, F. & REMUZZI, G. 2019. Estimated GFR: time for a critical appraisal. *Nat Rev Nephrol,* 15**,** 177-190.

POTTEL, H., HOSTE, L., DUBOURG, L., EBERT, N., SCHAEFFNER, E., ERIKSEN, B. O., MELSOM, T., LAMB, E. J., RULE, A. D., TURNER, S. T., GLASSOCK, R. J., DE SOUZA, V., SELISTRE, L., MARIAT, C., MARTENS, F. & DELANAYE, P. 2016. An estimated glomerular filtration rate equation for the full age spectrum. *Nephrol Dial Transplant,* 31**,** 798-806.

PREISS, D. J., GODBER, I. M., LAMB, E. J., DALTON, R. N. & GUNN, I. R. 2007. The influence of a cooked-meat meal on estimated glomerular filtration rate. *Ann Clin Biochem,* 44**,** 35-42.

PRICE, A. H., WEIR, C. J., WELSH, P., MCLACHLAN, S., STRACHAN, M. W. J., SATTAR, N. & PRICE, J. F. 2017. Comparison of non-traditional biomarkers, and combinations of biomarkers, for vascular risk prediction in people with type 2 diabetes: The Edinburgh Type 2 Diabetes Study. *Atherosclerosis,* 264**,** 67-73.

PRICE, J. F., REYNOLDS, R. M., MITCHELL, R. J., WILLIAMSON, R. M., FOWKES, F. G., DEARY, I. J., LEE, A. J., FRIER, B. M., HAYES, P. C. & STRACHAN, M. W. 2008. The Edinburgh Type 2 Diabetes Study: study protocol. *BMC Endocr Disord,* 8**,** 18.

PUAVILAI, G., CHANPRASERTYOTIN, S. & SRIPHRAPRADAENG, A. 1999. Diagnostic criteria for diabetes mellitus and other categories of glucose intolerance: 1997 criteria by the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus (ADA), 1998 WHO consultation criteria, and 1985 WHO criteria. World Health Organization. *Diabetes Res Clin Pract,* 44**,** 21-6.

PUGLIESE, G., PENNO, G., NATALI, A., BARUTTA, F., DI PAOLO, S., REBOLDI, G., GESUALDO, L., DE NICOLA, L., ITALIAN DIABETES, S. & THE ITALIAN SOCIETY OF, N. 2020. Diabetic kidney disease: new clinical and therapeutic issues. Joint position statement of the Italian Diabetes Society and the Italian Society of Nephrology on "The natural history of diabetic kidney disease and treatment of hyperglycemia in patients with type 2 diabetes and impaired renal function". *J Nephrol,* 33**,** 9-35.

RADER, D. J., HOEG, J. M. & BREWER, H. B., JR. 1994. Quantitation of plasma apolipoproteins in the primary and secondary prevention of coronary artery disease. *Ann Intern Med,* 120**,** 1012-25.

RAHMAN, M., YANG, W., AKKINA, S., ALPER, A., ANDERSON, A. H., APPEL, L. J., HE, J., RAJ, D. S., SCHELLING, J., STRAUSS, L., TEAL, V., RADER, D. J. & INVESTIGATORS, C. S. 2014. Relation of serum lipids and lipoproteins with progression of CKD: The CRIC study. *Clin J Am Soc Nephrol,* 9**,** 1190-8.

RAMSPEK, C. L., DE JONG, Y., DEKKER, F. W. & VAN DIEPEN, M. 2020. Towards the best kidney failure prediction tool: a systematic review and selection aid. *Nephrol Dial Transplant,* 35**,** 1527-1538.

REILLY, B. M. & EVANS, A. T. 2006. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med,* 144**,** 201-9.

RETNAKARAN, R., CULL, C. A., THORNE, K. I., ADLER, A. I., HOLMAN, R. R. & GROUP, U. S. 2006. Risk factors for renal dysfunction in type 2 diabetes: U.K. Prospective Diabetes Study 74. *Diabetes,* 55**,** 1832-9.

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. 2013. A combined epidemiologic and metabolomic approach improves CKD prediction. *Journal of the American Society of Nephrology : JASN,* 24**,** 1330-8.

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. 2016. Metabolomics of Chronic Kidney Disease Progression: A Case-Control Analysis in the Chronic Renal Insufficiency Cohort Study. *Am J Nephrol,* 43**,** 366-74.

RICCIARDI, C. A. & GNUDI, L. 2021. Kidney disease in diabetes: From mechanisms to clinical presentation and treatment strategies. *Metabolism,* 124**,** 154890.

RILEY, R. D., HAYDEN, J. A., STEYERBERG, E. W., MOONS, K. G., ABRAMS, K., KYZAS, P. A., MALATS, N., BRIGGS, A., SCHROTER, S., ALTMAN, D. G., HEMINGWAY, H. & GROUP, P. 2013. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLoS Med,* 10**,** e1001380.

ROOZENBEEK, B., MAAS, A. I., LINGSMA, H. F., BUTCHER, I., LU, J., MARMAROU, A., MCHUGH, G. S., WEIR, J., MURRAY, G. D., STEYERBERG, E. W. & GROUP, I. S. 2009. Baseline characteristics and statistical power in randomized controlled trials: selection, prognostic targeting, or covariate adjustment? *Crit Care Med,* 37**,** 2683-90.

ROYSTON, P. & ALTMAN, D. G. 2013. External validation of a Cox prognostic model: principles and methods. *BMC Med Res Methodol,* 13**,** 33.

RUGGENENTI, P., PORRINI, E. L., GASPARI, F., MOTTERLINI, N., CANNATA, A., CARRARA, F., CELLA, C., FERRARI, S., STUCCHI, N., PARVANOVA, A., ILIEV, I., DODESINI, A. R., TREVISAN, R., BOSSI, A., ZALETEL, J., REMUZZI, G. & INVESTIGATORS, G. F. R. S. 2012. Glomerular hyperfiltration and renal disease progression in type 2 diabetes. *Diabetes Care,* 35**,** 2061-8.

RUSSO, G. T., DE COSMO, S., VIAZZI, F., MIRIJELLO, A., CERIELLO, A., GUIDA, P., GIORDA, C., CUCINOTTA, D., PONTREMOLI, R., FIORETTO, P. & GROUP, A. M.-A. S. 2018. Diabetic kidney disease in the elderly: prevalence and clinical correlates. *BMC Geriatr,* 18**,** 38.

RUSSO, G. T., DE COSMO, S., VIAZZI, F., PACILLI, A., CERIELLO, A., GENOVESE, S., GUIDA, P., GIORDA, C., CUCINOTTA, D., PONTREMOLI, R., FIORETTO, P. & GROUP, A. M.-A. S. 2016. Plasma Triglycerides and HDL-C Levels Predict the Development of Diabetic Kidney Disease in Subjects With Type 2 Diabetes: The AMD Annals Initiative. *Diabetes Care,* 39**,** 2278-2287.

SACKS, F. M., HERMANS, M. P., FIORETTO, P., VALENSI, P., DAVIS, T., HORTON, E., WANNER, C., AL-RUBEAAN, K., ARONSON, R., BARZON, I., BISHOP, L., BONORA, E., BUNNAG, P., CHUANG, L. M., DEEROCHANAWONG, C., GOLDENBERG, R., HARSHFIELD, B., HERNANDEZ, C., HERZLINGER-BOTEIN, S., ITOH, H., JIA, W., JIANG, Y. D., KADOWAKI, T., LARANJO, N., LEITER, L., MIWA, T., ODAWARA, M., OHASHI, K., OHNO, A., PAN, C., PAN, J., PEDRO-BOTET, J., REINER, Z., ROTELLA, C. M., SIMO, R., TANAKA, M., TEDESCHI-REINER, E., TWUM-BARIMA, D., ZOPPINI, G. & CAREY, V. J. 2014. Association between plasma triglycerides and high-density lipoprotein cholesterol and microvascular kidney disease and retinopathy in type 2 diabetes mellitus: a global case-control study in 13 countries. *Circulation,* 129**,** 999-1008.

SAEEDI, P., PETERSOHN, I., SALPEA, P., MALANDA, B., KARURANGA, S., UNWIN, N., COLAGIURI, S., GUARIGUATA, L., MOTALA, A. A., OGURTSOVA, K., SHAW, J. E., BRIGHT, D., WILLIAMS, R. & COMMITTEE, I. D. F. D. A. 2019. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9(th) edition. *Diabetes Res Clin Pract,* 157**,** 107843.

SAMARASEKERA, E. J., CLARK, C. E., KAUR, S., PATEL, R. S., MILLS, J. & GUIDELINE, C. 2023. Cardiovascular disease risk assessment and reduction: summary of updated NICE guidance. *BMJ,* 381**,** 1028.

SAMUELSSON, O., MULEC, H., KNIGHT-GIBSON, C., ATTMAN, P. O., KRON, B., LARSSON, R., WEISS, L., WEDEL, H. & ALAUPOVIC, P. 1997. Lipoprotein abnormalities are associated with increased rate of progression of human chronic renal insufficiency. *Nephrol Dial Transplant,* 12**,** 1908-15.

SARAN, R., LI, Y., ROBINSON, B., AYANIAN, J., BALKRISHNAN, R., BRAGG-GRESHAM, J., CHEN, J. T., COPE, E., GIPSON, D., HE, K., HERMAN, W., HEUNG, M., HIRTH, R. A., JACOBSEN, S. S., KALANTAR-ZADEH, K., KOVESDY, C. P., LEICHTMAN, A. B., LU, Y., MOLNAR, M. Z., MORGENSTERN, H., NALLAMOTHU, B., O'HARE, A. M., PISONI, R., PLATTNER, B., PORT, F. K., RAO, P., RHEE, C. M., SCHAUBEL, D. E., SELEWSKI, D. T., SHAHINIAN, V., SIM, J. J., SONG, P., STREJA, E., KURELLA TAMURA, M., TENTORI, F., EGGERS, P. W., AGODOA, L. Y. & ABBOTT, K. C. 2015. US Renal Data System 2014 Annual Data Report: Epidemiology of Kidney Disease in the United States. *Am J Kidney Dis,* 66**,** Svii, S1-305.

SARMA, S., SOCKALINGAM, S. & DASH, S. 2021. Obesity as a multisystem disease: Trends in obesity rates and obesity-related complications. *Diabetes, Obesity and Metabolism,* 23**,** 3-16.

SAUCEDO, A. L., PERALES-QUINTANA, M. M., PANIAGUA-VEGA, D., SANCHEZ-MARTINEZ, C., CORDERO-PEREZ, P. & MINSKY, N. W. 2018. Chronic Kidney Disease and the Search for New Biomarkers for Early Diagnosis. *Curr Med Chem,* 25**,** 3719-3747.

SCHAEFER, F., TRACHTMAN, H., WUHL, E., KIRCHNER, M., HAYEK, S. S., ANARAT, A., DUZOVA, A., MIR, S., PARIPOVIC, D., YILMAZ, A., LUGANI, F., ARBEITER, K., LITWIN, M., OH, J., MATTEUCCI, M. C., GELLERMANN, J., WYGODA, S., JANKAUSKIENE, A., KLAUS, G., DUSEK, J., TESTA, S., ZUROWSKA, A., CALDAS AFONSO, A., TRACY, M., WEI, C., SEVER, S., SMOYER, W., REISER, J., CONSORTIUM, E. T. & THE, C. S. G. 2017. Association of Serum Soluble Urokinase Receptor Levels With Progression of Kidney Disease in Children. *JAMA Pediatr,* 171**,** e172914.

SCHAEFFNER, E. S., EBERT, N., DELANAYE, P., FREI, U., GAEDEKE, J., JAKOB, O., KUHLMANN, M. K., SCHUCHARDT, M., TOLLE, M., ZIEBIG, R., VAN DER GIET, M. & MARTUS, P. 2012. Two novel equations to estimate kidney function in persons aged 70 years or older. *Ann Intern Med,* 157**,** 471-81.

SCHANSTRA, J. P., ZURBIG, P., ALKHALAF, A., ARGILES, A., BAKKER, S. J., BEIGE, J., BILO, H. J., CHATZIKYRKOU, C., DAKNA, M., DAWSON, J., DELLES, C., HALLER, H., HAUBITZ, M., HUSI, H., JANKOWSKI, J., JERUMS, G., KLEEFSTRA, N., KUZNETSOVA, T., MAAHS, D. M., MENNE, J., MULLEN, W., ORTIZ, A., PERSSON, F., ROSSING, P., RUGGENENTI, P., RYCHLIK, I., SERRA, A. L., SIWY, J., SNELL-BERGEON, J., SPASOVSKI, G., STAESSEN, J. A., VLAHOU, A., MISCHAK, H. & VANHOLDER, R. 2015. Diagnosis and Prediction of CKD Progression by Assessment of Urinary Peptides. *J Am Soc Nephrol,* 26**,** 1999-2010.

SCHEI, J., STEFANSSON, V. T., MATHISEN, U. D., ERIKSEN, B. O., SOLBU, M. D., JENSSEN, T. G. & MELSOM, T. 2016. Residual Associations of Inflammatory Markers with eGFR after Accounting for Measured GFR in a Community-Based Cohort without CKD. *Clin J Am Soc Nephrol,* 11**,** 280-6.

SCHMITZ, A., VAETH, M. & MOGENSEN, C. E. 1994. Systolic blood pressure relates to the rate of progression of albuminuria in NIDDM. *Diabetologia,* 37**,** 1251-8.

SCHOTTKER, B., HERDER, C., MULLER, H., BRENNER, H. & ROTHENBACHER, D. 2012. Clinical utility of creatinine- and cystatin C-based definition of renal function for risk prediction of primary cardiovascular events in patients with diabetes. *Diabetes Care,* 35**,** 879-86.

SEKULA, P., GOEK, O. N., QUAYE, L., BARRIOS, C., LEVEY, A. S., ROMISCH-MARGL, W., MENNI, C., YET, I., GIEGER, C., 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., KASTENMULLER, G. & KOTTGEN, A. 2016. A metabolome-wide association study of kidney function and disease in the general population. *Journal of the American Society of Nephrology,* 27**,** 1175-1188.

SEN, T., LI, J., NEUEN, B. L., NEAL, B., ARNOTT, C., PARIKH, C. R., COCA, S. G., PERKOVIC, V., MAHAFFEY, K. W., YAVIN, Y., ROSENTHAL, N., HANSEN, M. K. & HEERSPINK, H. J. L. 2021. Effects of the SGLT2 inhibitor canagliflozin on plasma biomarkers TNFR-1, TNFR-2 and KIM-1 in the CANVAS trial. *Diabetologia,* 64**,** 2147-2158.

SHAH, G. M., LIN, Z. L., KAMANNA, V. S., PAI, R., BASSA, B., JIN, F. Y., ROH, D. D., KASHYAP, M. L. & KIRSCHENBAUM, M. A. 1996. Effect of serum subfractions from peritoneal dialysis patients on Hep-G2 cell apolipoprotein A-I and B metabolism. *Kidney Int,* 50**,** 2079-87.

SLIEKER, R. C., VAN DER HEIJDEN, A., SIDDIQUI, M. K., LANGENDOEN-GORT, M., NIJPELS, G., HERINGS, R., FEENSTRA, T. L., MOONS, K. G. M., BELL, S., ELDERS, P. J., T HART, L. M. & BEULENS, J. W. J. 2021. Performance of prediction models for nephropathy in people with type 2 diabetes: systematic review and external validation study. *BMJ,* 374**,** n2134.

SNIDERMAN, A. D. & FARAJ, M. 2007. Apolipoprotein B, apolipoprotein A-I, insulin resistance and the metabolic syndrome. *Curr Opin Lipidol,* 18**,** 633-7.

SOININEN, P., KANGAS, A. J., WURTZ, P., SUNA, T. & ALA-KORPELA, M. 2015. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circ Cardiovasc Genet,* 8**,** 192-206.

SOLINI, A., MANCA, M. L., PENNO, G., PUGLIESE, G., COBB, J. E. & FERRANNINI, E. 2016. Prediction of Declining Renal Function and Albuminuria in Patients With Type 2 Diabetes by Metabolomics. *Journal of Clinical Endocrinology & Metabolism,* 101**,** 696-704.

SONG, M., XU, B. P., LIANG, Q., WEI, Y., SONG, Y., CHEN, P., ZHOU, Z., ZHANG, N., HE, Q., LIU, L., LIU, T., ZHANG, K., HU, C., WANG, B., XU, X. & SHI, H. 2021. Association of serum choline levels and all-cause mortality risk in adults with hypertension: a nested case-control study. *Nutr Metab (Lond),* 18**,** 108.

STANIFER, J. W., KILONZO, K., WANG, D., SU, G., MAO, W., ZHANG, L., ZHANG, L., NAYAK-RAO, S. & MIRANDA, J. J. 2017. Traditional Medicines and Kidney Disease in Low- and Middle-Income Countries: Opportunities and Challenges. *Semin Nephrol,* 37**,** 245-259.

STEINBRENNER, I., KOTSIS, F., SCHLOSSER, P., STOCKMANN, H., MOHNEY, R. P., SCHMID, M., OEFNER, P. J., ECKARDT, K. U., KOTTGEN, A., SEKULA, P., MEISELBACH, H., SCHNEIDER, M. P., SCHIFFER, M., PROKOSCH, H. U., BARTHLEIN, B., BECK, A., REIS, A., EKICI, A. B., BECKER, S., BECKER-GROSSPITSCH, D., ALBERTH-SCHMIDT, U., HAUSKNECHT, B., WEIGEL, A., WALZ, G., SCHULTHEISS, U. T., MEDER, S., MITSCH, E., REINHARD, U., FLOEGE, J., SARITAS, T., SCHAEFFNER, E., BAID-AGRAWAL, S., THEISEN, K., HALLER, H., MENNE, J., ZEIER, M., SOMMERER, C., THEILINGER, J., WOLF, G., BUSCH, M., PAUL, R., SITTER, T., WANNER, C., KRANE, V., BORNER-KLEIN, A., BAUER, B., KRONENBERG, F., RASCHENBERGER, J., KOLLERITS, B., FORER, L., SCHONHERR, S., WEISSENSTEINER, H., OEFNER, P., GRONWALD, W. & NADAL, J. 2021. Urine Metabolite Levels, Adverse Kidney Outcomes, and Mortality in CKD Patients: A Metabolome-wide Association Study. *American Journal of Kidney Diseases,* 78(5)**,** 669-677.e1.

STEUBL, D., BLOCK, M., HERBST, V., NOCKHER, W. A., SCHLUMBERGER, W., KEMMNER, S., BACHMANN, Q., ANGERMANN, S., WEN, M., HEEMANN, U., RENDERS, L., GARIMELLA, P. S. & SCHERBERICH, J. 2019. Urinary uromodulin independently predicts end-stage renal disease and rapid kidney function decline in a cohort of chronic kidney disease patients. *Medicine (Baltimore),* 98**,** e15808.

STEVENS, L. A., CORESH, J., FELDMAN, H. I., GREENE, T., LASH, J. P., NELSON, R. G., RAHMAN, M., DEYSHER, A. E., ZHANG, Y. L., SCHMID, C. H. & LEVEY, A. S. 2007. Evaluation of the modification of diet in renal disease study equation in a large diverse population. *J Am Soc Nephrol,* 18**,** 2749-57.

STEVENS, L. A., CORESH, J., GREENE, T. & LEVEY, A. S. 2006. Assessing kidney function--measured and estimated glomerular filtration rate. *N Engl J Med,* 354**,** 2473-83.

STEYERBERG, E. W. 2019. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating,* Switzerland, Springer Cham.

STEYERBERG, E. W., MOONS, K. G., VAN DER WINDT, D. A., HAYDEN, J. A., PEREL, P., SCHROTER, S., RILEY, R. D., HEMINGWAY, H., ALTMAN, D. G. & GROUP, P. 2013. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med,* 10**,** e1001381.

SU, D., CHEN, J., DU, S., KIM, H., YU, B., WONG, K. E., BOERWINKLE, E. & REBHOLZ, C. M. 2023. Metabolomic Markers of Ultra-Processed Food and Incident CKD. *Clin J Am Soc Nephrol,* 18**,** 327-336.

SUN, H., SAEEDI, P., KARURANGA, S., PINKEPANK, M., OGURTSOVA, K., DUNCAN, B. B., STEIN, C., BASIT, A., CHAN, J. C. N., MBANYA, J. C., PAVKOV, M. E., RAMACHANDARAN, A., WILD, S. H., JAMES, S., HERMAN, W. H., ZHANG, P., BOMMER, C., KUO, S., BOYKO, E. J. & MAGLIANO, D. J. 2022. IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. *Diabetes Res Clin Pract,* 183**,** 109119.

SWARTLING, O., RYDELL, H., STENDAHL, M., SEGELMARK, M., TROLLE LAGERROS, Y. & EVANS, M. 2021. CKD Progression and Mortality Among Men and Women: A Nationwide Study in Sweden. *Am J Kidney Dis,* 78**,** 190-199 e1.

T HART, L. M., VOGELZANGS, N., MOOK-KANAMORI, D. O., BRAHIMAJ, A., NANO, J., VAN DER HEIJDEN, A., WILLEMS VAN DIJK, K., SLIEKER, R. C., STEYERBERG, E. W., IKRAM, M. A., BEEKMAN, M., BOOMSMA, D. I., VAN DUIJN, C. M., SLAGBOOM, P. E., STEHOUWER, C. D. A., SCHALKWIJK, C. G., ARTS, I. C. W., DEKKER, J. M., DEHGHAN, A., MUKA, T., VAN DER KALLEN, C. J. H., NIJPELS, G. & VAN GREEVENBROEK, M. M. J. 2018. Blood Metabolomic Measures Associate With Present and Future Glycemic Control in Type 2 Diabetes. *J Clin Endocrinol Metab,* 103**,** 4569-4579.

TAAL, M. W. & BRENNER, B. M. 2006. Predicting initiation and progression of chronic kidney disease: Developing renal risk scores. *Kidney Int,* 70**,** 1694-705.

TALLO, D. 2016. The British National Formulary. *Nurs Stand,* 31**,** 64-65.

TALMUD, P. J., HAWE, E., MILLER, G. J. & HUMPHRIES, S. E. 2002. Nonfasting apolipoprotein B and triglyceride levels as a useful predictor of coronary heart disease risk in middle-aged UK men. *Arterioscler Thromb Vasc Biol,* 22**,** 1918-23.

TANAKA, Y., ATSUMI, Y., MATSUOKA, K., ONUMA, T., TOHJIMA, T. & KAWAMORI, R. 1998. Role of glycemic control and blood pressure in the development and progression of nephropathy in elderly Japanese NIDDM patients. *Diabetes Care,* 21**,** 116-20.

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., MARKS, A., NADKARNI, G. N., NAVANEETHAN, S. D., NELSON, R. G., TITZE, S., SARNAK, M. J., STENGEL, B., WOODWARD, M., ISEKI, K. & CONSORTIUM, C. K. D. P. 2016. Multinational Assessment of Accuracy of Equations for Predicting Risk of Kidney Failure: A Meta-analysis. *JAMA,* 315**,** 164-74.

TANGRI, N., STEVENS, L. A., GRIFFITH, J., TIGHIOUART, H., DJURDJEV, O., NAIMARK, D., LEVIN, A. & LEVEY, A. S. 2011. A predictive model for progression of chronic kidney disease to kidney failure. *JAMA,* 305**,** 1553-9.

TAVARES, G., VENTURINI, G., PADILHA, K., ZATZ, R., PEREIRA, A. C., THADHANI, R. I., RHEE, E. P. & TITAN, S. M. O. 2018. 1,5-Anhydroglucitol predicts CKD progression in macroalbuminuric diabetic kidney disease: results from non-targeted metabolomics. *Metabolomics,* 14(4) (no pagination).

TENG, J., DWYER, K. M., HILL, P., SEE, E., EKINCI, E. I., JERUMS, G. & MACISAAC, R. J. 2014. Spectrum of renal disease in diabetes. *Nephrology (Carlton),* 19**,** 528-36.

TERVAERT, T. W., MOOYAART, A. L., AMANN, K., COHEN, A. H., COOK, H. T., DRACHENBERG, C. B., FERRARIO, F., FOGO, A. B., HAAS, M., DE HEER, E., JOH, K., NOEL, L. H., RADHAKRISHNAN, J., SESHAN, S. V., BAJEMA, I. M., BRUIJN, J. A. & RENAL PATHOLOGY, S. 2010. Pathologic classification of diabetic nephropathy. *J Am Soc Nephrol,* 21**,** 556-63.

THOMAS, M. C., BROWNLEE, M., SUSZTAK, K., SHARMA, K., JANDELEIT-DAHM, K. A., ZOUNGAS, S., ROSSING, P., GROOP, P. H. & COOPER, M. E. 2015. Diabetic kidney disease. *Nat Rev Dis Primers,* 1**,** 15018.

THOMAS, M. C., COOPER, M. E. & ZIMMET, P. 2016. Changing epidemiology of type 2 diabetes mellitus and associated chronic kidney disease. *Nat Rev Nephrol,* 12**,** 73-81.

THOMAS, M. C., MORAN, J. L., HARJUTSALO, V., THORN, L., WADEN, J., SARAHEIMO, M., TOLONEN, N., LEIVISKA, J., JULA, A., FORSBLOM, C., GROOP, P. H. & FINNDIANE STUDY, G. 2012. Hyperfiltration in type 1 diabetes: does it exist and does it matter for nephropathy? *Diabetologia,* 55**,** 1505-13.

THOMAS, M. C., ROSENGARD-BARLUND, M., MILLS, V., RONNBACK, M., THOMAS, S., FORSBLOM, C., COOPER, M. E., TASKINEN, M. R., VIBERTI, G. & GROOP, P. H. 2006a. Serum lipids and the progression of nephropathy in type 1 diabetes. *Diabetes Care,* 29**,** 317-22.

THOMAS, M. C., WEEKES, A. J., BROADLEY, O. J., COOPER, M. E. & MATHEW, T. H. 2006b. The burden of chronic kidney disease in Australian patients with type 2 diabetes (the NEFRON study). *Med J Aust,* 185**,** 140-4.

THOMSON, S. C., DENG, A., BAO, D., SATRIANO, J., BLANTZ, R. C. & VALLON, V. 2001. Ornithine decarboxylase, kidney size, and the tubular hypothesis of glomerular hyperfiltration in experimental diabetes. *J Clin Invest,* 107**,** 217-24.

THURAISINGHAM, R. C., NOTT, C. A., DODD, S. M. & YAQOOB, M. M. 2000. Increased nitrotyrosine staining in kidneys from patients with diabetic nephropathy. *Kidney Int,* 57**,** 1968-72.

TIBSHIRANI, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological),* 58.1 (1996): 267-288.

TILLIN, T., HUGHES, A. D., WANG, Q., WURTZ, P., ALA-KORPELA, M., SATTAR, N., FOROUHI, N. G., GODSLAND, I. F., EASTWOOD, S. V., MCKEIGUE, P. M. & CHATURVEDI, N. 2015. Diabetes risk and amino acid profiles: cross-sectional and prospective analyses of ethnicity, amino acids and

diabetes in a South Asian and European cohort from the SABRE (Southall And Brent REvisited) Study. *Diabetologia,* 58**,** 968-79.

TITAN, S. M., PECOITS-FILHO, R., BARRETO, S. M., LOPES, A. A., BENSENOR, I. J. & LOTUFO, P. A. 2017. GlycA, a marker of protein glycosylation, is related to albuminuria and estimated glomerular filtration rate: the ELSA-Brasil study. *BMC Nephrol,* 18**,** 367.

TITAN, S. M., VENTURINI, G., PADILHA, K., GOULART, A. C., LOTUFO, P. A., BENSENOR, I. J., KRIEGER, J. E., THADHANI, R. I., RHEE, E. P. & PEREIRA, A. C. 2019. Metabolomics biomarkers and the risk of overall mortality and ESRD in CKD: Results from the PRoGREDIR cohort. *PLoS ONE,* 14(3) (no pagination).

TOFTE, N., SUVITAIVAL, T., AHONEN, L., WINTHER, S. A., THEILADE, S., FRIMODT-MOLLER, M., AHLUWALIA, T. S. & ROSSING, P. 2019a. Lipidomic analysis reveals sphingomyelin and phosphatidylcholine species associated with renal impairment and all-cause mortality in type 1 diabetes. *Scientific reports,* 9(1)**,** 16398.

TOFTE, N., SUVITAIVAL, T., TROST, K., MATTILA, I. M., THEILADE, S., WINTHER, S. A., AHLUWALIA, T. S., FRIMODT-MOLLER, M., LEGIDO-QUIGLEY, C. & ROSSING, P. 2019b. Metabolomic Assessment Reveals Alteration in Polyols and Branched Chain Amino Acids Associated With Present and Future Renal Impairment in a Discovery Cohort of 637 Persons With Type 1 Diabetes. *Frontiers in Endocrinology,* 10.

TOFTE, N., VOGELZANGS, N., MOOK-KANAMORI, D., BRAHIMAJ, A., NANO, J., AHMADIZAR, F., VAN DIJK, K. W., FRIMODT-MOLLER, M., ARTS, I., BEULENS, J. W. J., RUTTERS, F., VAN DER HEIJDEN, A. A., KAVOUSI, M., STEHOUWER, C. D. A., NIJPELS, G., VAN GREEVENBROEK, M. M. J., VAN DER KALLEN, C. J. H., ROSSING, P., AHLUWALIA, T. S. & T HART, L. M. 2020. Plasma Metabolomics Identifies Markers of Impaired Renal Function: A Meta-analysis of 3089 Persons with Type 2 Diabetes. *J Clin Endocrinol Metab,* 105.

TOLL, D. B., JANSSEN, K. J., VERGOUWE, Y. & MOONS, K. G. 2008. Validation, updating and impact of clinical prediction rules: a review. *J Clin Epidemiol,* 61**,** 1085-94.

TONNEIJCK, L., MUSKIET, M. H., SMITS, M. M., VAN BOMMEL, E. J., HEERSPINK, H. J., VAN RAALTE, D. H. & JOLES, J. A. 2017. Glomerular Hyperfiltration in Diabetes: Mechanisms, Clinical Significance, and Treatment. *J Am Soc Nephrol,* 28**,** 1023-1039.

TRISCHITTA, V., MASTROIANNO, M., SCARALE, M. G., PREHN, C., SALVEMINI, L., FONTANA, A., ADAMSKI, J., SCHENA, F. P., COSMO, S. D., COPETTI, M. & MENZAGHI, C. 2023. Circulating metabolites improve the prediction of renal impairment in patients with type 2 diabetes. *BMJ Open Diabetes Research and Care,* 11(5) (no pagination).

TSAI, C. W., GRAMS, M. E., INKER, L. A., CORESH, J. & SELVIN, E. 2014. Cystatin C- and creatinine-based estimated glomerular filtration rate, vascular disease, and mortality in persons with diabetes in the U.S. *Diabetes Care,* 37**,** 1002-8.

TSAI, W. C., WU, H. Y., PENG, Y. S., KO, M. J., WU, M. S., HUNG, K. Y., WU, K. D., CHU, T. S. & CHIEN, K. L. 2016. Risk Factors for Development and Progression of Chronic Kidney Disease: A Systematic Review and Exploratory Meta-Analysis. *Medicine (Baltimore),* 95**,** e3013.

TSENG, L. N., TSENG, Y. H., JIANG, Y. D., CHANG, C. H., CHUNG, C. H., LIN, B. J., CHUANG, L. M., TAI, T. Y. & SHEU, W. H. 2012. Prevalence of hypertension and dyslipidemia and their associations with micro- and macrovascular diseases in patients with diabetes in Taiwan: an analysis of nationwide data for 2000-2009. *J Formos Med Assoc,* 111**,** 625-36.

TSUDA, A., ISHIMURA, E., UEDONO, H., YASUMOTO, M., ICHII, M., NAKATANI, S., MORI, K., UCHIDA, J., EMOTO, M., NAKATANI, T. & INABA, M. 2016. Comparison of the Estimated Glomerular Filtration Rate (eGFR) in Diabetic Patients, Non-Diabetic Patients and Living Kidney Donors. *Kidney Blood Press Res,* 41**,** 40-7.

TUTTLE, K. R., BAKRIS, G. L., BILOUS, R. W., CHIANG, J. L., DE BOER, I. H., 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. 2014. Diabetic kidney disease: a report from an ADA Consensus Conference. *Am J Kidney Dis,* 64**,** 510-33.

UEHARA-WATANABE, N., OKUNO-OZEKI, N., MINAMIDA, A., NAKAMURA, I., NAKATA, T., NAKAI, K., YAGI-TOMITA, A., IDA, T., IKEDA, K., KITANI, T., YAMASHITA, N., KAMEZAKI, M., KIRITA, Y., MATOBA, S., TAMAGAKI, K. & KUSABA, T. 2022. Direct evidence of proximal tubular proliferation in early diabetic nephropathy. *Sci Rep,* 12**,** 778.

UFNAL, M., ZADLO, A. & OSTASZEWSKI, R. 2015. TMAO: A small molecule of great expectations. *Nutrition,* 31**,** 1317-23.

VALLON, V. & KOMERS, R. 2011. Pathophysiology of the diabetic kidney. *Compr Physiol,* 1**,** 1175-232.

VALLON, V., RICHTER, K., BLANTZ, R. C., THOMSON, S. & OSSWALD, H. 1999. Glomerular hyperfiltration in experimental diabetes mellitus: potential role of tubular reabsorption. *J Am Soc Nephrol,* 10**,** 2569-76.

VAN DEN BERG, R. A., HOEFSLOOT, H. C., WESTERHUIS, J. A., SMILDE, A. K. & VAN DER WERF, M. J. 2006. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics,* 7**,** 142.

VAN DER BURGH, A. C., GEURTS, S., AHMAD, S., IKRAM, M. A., CHAKER, L., FERRARO, P. M. & GHANBARI, M. 2024. Circulating metabolites associated with kidney function decline and incident CKD: a multi-platform population-based study. *Clinical Kidney Journal,* 17(1) (no pagination).

VAN DER VORST, E. P. C. 2020. High-Density Lipoproteins and Apolipoprotein A1. *Subcell Biochem,* 94**,** 399-420.

VAN RIJN, M. H. C., VAN DE LUIJTGAARDEN, M., VAN ZUILEN, A. D., BLANKESTIJN, P. J., WETZELS, J. F. M., DEBRAY, T. P. A. & VAN DEN BRAND, J. 2021. Prognostic models for chronic kidney disease: a systematic review and external validation. *Nephrol Dial Transplant,* 36**,** 1837-1850.

VAZIRI, N. D. 2006. Dyslipidemia of chronic renal failure: the nature, mechanisms, and potential consequences. *Am J Physiol Renal Physiol,* 290**,** F262-72.

VAZIRI, N. D., DENG, G. & LIANG, K. 1999. Hepatic HDL receptor, SR-B1 and Apo A-I expression in chronic renal failure. *Nephrol Dial Transplant,* 14**,** 1462-6.

VERBEKE, F., LINDLEY, E., VAN BORTEL, L., VANHOLDER, R., LONDON, G., COCHAT, P., WIECEK, A., FOUQUE, D. & VAN BIESEN, W. 2014. A European Renal Best Practice (ERBP) position statement on the Kidney Disease: Improving Global Outcomes (KDIGO) clinical practice guideline for the management of blood pressure in non-dialysis-dependent chronic kidney disease: an endorsement with some caveats for real-life application. *Nephrol Dial Transplant,* 29**,** 490-6.

VIGNOLI, A., GHINI, V., MEONI, G., LICARI, C., TAKIS, P. G., TENORI, L., TURANO, P. & LUCHINAT, C. 2019. High-Throughput Metabolomics by 1D NMR. *Angew Chem Int Ed Engl,* 58**,** 968-994.

WALLDIUS, G., JUNGNER, I., HOLME, I., AASTVEIT, A. H., KOLAR, W. & STEINER, E. 2001. High apolipoprotein B, low apolipoprotein A-I, and improvement in the prediction of fatal myocardial infarction (AMORIS study): a prospective study. *Lancet,* 358**,** 2026-33.

WANG, F., SUN, L., SUN, Q., LIANG, L., GAO, X., LI, R., PAN, A., LI, H., DENG, Y., HU, F. B., WU, J., ZENG, R. & LIN, X. 2018. Associations of plasma amino acid and acylcarnitine profiles with incident reduced glomerular filtration rate. *Clinical Journal of the American Society of Nephrology,* 13(4)**,** 560-568.

WANG, T. J., LARSON, M. G., VASAN, R. S., CHENG, S., RHEE, E. P., MCCABE, E., LEWIS, G. D., FOX, C. S., JACQUES, P. F., FERNANDEZ, C., O'DONNELL, C. J., CARR, S. A., MOOTHA, V. K., FLOREZ, J. C., SOUZA, A., MELANDER, O., CLISH, C. B. & GERSZTEN, R. E. 2011. Metabolite profiles and the risk of developing diabetes. *Nat Med,* 17**,** 448-53.

WANNER, C., KRANE, V., MARZ, W., OLSCHEWSKI, M., MANN, J. F., RUF, G., RITZ, E., GERMAN, D. & DIALYSIS STUDY, I. 2005. Atorvastatin in patients with type 2 diabetes mellitus undergoing hemodialysis. *N Engl J Med,* 353**,** 238-48.

WELSH, P., RANKIN, N., LI, Q., MARK, P. B., WURTZ, P., ALA-KORPELA, M., MARRE, M., POULTER, N., HAMET, P., CHALMERS, J., WOODWARD, M. & SATTAR, N. 2018. Circulating amino acids and

the risk of macrovascular, microvascular and mortality outcomes in individuals with type 2 diabetes: results from the ADVANCE trial. *Diabetologia,* 61**,** 1581-1591.

WEN, D., ZHENG, Z., SURAPANENI, A., YU, B., ZHOU, L., ZHOU, W., XIE, D., SHOU, H., AVILA-PACHECO, J., KALIM, S., HE, J., HSU, C. Y., PARSA, A., RAO, P., SONDHEIMER, J., TOWNSEND, R., WAIKAR, S. S., REBHOLZ, C. M., DENBURG, M. R., KIMMEL, P. L., VASAN, R. S., CLISH, C. B., CORESH, J., FELDMAN, H. I., GRAMS, M. E. & RHEE, E. P. 2022. Metabolite profiling of CKD progression in the chronic renal insufficiency cohort study. *JCI Insight,* 7(20) (no pagination).

WERNER, E. R., BLAU, N. & THONY, B. 2011. Tetrahydrobiopterin: biochemistry and pathophysiology. *Biochem J,* 438**,** 397-414.

WHITE, C. A., ALLEN, C. M., AKBARI, A., COLLIER, C. P., HOLLAND, D. C., DAY, A. G. & KNOLL, G. A. 2019. Comparison of the new and traditional CKD-EPI GFR estimation equations with urinary inulin clearance: A study of equation performance. *Clin Chim Acta,* 488**,** 189-195.

WILSON, P. W., CASTELLI, W. P. & KANNEL, W. B. 1987. Coronary risk prediction in adults (the Framingham Heart Study). *Am J Cardiol,* 59**,** 91G-94G.

WISEMAN, M. J., MANGILI, R., ALBERETTO, M., KEEN, H. & VIBERTI, G. 1987. Glomerular response mechanisms to glycemic changes in insulin-dependent diabetics. *Kidney Int,* 31**,** 1012-8.

WISHART, D. S., GUO, A., OLER, E., WANG, F., ANJUM, A., PETERS, H., DIZON, R., SAYEEDA, Z., TIAN, S., LEE, B. L., BERJANSKII, M., MAH, R., YAMAMOTO, M., JOVEL, J., TORRES-CALZADA, C., HIEBERT-GIESBRECHT, M., LUI, V. W., VARSHAVI, D., VARSHAVI, D., ALLEN, D., ARNDT, D., KHETARPAL, N., SIVAKUMARAN, A., HARFORD, K., SANFORD, S., YEE, K., CAO, X., BUDINSKI, Z., LIIGAND, J., ZHANG, L., ZHENG, J., MANDAL, R., KARU, N., DAMBROVA, M., SCHIOTH, H. B., GREINER, R. & GAUTAM, V. 2022. HMDB 5.0: the Human Metabolome Database for 2022. *Nucleic Acids Res,* 50**,** D622-D631.

WORLD MEDICAL, A. 2001. World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects. *Bull World Health Organ,* 79**,** 373-4.

WURTZ, P., HAVULINNA, A. S., SOININEN, P., TYNKKYNEN, T., PRIETO-MERINO, D., TILLIN, T., GHORBANI, A., ARTATI, A., WANG, Q., TIAINEN, M., KANGAS, A. J., KETTUNEN, J., KAIKKONEN, J., MIKKILA, V., JULA, A., KAHONEN, M., LEHTIMAKI, T., LAWLOR, D. A., GAUNT, T. R., HUGHES, A. D., SATTAR, N., ILLIG, T., ADAMSKI, J., WANG, T. J., PEROLA, M., RIPATTI, S., VASAN, R. S., RAITAKARI, O. T., GERSZTEN, R. E., CASAS, J. P., CHATURVEDI, N., ALA-KORPELA, M. & SALOMAA, V. 2015. Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. *Circulation,* 131**,** 774-85.

WURTZ, P., KANGAS, A. J., SOININEN, P., LAWLOR, D. A., DAVEY SMITH, G. & ALA-KORPELA, M. 2017. Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Large-Scale Epidemiology: A Primer on -Omic Technologies. *Am J Epidemiol,* 186**,** 1084-1096.

XIE, Y., BOWE, B., MOKDAD, A. H., XIAN, H., YAN, Y., LI, T., MADDUKURI, G., TSAI, C. Y., FLOYD, T. & AL-ALY, Z. 2018. Analysis of the Global Burden of Disease study highlights the global, regional, and national trends of chronic kidney disease epidemiology from 1990 to 2016. *Kidney Int,* 94**,** 567-581.

XU, J. Z., GARRETT, M. E., SOLDANO, K. L., CHEN, S. T., CLISH, C. B., ASHLEY-KOCH, A. E. & TELEN, M. J. 2018. Clinical and metabolomic risk factors associated with rapid renal function decline in sickle cell disease. *American Journal of Hematology,* 93(12)**,** 1451-1460.

YANG, M., FOX, C. H., VASSALOTTI, J. & CHOI, M. 2011. Complications of progression of CKD. *Adv Chronic Kidney Dis,* 18**,** 400-5.

YANG, Y., YANG, B., LIU, B., LIANG, Y., LUO, Q., ZHAO, Z., LIU, Z., ZENG, Q. & XIONG, C. 2023. Circulating choline levels are associated with prognoses in patients with pulmonary hypertension: a cohort study. *BMC Pulm Med,* 23**,** 313.

YAPA, H. E., CHAMBERS, S., PURTELL, L. & BONNER, A. 2023. Impact of chronic kidney disease on everyday life: A descriptive qualitative study. *J Ren Care*.

YOUDEN, W. J. 1950. Index for rating diagnostic tests. *Cancer,* 3**,** 32-5.

YOUNG, G. A. & PARSONS, F. M. 1973. Impairment of phenylalanine hydroxylation in chronic renal insufficiency. *Clin Sci,* 45**,** 89-97.

YU, B., ZHENG, Y., NETTLETON, J. A., ALEXANDER, D., CORESH, J. & BOERWINKLE, E. 2014. Serum metabolomic profiling and incident ckd among african americans. *Clinical Journal of the American Society of Nephrology,* 9**,** 1410-1417.

YUI, Y., AOYAMA, T., MORISHITA, H., TAKAHASHI, M., TAKATSU, Y. & KAWAI, C. 1988. Serum prostacyclin stabilizing factor is identical to apolipoprotein A-I (Apo A-I). A novel function of Apo A-I. *J Clin Invest,* 82**,** 803-7.

ZACHARIAS, H. U., ALTENBUCHINGER, M., SCHULTHEISS, U. T., SAMOL, C., KOTSIS, F., POGUNTKE, I., SEKULA, P., KRUMSIEK, J., KOTTGEN, A., SPANG, R., OEFNER, P. J. & GRONWALD, W. 2019. A Novel Metabolic Signature To Predict the Requirement of Dialysis or Renal Transplantation in Patients with Chronic Kidney Disease. *J Proteome Res,* 18**,** 1796-1805.

ZAGHLOOL, S. B., HALAMA, A., STEPHAN, N., GUDMUNDSDOTTIR, V., GUDNASON, V., JENNINGS, L. L., THANGAM, M., AHLQVIST, E., MALIK, R. A., ALBAGHA, O. M. E., ABOU-SAMRA, A. B. & SUHRE, K. 2022. Metabolic and proteomic signatures of type 2 diabetes subtypes in an Arab population. *Nat Commun,* 13**,** 7121.

ZENG, W., BEYENE, H. B., KUOKKANEN, M., MIAO, G., MAGLIANO, D. J., UMANS, J. G., FRANCESCHINI, N., COLE, S. A., MICHAILIDIS, G., LEE, E. T., HOWARD, B. V., FIEHN, O., CURRAN, J. E., BLANGERO, J., MEIKLE, P. J. & ZHAO, J. 2022. Lipidomic profiling in the Strong Heart Study identified American Indians at risk of chronic kidney disease. *Kidney International,* 102(5)**,** 1154-1166.

ZEWINGER, S., SPEER, T., KLEBER, M. E., SCHARNAGL, H., WOITAS, R., LEPPER, P. M., PFAHLER, K., SEILER, S., HEINE, G. H., MARZ, W., SILBERNAGEL, G. & FLISER, D. 2014. HDL cholesterol is not associated with lower mortality in patients with kidney dysfunction. *J Am Soc Nephrol,* 25**,** 1073-82.

ZHANG, J., FUHRER, T., YE, H., KWAN, B., MONTEMAYOR, D., TUMOVA, J., DARSHI, M., AFSHINNIA, F., SCIALLA, J. J., ANDERSON, A., PORTER, A. C., TALIERCIO, J. J., RINCON-CHOLES, H., RAO, P., XIE, D., FELDMAN, H., SAUER, U., SHARMA, K. & NATARAJAN, L. 2022. High-Throughput Metabolomics and Diabetic Kidney Disease Progression: Evidence from the Chronic Renal Insufficiency (CRIC) Study. *American Journal of Nephrology,* 53(2-3)**,** 215-225.

ZHANG, S., LI, X., LUO, H., FANG, Z. Z. & AI, H. 2020a. Role of aromatic amino acids in pathogeneses of diabetic nephropathy in Chinese patients with type 2 diabetes. *J Diabetes Complications,* 34**,** 107667.

ZHANG, Y., ZHANG, S. & WANG, G. 2015. Metabolomic biomarkers in diabetic kidney diseases--A systematic review. *J Diabetes Complications,* 29**,** 1345-51.

ZHANG, Y. B., SHENG, L. T., WEI, W., GUO, H., YANG, H., MIN, X., GUO, K., YANG, K., ZHANG, X., HE, M., WU, T. & PAN, A. 2020b. Association of blood lipid profile with incident chronic kidney disease: A Mendelian randomization study. *Atherosclerosis,* 300**,** 19-25.

ZHENG, Y., LEY, S. H. & HU, F. B. 2018. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. *Nat Rev Endocrinol,* 14**,** 88-98.

ZHENG, Y., LI, Y., QI, Q., HRUBY, A., MANSON, J. E., WILLETT, W. C., WOLPIN, B. M., HU, F. B. & QI, L. 2016. Cumulative consumption of branched-chain amino acids and incidence of type 2 diabetes. *Int J Epidemiol,* 45**,** 1482-1492.

ZHU, H., BAI, M., XIE, X., WANG, J., WENG, C., DAI, H., CHEN, J., HAN, F. & LIN, W. 2022. Impaired Amino Acid Metabolism and Its Correlation with Diabetic Kidney Disease Progression in Type 2 Diabetes Mellitus. *Nutrients,* 14.

ZOCCALI, C., MALLAMACI, F., HALIMI, J. M., ROSSIGNOL, P., SARAFIDIS, P., DE CATERINA, R., GIUGLIANO, R. & ZANNAD, F. 2023. From Cardiorenal Syndrome to Chronic Cardiovascular and Kidney Disorder: A Conceptual Transition. *Clin J Am Soc Nephrol*.

ZURBIG, P., JERUMS, G., HOVIND, P., MACISAAC, R. J., MISCHAK, H., NIELSEN, S. E., PANAGIOTOPOULOS, S., PERSSON, F. & ROSSING, P. 2012. Urinary proteomics for early diagnosis in diabetic nephropathy. *Diabetes,* 61**,** 3304-13.

Appendix A- publications and presentations

This section contains the titles and abstracts for the work I published during my PhD studies with my supervisors and colleagues.

Publication(s):

Exploration of metabolomic markers associated with declining kidney function in people with type 2 diabetes mellitus (2024)

Authors: Justina Krasauskaite, Bryan Conway, Christopher Weir, Zhe Huang, Jackie Price

Abstract

Background: Metabolomics, the study of small molecules in biological systems, can provide valuable insights into kidney dysfunction in people with type 2 diabetes mellitus (T2DM), but prospective studies are scarce. We investigated the association between metabolites and kidney function decline in people with T2DM.

Methods: The Edinburgh Type 2 Diabetes Study, a population-based cohort of 1066 men and women aged 60-75 years with T2DM. We measured 149 serum metabolites at baseline and investigated individual associations with baseline estimated glomerular filtration rate (eGFR), incident chronic kidney disease (CKD, eGFR <60ml/min/(1.73m)2) and decliner status (5% eGFR decline per year).

Results: At baseline, mean eGFR was 77.5ml/min/(1.73m)2, N=1058) and 216 individuals had evidence of CKD. Of those without CKD, 155 developed CKD over a median 7-year follow-up. Eighty-eight metabolites were significantly associated with baseline eGFR (β range −4.08 to 3.92; PFDR<0.001). Very low density lipoproteins, triglycerides, amino acids (AAs), glycoprotein acetyls, and fatty acids showed inverse associations, while cholesterol and phospholipids in high density lipoproteins exhibited positive associations. AA isoleucine, apolipoprotein A1 (ApoA1) and total cholines were not only associated with baseline kidney measures (PFDR<0.05), but also showed stable, nominally significant association with incident CKD and decline.

Conclusion: Our study revealed widespread changes within the metabolomic profile of CKD, particularly in lipoproteins and their lipid compounds. We identified a smaller number of individual metabolites which specifically associated with kidney function decline. Replication studies are needed to confirm the longitudinal findings and explore if metabolic signals at baseline can predict kidney decline.

Combining serum metabolomic profiles with traditional risk factors improves 10-year cardiovascular risk prediction in people with type 2 diabetes (2023)

Authors: Zhe Huang, Lucija Klaric, Justina Krasauskaite, Wardah Khalid, Mark W J Strachan, James F Wilson, Jackie F Price

Abstract

Aims: To identify a group of metabolites associated with incident cardiovascular disease (CVD) in people with type 2 diabetes and assess its predictive performance over-and-above a current CVD risk score (QRISK3).

Methods and results: A panel of 228 serum metabolites was measured at baseline in 1066 individuals with type 2 diabetes (Edinburgh Type 2 Diabetes Study) who were then followed up for CVD over the subsequent 10 years. We applied 100 repeats of Cox least absolute shrinkage and selection operator to select metabolites with frequency >90% as components for a metabolites-based risk score (MRS). The predictive performance of the MRS was assessed in relation to a reference model that was based on QRISK3 plus prevalent CVD and statin use at baseline. Of 1021 available individuals, 255 (25.0%) developed CVD (median follow-up: 10.6 years). Twelve metabolites relating to fluid balance, ketone bodies, amino acids, fatty acids, glycolysis, and lipoproteins were selected to construct the MRS that showed positive association with 10-year cardiovascular risk following adjustment for traditional risk

factors [hazard ratio (HR) 2.67; 95% confidence interval (CI) 1.96, 3.64]. The c-statistic was 0.709 (95%CI 0.679, 0.739) for the reference model alone, increasing slightly to 0.728 (95%CI 0.700, 0.757) following addition of the MRS. Compared with the reference model, the net reclassification index and integrated discrimination index for the reference model plus the MRS were 0.362 (95%CI 0.179, 0.506) and 0.041 (95%CI 0.020, 0.071), respectively.

Conclusion: Metabolomics data might improve predictive performance of current CVD risk scores based on traditional risk factors in people with type 2 diabetes. External validation is warranted to assess the generalizability of improved CVD risk prediction using the MRS.

Serum metabolomic profiles associated with subclinical and clinical cardiovascular phenotypes in people with type 2 diabetes (2022)

Authors: Zhe Huang, Lucija Klaric, Justina Krasauskaite, Stela McLachlan, Mark W. J. Strachan, James F. Wilson & Jackie F. Price

Abstract

Background: Atherosclerotic cardiovascular diseases (CVD) is the leading cause of death in diabetes, but the full range of biomarkers reflecting atherosclerotic burden and CVD risk in people with diabetes is unknown. Metabolomics may help identify novel biomarkers potentially involved in development of atherosclerosis. We investigated the serum metabolomic profile of subclinical atherosclerosis, measured using ankle brachial index (ABI), in people with type 2 diabetes, compared with the profile for symptomatic CVD in the same population.

Methods: The Edinburgh Type 2 Diabetes Study is a cohort of 1,066 individuals with type 2 diabetes. ABI was measured at baseline, years 4 and 10, with cardiovascular events assessed at baseline and during 10 years of follow-up. A panel of 228 metabolites was measured at baseline using nuclear magnetic resonance spectrometry, and their association with both ABI and prevalent CVD was explored using univariate regression models and least absolute shrinkage and selection operator (LASSO). Metabolites associated with baseline ABI were further explored for association with follow-up ABI and incident CVD.

Results: Mean (standard deviation, SD) ABI at baseline was 0.97 (0.18, N = 1025), and prevalence of CVD was 35.0%. During 10-year follow-up, mean (SD) change in ABI was + 0.006 (0.178, n = 436), and 257 CVD events occurred. Lactate, glycerol, creatinine and glycoprotein acetyls levels were associated with baseline ABI in both univariate regression [βs (95% confidence interval, CI) ranged from − 0.025 (− 0.036, − 0.015) to − 0.023 (− 0.034, − 0.013), all $p < 0.0002$] and LASSO analysis. The associations remained nominally significant after adjustment for major vascular risk factors. In prospective analyses, lactate was nominally associated with ABI measured at years 4 and 10 after adjustment for baseline ABI. The four ABI-associated metabolites were all positively associated with prevalent CVD [odds ratios (ORs) ranged from 1.29 (1.13, 1.47) to 1.49 (1.29, 1.74), all $p < 0.0002$], and they were also positively associated with incident CVD [ORs (95% CI) ranged from 1.19 (1.02, 1.39) to 1.35 (1.17, 1.56), all $p < 0.05$].

Conclusions: Serum metabolites relating to glycolysis, fluid balance and inflammation were independently associated with both a marker of subclinical atherosclerosis and with symptomatic CVD in people with type 2 diabetes. Additional investigation is warranted to determine their roles as possible etiological and/or predictive biomarkers for atherosclerotic CVD.

Conference abstracts

58 th EASD Annual Meeting of the European Association for the Study of Diabetes

Identification of markers for predicting the onset of chronic kidney disease in older people with type 2 diabetes by metabolomic profiling: Edinburgh Type 2 Diabetes Study

J. Krasauskaite, B.R. Conway, C.J. Weir, Z. Huang, J.F. Price

Background and aims: Renal disease affects a large proportion of people with type 2 diabetes and it is associated with excess morbidity/ mortality. While well-established clinical biomarkers, namely estimated glomerular filtration rate (eGFR) and albuminuria are used in routine screening, these markers do not explain all of the risk. Hence, the search for new markers is a high priority. Metabolomics may reveal novel markers of chronic kidney disease (CKD) that could aid identification of patients at higher risk of renal impairment and improve risk prediction of incident CKD. We aimed to identify significant associations between metabolites and the clinically relevant outcome of incident CKD in a Scottish population of older people with type 2 diabetes and to evaluate the ability of metabolites to predict CKD onset.

Materials and methods: The Edinburgh type 2 diabetes Study (ET2DS) is a population-based cohort of 1,058 adults (49% female) with type 2 diabetes, aged 60-75 years. Nightingale metabolomic platform was used to measure 149 serum metabolite concentrations at baseline. Kidney function was determined by eGFR, calculated using the CKD-EPI equation. Incident CKD was defined as 2 of 3 eGFR records <60mL/min/1.73 m2 during follow-up. An initial multivariable-adjusted discovery screen considered the correlation between each metabolite and baseline eGFR (adjusted for age and sex). Metabolites that were significantly associated with eGFR were then related to incident CKD events in logistic regression analysis adjusted for known clinical risk factors. Risk prediction analysis involved refitting a published risk prediction model for incident CKD to evaluate the complementary value of significant metabolites.

Results: There were 823 participants in ET2DS with no CKD based on eGFR records at baseline and 217 (26%) experienced new onset CKD during follow-up (median= 6.8 years [IQR 0.9- 7.6]). Corrected for multiple testing, 68 metabolites were significantly associated with baseline eGRF (Bonferroni corrected p<0.00034). Of these, only amino acid phenylalanine (Phe) was significantly associated with incident CKD after adjustment for known clinical risk factors (OR 0.73 [95% CI 0.60- 0.89], p=0.002). Phe was added to the published risk prediction model containing the clinical variables (eGFR, age, sex, BMI, use of diabetes medications, cardiovascular disease history, smoking, hypertension, HbA1c, albumin-to-creatinine ratio). Phe remained significant in this model and higher levels of Phe reduced risk of CKD onset (HR 0.80, [95% CI 0.68-0.93] per unit of SD, p-value= 0.004). However, Phe yielded only a small improvement in risk prediction (original model concordance (c)-statistic 0.81 [95% CI 0.79-0.84], model +Phe c-statistic 0.82 [95% CI 0.79-0.84].

Conclusion: Amino acid Phe was associated with incident CKD in people with type 2 diabetes, although, it did not improve an already well performing risk prediction model. It is possible to hypothesise that a more sophisticated multivariable analysis may reveal a combination of metabolites associated with CKD onset that together may improve the risk prediction.

Diabetes UK, 2022

Associations between metabolomic profiles and chronic kidney disease in older people with type 2 diabetes: The Edinburgh Type 2 Diabetes Study (ET2DS)

J. Krasauskaite; B. R. Conway; C. J. Weir; J. F. Price

Aims: One of the most prevalent co-morbidities of type 2 diabetes is chronic kidney disease (CKD), with 18-30% prevalence. Metabolomics may reveal novel markers of CKD that could aid identification of patients at higher risk, but studies in type 2 diabetes populations are scarce. This discovery study aimed to identify metabolomic markers associated with CKD in a Scottish population of older people with type 2 diabetes.

Methods: The ET2DS is population-based cohort of people with type 2 diabetes, aged 60-75 years. Nightingale metabolomic platform measured 149 serum metabolites in 1,058 adults (48.7% female). Kidney function was determined by estimated glomerular filtration rate, eGFR, calculated using CKD-EPI equation. CKD at baseline was defined as persistently reduced eGFR <60mL/min/1.73 m2 or albuminuria on ≥ 2 of 3 samples taken within 2 years prior to baseline. Univariable analysis considered each metabolite independently and associations were determined using linear regression for eGFR and logistic regression for CKD.

Results: At ET2DS baseline, 352 (33%) participants had evidence of CKD. 38 metabolites were significantly associated with the outcomes after adjustment for age, sex and significance threshold correction for multiple comparisons. Results revealed subtle but significant changes in lipid subclasses, glycoprotein acetyls (inflammation marker) and glycolysis metabolites in participants with CKD compared to those without.

Summary: Assessment of metabolomic profiles in the ET2DS identified a substantial number of metabolites with altered serum levels in people with existing CKD. These results encourage further multivariable analysis and evaluation of metabolomic markers as predictors of future kidney function decline in unaffected individuals.

Appendix B- supplementary material for results reported on ET2DS

Table S5-1. Description of metabolites measured in ET2DS (n=1058).

HDL- high density lipoprotein; IDL- intermediate density lipoprotein; LDL- low density lipoprotein; VLDL – very low density lipoprotein; SD- standard deviation; g/l- grams per liter; mol/l- moles epr litre; mmol/l – milimoles per litre.

Variable name	Variable description	Subclass	Units	Median	Mean (SD)	NA, n	NULL, n
XXL_VLDL_P	Concentration of chylomicrons and extremely large VLDL particles	Lipoprotein VLDL	mol/l	0	0.000 (0.000)	0	265
XXL_VLDL_L	Total lipids in chylomicrons and extremely large VLDL	Lipoprotein VLDL	mmol/l	0.015	0.022 (0.028)	0	265
XXL_VLDL_PL	Phospholipids in chylomicrons and extremely large VLDL	Lipoprotein VLDL	mmol/l	0.001	0.002 (0.004)	0	265
XXL_VLDL_C	Total cholesterol in chylomicrons and extremely large VLDL	Lipoprotein VLDL	mmol/l	0.003	0.004 (0.005)	0	265
XXL_VLDL_CE	Cholesterol esters in chylomicrons and extremely large VLDL	Lipoprotein VLDL	mmol/l	0.002	0.002 (0.003)	0	265
XXL_VLDL_FC	Free cholesterol in chylomicrons and extremely large VLDL	Lipoprotein VLDL	mmol/l	0.001	0.002 (0.002)	0	265
XXL_VLDL_TG	Triglycerides in chylomicrons and extremely large VLDL	Lipoprotein VLDL	mmol/l	0.01	0.016 (0.020)	0	265
XL_VLDL_P	Concentration of very large VLDL particles	Lipoprotein VLDL	mol/l	0	0.000 (0.000)	0	243
XL_VLDL_L	Total lipids in very large VLDL	Lipoprotein VLDL	mmol/l	0.051	0.074 (0.084)	0	243
XL_VLDL_PL	Phospholipids in very large VLDL	Lipoprotein VLDL	mmol/l	0.007	0.011 (0.014)	0	243
XL_VLDL_C	Total cholesterol in very large VLDL	Lipoprotein VLDL	mmol/l	0.008	0.012 (0.016)	0	243
XL_VLDL_CE	Cholesterol esters in very large VLDL	Lipoprotein VLDL	mmol/l	0.005	0.007 (0.008)	0	243

XL_VLDL_FC	Free cholesterol in very large VLDL	Lipoprotein VLDL	mmol/l	0.003	0.005 (0.007)	0	243
XL_VLDL_TG	Triglycerides in very large VLDL	Lipoprotein VLDL	mmol/l	0.036	0.050 (0.055)	0	243
L_VLDL_P	Concentration of large VLDL particles	Lipoprotein VLDL	mol/l	0	0.000 (0.000)	0	57
L_VLDL_L	Total lipids in large VLDL	Lipoprotein VLDL	mmol/l	0.29	0.359 (0.288)	0	57
L_VLDL_PL	Phospholipids in large VLDL	Lipoprotein VLDL	mmol/l	0.053	0.066 (0.052)	0	57
L_VLDL_C	Total cholesterol in large VLDL	Lipoprotein VLDL	mmol/l	0.062	0.077 (0.062)	0	57
L_VLDL_CE	Cholesterol esters in large VLDL	Lipoprotein VLDL	mmol/l	0.036	0.042 (0.030)	0	57
L_VLDL_FC	Free cholesterol in large VLDL	Lipoprotein VLDL	mmol/l	0.026	0.034 (0.033)	0	57
L_VLDL_TG	Triglycerides in large VLDL	Lipoprotein VLDL	mmol/l	0.174	0.217 (0.175)	0	57
M_VLDL_P	Concentration of medium VLDL particles	Lipoprotein VLDL	mol/l	0	0.000 (0.000)	0	0
M_VLDL_L	Total lipids in medium VLDL	Lipoprotein VLDL	mmol/l	0.636	0.719 (0.411)	0	0
M_VLDL_PL	Phospholipids in medium VLDL	Lipoprotein VLDL	mmol/l	0.128	0.144 (0.077)	0	0
M_VLDL_C	Total cholesterol in medium VLDL	Lipoprotein VLDL	mmol/l	0.165	0.184 (0.096)	0	0
M_VLDL_CE	Cholesterol esters in medium VLDL	Lipoprotein VLDL	mmol/l	0.092	0.102 (0.047)	0	0
M_VLDL_FC	Free cholesterol in medium VLDL	Lipoprotein VLDL	mmol/l	0.072	0.082 (0.051)	0	0
M_VLDL_TG	Triglycerides in medium VLDL	Lipoprotein VLDL	mmol/l	0.339	0.391 (0.242)	0	0
S_VLDL_P	Concentration of small VLDL particles	Lipoprotein VLDL	mol/l	0	0.000 (0.000)	0	0
S_VLDL_L	Total lipids in small VLDL	Lipoprotein VLDL	mmol/l	0.617	0.653 (0.235)	0	0
S_VLDL_PL	Phospholipids in small VLDL	Lipoprotein VLDL	mmol/l	0.15	0.157 (0.048)	0	0
S_VLDL_C	Total cholesterol in small VLDL	Lipoprotein VLDL	mmol/l	0.192	0.200 (0.067)	0	0
S_VLDL_CE	Cholesterol esters in small VLDL	Lipoprotein VLDL	mmol/l	0.106	0.111 (0.040)	0	0
S_VLDL_FC	Free cholesterol in small VLDL	Lipoprotein VLDL	mmol/l	0.084	0.088 (0.032)	0	0
S_VLDL_TG	Triglycerides in small VLDL	Lipoprotein VLDL	mmol/l	0.276	0.297 (0.131)	0	0
XS_VLDL_P	Concentration of very small VLDL particles	Lipoprotein VLDL	mol/l	0	0.000 (0.000)	0	2
XS_VLDL_L	Total lipids in very small VLDL	Lipoprotein VLDL	mmol/l	0.428	0.442 (0.113)	0	2
XS_VLDL_PL	Phospholipids in very small VLDL	Lipoprotein VLDL	mmol/l	0.133	0.138 (0.037)	0	2
XS_VLDL_C	Total cholesterol in very small VLDL	Lipoprotein VLDL	mmol/l	0.175	0.180 (0.054)	0	2
XS_VLDL_CE	Cholesterol esters in very small VLDL	Lipoprotein VLDL	mmol/l	0.11	0.113 (0.039)	0	2
XS_VLDL_FC	Free cholesterol in very small VLDL	Lipoprotein VLDL	mmol/l	0.065	0.067 (0.019)	0	2

XS_VLDL_TG	Triglycerides in very small VLDL	Lipoprotein VLDL	mmol/l	0.117	0.124 (0.043)	0	2
IDL_P	Concentration of IDL particles	Lipoprotein IDL	mol/l	0	0.000 (0.000)	0	0
IDL_L	Total lipids in IDL	Lipoprotein IDL	mmol/l	0.825	0.853 (0.234)	0	0
IDL_PL	Phospholipids in IDL	Lipoprotein IDL	mmol/l	0.238	0.244 (0.062)	0	0
IDL_C	Total cholesterol in IDL	Lipoprotein IDL	mmol/l	0.475	0.490 (0.156)	0	0
IDL_CE	Cholesterol esters in IDL	Lipoprotein IDL	mmol/l	0.331	0.344 (0.109)	0	0
IDL_FC	Free cholesterol in IDL	Lipoprotein IDL	mmol/l	0.143	0.146 (0.049)	0	0
IDL_TG	Triglycerides in IDL	Lipoprotein IDL	mmol/l	0.113	0.119 (0.035)	0	0
L_LDL_P	Concentration of large LDL particles	Lipoprotein LDL	mol/l	0	0.000 (0.000)	0	0
L_LDL_L	Total lipids in large LDL	Lipoprotein LDL	mmol/l	0.98	1.011 (0.290)	0	0
L_LDL_PL	Phospholipids in large LDL	Lipoprotein LDL	mmol/l	0.264	0.270 (0.060)	0	0
L_LDL_C	Total cholesterol in large LDL	Lipoprotein LDL	mmol/l	0.626	0.646 (0.212)	0	0
L_LDL_CE	Cholesterol esters in large LDL	Lipoprotein LDL	mmol/l	0.444	0.459 (0.160)	0	0
L_LDL_FC	Free cholesterol in large LDL	Lipoprotein LDL	mmol/l	0.183	0.187 (0.053)	0	0
L_LDL_TG	Triglycerides in large LDL	Lipoprotein LDL	mmol/l	0.09	0.095 (0.029)	0	0
M_LDL_P	Concentration of medium LDL particles	Lipoprotein LDL	mol/l	0	0.000 (0.000)	0	0
M_LDL_L	Total lipids in medium LDL	Lipoprotein LDL	mmol/l	0.567	0.586 (0.179)	0	0
M_LDL_PL	Phospholipids in medium LDL	Lipoprotein LDL	mmol/l	0.165	0.169 (0.035)	0	0
M_LDL_C	Total cholesterol in medium LDL	Lipoprotein LDL	mmol/l	0.356	0.369 (0.136)	0	0
M_LDL_CE	Cholesterol esters in medium LDL	Lipoprotein LDL	mmol/l	0.246	0.258 (0.112)	0	0
M_LDL_FC	Free cholesterol in medium LDL	Lipoprotein LDL	mmol/l	0.109	0.111 (0.025)	0	0
M_LDL_TG	Triglycerides in medium LDL	Lipoprotein LDL	mmol/l	0.046	0.048 (0.015)	0	0
S_LDL_P	Concentration of small LDL particles	Lipoprotein LDL	mol/l	0	0.000 (0.000)	0	1
S_LDL_L	Total lipids in small LDL	Lipoprotein LDL	mmol/l	0.36	0.371 (0.112)	0	1
S_LDL_PL	Phospholipids in small LDL	Lipoprotein LDL	mmol/l	0.119	0.121 (0.025)	0	1
S_LDL_C	Total cholesterol in small LDL	Lipoprotein LDL	mmol/l	0.211	0.218 (0.083)	0	1
S_LDL_CE	Cholesterol esters in small LDL	Lipoprotein LDL	mmol/l	0.149	0.156 (0.068)	0	1
S_LDL_FC	Free cholesterol in small LDL	Lipoprotein LDL	mmol/l	0.061	0.062 (0.016)	0	1
S_LDL_TG	Triglycerides in small LDL	Lipoprotein LDL	mmol/l	0.03	0.032 (0.011)	0	1

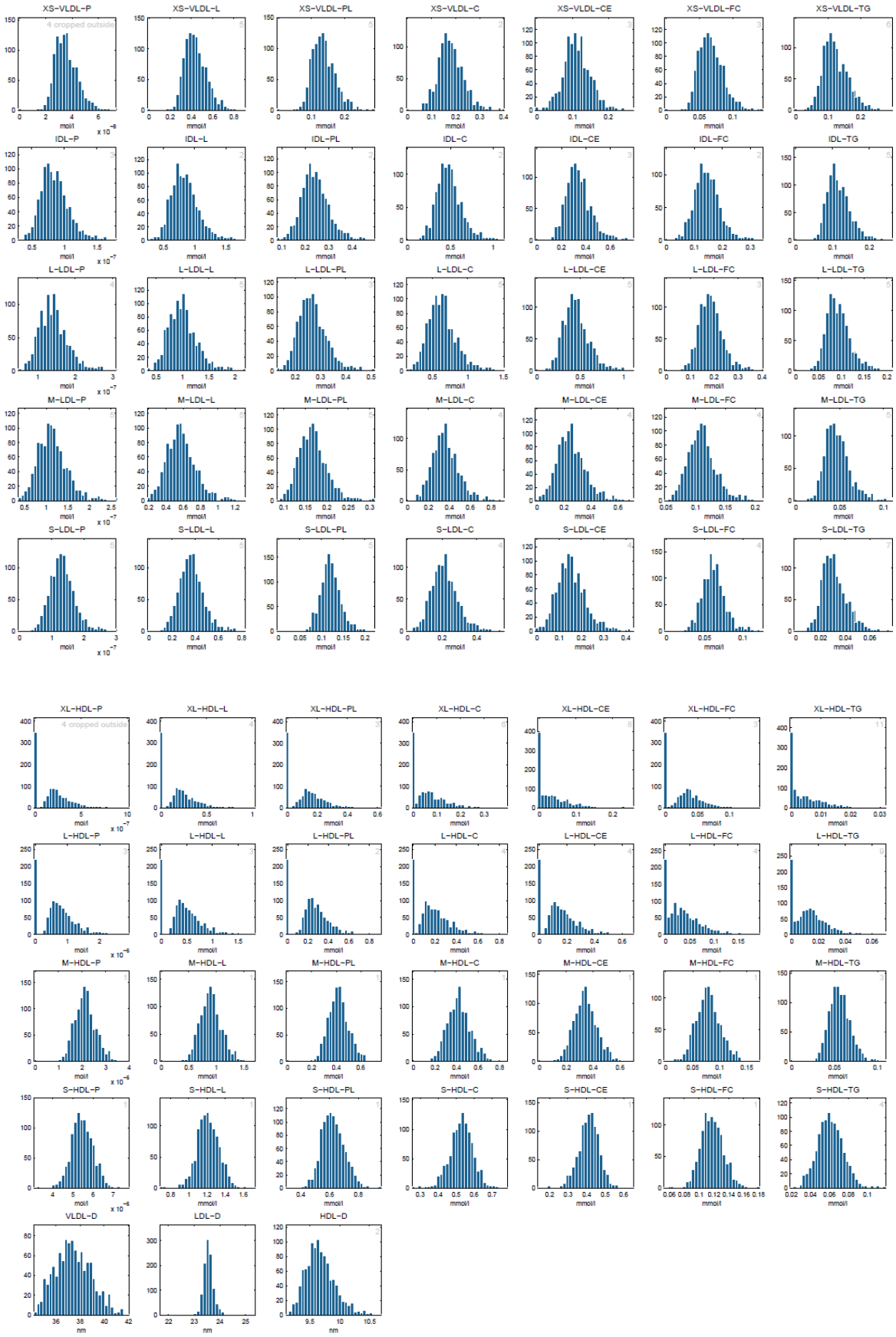
XL_HDL_P	Concentration of very large HDL particles	Lipoprotein HDL	mol/l	0	0.000 (0.000)	0	347
XL_HDL_L	Total lipids in very large HDL	Lipoprotein HDL	mmol/l	0.18	0.202 (0.204)	0	347
XL_HDL_PL	Phospholipids in very large HDL	Lipoprotein HDL	mmol/l	0.124	0.127 (0.122)	0	347
XL_HDL_C	Total cholesterol in very large HDL	Lipoprotein HDL	mmol/l	0.05	0.069 (0.081)	0	347
XL_HDL_CE	Cholesterol esters in very large HDL	Lipoprotein HDL	mmol/l	0.021	0.040 (0.053)	0	347
XL_HDL_FC	Free cholesterol in very large HDL	Lipoprotein HDL	mmol/l	0.029	0.030 (0.029)	0	347
XL_HDL_TG	Triglycerides in very large HDL	Lipoprotein HDL	mmol/l	0.003	0.005 (0.007)	0	347
L_HDL_P	Concentration of large HDL particles	Lipoprotein HDL	mol/l	0	0.000 (0.000)	0	221
L_HDL_L	Total lipids in large HDL	Lipoprotein HDL	mmol/l	0.42	0.445 (0.341)	0	221
L_HDL_PL	Phospholipids in large HDL	Lipoprotein HDL	mmol/l	0.243	0.239 (0.166)	0	221
L_HDL_C	Total cholesterol in large HDL	Lipoprotein HDL	mmol/l	0.164	0.192 (0.168)	0	221
L_HDL_CE	Cholesterol esters in large HDL	Lipoprotein HDL	mmol/l	0.132	0.153 (0.130)	0	221
L_HDL_FC	Free cholesterol in large HDL	Lipoprotein HDL	mmol/l	0.031	0.039 (0.038)	0	221
L_HDL_TG	Triglycerides in large HDL	Lipoprotein HDL	mmol/l	0.012	0.014 (0.014)	0	221
M_HDL_P	Concentration of medium HDL particles	Lipoprotein HDL	mol/l	0	0.000 (0.000)	0	1
M_HDL_L	Total lipids in medium HDL	Lipoprotein HDL	mmol/l	0.89	0.898 (0.196)	0	1
M_HDL_PL	Phospholipids in medium HDL	Lipoprotein HDL	mmol/l	0.416	0.420 (0.084)	0	1
M_HDL_C	Total cholesterol in medium HDL	Lipoprotein HDL	mmol/l	0.419	0.422 (0.109)	0	1
M_HDL_CE	Cholesterol esters in medium HDL	Lipoprotein HDL	mmol/l	0.34	0.343 (0.087)	0	1
M_HDL_FC	Free cholesterol in medium HDL	Lipoprotein HDL	mmol/l	0.078	0.079 (0.023)	0	1
M_HDL_TG	Triglycerides in medium HDL	Lipoprotein HDL	mmol/l	0.055	0.056 (0.013)	0	1
S_HDL_P	Concentration of small HDL particles	Lipoprotein HDL	mol/l	0	0.000 (0.000)	0	0
S_HDL_L	Total lipids in small HDL	Lipoprotein HDL	mmol/l	1.21	1.213 (0.126)	0	0
S_HDL_PL	Phospholipids in small HDL	Lipoprotein HDL	mmol/l	0.619	0.623 (0.082)	0	0
S_HDL_C	Total cholesterol in small HDL	Lipoprotein HDL	mmol/l	0.532	0.528 (0.066)	0	0
S_HDL_CE	Cholesterol esters in small HDL	Lipoprotein HDL	mmol/l	0.414	0.411 (0.059)	0	0
S_HDL_FC	Free cholesterol in small HDL	Lipoprotein HDL	mmol/l	0.116	0.118 (0.015)	0	0
S_HDL_TG	Triglycerides in small HDL	Lipoprotein HDL	mmol/l	0.061	0.061 (0.014)	0	0
VLDL_D	Mean diameter for VLDL particles	Lipoprotein particle sizes	nm	37.36	37.437 (1.476)	0	0

LDL_D	Mean diameter for LDL particles	Lipoprotein particle sizes	nm	23.55	23.560 (0.197)	0	0
HDL_D	Mean diameter for HDL particles	Lipoprotein particle sizes	nm	9.651	9.678 (0.249)	0	0
Serum_C	Serum total cholesterol	Cholesterol	mmol/l	3.529	3.605 (0.794)	0	0
VLDL_C	Total cholesterol in VLDL	Cholesterol	mmol/l	0.612	0.658 (0.247)	0	0
Remnant_C	Remnant cholesterol (non-HDL, non-LDL -cholesterol)	Cholesterol	mmol/l	1.093	1.148 (0.333)	0	0
LDL_C	Total cholesterol in LDL	Cholesterol	mmol/l	1.19	1.234 (0.428)	0	0
HDL_C	Total cholesterol in HDL	Cholesterol	mmol/l	1.184	1.224 (0.325)	0	0
HDL2_C	Total cholesterol in HDL2	Cholesterol	mmol/l	0.736	0.770 (0.308)	0	1
HDL3_C	Total cholesterol in HDL3	Cholesterol	mmol/l	0.452	0.454 (0.030)	0	0
EstC	Esterified cholesterol	Cholesterol	mmol/l	2.508	2.568 (0.580)	1	0
FreeC	Free cholesterol	Cholesterol	mmol/l	1.013	1.039 (0.227)	1	0
Serum_TG	Serum total triglycerides	Glycerides & phospholipid	mmol/l	1.393	1.532 (0.715)	0	0
VLDL_TG	Triglycerides in VLDL	Glycerides & phospholipid	mmol/l	0.962	1.098 (0.646)	0	0
LDL_TG	Triglycerides in LDL	Glycerides & phospholipid	mmol/l	0.167	0.174 (0.054)	0	0
HDL_TG	Triglycerides in HDL	Glycerides & phospholipid	mmol/l	0.136	0.141 (0.034)	0	0
TotPG	Total phosphoglycerides	Glycerides & phospholipid	mmol/l	1.661	1.684 (0.334)	1	0
PC	Phosphatidylcholine and other cholines	Glycerides & phospholipid	mmol/l	1.765	1.790 (0.310)	1	0
SM	Sphingomyelins	Glycerides & phospholipid	mmol/l	0.403	0.409 (0.072)	1	0
TotCho	Total cholines	Glycerides & phospholipid	mmol/l	2.094	2.119 (0.340)	1	0
ApoA1	Apolipoprotein A-I	Apolipoprotein	g/l	1.361	1.381 (0.190)	0	0

ApoB	Apolipoprotein B	Apolipoprotein	g/l	0.762	0.793 (0.186)	0	0
ApoB/ApoA1	Ratio of Apolipoprotein B to Apolipoprotein A-I	Apolipoprotein		0.567	0.582 (0.142)	0	0
TotFA	Total fatty acids	Fatty acids	mmol/l	9.548	9.896 (2.254)	1	0
UnSat	Estimated degree of unsaturation	Fatty acids		1.22	1.224 (0.093)	1	0
DHA	22:6, docosahexaenoic acid	Fatty acids	mmol/l	0.188	0.194 (0.060)	1	0
LA	18:2, linoleic acid	Fatty acids	mmol/l	2.475	2.515 (0.522)	1	0
FAw3	Omega-3 fatty acids	Fatty acids	mmol/l	0.545	0.568 (0.145)	1	0
FAw6	Omega-6 fatty acids	Fatty acids	mmol/l	3.1	3.133 (0.605)	1	0
PUFA	Polyunsaturated fatty acids	Fatty acids	mmol/l	3.661	3.701 (0.696)	1	0
MUFA	Monounsaturated fatty acids; 16:1, 18:1	Fatty acids	mmol/l	2.435	2.593 (0.910)	1	0
SFA	Saturated fatty acids	Fatty acids	mmol/l	3.462	3.601 (0.921)	1	0
DHA_FA		Fatty acids		1.940	1.995 (0.574)	1	0
LA_FA		Fatty acids		25.560	25.810 (4.069)	1	0
FAw3_FA		Fatty acids		5.707	5.832 1.283)	1	0
FAw6_FA		Fatty acids		31.940	32.150 (4.485)	1	0
PUFA_FA		Fatty acids		37.680	37.982 (5.067)	1	0
MUFA_FA		Fatty acids		25.910	25.747 (4.434)	1	0
SFA_FA		Fatty acids		36.290	36.272 (2.563)	1	0
Glc	Glucose	Glycolysis-related	mmol/l	6.226	6.498 (1.682)	1	0
Lac	Lactate	Glycolysis-related	mmol/l	1.356	1.437 (0.438)	0	0
Pyr	Pyruvate	Glycolysis-related	mmol/l	0.063	0.071 (0.036)	21	0
Cit	Citrate	Glycolysis-related	mmol/l	0.136	0.138 (0.023)	0	0
Glol	Glycerol	Glycolysis-related	mmol/l	0.09	0.095 (0.030)	3	0
Ala	Alanine	Amino acids	mmol/l	0.344	0.350 (0.057)	0	0
Gln	Glutamine	Amino acids	mmol/l	0.434	0.436 (0.062)	0	0
Gly	Glycine	Amino acids	mmol/l	0.249	0.256 (0.044)	0	0
His	Histidine	Amino acids	mmol/l	0.052	0.052 (0.009)	1	0
Ile	Isoleucine	Amino acids	mmol/l	0.071	0.073 (0.017)	0	0
Leu	Leucine	Amino acids	mmol/l	0.077	0.079 (0.016)	0	0

Val	Valine	Amino acids	mmol/l	0.179	0.180 (0.033)	0	0
Phe	Phenylalanine	Amino acids	mmol/l	0.08	0.081 (0.012)	0	0
Tyr	Tyrosine	Amino acids	mmol/l	0.053	0.055 (0.013)	1	0
Ace	Acetate	Ketone bodies	mmol/l	0.037	0.039 (0.011)	0	0
AcAce	Acetoacetate	Ketone bodies	mmol/l	0.04	0.047 (0.027)	0	0
bOHBut	3-hydroxybutyrate	Ketone bodies	mmol/l	0.121	0.148 (0.098)	3	0
Crea	Creatinine	Fluid balance	mmol/l	0.063	0.068 (0.022)	1	0
Alb	Albumin	Fluid balance	signal area	0.095	0.095 (0.006)	0	0
Gp	Glycoprotein acetyls, mainly a1-acid glycoprotein	Inflammation	mmol/l	1.383	1.421 (0.282)	0	0

Figure S5.1. Histograms of individual metabolites measured in ET2DS (full names can be found in the Appendix B, Table S5.1



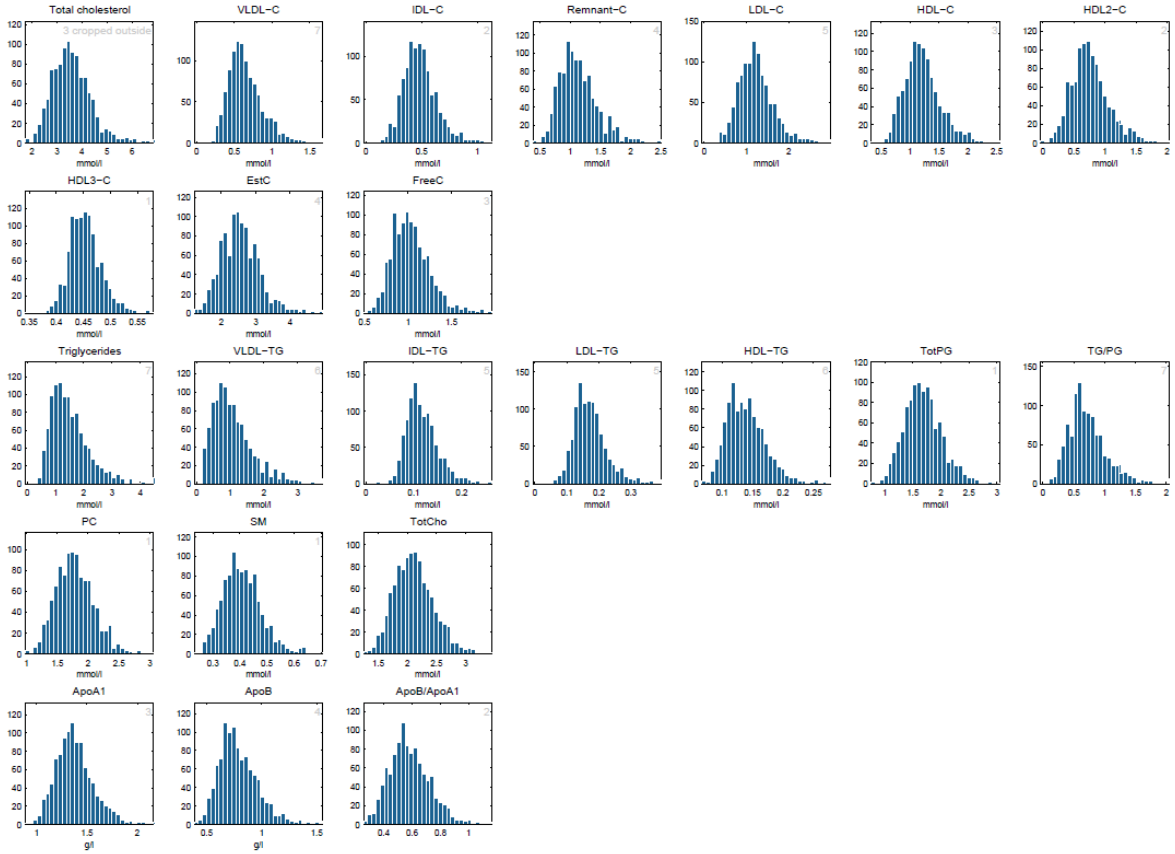


Table S5-2 Summary of associations between each metabolite and baseline eGFR in ET2DS (N=1058). Sorted by P-value (smallest to largest).

Baseline eGFR- age and sex adjusted model.

Metabolite	coeff	CIL	CIU	p	q	group
HDL2_C	4.51	3.43	5.59	8.13E-16	5.39E-14	Lipoprotein particle sizes
HDL_C	4.52	3.43	5.61	1.08E-15	5.39E-14	Lipoprotein particle sizes
M_HDL_CE	4.39	3.31	5.46	3.95E-15	1.47E-13	HDL
XS_VLDL_TG	-4.30	-5.37	-3.24	5.00E-15	1.49E-13	VLDL
M_HDL_C	4.33	3.24	5.41	1.12E-14	2.78E-13	HDL
Gp	-4.21	-5.27	-3.15	1.46E-14	3.11E-13	Inflamation-related
Glol	-4.52	-5.67	-3.37	3.02E-14	5.63E-13	Glycolysis-related
S_VLDL_FC	-4.10	-5.16	-3.03	8.66E-14	1.43E-12	VLDL
S_VLDL_C	-4.08	-5.15	-3.02	1.16E-13	1.73E-12	VLDL
IDL_TG	-4.14	-5.22	-3.05	1.69E-13	2.29E-12	IDL
S_VLDL_PL	-4.02	-5.08	-2.95	2.74E-13	3.40E-12	VLDL
S_VLDL_L	-3.91	-4.97	-2.84	1.13E-12	1.30E-11	VLDL
M_HDL_L	4.02	2.92	5.11	1.32E-12	1.41E-11	HDL
S_VLDL_P	-3.84	-4.91	-2.78	2.71E-12	2.69E-11	VLDL
M_HDL_P	3.94	2.85	5.04	3.42E-12	3.19E-11	HDL
ApoA1	3.93	2.83	5.03	4.71E-12	4.13E-11	Apolipoproteins
M_HDL_FC	3.86	2.76	4.97	9.94E-12	8.23E-11	HDL
M_HDL_PL	3.83	2.72	4.94	1.94E-11	1.52E-10	HDL

Cit	-3.72	-4.83	-2.61	7.48E-11	5.57E-10	Glycolysis-related
S_VLDL_TG	-3.51	-4.58	-2.44	1.98E-10	1.40E-09	VLDL
S_VLDL_CE	-3.51	-4.59	-2.44	2.09E-10	1.41E-09	VLDL
S_HDL_TG	-3.47	-4.54	-2.40	2.83E-10	1.83E-09	HDL
VLDL_C	-3.46	-4.53	-2.38	3.78E-10	2.34E-09	Lipoprotein particle sizes
XS_VLDL_P	-3.53	-4.63	-2.42	5.28E-10	3.15E-09	VLDL
S_LDL_TG	-3.42	-4.50	-2.33	8.42E-10	4.82E-09	LDL
HDL_D	3.42	2.31	4.53	2.03E-09	1.12E-08	Lipoprotein particle sizes
LDL_TG	-3.30	-4.41	-2.20	5.40E-09	2.87E-08	Lipoprotein particle sizes
XS_VLDL_L	-3.32	-4.44	-2.21	5.62E-09	2.89E-08	VLDL
L_LDL_TG	-3.26	-4.37	-2.15	1.08E-08	5.38E-08	LDL
XS_VLDL_FC	-3.24	-4.35	-2.13	1.46E-08	7.04E-08	VLDL
Serum_TG	-3.07	-4.15	-2.00	2.63E-08	1.21E-07	Lipoprotein particle sizes
L_HDL_CE	3.17	2.06	4.29	2.79E-08	1.21E-07	HDL
L_HDL_P	3.18	2.06	4.29	2.82E-08	1.21E-07	HDL
L_HDL_L	3.18	2.06	4.29	2.83E-08	1.21E-07	HDL
L_HDL_PL	3.16	2.05	4.28	3.16E-08	1.31E-07	HDL
L_HDL_C	3.13	2.02	4.24	4.26E-08	1.72E-07	HDL
Phe	-2.99	-4.06	-1.92	5.64E-08	2.21E-07	Amino acids
L_VLDL_CE	-2.97	-4.04	-1.90	6.94E-08	2.65E-07	VLDL
M_LDL_TG	-3.01	-4.11	-1.90	1.11E-07	4.13E-07	LDL
M_VLDL_FC	-2.87	-3.94	-1.79	2.06E-07	7.49E-07	VLDL
VLDL_TG	-2.86	-3.94	-1.78	2.27E-07	8.04E-07	Lipoprotein particle sizes
L_VLDL_C	-2.81	-3.88	-1.73	3.51E-07	1.22E-06	VLDL
L_HDL_FC	2.88	1.77	3.99	4.29E-07	1.43E-06	HDL
M_VLDL_C	-2.79	-3.87	-1.71	4.33E-07	1.43E-06	VLDL
M_VLDL_PL	-2.73	-3.81	-1.65	8.18E-07	2.65E-06	VLDL
Ile	-2.76	-3.86	-1.66	1.03E-06	3.28E-06	Amino acids
M_VLDL_L	-2.68	-3.76	-1.60	1.25E-06	3.87E-06	VLDL
M_VLDL_P	-2.66	-3.74	-1.58	1.56E-06	4.75E-06	VLDL
S_HDL_C	2.62	1.53	3.70	2.45E-06	7.31E-06	HDL
L_VLDL_FC	-2.58	-3.65	-1.50	2.92E-06	8.53E-06	VLDL
M_VLDL_TG	-2.59	-3.67	-1.51	3.03E-06	8.69E-06	VLDL
M_VLDL_CE	-2.57	-3.64	-1.49	3.40E-06	9.56E-06	VLDL
L_VLDL_L	-2.49	-3.57	-1.41	6.79E-06	1.87E-05	VLDL
MUFA	-2.51	-3.60	-1.41	7.34E-06	1.99E-05	Fatty acids
Gly	-2.58	-3.71	-1.45	8.29E-06	2.20E-05	Amino acids
L_VLDL_P	-2.46	-3.54	-1.38	8.68E-06	2.27E-05	VLDL
L_VLDL_PL	-2.44	-3.52	-1.36	1.01E-05	2.60E-05	VLDL
S_HDL_CE	2.40	1.32	3.49	1.50E-05	3.78E-05	HDL
XXL_VLDL_CE	-2.37	-3.45	-1.29	1.82E-05	4.48E-05	VLDL
L_VLDL_TG	-2.37	-3.45	-1.29	1.83E-05	4.48E-05	VLDL
XXL_VLDL_C	-2.35	-3.43	-1.27	2.17E-05	5.21E-05	VLDL
XXL_VLDL_FC	-2.34	-3.42	-1.26	2.34E-05	5.55E-05	VLDL
S_HDL_L	2.35	1.26	3.44	2.52E-05	5.86E-05	HDL
XXL_VLDL_L	-2.32	-3.40	-1.24	2.73E-05	6.25E-05	VLDL

XXL_VLDL_P	-2.31	-3.39	-1.23	2.83E-05	6.40E-05	VLDL
XXL_VLDL_PL	-2.31	-3.39	-1.23	3.03E-05	6.74E-05	VLDL
TotCho	2.42	1.28	3.56	3.45E-05	7.56E-05	Lipoprotein particle sizes
XXL_VLDL_TG	-2.24	-3.32	-1.16	4.97E-05	0.000107	VLDL
S_HDL_P	2.19	1.10	3.29	8.64E-05	0.000184	HDL
XL_VLDL_CE	-2.15	-3.23	-1.06	0.000106	0.000223	VLDL
L_HDL_TG	2.23	1.10	3.35	0.000115	0.000238	HDL
XL_VLDL_C	-2.13	-3.21	-1.04	0.000125	0.000254	VLDL
XL_VLDL_FC	-2.10	-3.18	-1.01	0.000155	0.000311	VLDL
S_HDL_FC	2.10	1.01	3.19	0.000169	0.000335	HDL
S_HDL_PL	2.08	0.99	3.17	0.000198	0.000389	HDL
UnSat	2.05	0.97	3.13	0.000215	0.000411	Fatty acids
Remnant_C	-2.08	-3.18	-0.98	0.000215	0.000411	Lipoprotein particle sizes
XL_VLDL_L	-2.03	-3.12	-0.95	0.000242	0.000456	VLDL
XL_VLDL_P	-2.02	-3.11	-0.94	0.000263	0.000489	VLDL
XL_VLDL_TG	-2.00	-3.09	-0.92	0.000297	0.000546	VLDL
XL_VLDL_PL	-1.97	-3.05	-0.88	0.000394	0.000717	VLDL
HDL_TG	-1.93	-3.02	-0.83	0.000591	0.001054	Lipoprotein particle sizes
XS_VLDL_PL	-1.97	-3.09	-0.85	0.000594	0.001054	VLDL
PC	1.99	0.85	3.13	0.000648	0.001135	Lipoprotein particle sizes
ApoB	-1.87	-2.97	-0.78	0.000781	0.001354	Apolipoproteins
EstC	1.90	0.79	3.01	0.00085	0.001455	Lipoprotein particle sizes
XS_VLDL_C	-1.76	-2.89	-0.62	0.00239	0.004046	VLDL
FAw3	1.62	0.51	2.74	0.00434	0.007266	Fatty acids
L_LDL_FC	1.47	0.36	2.57	0.009187	0.01521	LDL
Glc	1.36	0.28	2.44	0.014006	0.022926	Glycolysis-related
VLDL_D	-1.37	-2.47	-0.28	0.014156	0.022926	Lipoprotein particle sizes
XL_HDL_FC	1.38	0.28	2.49	0.014405	0.023079	HDL
XL_HDL_P	1.37	0.27	2.48	0.014964	0.023559	HDL
Serum_C	1.39	0.27	2.51	0.015021	0.023559	Lipoprotein particle sizes
XL_HDL_L	1.37	0.26	2.48	0.015187	0.023572	HDL
IDL_CE	1.35	0.24	2.46	0.017608	0.026927	IDL
XL_HDL_C	1.34	0.23	2.44	0.01773	0.026927	HDL
M_LDL_FC	1.33	0.23	2.42	0.017891	0.026927	LDL
SM	1.37	0.22	2.53	0.0201	0.029948	Lipoprotein particle sizes
Alb	1.28	0.19	2.38	0.021571	0.031822	Fluid balance
M_HDL_TG	-1.22	-2.32	-0.12	0.029201	0.042656	HDL
IDL_C	1.23	0.12	2.35	0.03033	0.043875	IDL
L_LDL_PL	1.21	0.10	2.32	0.032225	0.046169	LDL

Fully adjusted model.

Metabolite	p	coeff	CIL	CIU	q	group
Crea	0	-17.40	-17.88	-16.92	0	Fluid balance
XS_VLDL_TG	1.07E-13	-3.95	-4.97	-2.92	7.04E-12	VLDL
S_VLDL_C	1.42E-13	-3.92	-4.94	-2.89	7.04E-12	VLDL

S_VLDL_FC	9.35E-13	-3.78	-4.81	-2.75	3.48E-11	VLDL
S_VLDL_PL	2.06E-12	-3.73	-4.76	-2.70	6.14E-11	VLDL
IDL_TG	5.51E-12	-3.77	-4.83	-2.71	1.37E-10	IDL
S_VLDL_L	2.24E-11	-3.56	-4.59	-2.53	4.76E-10	VLDL
S_VLDL_CE	3.29E-11	-3.52	-4.55	-2.49	6.12E-10	VLDL
S_VLDL_P	6.72E-11	-3.47	-4.51	-2.44	1.11E-09	VLDL
XS_VLDL_P	1.55E-10	-3.47	-4.53	-2.42	2.31E-09	VLDL
HDL2_C	2.35E-10	3.52	2.44	4.59	3.19E-09	Lipoprotein particle sizes
HDL_C	3.52E-10	3.53	2.44	4.62	4.36E-09	Lipoprotein particle sizes
XS_VLDL_FC	4.76E-10	-3.38	-4.44	-2.33	5.46E-09	VLDL
Gp	9.94E-10	-3.32	-4.38	-2.26	9.60E-09	Inflamation-related
XS_VLDL_L	9.97E-10	-3.33	-4.38	-2.27	9.60E-09	VLDL
S_LDL_TG	1.03E-09	-3.28	-4.32	-2.23	9.60E-09	LDL
VLDL_C	1.93E-09	-3.20	-4.24	-2.17	1.70E-08	Lipoprotein particle sizes
M_HDL_CE	3.36E-09	3.29	2.21	4.38	2.78E-08	HDL
Cit	4.03E-09	-3.29	-4.37	-2.20	3.16E-08	Glycolysis-related
LDL_TG	8.14E-09	-3.16	-4.23	-2.09	6.07E-08	Lipoprotein particle sizes
M_HDL_C	8.82E-09	3.22	2.13	4.30	6.26E-08	HDL
S_VLDL_TG	1.01E-08	-3.06	-4.10	-2.02	6.81E-08	VLDL
L_LDL_TG	1.93E-08	-3.10	-4.17	-2.02	1.25E-07	LDL
S_HDL_TG	3.95E-08	-2.94	-3.98	-1.90	2.45E-07	HDL
M_LDL_TG	9.10E-08	-2.93	-3.99	-1.86	5.43E-07	LDL
Glol	1.41E-07	-3.29	-4.51	-2.07	8.05E-07	Glycolysis-related
M_HDL_L	3.20E-07	2.88	1.78	3.98	1.76E-06	HDL
M_HDL_P	6.11E-07	2.81	1.71	3.91	3.25E-06	HDL
ApoA1	7.33E-07	2.80	1.69	3.90	3.77E-06	Apolipoproteins
Serum_TG	9.18E-07	-2.63	-3.67	-1.58	4.56E-06	Lipoprotein particle sizes
HDL_D	1.23E-06	2.70	1.62	3.79	5.90E-06	Lipoprotein particle sizes
M_HDL_FC	1.86E-06	2.69	1.59	3.79	8.67E-06	HDL
M_HDL_PL	1.95E-06	2.69	1.59	3.79	8.78E-06	HDL
L_VLDL_CE	2.68E-06	-2.52	-3.56	-1.47	1.17E-05	VLDL
Gly	2.79E-06	-2.62	-3.72	-1.53	1.19E-05	Amino acids
M_VLDL_C	3.72E-06	-2.47	-3.51	-1.43	1.54E-05	VLDL
M_VLDL_FC	4.28E-06	-2.46	-3.50	-1.41	1.73E-05	VLDL
L_HDL_CE	6.78E-06	2.51	1.42	3.59	2.66E-05	HDL
VLDL_TG	7.07E-06	-2.41	-3.46	-1.36	2.70E-05	Lipoprotein particle sizes
M_VLDL_CE	8.25E-06	-2.37	-3.41	-1.33	3.04E-05	VLDL
MUFA	8.37E-06	-2.39	-3.44	-1.34	3.04E-05	Fatty acids
L_HDL_L	9.71E-06	2.47	1.38	3.56	3.38E-05	HDL
L_HDL_P	9.75E-06	2.47	1.38	3.57	3.38E-05	HDL
L_HDL_C	1.04E-05	2.45	1.37	3.54	3.52E-05	HDL

M_VLDL_PL	1.11E-05	-2.35	-3.40	-1.31	3.69E-05	VLDL
L_HDL_PL	1.22E-05	2.44	1.35	3.53	3.94E-05	HDL
L_VLDL_C	1.29E-05	-2.34	-3.39	-1.29	4.08E-05	VLDL
XS_VLDL_PL	1.84E-05	-2.34	-3.40	-1.27	5.71E-05	VLDL
M_VLDL_L	2.14E-05	-2.28	-3.32	-1.23	6.50E-05	VLDL
M_VLDL_P	2.76E-05	-2.25	-3.30	-1.20	8.22E-05	VLDL
Remnant_C	2.97E-05	-2.25	-3.30	-1.20	8.69E-05	Lipoprotein particle sizes
Phe	3.76E-05	-2.21	-3.26	-1.16	0.000108	Amino acids
Ile	5.34E-05	-2.24	-3.33	-1.16	0.00015	Amino acids
M_VLDL_TG	6.17E-05	-2.15	-3.20	-1.10	0.00017	VLDL
L_HDL_FC	8.02E-05	2.19	1.10	3.28	0.000217	HDL
L_VLDL_FC	9.95E-05	-2.09	-3.14	-1.04	0.000265	VLDL
L_VLDL_L	0.000184	-2.01	-3.06	-0.96	0.00048	VLDL
ApoB	0.000195	-2.00	-3.04	-0.95	0.000501	Apolipoproteins
L_VLDL_PL	0.000219	-1.99	-3.04	-0.94	0.000554	VLDL
L_VLDL_P	0.000229	-1.98	-3.03	-0.93	0.000569	VLDL
XS_VLDL_C	0.000365	-1.96	-3.03	-0.88	0.000892	VLDL
XXL_VLDL_CE	0.000383	-1.92	-2.97	-0.86	0.000922	VLDL
L_VLDL_TG	0.00045	-1.89	-2.94	-0.83	0.001065	VLDL
XXL_VLDL_C	0.000497	-1.88	-2.94	-0.83	0.001156	VLDL
XXL_VLDL_PL	0.000622	-1.85	-2.91	-0.79	0.001426	VLDL
XXL_VLDL_L	0.000632	-1.85	-2.91	-0.79	0.001426	VLDL
XXL_VLDL_P	0.000653	-1.84	-2.90	-0.79	0.001448	VLDL
XXL_VLDL_FC	0.000661	-1.84	-2.90	-0.78	0.001448	VLDL
HDL_TG	0.000882	-1.77	-2.82	-0.73	0.001905	Lipoprotein particle sizes
XXL_VLDL_TG	0.001093	-1.77	-2.82	-0.71	0.002326	VLDL
S_HDL_C	0.00213	1.66	0.60	2.72	0.00447	HDL
XL_VLDL_CE	0.002366	-1.64	-2.69	-0.58	0.004897	VLDL
UnSat	0.002559	1.63	0.57	2.69	0.005223	Fatty acids
L_HDL_TG	0.002626	1.68	0.59	2.78	0.005282	HDL
XL_VLDL_C	0.002659	-1.62	-2.68	-0.56	0.005282	VLDL
XL_VLDL_FC	0.003184	-1.59	-2.65	-0.54	0.006243	VLDL
XL_VLDL_L	0.004081	-1.55	-2.60	-0.49	0.007896	VLDL
XL_VLDL_P	0.004315	-1.54	-2.59	-0.48	0.008242	VLDL
XL_VLDL_TG	0.004558	-1.53	-2.58	-0.47	0.008597	VLDL
S_HDL_CE	0.005204	1.51	0.45	2.56	0.009693	HDL
XL_VLDL_PL	0.005844	-1.49	-2.54	-0.43	0.01075	VLDL
S_HDL_L	0.00609	1.49	0.43	2.55	0.011065	HDL
Glc	0.006751	1.74	0.48	3.00	0.012119	Glycolysis-related
S_HDL_PL	0.006923	1.46	0.40	2.52	0.01228	HDL
S_HDL_FC	0.009074	1.41	0.35	2.47	0.015907	HDL
M_HDL_TG	0.00979	-1.38	-2.43	-0.33	0.016962	HDL
S_HDL_P	0.012062	1.36	0.30	2.42	0.020657	HDL

TotFA	0.027866	-1.19	-2.25	-0.13	0.047183	Fatty acids
XL_HDL_FC	0.029395	1.17	0.12	2.23	0.049212	HDL

Table S5-3. Summary of significant associations between each metabolite and baseline eGFR with additional adjustment for lipid-lowering medication use in ET2DS (N=1058). Sorted by P-value (smallest to largest).

Metabolite	p	coeff	StandardError	CIL	CIU	fdr
Crea	0.00E+00	-17.40	0.24	-17.88	-16.92	0.00E+00
S_VLDL_C	3.19E-14	-4.06	0.53	-5.10	-3.03	2.38E-12
XS_VLDL_TG	7.01E-14	-3.98	0.52	-5.01	-2.95	3.48E-12
S_VLDL_FC	4.16E-13	-3.85	0.52	-4.87	-2.82	1.55E-11
IDL_TG	3.27E-12	-3.81	0.54	-4.87	-2.75	8.13E-11
S_VLDL_PL	9.83E-13	-3.79	0.52	-4.82	-2.76	2.93E-11
S_VLDL_CE	5.75E-12	-3.73	0.54	-4.78	-2.68	1.22E-10
S_VLDL_L	1.05E-11	-3.62	0.53	-4.66	-2.59	1.96E-10
XS_VLDL_P	4.90E-11	-3.60	0.54	-4.66	-2.54	7.30E-10
S_VLDL_P	3.35E-11	-3.53	0.53	-4.57	-2.50	5.55E-10
XS_VLDL_L	3.07E-10	-3.47	0.55	-4.54	-2.40	3.81E-09
XS_VLDL_FC	2.36E-10	-3.47	0.54	-4.53	-2.40	3.19E-09
Gp	6.13E-10	-3.36	0.54	-4.42	-2.31	5.38E-09
S_LDL_TG	4.24E-10	-3.36	0.53	-4.41	-2.32	4.52E-09
VLDL_C	4.89E-10	-3.36	0.53	-4.41	-2.31	4.86E-09
Glol	8.48E-08	-3.35	0.62	-4.56	-2.13	4.86E-07
Cit	3.68E-09	-3.29	0.55	-4.38	-2.21	2.89E-08
LDL_TG	4.00E-09	-3.24	0.55	-4.31	-2.17	2.98E-08
L_LDL_TG	1.04E-08	-3.17	0.55	-4.25	-2.09	6.77E-08
S_VLDL_TG	6.23E-09	-3.11	0.53	-4.15	-2.07	4.42E-08
M_LDL_TG	4.97E-08	-3.00	0.55	-4.07	-1.93	2.96E-07
S_HDL_TG	2.89E-08	-2.97	0.53	-4.02	-1.93	1.80E-07
Serum_TG	5.29E-07	-2.69	0.53	-3.73	-1.64	2.82E-06
Gly	2.95E-06	-2.62	0.56	-3.72	-1.53	1.19E-05
L_VLDL_CE	1.34E-06	-2.60	0.53	-3.65	-1.55	6.44E-06
M_VLDL_C	1.81E-06	-2.56	0.53	-3.60	-1.51	8.19E-06
M_VLDL_FC	2.56E-06	-2.52	0.53	-3.56	-1.47	1.06E-05
M_VLDL_CE	3.06E-06	-2.52	0.54	-3.57	-1.46	1.20E-05
Remnant_C	6.18E-06	-2.51	0.55	-3.60	-1.43	2.25E-05
XS_VLDL_PL	7.97E-06	-2.47	0.55	-3.55	-1.39	2.70E-05
VLDL_TG	4.46E-06	-2.47	0.53	-3.52	-1.42	1.70E-05
MUFA	6.14E-06	-2.43	0.53	-3.48	-1.38	2.25E-05
L_VLDL_C	6.94E-06	-2.42	0.54	-3.47	-1.37	2.40E-05
M_VLDL_PL	6.82E-06	-2.41	0.53	-3.46	-1.37	2.40E-05
M_VLDL_L	1.34E-05	-2.33	0.53	-3.38	-1.29	4.31E-05
M_VLDL_P	1.76E-05	-2.30	0.53	-3.35	-1.26	5.15E-05
Ile	4.22E-05	-2.28	0.55	-3.37	-1.19	1.19E-04

ApoB	5.95E-05	-2.20	0.55	-3.27	-1.13	1.58E-04
M_VLDL_TG	4.24E-05	-2.20	0.54	-3.25	-1.15	1.19E-04
Phe	4.62E-05	-2.18	0.53	-3.23	-1.14	1.27E-04
L_VLDL_FC	5.36E-05	-2.17	0.54	-3.22	-1.12	1.45E-04
XS_VLDL_C	1.93E-04	-2.09	0.56	-3.18	-0.99	4.72E-04
L_VLDL_L	1.12E-04	-2.08	0.54	-3.13	-1.03	2.87E-04
L_VLDL_PL	1.30E-04	-2.06	0.54	-3.11	-1.01	3.28E-04
L_VLDL_P	1.41E-04	-2.05	0.54	-3.10	-1.00	3.49E-04
XXL_VLDL_CE	2.15E-04	-2.02	0.54	-3.08	-0.95	5.17E-04
XXL_VLDL_C	2.91E-04	-1.97	0.54	-3.04	-0.91	6.77E-04
L_VLDL_TG	2.86E-04	-1.95	0.54	-3.00	-0.90	6.77E-04
XXL_VLDL_PL	3.70E-04	-1.94	0.54	-3.00	-0.87	8.48E-04
XXL_VLDL_L	3.78E-04	-1.94	0.54	-3.00	-0.87	8.53E-04
XXL_VLDL_P	3.91E-04	-1.93	0.54	-3.00	-0.87	8.70E-04
XXL_VLDL_FC	4.07E-04	-1.92	0.54	-2.99	-0.86	8.91E-04
XXL_VLDL_TG	6.91E-04	-1.85	0.54	-2.91	-0.78	1.47E-03
HDL_TG	5.82E-04	-1.84	0.53	-2.89	-0.79	1.26E-03
XL_VLDL_CE	1.54E-03	-1.72	0.54	-2.78	-0.66	3.24E-03
XL_VLDL_C	1.72E-03	-1.70	0.54	-2.76	-0.64	3.57E-03
XL_VLDL_FC	2.03E-03	-1.68	0.54	-2.74	-0.61	4.08E-03
XL_VLDL_L	2.79E-03	-1.62	0.54	-2.68	-0.56	5.48E-03
XL_VLDL_P	2.97E-03	-1.61	0.54	-2.67	-0.55	5.76E-03
XL_VLDL_TG	3.19E-03	-1.60	0.54	-2.66	-0.54	6.09E-03
XL_VLDL_PL	3.96E-03	-1.56	0.54	-2.62	-0.50	7.37E-03
M_HDL_TG	6.89E-03	-1.45	0.54	-2.50	-0.40	1.21E-02
TotFA	1.85E-02	-1.28	0.54	-2.34	-0.21	3.14E-02
S_HDL_P	1.07E-02	1.40	0.55	0.33	2.47	1.85E-02
S_HDL_CE	5.35E-03	1.50	0.54	0.45	2.56	9.60E-03
S_HDL_FC	6.20E-03	1.52	0.55	0.43	2.60	1.10E-02
S_HDL_L	5.26E-03	1.53	0.55	0.46	2.60	9.56E-03
S_HDL_PL	5.23E-03	1.54	0.55	0.46	2.62	9.56E-03
Glc	1.13E-02	1.62	0.64	0.37	2.88	1.93E-02
L_HDL_TG	3.33E-03	1.64	0.56	0.55	2.74	6.27E-03
S_HDL_C	2.09E-03	1.67	0.54	0.61	2.73	4.16E-03
UnSat	1.95E-03	1.68	0.54	0.62	2.75	3.98E-03
L_HDL_FC	1.11E-04	2.15	0.55	1.06	3.23	2.87E-04
L_HDL_PL	1.74E-05	2.40	0.56	1.31	3.49	5.15E-05
L_HDL_C	1.46E-05	2.42	0.55	1.33	3.50	4.43E-05
L_HDL_L	1.38E-05	2.43	0.56	1.34	3.53	4.31E-05
L_HDL_P	1.39E-05	2.43	0.56	1.34	3.53	4.31E-05
L_HDL_CE	9.52E-06	2.47	0.55	1.38	3.56	3.15E-05
M_HDL_PL	2.50E-06	2.67	0.56	1.56	3.78	1.06E-05
M_HDL_FC	2.06E-06	2.69	0.56	1.58	3.79	9.02E-06
HDL_D	1.69E-06	2.69	0.56	1.59	3.79	7.86E-06

ApoA1	1.34E-06	2.74	0.56	1.64	3.85	6.44E-06
M_HDL_P	7.76E-07	2.79	0.56	1.69	3.89	3.99E-06
M_HDL_L	3.98E-07	2.86	0.56	1.76	3.97	2.19E-06
M_HDL_C	9.58E-09	3.22	0.56	2.13	4.31	6.49E-08
M_HDL_CE	3.59E-09	3.30	0.55	2.21	4.39	2.89E-08
HDL2_C	3.42E-10	3.49	0.55	2.41	4.57	3.92E-09
HDL_C	5.53E-10	3.50	0.56	2.40	4.59	5.15E-09

Table S5-4. Summary of associations between each metabolite and baseline CKD in ET2DS (N=1058). Sorted by P-value (smallest to largest).

Age and sex adjusted model

metabolite	or	CIL	CIU	p	fdr	group
HDL_C	0.55	0.46	0.66	1.79E-11	9.28E-10	Lipoprotein particle sizes
XS_VLDL_TG	1.85	1.55	2.22	1.87E-11	9.28E-10	VLDL
M_HDL_CE	0.59	0.50	0.69	6.79E-11	2.53E-09	HDL
HDL2_C	0.57	0.48	0.68	8.72E-11	2.60E-09	Lipoprotein particle sizes
M_HDL_C	0.59	0.50	0.69	1.05E-10	2.61E-09	HDL
Gp	1.67	1.43	1.96	1.83E-10	3.89E-09	Inflamation-related
S_VLDL_FC	1.76	1.48	2.10	3.69E-10	6.88E-09	VLDL
S_VLDL_PL	1.71	1.44	2.03	9.56E-10	1.58E-08	VLDL
S_VLDL_C	1.71	1.44	2.04	1.22E-09	1.83E-08	VLDL
ApoA1	0.58	0.49	0.69	1.40E-09	1.90E-08	Apolipoproteins
M_HDL_L	0.61	0.51	0.71	2.13E-09	2.65E-08	HDL
S_VLDL_L	1.68	1.42	1.99	2.42E-09	2.77E-08	VLDL
IDL_TG	1.66	1.41	1.97	2.73E-09	2.91E-08	IDL
Glol	1.69	1.42	2.02	3.40E-09	3.38E-08	Glycolysis-related
M_HDL_P	0.61	0.52	0.72	4.02E-09	3.69E-08	HDL
S_VLDL_P	1.66	1.41	1.97	4.21E-09	3.69E-08	VLDL
M_HDL_FC	0.63	0.53	0.73	7.20E-09	5.96E-08	HDL
Cit	1.63	1.38	1.94	9.20E-09	7.21E-08	Glycolysis-related
M_HDL_PL	0.62	0.52	0.73	1.39E-08	1.04E-07	HDL
S_VLDL_TG	1.59	1.35	1.88	6.89E-08	4.89E-07	VLDL
HDL_D	0.61	0.51	0.73	1.27E-07	8.59E-07	Lipoprotein particle sizes
XS_VLDL_P	1.57	1.33	1.87	2.07E-07	1.34E-06	VLDL
VLDL_C	1.53	1.30	1.80	2.37E-07	1.47E-06	Lipoprotein particle sizes
S_VLDL_CE	1.55	1.31	1.85	4.81E-07	2.87E-06	VLDL
S_HDL_TG	1.53	1.30	1.81	5.23E-07	3.00E-06	HDL
S_LDL_TG	1.51	1.28	1.79	1.35E-06	7.37E-06	LDL
XS_VLDL_L	1.52	1.28	1.80	1.38E-06	7.37E-06	VLDL
L_HDL_L	0.68	0.58	0.80	1.58E-06	7.94E-06	HDL
L_HDL_P	0.68	0.58	0.80	1.60E-06	7.94E-06	HDL
L_HDL_CE	0.68	0.58	0.80	1.69E-06	8.13E-06	HDL
L_HDL_PL	0.69	0.59	0.80	1.84E-06	8.57E-06	HDL
L_HDL_C	0.68	0.59	0.80	2.01E-06	9.06E-06	HDL
Serum_TG	1.47	1.25	1.72	2.39E-06	1.05E-05	Lipoprotein particle sizes

XS_VLDL_FC	1.51	1.27	1.80	3.15E-06	1.34E-05	VLDL
L_HDL_FC	0.70	0.60	0.82	6.54E-06	2.71E-05	HDL
LDL_TG	1.45	1.23	1.71	8.70E-06	3.50E-05	Lipoprotein particle sizes
M_VLDL_FC	1.45	1.23	1.72	9.67E-06	3.79E-05	VLDL
VLDL_TG	1.44	1.23	1.70	1.07E-05	4.08E-05	Lipoprotein particle sizes
L_LDL_TG	1.45	1.23	1.71	1.16E-05	4.32E-05	LDL
M_VLDL_C	1.41	1.21	1.66	2.06E-05	7.45E-05	VLDL
L_VLDL_CE	1.48	1.24	1.78	2.10E-05	7.45E-05	VLDL
M_VLDL_PL	1.41	1.20	1.66	2.63E-05	9.12E-05	VLDL
M_VLDL_L	1.40	1.20	1.65	3.73E-05	0.000126	VLDL
M_VLDL_P	1.40	1.19	1.65	4.44E-05	0.000145	VLDL
L_VLDL_C	1.45	1.22	1.74	4.49E-05	0.000145	VLDL
S_HDL_C	0.74	0.63	0.85	5.35E-05	0.00017	HDL
TotCho	0.71	0.60	0.84	5.47E-05	0.00017	Lipoprotein particle sizes
M_VLDL_TG	1.39	1.18	1.64	7.23E-05	0.00022	VLDL
M_LDL_TG	1.39	1.18	1.64	8.65E-05	0.000258	LDL
Phe	1.36	1.16	1.58	9.20E-05	0.000269	Amino acids
Ile	1.37	1.17	1.62	0.000114	0.000327	Amino acids
M_VLDL_CE	1.36	1.16	1.59	0.00013	0.000365	VLDL
S_HDL_CE	0.75	0.65	0.87	0.000161	0.000443	HDL
L_VLDL_FC	1.43	1.19	1.74	0.000184	0.000498	VLDL
L_HDL_TG	0.75	0.64	0.88	0.000263	0.000701	HDL
L_VLDL_L	1.37	1.16	1.63	0.000278	0.000717	VLDL
MUFA	1.35	1.15	1.58	0.000279	0.000717	Fatty acids
L_VLDL_P	1.37	1.16	1.63	0.00033	0.000833	VLDL
L_VLDL_PL	1.36	1.15	1.62	0.000344	0.000855	VLDL
XXL_VLDL_FC	1.36	1.15	1.62	0.000424	0.001036	VLDL
S_HDL_L	0.76	0.65	0.89	0.000471	0.001131	HDL
XXL_VLDL_PL	1.35	1.14	1.61	0.000543	0.001284	VLDL
XXL_VLDL_CE	1.36	1.15	1.62	0.000565	0.001296	VLDL
L_VLDL_TG	1.35	1.14	1.61	0.000571	0.001296	VLDL
XXL_VLDL_C	1.36	1.14	1.62	0.000574	0.001296	VLDL
Gly	1.32	1.12	1.54	0.000612	0.001361	Amino acids
EstC	0.76	0.64	0.89	0.000624	0.001368	Lipoprotein particle sizes
XXL_VLDL_L	1.35	1.14	1.61	0.000668	0.001442	VLDL
XXL_VLDL_P	1.35	1.14	1.61	0.000697	0.001483	VLDL
XL_VLDL_CE	1.35	1.14	1.61	0.000793	0.001664	VLDL
XL_VLDL_C	1.34	1.14	1.61	0.000822	0.001701	VLDL
XL_VLDL_FC	1.34	1.13	1.60	0.000848	0.001732	VLDL
XXL_VLDL_TG	1.33	1.13	1.59	0.001035	0.002083	VLDL
S_HDL_P	0.78	0.67	0.90	0.001062	0.00211	HDL
PC	0.76	0.65	0.90	0.001122	0.0022	Lipoprotein particle sizes
XL_VLDL_L	1.32	1.12	1.58	0.001381	0.002672	VLDL
XL_VLDL_P	1.32	1.12	1.57	0.001489	0.002844	VLDL
XL_VLDL_TG	1.31	1.11	1.57	0.001716	0.00322	VLDL
XL_VLDL_PL	1.31	1.11	1.57	0.001729	0.00322	VLDL

S_HDL_PL	0.80	0.68	0.93	0.003582	0.006588	HDL
S_HDL_FC	0.80	0.69	0.93	0.003977	0.007226	HDL
Remnant_C	1.25	1.07	1.47	0.004807	0.008629	Lipoprotein particle sizes
XS_VLDL_PL	1.26	1.07	1.49	0.00651	0.011548	VLDL
L_LDL_FC	0.81	0.69	0.94	0.00696	0.0122	LDL
IDL_CE	0.81	0.69	0.94	0.007136	0.012364	IDL
Serum_C	0.81	0.68	0.94	0.00832	0.01425	Lipoprotein particle sizes
XL_HDL_P	0.81	0.70	0.95	0.00879	0.014884	HDL
XL_HDL_L	0.81	0.70	0.95	0.009038	0.015131	HDL
SM	0.80	0.68	0.95	0.009156	0.015158	Lipoprotein particle sizes
ApoB	1.23	1.05	1.43	0.009419	0.015422	Apolipoproteins
XL_HDL_C	0.81	0.70	0.95	0.009543	0.015456	HDL
XL_HDL_FC	0.82	0.70	0.95	0.010037	0.016081	HDL
HDL_TG	1.22	1.05	1.43	0.010557	0.016734	Lipoprotein particle sizes
UnSat	0.82	0.70	0.95	0.011139	0.017471	Fatty acids
M_LDL_FC	0.82	0.70	0.96	0.011567	0.017953	LDL
IDL_C	0.82	0.70	0.96	0.012533	0.019252	IDL
XL_HDL_CE	0.83	0.71	0.97	0.017036	0.025901	HDL
L_LDL_PL	0.82	0.70	0.97	0.017252	0.025966	LDL
XL_HDL_PL	0.84	0.72	0.98	0.024633	0.036525	HDL
XS_VLDL_C	1.22	1.03	1.45	0.024758	0.036525	VLDL
S_LDL_PL	0.84	0.72	0.98	0.026803	0.038935	LDL
VLDL_D	1.19	1.02	1.39	0.026915	0.038935	Lipoprotein particle sizes

Fully adjusted model.

metabolite	p	or	CIL	CIU	fdr	group
Crea	2.59E-34	112.82	55.50	253.90	3.86E-32	Fluid balance
XS_VLDL_TG	9.58E-10	1.80	1.50	2.19	7.14E-08	VLDL
S_VLDL_C	3.91E-09	1.75	1.46	2.11	1.55E-07	VLDL
S_VLDL_FC	4.16E-09	1.75	1.46	2.11	1.55E-07	VLDL
S_VLDL_PL	8.25E-09	1.71	1.43	2.05	2.46E-07	VLDL
S_VLDL_L	3.31E-08	1.67	1.39	2.00	8.22E-07	VLDL
S_VLDL_P	6.36E-08	1.64	1.38	1.98	1.35E-06	VLDL
IDL_TG	1.38E-07	1.62	1.36	1.94	2.57E-06	IDL
HDL_C	1.64E-07	0.61	0.51	0.73	2.72E-06	Lipoprotein particle sizes
S_VLDL_CE	4.46E-07	1.60	1.34	1.93	6.64E-06	VLDL
XS_VLDL_P	6.81E-07	1.58	1.32	1.90	9.22E-06	VLDL
HDL2_C	7.61E-07	0.64	0.53	0.76	9.45E-06	Lipoprotein particle sizes
Gp	8.69E-07	1.53	1.30	1.82	9.96E-06	Inflamation-related
Cit	1.11E-06	1.57	1.31	1.89	1.18E-05	Glycolysis-related
S_VLDL_TG	1.32E-06	1.55	1.30	1.86	1.23E-05	VLDL
VLDL_C	1.32E-06	1.54	1.29	1.84	1.23E-05	Lipoprotein particle sizes
M_HDL_CE	1.59E-06	0.66	0.55	0.78	1.39E-05	HDL
XS_VLDL_FC	2.29E-06	1.55	1.30	1.87	1.90E-05	VLDL
M_HDL_C	2.53E-06	0.66	0.56	0.79	1.98E-05	HDL
XS_VLDL_L	2.93E-06	1.54	1.29	1.84	2.18E-05	VLDL

S_LDL_TG	5.60E-06	1.51	1.27	1.80	3.97E-05	LDL
ApoA1	1.10E-05	0.66	0.55	0.79	7.43E-05	Apolipoproteins
S_HDL_TG	1.35E-05	1.48	1.24	1.78	8.75E-05	HDL
M_HDL_L	2.61E-05	0.69	0.58	0.82	1.62E-04	HDL
HDL_D	3.27E-05	0.67	0.55	0.81	1.93E-04	Lipoprotein particle sizes
Serum_TG	3.38E-05	1.44	1.21	1.71	1.93E-04	Lipoprotein particle sizes
LDL_TG	3.82E-05	1.45	1.21	1.73	2.11E-04	Lipoprotein particle sizes
M_HDL_P	4.04E-05	0.69	0.58	0.82	2.15E-04	HDL
L_LDL_TG	6.04E-05	1.44	1.20	1.72	3.11E-04	LDL
M_VLDL_FC	6.79E-05	1.44	1.21	1.72	3.37E-04	VLDL
M_HDL_FC	7.77E-05	0.71	0.60	0.84	3.73E-04	HDL
M_VLDL_C	8.40E-05	1.41	1.19	1.68	3.91E-04	VLDL
M_HDL_PL	8.76E-05	0.70	0.59	0.84	3.96E-04	HDL
VLDL_TG	1.10E-04	1.41	1.19	1.69	4.82E-04	Lipoprotein particle sizes
L_VLDL_CE	1.23E-04	1.47	1.21	1.79	5.23E-04	VLDL
M_VLDL_PL	1.56E-04	1.40	1.18	1.66	6.47E-04	VLDL
L_HDL_CE	2.05E-04	0.73	0.61	0.86	8.22E-04	HDL
M_LDL_TG	2.10E-04	1.39	1.17	1.66	8.22E-04	LDL
M_VLDL_CE	2.19E-04	1.37	1.16	1.63	8.30E-04	VLDL
MUFA	2.39E-04	1.38	1.16	1.64	8.30E-04	Fatty acids
L_HDL_L	2.40E-04	0.73	0.61	0.86	8.30E-04	HDL
L_HDL_C	5.74E-04	0.73	0.61	0.86	8.30E-04	HDL
M_VLDL_L	2.40E-03	1.38	1.17	1.65	8.30E-04	VLDL
L_HDL_P	3.44E-03	0.73	0.61	0.86	8.30E-04	HDL
L_HDL_PL	6.61E-03	0.73	0.62	0.87	9.31E-04	HDL
L_VLDL_C	6.95E-03	1.43	1.18	1.74	9.31E-04	VLDL
M_VLDL_P	7.39E-03	1.38	1.16	1.64	9.31E-04	VLDL
M_VLDL_TG	8.91E-03	1.37	1.15	1.63	1.57E-03	VLDL
Glol	9.45E-03	1.42	1.16	1.73	1.75E-03	Glycolysis-related
L_HDL_FC	1.03E-02	0.75	0.63	0.88	1.85E-03	HDL
L_VLDL_FC	1.14E-02	1.40	1.15	1.73	3.06E-03	VLDL
XS_VLDL_PL	1.19E-02	1.32	1.11	1.58	4.74E-03	VLDL
L_VLDL_L	1.48E-02	1.34	1.12	1.62	4.91E-03	VLDL
L_VLDL_PL	1.55E-02	1.34	1.12	1.61	5.35E-03	VLDL
L_VLDL_P	1.56E-02	1.34	1.12	1.61	5.58E-03	VLDL
Gly	1.64E-02	1.30	1.10	1.54	6.39E-03	Amino acids
XXL_VLDL_CE	1.96E-02	1.33	1.11	1.62	6.49E-03	VLDL
Ile	2.27E-02	1.31	1.10	1.57	6.49E-03	Amino acids
Remnant_C	2.50E-02	1.30	1.10	1.54	6.49E-03	Lipoprotein particle sizes
XXL_VLDL_FC	2.53E-02	1.33	1.11	1.61	6.59E-03	VLDL
XXL_VLDL_PL	0.002806	1.33	1.11	1.60	6.75E-03	VLDL
XXL_VLDL_C	0.002807	1.33	1.11	1.61	6.75E-03	VLDL
L_VLDL_TG	0.003436	1.32	1.10	1.59	8.13E-03	VLDL
XXL_VLDL_L	0.003518	1.32	1.10	1.60	8.19E-03	VLDL

XXL_VLDL_P	0.003666	1.32	1.10	1.59	8.40E-03	VLDL
XXL_VLDL_TG	0.005295	1.30	1.09	1.58	1.20E-02	VLDL
XL_VLDL_CE	0.005745	1.30	1.08	1.57	1.28E-02	VLDL
XL_VLDL_C	0.005883	1.30	1.08	1.57	1.29E-02	VLDL
XL_VLDL_FC	0.005956	1.30	1.08	1.57	1.29E-02	VLDL
ApoB	0.006613	1.26	1.07	1.48	1.41E-02	Apolipoproteins
L_HDL_TG	0.006947	0.79	0.67	0.94	1.46E-02	HDL
S_HDL_C	0.007392	0.81	0.69	0.94	1.53E-02	HDL
Phe	0.008138	1.25	1.06	1.47	1.66E-02	Amino acids
XL_VLDL_L	0.008906	1.28	1.07	1.54	1.79E-02	VLDL
XL_VLDL_P	0.009446	1.28	1.07	1.54	1.88E-02	VLDL
XL_VLDL_PL	0.010335	1.27	1.06	1.54	2.02E-02	VLDL
XL_VLDL_TG	0.010442	1.27	1.06	1.54	2.02E-02	VLDL
HDL_TG	0.011367	1.24	1.05	1.46	2.17E-02	Lipoprotein particle sizes
S_HDL_CE	0.011928	0.82	0.70	0.96	2.25E-02	HDL
XL_HDL_C	0.014809	0.81	0.69	0.96	2.76E-02	HDL
XL_HDL_P	0.015454	0.81	0.69	0.96	2.83E-02	HDL
XL_HDL_L	0.015554	0.81	0.69	0.96	2.83E-02	HDL
XL_HDL_FC	0.016369	0.82	0.69	0.96	2.94E-02	HDL
XS_VLDL_C	0.019646	1.23	1.04	1.48	3.48E-02	VLDL
XL_HDL_CE	0.022694	0.82	0.70	0.97	3.98E-02	HDL
TotCho	0.024954	0.82	0.68	0.97	4.32E-02	Lipoprotein particle sizes
S_HDL_L	0.025264	0.83	0.71	0.98	4.33E-02	HDL

Table S5-5. Summary of associations between each metabolite and baseline uACR in ET2DS (N=1058). Sorted by P-value (smallest to largest).

Age and sex-adjusted model

metabolite	coeff	StandardError	CIL	CIU	p	fdr	group
Gp	0.15	0.03	0.10	0.21	1.29E-08	1.92E-06	Inflammation-related
Crea	0.15	0.03	0.09	0.21	2.05E-07	1.53E-05	Fluid balance
M_HDL_FC	-0.13	0.03	-0.19	-0.08	1.64E-06	8.13E-05	HDL
HDL2_C	-0.13	0.03	-0.18	-0.07	3.55E-06	1.32E-04	Lipoprotein particle sizes
M_HDL_C	-0.13	0.03	-0.18	-0.07	4.51E-06	1.34E-04	HDL
M_HDL_CE	-0.12	0.03	-0.18	-0.07	7.04E-06	1.75E-04	HDL
Tyr	-0.12	0.03	-0.17	-0.07	9.46E-06	1.83E-04	Amino acids
HDL_C	-0.12	0.03	-0.18	-0.07	9.84E-06	1.83E-04	Lipoprotein particle sizes
M_HDL_L	-0.12	0.03	-0.17	-0.07	1.41E-05	2.34E-04	HDL
M_HDL_P	-0.12	0.03	-0.17	-0.06	2.08E-05	3.06E-04	HDL

M_HDL_PL	-0.12	0.03	-0.17	-0.06	2.26E-05	3.06E-04	HDL
Glol	0.12	0.03	0.06	0.18	5.00E-05	6.21E-04	Glycolysis-related
L_HDL_PL	-0.10	0.03	-0.15	-0.05	3.47E-04	3.62E-03	HDL
L_HDL_L	-0.10	0.03	-0.15	-0.04	3.77E-04	3.62E-03	HDL
L_HDL_P	-0.10	0.03	-0.15	-0.04	3.81E-04	3.62E-03	HDL
L_HDL_FC	-0.10	0.03	-0.15	-0.04	3.88E-04	3.62E-03	HDL
L_HDL_C	-0.10	0.03	-0.15	-0.04	4.67E-04	4.09E-03	HDL
L_HDL_CE	-0.10	0.03	-0.15	-0.04	5.06E-04	4.18E-03	HDL
ApoA1	-0.09	0.03	-0.15	-0.04	9.37E-04	7.35E-03	Apolipoproteins
HDL_D	-0.09	0.03	-0.14	-0.03	1.87E-03	1.40E-02	Lipoprotein particle sizes
IDL_TG	0.08	0.03	0.03	0.14	2.19E-03	1.56E-02	IDL
L_HDL_TG	-0.09	0.03	-0.14	-0.03	2.40E-03	1.63E-02	HDL
Serum_TG	0.08	0.03	0.03	0.13	2.77E-03	1.79E-02	Lipoprotein particle sizes
VLDL_C	0.08	0.03	0.03	0.13	3.93E-03	2.39E-02	Lipoprotein particle sizes
S_HDL_FC	-0.08	0.03	-0.13	-0.02	4.02E-03	2.39E-02	HDL
S_VLDL_TG	0.08	0.03	0.02	0.13	4.17E-03	2.39E-02	VLDL
S_VLDL_P	0.08	0.03	0.02	0.13	4.44E-03	2.45E-02	VLDL
S_HDL_L	-0.08	0.03	-0.13	-0.02	4.86E-03	2.45E-02	HDL
M_VLDL_C	0.08	0.03	0.02	0.13	4.89E-03	2.45E-02	VLDL
S_VLDL_L	0.08	0.03	0.02	0.13	4.93E-03	2.45E-02	VLDL
M_VLDL_PL	0.08	0.03	0.02	0.13	5.22E-03	2.51E-02	VLDL
M_VLDL_L	0.07	0.03	0.02	0.13	5.56E-03	2.59E-02	VLDL
M_VLDL_P	0.07	0.03	0.02	0.13	5.77E-03	2.61E-02	VLDL
VLDL_TG	0.07	0.03	0.02	0.13	5.95E-03	2.61E-02	Lipoprotein particle sizes
S_HDL_C	-0.07	0.03	-0.13	-0.02	6.98E-03	2.97E-02	HDL
S_HDL_P	-0.07	0.03	-0.13	-0.02	7.48E-03	3.10E-02	HDL
M_VLDL_TG	0.07	0.03	0.02	0.12	7.83E-03	3.12E-02	VLDL
M_VLDL_CE	0.07	0.03	0.02	0.12	7.97E-03	3.12E-02	VLDL
S_HDL_TG	0.07	0.03	0.02	0.12	8.72E-03	3.33E-02	HDL

M_VLDL_FC	0.07	0.03	0.02	0.12	9.19E-03	3.41E-02	VLDL
S_HDL_PL	-0.07	0.03	-0.12	-0.02	9.37E-03	3.41E-02	HDL
S_LDL_TG	0.07	0.03	0.02	0.12	1.17E-02	4.15E-02	LDL
S_VLDL_FC	0.07	0.03	0.01	0.12	1.20E-02	4.17E-02	VLDL
ApoB	0.07	0.03	0.01	0.12	1.31E-02	4.42E-02	Apolipoproteins
Glc	0.07	0.03	0.01	0.12	1.40E-02	4.54E-02	Glycolysis-related
S_VLDL_PL	0.07	0.03	0.01	0.12	1.40E-02	4.54E-02	VLDL

Fully adjusted model

metabolite	coeff	StandardError	CIL	CIU	p	fdr	group
Tyr	-0.11	0.03	-0.16	-0.06	0.00	0.01	Amino acids
Gp	0.10	0.03	0.05	0.15	0.00	0.02	Inflammation-related
M_HDL_FC	-0.07	0.03	-0.12	-0.01	0.02	0.54	HDL
HDL2_C	-0.07	0.03	-0.12	-0.01	0.02	0.54	Lipoprotein particle sizes
HDL_C	-0.06	0.03	-0.12	-0.01	0.03	0.54	Lipoprotein particle sizes
L_HDL_FC	-0.05	0.03	-0.11	0.00	0.05	0.54	HDL
M_VLDL_CE	0.05	0.03	0.00	0.10	0.05	0.54	VLDL

Table S5-6. Summary of associations between each metabolite and baseline albuminuria status in ET2DS (N=1058). Sorted by P-value (smallest to largest).

Age and sex adjusted-model.

metabolite	or	CIL	CIU	p	fdr	group
Crea	1.79	1.53	2.11	1.42E-12	2.12E-10	Fluid balance
M_HDL_C	0.68	0.58	0.79	4.77E-07	1.57E-05	HDL
M_HDL_FC	0.68	0.58	0.79	5.17E-07	1.57E-05	HDL
M_HDL_L	0.67	0.57	0.78	5.93E-07	1.57E-05	HDL
M_HDL_PL	0.67	0.57	0.78	6.86E-07	1.57E-05	HDL
M_HDL_CE	0.68	0.58	0.79	6.86E-07	1.57E-05	HDL
M_HDL_P	0.67	0.57	0.78	7.37E-07	1.57E-05	HDL
HDL_C	0.70	0.59	0.82	1.23E-05	2.28E-04	Lipoprotein particle sizes
HDL2_C	0.72	0.62	0.83	1.45E-05	2.40E-04	Lipoprotein particle sizes
Gp	1.38	1.19	1.61	2.05E-05	3.05E-04	Inflamation-related

ApoA1	0.71	0.60	0.84	5.42E-05	7.34E-04	Apolipoproteins
S_HDL_L	0.74	0.63	0.86	8.86E-05	1.10E-03	HDL
S_HDL_P	0.74	0.63	0.86	1.32E-04	1.51E-03	HDL
S_HDL_PL	0.74	0.63	0.87	1.60E-04	1.70E-03	HDL
S_HDL_FC	0.75	0.64	0.87	2.05E-04	2.04E-03	HDL
Glol	1.36	1.15	1.61	3.78E-04	3.52E-03	Glycolysis-related
L_HDL_PL	0.77	0.66	0.90	7.09E-04	6.22E-03	HDL
Tyr	0.76	0.65	0.89	7.82E-04	6.48E-03	Amino acids
TotCho	0.76	0.64	0.89	8.84E-04	6.93E-03	Lipoprotein particle sizes
Alb	0.77	0.65	0.90	9.89E-04	7.32E-03	Fluid balance
L_HDL_P	0.78	0.67	0.90	1.10E-03	7.32E-03	HDL
S_HDL_C	0.78	0.68	0.91	1.10E-03	7.32E-03	HDL
L_HDL_L	0.78	0.67	0.90	1.13E-03	7.32E-03	HDL
L_HDL_FC	0.79	0.68	0.91	1.64E-03	1.02E-02	HDL
L_HDL_C	0.79	0.68	0.92	1.96E-03	1.17E-02	HDL
L_HDL_CE	0.79	0.68	0.92	2.08E-03	1.19E-02	HDL
PC	0.78	0.66	0.92	2.61E-03	1.44E-02	Lipoprotein particle sizes
L_HDL_TG	0.81	0.70	0.94	5.77E-03	3.07E-02	HDL
S_HDL_CE	0.82	0.71	0.94	6.11E-03	3.14E-02	HDL
IDL_TG	1.24	1.06	1.46	6.98E-03	3.46E-02	IDL
TotPG	0.81	0.69	0.94	7.19E-03	3.46E-02	Lipoprotein particle sizes
HDL_D	0.80	0.67	0.94	8.40E-03	3.91E-02	Lipoprotein particle sizes

Fully adjusted model

metabolite	or	CIL	CIU	p	fdr	group
Tyr	0.76	0.64	0.90	0.00	0.26	Amino acids
M_HDL_PL	0.82	0.69	0.98	0.03	0.77	HDL
M_HDL_FC	0.83	0.70	0.98	0.03	0.77	HDL
S_HDL_L	0.84	0.71	0.99	0.03	0.77	HDL
S_HDL_P	0.84	0.71	0.99	0.03	0.77	HDL
S_HDL_PL	0.84	0.71	0.99	0.04	0.77	HDL
S_HDL_FC	0.84	0.71	0.99	0.04	0.77	HDL

M_HDL_L	0.84	0.71	1.00	0.05	0.77	HDL
M_HDL_P	0.84	0.71	1.00	0.05	0.77	HDL

Table S5-7. Summary of associations between each metabolite and incident CKD in ET2DS (N=831). Sorted by P-value (smallest to largest).

Age sex-adjusted model

metabolite	coeff	or	CIL	CIU	p	fdr	group
Glc	0.29	1.34	1.12	1.61	0.00	0.18	Glycolysis-related
Ile	0.30	1.35	1.11	1.64	0.00	0.18	Amino acids
LA	-0.27	0.76	0.63	0.92	0.00	0.18	Fatty acids
Val	0.28	1.32	1.09	1.61	0.01	0.18	Amino acids
Crea	0.54	1.71	1.17	2.52	0.01	0.18	Fluid balance
bOHBut	0.24	1.28	1.07	1.53	0.01	0.19	Ketone bodies
FAw6	-0.25	0.78	0.65	0.94	0.01	0.21	Fatty acids
PUFA	-0.22	0.80	0.66	0.97	0.02	0.41	Fatty acids
Tyr	-0.21	0.81	0.67	0.97	0.02	0.41	Amino acids
Pyr	0.21	1.23	1.02	1.49	0.03	0.42	Glycolysis-related
EstC	-0.20	0.82	0.68	0.98	0.03	0.42	Lipoprotein particle sizes
TotCho	-0.21	0.81	0.67	0.98	0.03	0.42	Lipoprotein particle sizes
ApoA1	-0.20	0.82	0.67	0.99	0.04	0.49	Apolipoproteins
TotPG	-0.19	0.83	0.69	1.00	0.05	0.50	Lipoprotein particle sizes
Serum_C	-0.19	0.83	0.69	1.00	0.05	0.50	Lipoprotein particle sizes
Ace	-0.20	0.82	0.67	1.00	0.05	0.50	Ketone bodies

Fully adjusted model.

metabolite	or	CIL	CIU	p	fdr	group
LA	0.77	0.63	0.93	0.01	0.73	Fatty acids
Crea	1.56	1.05	2.35	0.03	0.73	Fluid balance
FAw6	0.81	0.66	0.99	0.04	0.73	Fatty acids
Phe	0.81	0.67	0.99	0.04	0.73	Amino acids

Table S5-8. Sensitivity analysis: Summary of associations between each metabolite and incident CKD in ET2DS (N=694). Sorted by P-value (smallest to largest).

Fully adjusted model

metabolite	coeff	or	CIL	CIU	p	fdr
LA	-0.28	0.76	0.60	0.95	0.02	0.34
L_LDL_CE	-0.25	0.78	0.62	0.96	0.02	0.34
L_LDL_C	-0.25	0.78	0.63	0.97	0.03	0.34

L_LDL_L	-0.25	0.78	0.62	0.98	0.03	0.34
L_LDL_P	-0.25	0.78	0.63	0.98	0.03	0.34
Crea	0.52	1.68	1.05	2.70	0.03	0.34
Ile	0.27	1.31	1.03	1.68	0.03	0.34
IDL_PL	-0.24	0.78	0.63	0.98	0.03	0.34
LDL_C	-0.23	0.79	0.64	0.99	0.04	0.34
IDL_L	-0.24	0.79	0.63	0.98	0.04	0.34
IDL_C	-0.24	0.79	0.63	0.99	0.04	0.34
IDL_CE	-0.24	0.79	0.63	0.99	0.04	0.34
IDL_P	-0.24	0.79	0.63	0.99	0.04	0.34
L_LDL_PL	-0.24	0.79	0.63	0.99	0.04	0.34
L_LDL_FC	-0.23	0.79	0.64	0.99	0.04	0.34
EstC	-0.23	0.79	0.63	0.99	0.04	0.34
M_LDL_P	-0.22	0.80	0.64	0.99	0.04	0.34
M_LDL_L	-0.22	0.80	0.64	1.00	0.05	0.34
M_LDL_CE	-0.21	0.81	0.66	1.00	0.05	0.34
IDL_FC	-0.22	0.80	0.64	1.00	0.05	0.34
M_LDL_C	-0.22	0.81	0.65	1.00	0.05	0.34

Table S5-9. Summary of associations between each metabolite and annual rate of eGFR change % in ET2DS (N=1058). Sorted by P-value (smallest to largest).

Age sex-adjusted model

metabolite	coeff	CIL	CIU	p	fdr	group
M_HDL_P	0.47	0.18	0.76	0.00	0.05	HDL
M_HDL_L	0.46	0.18	0.75	0.00	0.05	HDL
Val	-0.44	-0.71	-0.16	0.00	0.05	Amino acids
M_HDL_PL	0.45	0.17	0.74	0.00	0.05	HDL
M_HDL_CE	0.44	0.15	0.72	0.00	0.05	HDL
M_HDL_C	0.44	0.15	0.72	0.00	0.05	HDL
TotPG	0.43	0.15	0.71	0.00	0.05	Lipoprotein particle sizes
Glc	-0.41	-0.68	-0.14	0.00	0.05	Glycolysis-related

Fully adjusted model

metabolite	coeff	CIL	CIU	p	fdr	group
Val	-0.36	-0.63	-0.08	0.01	0.57	Amino acids
AcAce	-0.32	-0.59	-0.05	0.02	0.57	Ketone bodies
TotPG	0.32	0.04	0.60	0.03	0.57	Lipoprotein particle sizes
SFA	0.31	0.04	0.58	0.03	0.57	Fatty acids
FAw6	0.31	0.04	0.59	0.03	0.57	Fatty acids
LA	0.30	0.03	0.58	0.03	0.57	Fatty acids
M_HDL_P	0.31	0.01	0.60	0.04	0.57	HDL
M_HDL_PL	0.30	0.01	0.59	0.04	0.57	HDL

TotFA	0.28	0.01	0.55	0.05	0.57	Fatty acids
M_HDL_L	0.30	0.01	0.59	0.05	0.57	HDL
PUFA	0.29	0.01	0.57	0.05	0.57	Fatty acids
M_HDL_TG	0.28	0.00	0.55	0.05	0.57	HDL

Table A5-10. Summary of associations between each metabolite and rapid decliner status in ET2DS (N=1058). Sorted by P-value (smallest to largest).

Age and sex-adjusted model

metabolite	coeff	or	CIL	CIU	p	fdr
Glc	0.32	1.37	1.16	1.62	0.00	0.04
M_HDL_CE	-0.28	0.76	0.64	0.90	0.00	0.06
M_HDL_C	-0.28	0.76	0.64	0.90	0.00	0.06
M_HDL_L	-0.28	0.76	0.63	0.90	0.00	0.06
M_HDL_P	-0.28	0.76	0.64	0.90	0.00	0.06
M_HDL_PL	-0.27	0.76	0.64	0.91	0.00	0.06
M_HDL_FC	-0.25	0.78	0.66	0.93	0.00	0.09
TotPG	-0.25	0.78	0.65	0.93	0.01	0.12
ApoA1	-0.26	0.77	0.64	0.93	0.01	0.12
TotCho	-0.24	0.79	0.65	0.95	0.01	0.18
Gp	0.21	1.23	1.04	1.46	0.02	0.21
PC	-0.22	0.80	0.67	0.96	0.02	0.23
HDL_C	-0.21	0.81	0.68	0.98	0.03	0.31
SFA	-0.20	0.82	0.68	0.98	0.03	0.31
Ace	-0.20	0.82	0.68	0.98	0.03	0.34
bOHBut	0.18	1.19	1.01	1.41	0.04	0.34
S_HDL_L	-0.17	0.84	0.71	1.00	0.04	0.35
S_HDL_PL	-0.18	0.84	0.71	1.00	0.05	0.35
Ile	0.18	1.20	1.00	1.44	0.05	0.35
S_HDL_P	-0.17	0.84	0.71	1.00	0.05	0.35

Fully adjusted model

metabolite	coeff	or	CIL	CIU	p	fdr	group
Glc	0.26	1.30	1.05	1.61	0.02	1.00	Glycolysis-related
SFA	-0.21	0.81	0.67	0.98	0.03	1.00	Fatty acids
UnSat	0.19	1.21	1.00	1.46	0.05	1.00	Fatty acids
TotPG	-0.18	0.83	0.69	1.00	0.05	1.00	Lipoprotein particle sizes