

OULU 2024  
D 1771

ACTA UNIVERSITATIS OULUENSIS

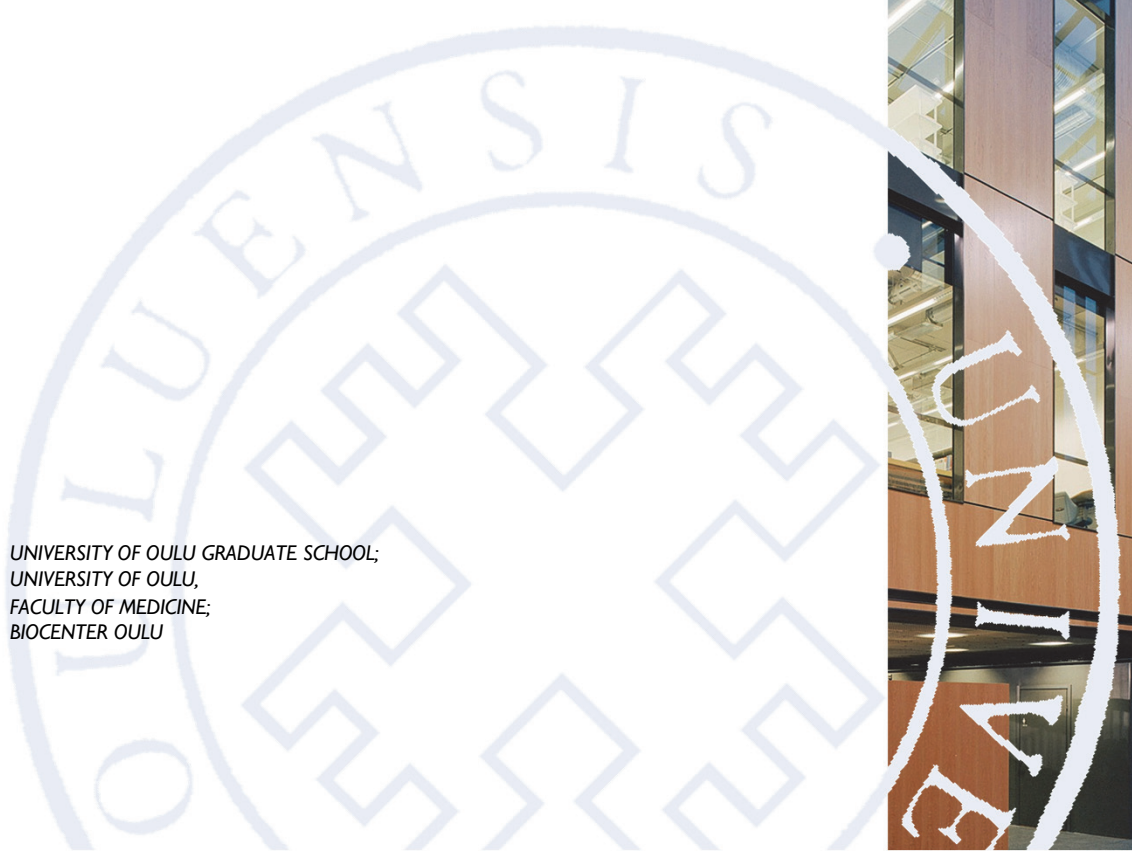
*Tianqi Li*

EPIDEMIOLOGICAL  
APPLICATIONS OF  
QUANTITATIVE URINARY  
NMR METABOLOMICS

UNIVERSITY OF OULU GRADUATE SCHOOL;  
UNIVERSITY OF OULU,  
FACULTY OF MEDICINE;  
BIOCENTER OULU

D

MEDICA





ACTA UNIVERSITATIS OULUENSIS  
D Medica 1771

*TIANQI LI*

**EPIDEMIOLOGICAL APPLICATIONS  
OF QUANTITATIVE URINARY NMR  
METABOLOMICS**

Academic dissertation to be presented with the assent of the Doctoral Programme Committee of Health and Biosciences of the University of Oulu for public defence in the Leena Palotie auditorium (101A) of the Faculty of Medicine (Aapistie 5 A), on 8 March 2024, at 12 noon

UNIVERSITY OF OULU, OULU 2024

Copyright © 2024  
Acta Univ. Oul. D 1771, 2024

Supervised by  
Professor Miika Ala-Korpela  
Professor Johannes Kettunen  
Doctor Pauli Ohukainen

Reviewed by  
Professor Karsten Suhre  
Associate Professor Juha Hulmi

Opponent  
Professor Daniel Monleon

ISBN 978-952-62-3986-6 (Paperback)  
ISBN 978-952-62-3987-3 (PDF)

ISSN 0355-3221 (Printed)  
ISSN 1796-2234 (Online)

Cover Design  
Raimo Ahonen

PUNAMUSTA  
TAMPERE 2024

## **Li, Tianqi, Epidemiological applications of quantitative urinary NMR metabolomics.**

University of Oulu Graduate School; University of Oulu, Faculty of Medicine; Biocenter Oulu

*Acta Univ. Oul. D 1771, 2024*

University of Oulu, P.O. Box 8000, FI-90014 University of Oulu, Finland

### ***Abstract***

Urinary metabolome is associated with glomerular filtration and molecular reabsorption processes in the kidneys, and reflects several key biochemical pathways related to the cardiometabolic state, gut microbial metabolic activity, and dietary characteristics. Thus, detailed quantitative data on urinary metabolites allow the assessment of renal function and its inference under various metabolic conditions. The development of the quantitative high-throughput urine NMR metabolomics platform allows simultaneous quantification of multiple urine metabolites, reflecting the metabolic changes of the body to endogenous and exogenous stimuli. Moreover, the urine NMR spectroscopy exhibits strong resilience and precision, as does the entire quantification process. This enables the platform to provide automated, high-throughput experiments for large-scale epidemiological applications in a cost-efficient manner. Currently, quantitative metabolic approaches with large enough numbers of individuals for appropriate epidemiological studies, targeting an improved understanding of urinary metabolites in health and as potential biomarkers of the risk of disease are few. The present thesis focuses on the epidemiological applications of quantitative urinary NMR metabolomics. Based on quantitative urinary metabolomics data from a high-throughput urine NMR metabolomics platform developed by our group, the thesis demonstrates the distribution characteristics of urinary metabolites at the population level, proposes a rationale for an appropriate normalization strategy for urinary metabolite concentrations in epidemiological analyses, and shows their associations with clinical and biochemical measurements in extensive population-based cohorts. The novel extensive quantitative urinary metabolomics data and the associations with clinically relevant measures support the application of urine metabolomics in large-scale epidemiological studies and provide new insights into renal function and related urinary biomarkers.

*Keywords:* biomarkers, body mass index, clinical measures, distribution, epidemiology, estimated glomerular filtration rate, kidneys, metabolism, metabolites, metabolomics, normalization, nuclear magnetic resonance, quantification, sex, smoking, the risk of disease, urine



## **Li, Tianqi, Kvantitatiivisen virtsan NMR-metabolomiikan epidemiologiset sovellukset.**

Oulun yliopiston tutkijakoulu; Oulun yliopisto, lääketieteellinen tiedekunta; Biocenter Oulu

*Acta Univ. Oul. D 1771, 2024*

Oulun yliopisto, PL 8000, 90014 Oulun yliopisto

### ***Tiivistelmä***

Virtsan metabolomi on kiinteästi sidoksissa munuaisten glomerulusten suodattamiseen ja molekyylien takaisinabsorptioprosessiin. Tämä ilmentää useiden keskeisten biokemiallisten reittien merkitystä, jotka vaikuttavat kardiometaboliseen tilaan, suoliston mikrobien aineenvaihduntaan sekä ruokavalion ominaisuuksiin. Siksi yksityiskohtaiset kvantitatiiviset tiedot virtsan metaboliiteista mahdollistavat munuaisten toiminnan arvioinnin ja sen analysoinnin erilaisissa metabolisissa olosuhteissa. Korkean suorituskyvyn virtsan NMR-metabolomiikan alustan kehittäminen mahdollistaa useiden virtsan metaboliittien samanaikaisen tarkan kvantitatiivisen analyysin. Nämä metaboliitit heijastavat kehon aineenvaihdunnallisia reaktioita endogeenisten ja eksogeenisten ärsykkeiden vaikutuksesta. Lisäksi sekä virtsan NMR-spektroskopia että koko kvantitatiivinen analyysiprosessi ovat osoittautuneet erittäin joustaviksi ja tarkoiksi, mikä mahdollistaa automatisoituja korkean suorituskyvyn kokeita. Näitä kokeita voidaan edelleen toteuttaa kustannustehokkaasti laajoissa epidemiologisissa tutkimuksissa. Tällä hetkellä on niukasti kvantitatiivisia metabolomiikkamenetelmiä, jotka soveltuvat suurimittaisiin epidemiologisiin tutkimuksiin, joiden tavoitteena on syventää ymmärrystä virtsan metabolomiikan terveysvaikutuksista ja sen potentiaalisista sairausriskin biomarkkereista. Tämä tutkielma keskittyy erityisesti kvantitatiivisen virtsan NMR-metabolomiikan epidemiologisiin sovelluksiin. Perustuen tutkimusryhmämme kehittämään korkean suorituskyvyn virtsan NMR-metabolomiikan alustaan, tämä tutkielma esittelee virtsan metaboliittien jakautumisen populaatiotasolla ja esittää perusteet tarkoituksenmukaiselle virtsan metaboliittipitoisuuksien normalisointistrategialle epidemiologisissa analyyseissä. Lisäksi tutkielma tutkii näiden metaboliittien yhteyksiä laajamittaisissa kohorttitutkimuksissa tehtyihin kliinisiin ja biokemiallisiin mittauksiin. Uudet laajat kvantitatiiviset virtsan metabolomiikkatiedot ja niiden liittyminen kliinisesti merkittäviin mittareihin tukevat virtsan metabolomiikan soveltamista laajoissa epidemiologisissa tutkimuksissa ja avaavat uusia näkökulmia munuaisten toimintaan sekä siihen liittyviin virtsan biomarkkereihin.

*Asiasanat:* aineenvaihdunta, arvioitu glomerulusten suodatusnopeus, biomarkerit, epidemiologia, jakautuminen, kliiniset mitat, kvantitatiivinen analyysi, metaboliitit, metabolomiikka, munuaiset, normalisointi, painoindeksi, sairauden riski, sukupuoli, tupakointi, virtsa, ydinmagneettinen resonanssi



*To all my loved ones*



## Acknowledgements

The work for thesis was carried out at the Unit of Population Health, Faculty of Medicine, University of Oulu, during 2018-2023. I acknowledge the financial support received from the University of Oulu Graduate School and Biocenter Oulu.

I want to convey my deepest appreciation to my supervisors Professor Mika Ala-Korpela, Professor Johannes Kettunen and Dr. Pauli Ohukainen, for their outstanding guidance, counsel, and unwavering support throughout this endeavor. Mika, I extend my thanks for introducing me to this captivating scientific field, allowing me to collaborate with exceptional individuals, and enabling me to broaden my horizons while steadily working towards self-improvement. Your passion for science is truly motivating, and your capacity to generate innovative ideas is truly remarkable. Johannes, it has been a pleasure collaborating with you; your enthusiastic and meticulous scientific approach, coupled with a diligent work ethic, has left a lasting impression on me. Pauli, I am grateful for your patience and guidance; your encouragement has allowed me to express my thoughts more freely.

I would like to thank Dr. Tuulia Tynkkynen and Andrei Ihanus, for their expertise and hard work in the NMR laboratories have not only provided these valuable metabolomics datasets but have also ensured their quality with precise quantitative molecular concentrations. Additionally, I appreciate Associate Professor Ville-Petteri Mäkinen for providing enthusiastic guidance and support in my projects, especially for helping me in statistical analysis.

I want to express my gratitude to Professor Jukka Hakkola, Docent Minna Ruddock and Docent Johanna Magga for their participation in my follow-up group, offering valuable feedback and support throughout my doctoral studies. I extend my gratitude to the pre-examiners, Professor Karsten Suhre from the Weill Cornell Medicine-Qatar and Associate Professor Juha Hulmi from the University of Jyväskylä, for their meticulous examination of the thesis and their positive, constructive comments that enhanced the quality of the work. I also warmly appreciate Professor Daniel Monleon from the University of Valencia for accepting the invitation and dedicating time to serve as the opponent in my defense.

I want to express my gratitude to all my colleagues at the Unit of Population Health, especially Olga Anufrieva. Thank you, Olga, for providing me with valuable advice on both life and work when I first arrived in Finland. The time we spent together attending fitness classes was the happiest during my early days in Finland. I also want to thank Minna, Eeva, Marita, and Jaakko. During the COVID-19 pandemic, we experienced a unique period of remote work, and your

companionship in virtual gatherings through online chat rooms helped me get through those lonely times. Additionally, I deeply appreciate Siyu Zhao. Thank you for sharing lunches and breaks, discussing the highs and lows of our research work. We've traveled across Europe together, leaving behind many unforgettable memories. Furthermore, I want to thank all my friends, especially Xiayun Yang and Saranya Palaniswamy; your companionship holds the dearest memories of my time here.

In closing, my deepest gratitude goes to my parents, Qiubo and Yuanpei, whose unwavering love and support have been the cornerstone of my journey. Their enduring support and belief in my aspirations have molded me into the person I am today. Mom and Dad, your wisdom and encouragement have consistently been my guiding lights, illuminating the path through challenges and triumphs alike. I am immensely thankful for your continuous trust and support. In critical moments of my life, regardless of the decisions I made, you encouraged me to pursue my dreams. It is your support and understanding that have brought me to where I am today. This achievement is not just mine; it is also a testament to the boundless love of my parents.

Oulu, 10.11.2023

Tianqi Li

## List of abbreviations and symbols

2-PY	N1-Methyl-2-pyridone-5-carboxamide
4-HPA	4-Hydroxyphenylacetate
ABS	Absolute concentrations
ACR	Albumin-creatinine ratio
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
ANOVA	Analysis of variance
ApoA-I	Apolipoprotein A-I
ApoB	Apolipoprotein B
ARIC	Atherosclerosis Risk in Communities
AST	Aspartate aminotransferase
AUC	Receiver operating curve
BMI	Body mass index
CE	Capillary electrophoresis
CI	Confidence interval
CKD	Chronic kidney disease
CRIC	Chronic Renal Insufficiency
CRP	C-reactive protein
CS	Constant sum
CSF	Cerebrospinal fluid
CTLS	Constrained total line shape
CV	Coefficients of variation
DBP	Diastolic blood pressure
DESeq2	Differential gene expression analysis based on the negative binomial distribution
DKD	Diabetic kidney disease
DM	Diabetes mellitus
DMA	Dimethylamine
DMG	N,N-dimethylglycine
DNA	Deoxyribonucleic acid
eGFR	Estimated glomerular filtration rate
ESI	Electrospray ionisation
ESKD	End-stage kidney disease
FDR	False discovery rate
FT4	Free thyroxine

GC	Gas chromatography
GFR	Glomerular filtration rate
GGT	Gamma-glutamyl transferase
GlycA	Glycoprotein acetyls
HbA1c	Haemoglobin A1c
HDL-C	High-density lipoprotein
HMDB	Human Metabolome Database
HOMA-IR	Homoeostasis model assessment of insulin resistance
HPHPA	3-(3-Hydroxyphenyl)-3-hydroxypropanoate
HR	Hazard ratio
hsCRP	High-sensitivity C-reactive protein
IEC	Ion exchange chromatography
IEM	Inborn errors of metabolism
IMS	Ion mobility spectrometry
IS	internal standard
IS-CREA	Normalization to creatinine concentration
IS-GLUC	Normalization to glucose concentration
IS-PSEURID	Normalization to pseudouridine concentration
IS-UREA	Normalization to urea concentration
KIDMED	Mediterranean Diet Quality Index in children and adolescents
LC	Liquid chromatography
LDL-C	Low-density lipoprotein cholesterol
MAP	mean arterial pressure
MS	Mass spectrometry
MVA	Multivariate statistical analysis
NAM	Nicotinamide
NANA	N-acetylneuraminic acid
NFBC	Northern Finland Birth Cohort
NMR	Nuclear magnetic resonance
NSAID	Nonsteroidal anti-inflammatory drugs
OCN	Osteocalcin
OPLS-DA	Orthogonal projections to latent structures discriminant analysis
OR	Odds ratio
PCA	Principal component analysis
PLS-DA	Partial least squares for discriminant analysis
PQN	Probabilistic quotient normalization

QTOFMS	Quadrupole-time-of-flight-mass spectrometry
RNA	Ribonucleic Acid
ROC	Receiver operating characteristic
SBP	Systolic blood pressure
SD	Standard deviation
SGLT1	Sodium-glucose cotransporter 1
SGLT2	Sodium-glucose cotransporter 2
SHBG	Sex hormone binding globulin
SHIP	Study of Health in Pomerania
T1D	Type 1 diabetes
T2D	Type 2 diabetes
TG	Triglyceride
TMAO	Trimethylamine N-oxide
TSH	Thyroid-stimulating hormone
TSP	3- (trimethylsilyl) propionic-2,2,3,3 acid D4
UK	United Kingdom
UPF	Ultra-processed foods
UPLC	Ultra-performance liquid chromatography
USA	United States of America
WBC	White blood cell count
WHR	Waist-to-hip ratio
YFS	Cardiovascular Risk in Young Finns Study



## List of original publications

This thesis is based on the following publications, which are referred throughout the text by their Roman numerals:

- I Li, T., Ihanus, A., Ohukainen, P., Järvelin, M. R., Kettunen, J., Mäkinen, V. P., Tynkkynen, T., & Ala-Korpela, M. (2022). There is always glucose in normal urine: unspecific excretion associated with serum glucose and glomerular filtration rate. *International Journal of Epidemiology*, 51(6), 2022–2025, doi: 10.1093/ije/dyac060
- II Li, T., Tynkkynen, T., Ihanus, A., Zhao, S., Mäkinen, V. P., & Ala-Korpela, M. (2022). Characteristics of Normalization Methods in Quantitative Urinary Metabolomics-Implications for Epidemiological Applications and Interpretations. *Biomolecules*, 12(7), doi: 10.3390/biom12070903
- III Li, T.<sup>†</sup>, Ihanus, A.<sup>†</sup>, Ohukainen, P., Järvelin, M. R., Kähönen, M., Kettunen, J., Raitakari, O.T., Lehtimäki, Terho., Mäkinen, V. P., Tynkkynen, T.<sup>‡</sup>, & Ala-Korpela, M.<sup>\*‡</sup>. (2023) Clinical and biochemical associations of urinary metabolites: Quantitative epidemiological approach on renal-cardiometabolic biomarkers. *International Journal of Epidemiology*, doi: 10.1093/ije/dyad162

<sup>†</sup>Joint first author

<sup>‡</sup>Joint last author

\* Corresponding author



# Contents

<b>Abstract</b>	
<b>Tiivistelmä</b>	
<b>Acknowledgements</b>	<b>9</b>
<b>List of abbreviations and symbols</b>	<b>11</b>
<b>List of original publications</b>	<b>15</b>
<b>Contents</b>	<b>17</b>
<b>1 Introduction</b>	<b>19</b>
<b>2 Urine metabolomics in epidemiology</b>	<b>21</b>
2.1 Opportunities and challenges in epidemiological metabolomics .....	21
2.2 Metabolomics .....	23
2.2.1 Human metabolome .....	25
2.2.2 Urine metabolome .....	26
2.2.3 Development of analytical technologies .....	29
2.3 Epidemiological application of urinary metabolomics .....	35
2.3.1 Urine sampling .....	35
2.3.2 Data pre-processing and statistical methods .....	36
2.3.3 Epidemiological application of urinary metabolomics .....	39
<b>3 Aim of study</b>	<b>55</b>
<b>4 Methods</b>	<b>57</b>
4.1 Epidemiological cohorts and study populations .....	57
4.1.1 Applied clinical and biochemical measures .....	60
4.1.2 Applied quantified urinary metabolomics data .....	66
4.2 Applied normalization methods .....	73
4.3 Statistical analyses .....	79
<b>5 Results and discussion</b>	<b>83</b>
5.1 Characteristics of urinary glucose at the population level (Study I) .....	83
5.1.1 Distribution of urinary glucose concentrations at the population level .....	83
5.1.2 Correlation and regression analyses of urinary glucose in NFBC1966 .....	84
5.1.3 Novel associations of urinary glucose with customary clinical and biochemical measures .....	87
5.1.4 Summary of Study I .....	88

5.2	Characteristics of normalization methods in quantitative urinary metabolomics (Study II).....	88
5.2.1	Intra-urinary metabolite associations and metabolic cluster analysis .....	89
5.2.2	Epidemiological Exemplars: BMI and MAP.....	95
5.2.3	Summary of Study II .....	98
5.3	Clinical and biochemical associations of urinary metabolites (Study III).....	100
5.3.1	Quantitative urinary metabolite abundance at the population level .....	101
5.3.2	Association clusters of urinary metabolites .....	106
5.3.3	Metabolite associations with BMI .....	110
5.3.4	Metabolite associations with smoking history.....	110
5.3.5	Unspecific renal excretion of amino acids.....	111
5.3.6	Summary of Study III.....	113
<b>6</b>	<b>Conclusions</b>	<b>117</b>
	<b>List of references</b>	<b>121</b>
	<b>Original publications</b>	<b>139</b>

# 1 Introduction

Advances in high-throughput omics technologies such as genomics, epigenetics, proteomics, and metabolomics offer a novel, systematic approach to understanding the etiology of diseases. (Würtz et al., 2017). Due to the identification and quantification of metabolites from body fluids, as well as cost-effective and high-throughput measurements of nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS), metabolomics are increasingly applied in epidemiological research (Dunn, Broadhurst, Atherton, et al., 2011). Biomarkers obtained from comprehensive expression profiling methods conducted on readily obtainable body fluids, such as serum (Würtz et al., 2017), urine (Tynkkynen et al., 2019), cerebrospinal fluid (CSF) (Holmes et al., 2006), or saliva (Galloway et al., 2016), have potentiality for risk prediction at the population level. Among these biological fluid types, urine and serum samples are the two most commonly used. Urine has various benefits: it is abundant, sterile and simple to collect (Tynkkynen et al., 2019). The molecular content of urine, physiologically connected to the glomerular filtration and molecular reabsorption processes in the kidneys, includes soluble waste products as well as excess water, sugar, among many others from the blood (Bouatra et al., 2013). Urinary metabolites reflect multiple key biochemical pathways in relation to renal function, gut microbial metabolism and food intake. Thus, considerable value of urinary metabolites in epidemiological applications has been recognized for quite some time (Bouatra et al., 2013; Elliott et al., 2015; Emwas et al., 2016; Holmes et al., 2008; Nicholson et al., 2011; Tynkkynen et al., 2019).

However, existing epidemiological studies with quantitative urinary metabolic data in sufficient number of individuals, aiming at improved understanding of urinary metabolites in health and as potential biomarkers of the risk of disease, are very few. NMR spectroscopy offers a comprehensive quantitative method for urine analysis (Emwas et al., 2016; Holmes et al., 2008; Li, Ihanus, et al., 2022) and holds promise for enabling cost-effective, fully automated high-throughput experimentation. This capability is crucial for conducting large-scale systems epidemiology studies. (Ala-Korpela & Davey Smith, 2016; Mutter et al., 2022; Tynkkynen et al., 2019). Our research group has released a setup for quantitative high-throughput analysis of urine, encompassing all essential aspects of sample preparation and NMR experimentation; initially conducted fully automated quantitative analyses of urine NMR spectra; and put forth an open access

quantitative pipeline for urine NMR metabolomics to support large-scale studies. (Tynkkynen et al., 2019).

The focus of this thesis is on epidemiological applications of quantitative urinary NMR metabolomics. Using population-level quantitative data of urinary metabolites by the high-throughput urine NMR metabolomics platform, the specific applications featured in this thesis are in three parts: 1) distribution characteristics of quantified urinary metabolites at the population level, 2) the reasoning behind selecting a suitable normalization strategy for urinary metabolite concentrations in epidemiological analyses, 3) and the associations between quantified urinary metabolite and commonly clinical and biochemical measurements in large-scale population-based cohorts.

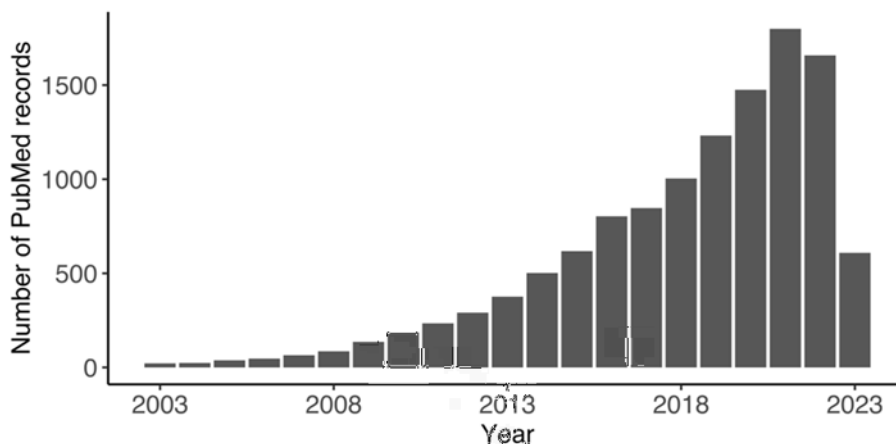
The present study extends the epidemiological scale of urine metabolomics to a new level. While the first two parts delve into key issues in the application of urine metabolomics in epidemiological studies, the third applies a comprehensive quantitative urine metabolomics study at an epidemiological scale, incorporating independent replications, providing extensive information on the clinical and biochemical associations of urinary metabolites to gain novel insights into renal function and associated disease biomarkers.

## 2 Urine metabolomics in epidemiology

Modern definitions of epidemiology refer to the study (scientific, systematic, and data-driven) of the distribution (who, when, and where), determinants of health and disease (pathophysiological, environmental, behavioural), and control of health problems (biological, social, economic, political, administrative, legal) in a defined population (Gerstman, 2013; Frérot et al., 2018). Epidemiological practice and the outcomes of epidemiological analysis play a pivotal role in managing population-based health. Firstly, epidemiology seeks to pinpoint the causes and associated risk factors of diseases with the aim of reducing morbidity and mortality from the disease. Second, it assesses extent of disease found in the community, which is essential for planning healthcare services, infrastructure, and the training of future healthcare providers. Third, it delves into the natural history and prognosis of diseases. Fourth, it evaluates both existing and newly developed strategies for prevention, treatment, and healthcare delivery models. Fifth, it forms the basis for crafting public policies related to environmental concerns, genetic factors, and other considerations pertaining to disease prevention and health promotion (Gordis, 2014).

### 2.1 Opportunities and challenges in epidemiological metabolomics

Recent advances in high-throughput technologies have made it possible to generate population-scale metabolomics and other "omics" data (Fearnley & Inouye, 2016). The application of metabolomics in epidemiological research is the systematic use of epidemiological strategies and principles to investigate a population-based variation in the human metabolome (Lasky-Su et al., 2021). This provides biological and mechanistic insights which are not usually available to classical epidemiological studies that rely on questionnaires or environmental exposure data (van Roekel et al., 2019). Over the past two decades, there has been a steady rise in the number of epidemiological studies with a focus on metabolomics (Fig. 1). A PubMed search up to May 2023 revealed 10 661 papers in related fields, of which 62% were in Published within the last five years, which might only represent only a portion of all studies in this research field, underscoring the growing scientific fascination in this area of investigation.



**Fig. 1. Yearly number of PubMed records containing both keywords related to epidemiology/population-based research and metabolomics in the period 2003–2023 (2023 number based on the following search terms ["epidemiologic" or "epidemiology" or "population-based" or "cohort" or "observational" or "cross-sectional" or "case-control"] AND ["metabolomic" or "metabolomics" or "metabonomics" or "metabolome" or "metabonomic" or "metabonome" or "metabolite profile" or "metabolic profile" or "metabolic signature" or "glycomics" or "glycomic" or "lipidomics" or "lipidomic"] up to May 08, 2023).**

Metabolomics epidemiology has been applied to characterizing exposures, detecting early markers of disease, improving disease diagnosis, following response to treatment or disease progression, and comprehending disease etiology (Dubin & Rhee, 2020; Li et al., 2020; Naylor et al., 2021; Shao et al., 2021; Soininen et al., 2015; van Roekel et al., 2019). However, the challenges in this field are at multiple scientific levels. Metabolomics has been around for some 25 years with rather common hype on the potentially revolutionary role of related methodologies in personalized risk assessment and precision medicine. As an extreme, the spectroscopy driven “diagnostic branch” of metabolomics has been fundamentally flawed for the whole time and this fallacy is partly persisting in the field (Bevilacqua & Bro, 2020; Kjeldahl & Bro, 2010). This is a vital misconception to denounce (Madsen et al., 2010; Xia et al., 2013), particularly for common (and polygenic) cardiometabolic diseases that relate to the focus of this thesis. In fact, over the years various reviews and other contributions have been published to elaborate the scientific fundamentals in this field (Ala-Korpela, 2007, 2008, 2016, 2018; Ala-Korpela et al., 2012; Ala-Korpela & Davey Smith, 2016; Elliott et al.,

2015; McMahon et al., 2017; Mutter et al., 2022; Rose, 1985; Salihovic et al., 2020; Smith, 2011; Soininen et al., 2015; Würtz et al., 2017).

In fact, the fundamental setting by Mother Nature is that all polygenic conditions (e.g. all common cardiometabolic diseases) are continuous traits. This reality is in a central role of making it impossible to find precise (metabolic or genetic) predictive biomarkers (Ala-Korpela & Holmes, 2020; Holmes & Davey Smith, 2019; McCarthy & Birney, 2021; Niiranen & Vasan, 2016; Plomin et al., 2009; Visscher et al., 2021; Ware, 2006). The scientific consequence is that epidemiology and public health are much more about populations than about individuals (Rose, 1985; Sniderman et al., 2018) – since small population level effects in causal disease biomarkers can result in pronounced changes in population health, we just cannot predict who the individuals are to be affected. A prominent example is the North Karelia Project, at a time (early 1970s) when a pioneering population-based lifestyle intervention study that led to over 80% reduction of cardiovascular disease mortality in three decades (Puska & Jainsi, 2020). These population-level prevention principles are now incorporated in the World Health Organisation strategies and recommendations (Waxman & World Health Assembly, 2004).

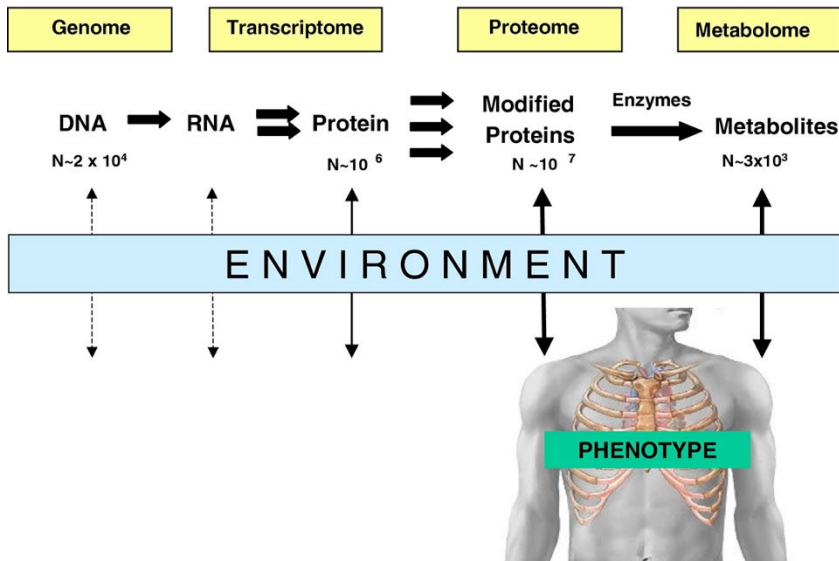
Overall, it should be noted that epidemiological metabolomics is the study on populations and there is an inevitable continuum between health and disease. Based on these principles, epidemiological metabolomics plays a positive role in health management and disease prevention of population. The next section will introduce the research scope of metabolomics.

## **2.2 Metabolomics**

Metabolomics is qualitative and quantitative profiling of multiple metabolic measures in biofluids, cells and tissues (Wishart et al., 2022). The metabolic measures usually include numerous small molecules (<1.5 kDa) metabolites. In addition, depending on the sample and detection method, for example, in serum NMR metabolomics, reliably detected measures such as lipoprotein, clinical lipids, apolipoprotein A-I and B are also included (Ala-Korpela et al., 2022; Soininen et al., 2015). The essential information flow in biologic systems is viewed as from DNA (genome) to RNA (transcriptome) to proteins (proteome) to metabolites (metabolome) (Dubin & Rhee, 2020). The phenotype emerges as a result of intricate interactions among components at all functional levels and their

interaction with the environment. These interactions are presented in Fig. 2 (Lewis et al., 2008).

While endogenous metabolites have a remarkable conservation across populations or species, the metabolome of an individual is not invariant. In fact, metabolomics is inherently sensitive to subtle changes in biological pathways caused by internal (i.e. physiological and intracellular) signalling and external (i.e. diet, medication, lifestyle) stimuli. Since metabolites can be sugars, lipids, amino acids, organic acids and nucleic acids generated through endogenous catabolism or anabolism, as well as exotic xenobiotic compounds from the environment (Wishart, 2019), this suggests that the metabolome is a particularly useful probe of individual phenotypes that offers molecular insights into the underlying mechanisms of various physiological conditions and pathophysiological processes (Johnson et al., 2016). Advances in quantified high-throughput technologies promote the application of metabolomics in large-scale epidemiological studies to understanding the etiology of various diseases and the ability to predict the risk of disease (Holmes et al., 2008; Tynkkynen et al., 2019; Wang et al., 2016; Würtz et al., 2012, 2017). Next, I will introduce the research content of human metabolome, the characteristics of urine metabolome and the development of high-throughput analysis technologies in the field of metabolomics.



**Fig. 2. Interactions of functional levels (genome, transcriptome, proteome, and metabolome) in biological systems** (Lewis et al., 2008). **Of note, the number of identified metabolites has increased by more than 7200% (Wishart et al., 2022) (in 2022) since this figure was originally published in 2008. Fig.2 is reprinted with permission from Elsevier under License ID: 5621931137328 (Lewis et al., 2008).**

### 2.2.1 Human metabolome

The human metabolome consists of all the molecules found in the human body. The Human Metabolome Database (HMDB) is the world's most extensive and comprehensive database dedicated to organism-specific metabolomics. The most recent version, HMDB 5.0, released in in 2022 contains 217 920 annotated metabolite entries and 1 581 537 unannotated derivatized metabolite entries (Wishart et al., 2022). Due to the wide range of physicochemical properties of the metabolome, multiple analytical techniques have been developed for comprehensive study of the metabolome. These include the use of NMR spectroscopy and MS often coupled with chromatography (Dunn, Broadhurst, Atherton, et al., 2011).

The human metabolome can be classified into the endogenous metabolome and the exogenous metabolome. The endogenous metabolome includes molecules produced by enzymes encoded by the genome or microbial genome. Since

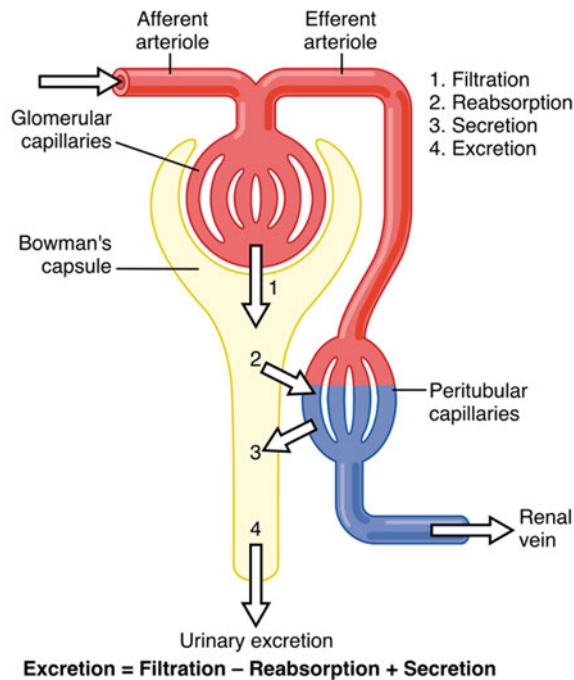
metabolic enzymes are generally highly conserved, the endogenous metabolome has a remarkable conservation (Peregrín-Alvarez et al., 2009). Also known as the exposome, the exogenous metabolome, is highly variable and is greatly influenced by all external and non-genetic factors, such as dietary patterns, exercise, medication and lifestyle (Bermingham et al., 2021; Rappaport et al., 2014; Yousri et al., 2014). Thus, the variation of the human metabolome could reflect the influence of genetic differences as well as environmental exposures.

Research of the human metabolome involves the collection, extraction, and analysis of a wide range of biological samples. These include blood, urine, CSF, saliva, sweat, hair, faeces and different tissues. Among these samples, serum (or plasma) and urine are the most common studied samples (Dunn, Broadhurst, Begley, et al., 2011), while urine samples are abundant, sterile and the collection of urine is non-invasive. In addition, compared to blood, urine is not affected by many homeostatic mechanisms and have a greater variety of endogenous metabolites, which could reflect the abnormality of circulation measures in the body and directly correlate with the external stimuli of the body (Bouatra et al., 2013). However, in immense contrast to systemic blood biomarkers that have been widely used in large-scale quantitative studies in epidemiology (Ala-Korpela et al., 2022; Würtz et al., 2017), existing studies applying quantitative urine metabolic methods at an epidemiological scale are rare.

### **2.2.2 Urine metabolome**

Urine is produced by the filtration of blood in the kidneys. Within each kidney, there are roughly 1 million nephrons, with each nephron comprising a spherical filtration unit known as the renal corpuscle, and a tubule that extends from renal corpuscle. The process of urine formation entails three steps: glomerular filtration, selective reabsorption, and tubular secretion (Fig. 3) (Hall, J. E. & Hall, 2020). As the site of blood filtration, the glomerulus filters water and small solutes from the blood, leaving blood cells and large proteins in circulation. The resulting filtrate contains both beneficial elements like amino acids, glucose, vitamins, electrolytes, hormones, ions, etc., as well as detrimental substances including metabolic waste products such as urea, creatinine, uric acid, ions, etc. When the filtrate moves through the renal tubules, the essential nutrients and water are reabsorbed into the capillaries. Tubular secretion transfers substances from the blood in the peritubular capillaries into the filtrate in the renal tubules, facilitating the active excretion of creatinine or excess  $H^+$  or excess  $K^+$  and other wastes out of the body. The tubular

filtrate ultimately becomes urine (Eaton & Pooler, 2018; Hall, J. E. & Hall, 2020), which contains high concentrations of urea, creatinine, ammonia, inorganic salts, organic acids, various water-soluble toxins, and pigment products of the breakdown of haemoglobin (Bouatra et al., 2013).



**Fig. 3. Processes of urine formation.** Fig.3 is reprinted from with permission from Elsevier Health Sciences under License ID: 1420343-1 (Hall, J. E. & Hall, 2020).

The kidney is generally considered the excretory organ of the organism, responsible for eliminating metabolic waste and various bioactive substances including hormones and many foreign substances, especially drugs. In addition, it also plays an vital role in maintaining body fluids and electrolyte balance, as well as regulating the production of red blood cells and vitamin D (Eaton & Pooler, 2018). A critical indicator of kidney performance is the glomerular filtration rate (GFR): the flow rate of filtered fluid through the kidney. In a normal young adult male, the GFR can be as high as 180 L per day, while the total volume of human plasma averages around 3 L, which means the entire plasma volume is filtered by the kidneys about 60 times a day. This substantial filtration capacity enables the kidneys to expel huge amounts of waste and precisely regulate the internal

environment's composition. The gold standard for measuring GFR is the inulin clearance, when the very accurate values are required for research and a number of clinical situations (Eaton & Pooler, 2018). As this method is cumbersome, the estimated glomerular filtration rate (eGFR) is a customary clinical way to approximate an individual's kidney function and is estimated based on age, sex, and serum creatinine (Inker et al., 2012). Creatinine arises from the breakdown of creatine, a process ongoing in our skeletal muscles, and it's consistently released into our bloodstream. Since this release is linked to skeletal muscle mass, in an individual with stable muscle mass, creatinine production remains steady. Despite being freely filtered and not reabsorbed, a small portion of creatinine is secreted by the proximal tubule. The creatinine found in urine comprises both a filtered and a secreted portion. Due to secretion, creatinine clearance is slightly greater than the GFR. Factors that contribute to variations of systemic creatinine concentrations, such as diet (Cai et al., 2020) and changes in muscle mass (Baxmann et al., 2008), could also affect the interpretation of eGFR-related results.

A decreased GFR indicates decreased kidney function, which can be caused by a range of kidney diseases such as acute kidney injury, chronic kidney disease (CKD) and renal failure (Levey & Coresh, 2012). Among these conditions, CKD has emerged as a prominent factor contributing to mortality and distress in the 21st century (Naghavi et al., 2015; Rhee & Kovesdy, 2015). CKD is a progressive disease that influences more than 10% of the general population worldwide, totalling more than 800 million people (Jager et al., 2019). CKD is more prevalent in older individuals, women, ethnic minorities, and those with diabetes and hypertension. The prevalence and devastating impact of CKD should prompt enhanced efforts to develop and implement effective preventive and therapeutic measures aimed at reducing the development of CKD and slowing its progression (Kovesdy, 2022). From the above description of urine formation process, it can be recognized that the molecular content of urine is physiologically related to the glomerular filtration and reabsorption process of the kidney. Urine metabolome reflects the kidney function, including various metabolites related to (patho)physiology and cardiometabolic conditions, microbial metabolic activities, and dietary characteristics. Thus, detailed quantitative data on urinary metabolites can supply direct molecular probes to estimate renal function and its inferences under various metabolic conditions.

The human urine metabolome database is an electronic database freely available at <http://www.urinemetabolome.ca>. It provides extensive data on approximately 3 100 small molecule metabolites present in human urine, along

with around 3 900 concentration values (Bouatra et al., 2013). Unlike the serum metabolome, which is highly conserved and stable in normal physiological conditions, the composition of the urine metabolome is complex and diverse and is easily changed due to stimulation by external factors such as the environment and diet. Clinical urinalysis involves evaluating the physical properties of urine, such as colour and clarity, microscopic examination, and chemical analysis using urine dipsticks (Haq & Patel, 2023). Test strips can readily measure urinary glucose, bilirubin, ketone bodies, leukocyte esterase, nitrates, haemoglobin, urobilinogen, specific gravity and protein (Bouatra et al., 2013). However, the aforementioned clinical urine measurements are limited by high detection limits of the methods, and quantitative urinary data detected by modern metabolomics techniques are rarely used appropriately in large-scale populations. Therefore, urine samples contain a wealth of underutilized metabolic information and have potential translational applications, including urinary metabolites that could provide help in assessing the risk of disease and diagnostics. For example, a cross-sectional study comparing small molecules in the urine of 469 lung cancer patients and 536 controls found elevated levels of creatine riboside and N-acetylneuraminic acid (NANA, a novel identified molecule) are associated with early lung cancer diagnosis and worse prognosis (Mathé et al., 2014). And in the follow-up study of the Framingham Heart Study (the mean follow-up time was 9.7 years), 193 individuals with CKD (eGFR<60 ml/min/1.73m<sup>2</sup>) emerged, compared with 193 control individuals, low urinary glycine and histidine were found to be associated with incident CKD (McMahon et al., 2017). Moreover, the Finnish Diabetic Nephropathy study, a prospective observational research initiative focusing on type 1 diabetes (T1D) patients (n=2 670), collected 24-hour urine samples. The present study investigated the associations between 51 urinary metabolites and the risk of diabetic nephropathy progression. Among these metabolites, seven, namely leucine, valine, isoleucine, pseudouridine, threonine, and citrate, were found to be associated with the overall progression. Additionally, 2-hydroxyisobutyric acid was linked not only to the overall progression but also to the progression from normoalbuminuria (Mutter et al., 2022).

### **2.2.3 Development of analytical technologies**

The application of metabolomics in large-scale epidemiological research requires efficient and advanced metabolic analysis techniques. The most common of such employed techniques are NMR spectroscopy and mass spectrometry. Table 1

summarizes the advantages and limitations of the application of NMR and MS in metabolomics (Babushok et al., 2007; Chan et al., 2011; Emwas et al., 2013, 2016, 2019; Emwas, 2015; Kind et al., 2009; Rainville et al., 2017; Tynkkynen et al., 2019). Additionally, other more specialized techniques are applied (for example, Fourier transform infrared spectroscopy and array electrochemical detection) in a number of cases (Lindon & Nicholson, 2008). Each analytical technique has its own strengths and weaknesses; the selection of method mainly depends on the focus of the research and the nature of the sample (Emwas, 2015). However, due to the great diversity in metabolite chemical properties and the wide dynamic range of metabolite concentrations, currently no single technique can capture complete metabolic information (Chan et al., 2011). A combination of multiple analytical methods contributes to improve metabolite coverage (Elliott et al., 2015; Kurbatova et al., 2020).

Because MS is highly sensitive (picomolar to femtomolar detection levels) and versatile, it has the ability to measure a wider range of metabolites (Laíns et al., 2019). MS works by forming positively or negatively charged species (ions) from analytes of interest. It separates ions based on their mass-to-charge ratio ( $m/z$ ) and detects ions (Dunn, Broadhurst, Atherton, et al., 2011). MS is usually coupled with separation techniques such as LC, GC, or CE. Biological samples containing complex metabolites can be separated by the separation techniques described above to significantly reduce their complexity before entering the mass spectrometer, thereby increasing the number of metabolites that can be detected (Johnson et al., 2016). Different types of biological samples have different sample preparation procedures. For example, serum or plasma samples have a complex matrix consisting of low molecular weight organic and inorganic chemicals (metabolites) and other high molecular weight species (proteins and RNA). Thus, sample preparations of them for LC or GC analysis involves a deproteinization step to remove high molecular weight species (Dunn, Broadhurst, Atherton, et al., 2011). In contrast, the urine samples obtained from healthy individuals contain very low protein and have simple preparation steps, usually involving dilution and analysis. However, since the high abundance of urea in urine is detrimental to GC-MS instrumentation and data quality, urine analysis by GC-MS should be performed after urease treatment removing high concentrations of urea (Dunn, Broadhurst, Atherton, et al., 2011). The higher resolution and dynamic range of MS make it widely used in the study of large-scale population metabolome (Ala-Korpela et al., 2022; Ala-Korpela & Davey Smith, 2016; Deelen et al., 2019; Huynh et al., 2019; Würtz et al., 2017). However, since MS only detects readily ionized metabolites,

more than 40% of the chemical substances cannot be observed by MS (Peregrín-Alvarez et al., 2009; Rappaport et al., 2014). Moreover, the use of additional separation methods, involving complex sample preparation, may result in changes in the metabolome that are not biologically relevant, which reduces reproducibility and increases cost and time of analysis by MS platforms (Marshall & Powers, 2017).

NMR spectroscopy is non-destructive, non-selective, highly reproducible, and easily quantifiable. It requires minimal sample handling and allows the routine identification of novel compounds (Emwas et al., 2019). NMR is a physical process that probes the intrinsic spin properties of atomic nuclei by strong constant magnetic field. NMR results rely on the specific magnetic properties of certain atomic nuclei, such as  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ , and  $^{31}\text{P}$  (Elbaz et al., 2015; Kaplan et al., 1990; Ye et al., 2009). NMR works with any kind of sample containing spin nuclei, and the most common types are  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectroscopy. The naturally high abundance and high intrinsic of  $^1\text{H}$  atoms make  $^1\text{H}$  NMR spectroscopy the most widely used. In the present thesis, NMR refers to  $^1\text{H}$  NMR spectroscopy unless otherwise indicated.

The high automation and exceptional reproducibility of NMR give this method many advantages over other analytical techniques in large-scale and long-term metabolomic studies such as epidemiological research (Emwas et al., 2019). Similar to MS, NMR-based metabolomics studies typically use biofluids as main sources of data, especially urine and serum or plasma, which are important in animal and human research (Lind et al., 2016). In NMR experiments, plasma is provided by collecting whole blood into commercially available anticoagulant-treated tubes followed by centrifugation. Alternatively, blood can be collected in an additive-free vacutainer, placed at a low temperature to coagulate, and centrifuged to result in serum (Beckonert et al., 2007). As for urine samples, these are collected into tubes containing antibacterial additives such as sodium azide and sodium fluoride to prevent possible bacterial contamination (Chetwynd et al., 2017; Khamis et al., 2017). During sample preparation of NMR, the urine is mixed with a phosphate buffer (pH=7.4) to stabilize the urinary pH, thereby avoiding chemical shift variations due to pH differences in NMR spectroscopy (Beckonert et al., 2007). Limited by low sensitivity, NMR can only detect moderate to high abundant metabolites ( $\geq 1\mu\text{M}$ ) (Khamis et al., 2017). However, its sensitivity could be improved by multiple scans (time), strong superconductive magnets cryo- and microprobes, and dynamic nuclear polarization (Emwas, 2015). NMR has the potential to provide fully automated high-throughput experiments in a cost-effective manner, which would be crucial for the application of urine metabolomics

in large-scale epidemiology. In this thesis, a quantitative high-throughput urine NMR metabolomics platform has been applied to offer a comprehensive quantitative approach for urine analysis (Tynkkynen et al., 2019). Next, this platform will be briefly introduced.

### *A quantitative high-throughput urine NMR metabolomics platform*

A quantitative high-throughput urine NMR metabolomics platform has been developed (and is under development) as a collaborative effort between the University of Oulu and the University of Eastern Finland (Tynkkynen et al., 2019). It includes all experimental details regarding sample preparation and NMR spectroscopy.

Urine sample preparation: The protocol for preparing urine samples involves the use of an automated liquid handler (PerkinElmer JANUS 8-tip Automated Workstation) with a capacity to process approximately 100 urine samples per hour. Initially, the frozen urine samples, stored at  $-80^{\circ}\text{C}$ , are thawed overnight in a refrigerator at  $+4^{\circ}\text{C}$ . They are then gently mixed (inverted 3–5 times) and centrifuged ( $3500 \times g$ , 5 min,  $+4^{\circ}\text{C}$ ). To address the substantial variability in pH, ionic strength, multivalent cations concentrations, and metabolite composition in urine samples, 70  $\mu\text{l}$  of the phosphate buffer (1.5 M potassium dihydrogen phosphate, 0.2% sodium azide, 5.8 mM TSP in deuterium oxide, pH 7.0) is added to 96 deep-well plates to minimize alterations in pH and ionic strength. TSP (2, 2, 3, 3-tetradeutero-3-(trimethylsilyl)-propionic acid) is employed as a reference for chemical shifts and internal concentration. Subsequently,  $2 \times 315 \mu\text{l}$  of centrifuged urine samples are transferred to deep well plates and mixed using aspiration and dispensing of samples ( $3 \times 450 \mu\text{l}$  by JANUS). The samples in the 96 deep well plate are then centrifuged ( $3500 \times g$ , 5 min,  $+4^{\circ}\text{C}$ ), after which 520  $\mu\text{l}$  of the centrifuged sample is moved to 5 mm 96 NMR tubes using JANUS. For a more detailed guide on sample preparation, refer to a previously published work (Tynkkynen et al., 2019). At present, this procedure is tailored exclusively for 5 mm NMR sample tubes. However, if sample volume becomes a potential constraint in future applications, there's flexibility to adjust the volume and alter the approach for 3 mm NMR sample tubes. From an analytical chemistry standpoint, any type of urine sample (i.e. spot urine, overnight or 24-hour collection) can be employed for analysis.

NMR measurements: NMR measurements are conducted utilizing a 600MHz Bruker NMR spectrometer fitted with a cryoprobe (Bruker Prodigy TCI 600 S3

H&F-C/N-D-05 Z) and an automatic cooled SampleJet sample changer. The use of a 600 MHz spectrometer helps to reduce, though not eliminate, signal overlap among urine metabolites. Standard water-suppressed measurements are employed in this process. With this hardware set-up, over 200 urine samples can be automatically processed and analysed within a 24-hour timeframe. Depending on the signal multiplicity and the number of protons contributing to the measured signal, typical detection limits under these experimental conditions are approximately 1  $\mu\text{mol/L}$  (or 0.2  $\mu\text{M/mM}$  creatinine). A comprehensive NMR measurement protocol and parameters can be found in earlier publications (Tynkkynen et al., 2019). The approach applied in developing this high-throughput quantitative urine NMR metabolomics platform was tailored for maximizing efficiency and cost-effectiveness.

The quantitative high-throughput urine NMR metabolomics platform lays the foundation for the urine metabolomics application at an epidemiological scale, which provides detailed quantitative data of urine metabolites with the potential to be utilized as molecular probes to assess kidney function and its corollaries in various metabolic conditions.

**Table 1. The characteristics of MS and NMR spectroscopy in metabolomics studies.**

Characteristics	MS	NMR spectroscopy
Sensitivity	High, detection limit reach nanomolar	Low (millimolar to micromolar) but can be improved with multiple scans, higher field strengths, cryo- and microprobes, and dynamic nuclear polarization.
Selectivity	Can be used for both selective and non-selective (targeted and nontargeted) analyses, depending on the sample preparation and separation methods.	Usually non-selective. A universal detector for all hydrogen-containing metabolites.
Reproducibility	Moderate.	Very high.
Sample volume	Low. Typically 10–100 $\mu\text{L}$ .	Relatively high. Usually 100–600 $\mu\text{L}$ , most NMR spectroscopists use 5 mm tubes, the less sample volume is possible if using 3 mm or even 1.7 mm tubes.

Characteristics	MS	NMR spectroscopy
Sample preparation	More demanding. Require different columns and optimization of ionization conditions, e.g. for GC-MS, preparing urine samples involves removing urea through urease treatment, precipitating proteins with methanol, and conducting trimethylsilyl derivatization (Chan et al., 2011).	Minimal sample preparation, can be automated, e.g. transfer the urine sample to an NMR tube and add a phosphate buffer (pH 7.0).
Sample recovery	Destructive. Samples cannot be recovered.	Non-destructive. Multiple analyses can be performed on the same sample, ease of sample recovery depends on selection of solvent and recovered sample can be stored for a long time.
Number of detectable metabolites	From dozens to thousands of metabolites, depending on the choice of ionization mode and mass analyser. e.g. Nearly ca. 20 000 features were seen for urine using LC-IMS-MS (Rainville et al., 2017).	From dozens to a few hundred metabolites, depending on spectral resolution and biological sample types. e.g. 122 urine metabolites were identified by the high-throughput urine NMR metabolomics platform (Tynkkynen et al., 2019).
Identification and quantification	Identified based on internal standards or mass spectral library, such as NIST (Babushok et al., 2007) and FiehnLib (Kind et al., 2009). Accurate quantification requires molecular class-specific standards. In untargeted experiments, most peaks cannot be identified and quantified.	Identify metabolites using chemical shifts and/or coupling patterns. e.g. quantifications rely on the CTLS fitting analysis tools and automated quantitative regression-based spectral analysis (Tynkkynen et al., 2019). Essentially all peaks can be identified and quantified. However, variation of chemical shifts is challenging.

Characteristics	MS	NMR spectroscopy
Sample analysis time	Longer than NMR, depending on different separation techniques before MS, e.g. for GC-MS, preparation of urine sample is time-consuming (depletion of urea and extraction: ~1.5 hour for 40 samples, derivatization: ~3.5 hour), and the time for data analysis: ~2 hour for 40 samples (depending on computer configuration and complexity of data) (Chan et al., 2011).	Fast. Sample can be analysed in one measurement. e.g. Analysing over 200 urine samples is feasible within approximately 10 hours for one 96 NMR tube rack during the NMR experimentation process (Tynkkynen et al., 2019).
Tissue samples	Requires tissue extraction.	Tissues can be analysed directly by high-resolution magic-angle sample spinning NMR.

## 2.3 Epidemiological application of urinary metabolomics

The potential of urine metabolites in epidemiology and translational medicine has been recognized for quite some time (Bouatra et al., 2013; Elliott et al., 2015; Emwas et al., 2016; Holmes et al., 2008; Tynkkynen et al., 2019). The molecular content of urine is physiologically related to glomerular filtration and molecular reabsorption processes in the kidney, and reflects several key biochemical pathways associated with cardiometabolic status, gut microbial metabolic activity, and dietary characteristics. Thus, detailed quantitative data on urinary metabolites allow the assessment of renal function and its inference under various metabolic conditions. However, most of the existing urinary metabolomics research for humans only involves limited sample sizes ranging from tens to hundreds of individuals. Many of these studies usually use analytical methods without a rigorous assessment, which is misleading in analysis and interpretation. In the following sections, the consideration of urine sampling, data pre-processing and statistical analysis in urine epidemiology will be introduced. In addition, published literature on appropriate applications of urine metabolomics in epidemiology will also be summarized.

### 2.3.1 Urine sampling

The composition of the urine metabolome is complex and diverse and is easily changed due to stimulation by external factors such as pharmacological effects,

dietary intake, and diurnal rhythms (Chan et al., 2011; Slupsky et al., 2007). When urine samples are collected, different collection timing can lead to considerable differences in the quality and quantity of urinary metabolites (Holmes et al., 2008; Lau et al., 2018). Urine samples are typically collected in three types: first morning void, spot urine, and 24-hour urine collection (Chetwynd et al., 2017). Ideally, the 24-hour collection represents a full circadian cycle, reducing the effects of any circadian variations, and the use of 24-hour urine collections would greatly reduce the dilution effect of urine samples. However, collecting 24-hour urine samples for clinical research is often not feasible (Kohler et al., 2016; Lains et al., 2019). On the one hand, the metabolome of 24-hour urine suffers from a lack of or variable cooling during sampling. Moreover, the urine metabolome in the sample collector could be changed by the influence of living, metabolic active cells or released enzymes. On the other hand, 24-hour urine collection is often prone to error due to a lack of supervision by medical experts (Liu et al., 2020). Spot urine is defined as urine collected at a particular time point of the day, and it might be used when a number of interventions have been taken such as diet or medication (Chetwynd et al., 2017; Holmes et al., 2008). Generally, spot urine samples should be avoided in metabolomics, since a variety of factors such as the collection time of day, physical activity, and diet consumption can highly affect the individual metabolite pattern in the spot urine sample (Liu et al., 2020). To minimize the aforementioned issues, first morning void is the preferred sample type in most cases (Lau et al., 2018; Li, Ihanus, et al., 2022; Titan et al., 2019; Tynkkynen et al., 2019), since it reduces the effects of diet or medication, particularly following an overnight fast. This type of sample is recommended by a white paper on metabolomics research from the perspective of a community (Kirwan et al., 2018). However, the perfect urine sample for biomarker research is hard to achieve. Therefore, due to the complexity and sensitivity of the urine metabolome, any novel urinary biomarkers should be treated with caution until further confirmation.

### ***2.3.2 Data pre-processing and statistical methods***

Appropriate data pre-processing and statistical analysis are crucial to obtain high-quality data in metabolomics research. Data pre-processing is necessary for metabolomics data before statistical analysis due to the variation in the data generated from metabolomics experiments, such as differences in the orders of magnitude between measured metabolite concentrations, technical variation, and heteroscedasticity (van den Berg et al., 2006). Data pre-processing steps typically

include normalization, missing value imputation, outlier detection and removal, scaling, and transformations (Dunn, Broadhurst, Atherton, et al., 2011; van den Berg et al., 2006).

A key issue in the analysis process of urine metabolomics data is the normalization step. Urine, as the waste biological fluid produced by the continuous filtration of plasma by the kidneys, is different from the plasma metabolome under strict physiological regulation, and its chemical properties are not controlled. For example, urine sample volumes and metabolite concentrations can be largely varied even within the same individual (Tynkkynen et al., 2019). Although the dilution effect of the urine sample can be greatly or partly reduced by collecting a 24-hour urine sample or a morning spot sample (Vart & Grams, 2016), these types of urine collections are usually not feasible. Therefore, an additional normalization step is required during data analysis.

Currently, a variety of normalization methods, including creatinine ratio, specific gravity, osmolality, urine volume, probabilistic quotient normalization (PQN), and area under the curve, are often applied to account for the dilution effects of urine samples (Cook et al., 2020; Wu & Li, 2016). The most widely accepted approach of normalization is referenced to urinary creatinine concentrations (Mathé et al., 2014; Mutter et al., 2022; Tynkkynen et al., 2019), which is based on relatively constant rate of creatinine excretion via glomerular filtration at a wide range of GFR (Waikar et al., 2010). Creatinine is freely filtered and not reabsorbed in the kidneys (Eaton & Pooler, 2018), and when calibrated by urinary creatinine, can overcome confounding from a urine volume variation. In fact, creatinine excretion is affected by sex, age, ethnicity, muscle mass and pathophysiological state (Forni Oagna et al., 2015; Waikar et al., 2010). Consequently, it could be beneficial to explore alternative normalization approaches. This would help assess the reliability of using creatinine as a reference across under different biological assumptions. Additionally, conducting a comparative analysis of diverse normalization methods, while considering factors like physiology, metabolism, and renal function, would contribute to a more nuanced understanding of the associations among urinary metabolites in epidemiological investigations.

Statistical analysis performed in the metabolomics includes univariate statistical analysis and multivariate statistical analysis (MVA). Univariate analysis accepts only one random variable at a time, such as analysis of variance (ANOVA), Student's t-test, and non-parametric Kruskal-Wallis (Dunn, Broadhurst, Atherton, et al., 2011). In contrast, MVA includes unsupervised and supervised methods (Emwas et al., 2013). Unsupervised techniques aim to analyse and cluster

unlabelled datasets to discover hidden patterns or data groupings without human intervention, such as principal component analysis (PCA) (Jolliffe & Cadima, 2016), hierarchical cluster analysis (Nielsen, 2016), and self-organizing map (Gao et al., 2019). PCA is one of the most commonly used unsupervised methods and it is typically utilized as a data overview tool to capture the most variability in datasets and to identify outliers or hidden biases in the research (Khamis et al., 2017). Supervised learning is defined by its use of labelled datasets to classify data or to predict outcomes. Supervised methods are broadly applied in the identification of biomarkers, classification, and prediction (Emwas et al., 2013). Among the most commonly used supervised techniques are discriminant analysis (i.e. linear discriminant analysis, partial least squares for discriminant analysis [PLS-DA], orthogonal projections to latent structures discriminant analysis [OPLS-DA]) (Wold et al., 2001), regression analysis (e.g. multiple linear regression, multiple logistic regression, principal component regression) (Antoniewicz et al., 2006), and the support vector machine (Cortes & Vapnik, 1995).

Multivariate approaches are particularly important for metabolomics, since a single biomarker will not be amply specific for a given condition by itself (Madsen et al., 2010). However, the common solution for multivariate methods (e.g. OPLS-DA) has long (mis)guided metabolomics studies, especially the overfitting of classification models (with a high number of variables and very small numbers of participants) and a scarcity of independent replications (Ala-Korpela, 2007, 2008, 2016; Ala-Korpela et al., 2012; Ala-Korpela & Davey Smith, 2016; Madsen et al., 2010; Soinen et al., 2015; Xia et al., 2013). Essentially, many of the limitations in the use of multivariate methods are the result of a combination of misinterpretation and uncritical push-the-button analyses. These simple classification methods, usually aimed at separating 'healthy' individuals from 'disease' individuals, are not optimal for metabolic understanding in epidemiology, and we should accept the inevitability of the health-disease continuum rather than being hampered by a clearly unattainable black and white diagnosis (Ala-Korpela et al., 2012; Kjeldahl & Bro, 2010). Regardless of the chosen method, both statistical and biological validation are critical. Additionally, the self-critical interpretation of statistical results, biological significance, and potential clinical application is important to metabolomics researchers (Ala-Korpela, 2018; Ala-Korpela & Davey Smith, 2016).

The abuse of dichotomies also manifests itself in the use of p values to determine whether a finding refutes or supports a scientific hypothesis. For example, outcomes are arbitrarily labelled as either 'statistically significant' or

'statistically non-significant' solely based on the p value (typically using the widely adopted threshold of  $p < 0.05$ ). However, it's important to note that all statistical measures, including both p values and confidence intervals, naturally exhibit variability from one study to another, sometimes to an unexpected degree. This means that random fluctuations alone can lead to substantial disparities in p values. Therefore, in the results section, precise p values should be provided, and regardless of what the statistics show, the possible explanations for the results should be proposed. Inferences should be scientific and factors such as study design, data quality, background evidence, and comprehension of underlying mechanisms are often more important than statistical measures such as p values or intervals (Amrhein et al., 2019; Sterne & Davey Smith, 2001).

### **2.3.3 Epidemiological application of urinary metabolomics**

#### *Literature research and exclusion criteria*

In this section, I present examples of the epidemiological application of urinary metabolomics that describe the relationship between urinary metabolites and general characteristics of population, as well as the risk of disease factors. Representative papers were selected from the search conducted in PubMed until 30 March 2023 with the following search terms ((urine OR urinary OR urine-based) AND (metabolite OR metabolites OR metabolome OR metabolomic OR metabolomics OR metabolic OR biomarker OR biomarkers)), and the records were restricted to studies in humans. The records were screened for eligible studies based on the following criteria. Research targeting individual molecules or individual groups of molecules with specific methods were excluded, only including studies on comprehensive metabolic profiling of multiple molecular pathways. Studies devoted to exogenous substances were also excluded; we only retained studies focusing on endogenous metabolites. Finally, because few urinary metabolomics studies meet the requirements of absolute metabolite quantification and the large sample size for appropriate epidemiological applications, I chose 1 000 individuals as the inclusion threshold based on recent power calculations (Nicholson et al., 2011). Table 2 summarizes the large-scale applications of quantitative urine metabolomics (n=20) that include around or more than 1 000 individuals with absolute quantification of urinary metabolites.

**Table 2. Large-scale applications of quantitative urine metabolomics.**

Focus	Number of Study Participants	Sample type	Analytical method	Normalization	References
Human metabolic phenotype diversity	n=4 630 from the INTERMAP study: China (n=832), Japan (n=1 138), UK (n=496), and USA (n=2 164).	24-hour urine	<sup>1</sup> H-NMR (600 MHz)	PQN	(Holmes et al., 2008)
Lung cancer	n=1 005 (469 cases and 536 controls) for training set and n=158 (80 cases and 78 controls) for validation set; n=48 for tissue set.	Spot urine and tissue	UPLC-ESI-QTOFMS	Creatinine	(Mathé et al., 2014)
Inborn errors of metabolism	n=989 neonates recruited from 14 birth centres from all over Turkey.	Urine samples obtained within the first week of life using a standard paediatric urine collection bag	<sup>1</sup> H-NMR (500 MHz)	Creatinine	(Aygen et al., 2014)
Sex differences related with risk of diabetes	n=2 709 from the SHIP study.	Spot urine	<sup>1</sup> H-NMR (400 MHz)	Creatinine	(Friedrich et al., 2015)
Adiposity	n=2 324 from the INTERMAP study: USA (n=1 880) and UK (n=444).	24-hour urine	<sup>1</sup> H-NMR (600 MHz) and IEC	PQN	(Elliott et al., 2015)
Biological age	n=3 611 from the SHIP-0 for discovery and n=850 from SHIP-TREND for replication.	Spot urine	<sup>1</sup> H-NMR (400 MHz)	PQN and Creatinine	(Hertel et al., 2016)
Inborn errors of metabolism	n=23 140 cases between across the Indian.	Spot urine	GC-MS	Creatinine	(Hampe et al., 2017)
Thyroid function	n=5 620 from the Inter99 study and n=3 788 from the Health2006/Health2008 study	Urine	<sup>1</sup> H-NMR (400 MHz)	Creatinine	(Friedrich et al., 2017)

Focus	Number of Study Participants	Sample type	Analytical method	Normalization	References
Chronic kidney disease	n=386 from the Framingham Heart Study (193 cases, 193 controls) for discovery and n=998 from the ARIC study for replication.	Spot morning urine	LC-MS	Creatinine	(McMahon et al., 2017)
Serum osteocalcin action in non-diabetic	n=931 from SHIP-TREND study.	Spot urine and plasma	LC-MS/MS	PQN	(Entenmann et al., 2017)
Chronic low-grade inflammation	n=925 from SHIP-TREND study.	Spot urine and plasma	<sup>1</sup> H-NMR (400 MHz) and LC-MS/MS	PQN	(Pietzner et al., 2017)
Free thyroxine	n=952 from SHIP-TREND study.	Spot urine and plasma	<sup>1</sup> H-NMR (400 MHz) and LC-MS/MS	PQN	(Lange et al., 2018)
Glucose homeostasis	n=3 986 at both baseline and 5-year follow-up from Inter99 study.	Spot urine	<sup>1</sup> H-NMR (400 MHz)	PQN	(Friedrich et al., 2018)
Children metabolome	n=1 192 children from the HELIX study: UK (n=199), France (n=157), Spain (n=207), Lithuania (n=201), Norway (n=229) and Greece (n=199).	Two urine samples, collected at last night-time and first morning voids before the clinical examination, were pooled as a representative sample of the last 24 hour, and serum	<sup>1</sup> H-NMR (600 MHz) and LC-MS/MS	PQN	(Lau et al., 2018)
High-throughput urine NMR metabolomics platform	n=995 from the NFBC1966 cohort.	Spot morning urine and serum	<sup>1</sup> H-NMR (600 MHz)	Creatinine	(Tynkkynen et al., 2019)
Fish consumption and cardiovascular risk	n=4 680 from the INTERMAP study.	24-hour urine	<sup>1</sup> H-NMR (600 MHz)	PQN	(Gibson et al., 2020)

Focus	Number of Study Participants	Sample type	Analytical method	Normalization	References
Type 2 diabetes	n=789 from the PIVUS study (108 prevalent cases of T2D) and n=635 from the ULSAM study (89 cases of prevalent T2D).	24-hour urine	UPLC-QTOFMS	PQN	(Salihovic et al., 2020)
Diabetic nephropathy in type 1 diabetes	n=2 670 individuals with type 1 diabetes from the FinnDiane study.	24-hour urine	<sup>1</sup> H-NMR (600 MHz)	Creatinine	(Mutter et al., 2022)
Diet quality in European children	n=1 147 children from the HELIX project.	Two urine samples, collected at last night-time and first morning voids before the clinical examination, were pooled as a representative sample of the last 24 hour.	<sup>1</sup> H-NMR (600 MHz)	Creatinine	(Stratakis et al., 2022)
Diabetic kidney disease progression	n=995 participants with diabetes from the CRIC study.	24-hour urine	Flow Injection-TOF MS	Creatinine	(Zhang et al., 2022)

## *Non-disease association studies*

**General characteristics:** A total of six studies on the associations between general characteristics of the population and urinary metabolites are enrolled, including the effects of blood pressure (Holmes et al., 2008; Li, Tynkkynen, et al., 2022), body mass index (BMI) (Elliott et al., 2015; Lau et al., 2018; Li, Tynkkynen, et al., 2022; Tynkkynen et al., 2019), age (Hertel et al., 2016; Lau et al., 2018), and sex (Lau et al., 2018) on the metabolic profile of urine.

In the INTERMAP study (Stamler et al., 2003), a large cross-sectional research involved 4 630 men and women aged between 40 and 59 from 17 population samples in four countries (China, n=832; Japan, n=1 138; the UK, n=496; and the USA, n=2 164.). Four discriminatory metabolites (alanine, formate, hippurate and N-methylnicotinate) were quantified from the NMR spectra of 24-hour urine specimens. In multiple linear regression models (model 1 is adjusted for sex, age, BMI, sample, supplement use, special diet, cardiovascular disease or diabetes mellitus diagnosis, family history of high blood pressure, physical activity; model 2 is model 1 without BMI; model 3 is model 1 plus 7-day alcohol [g/24-hour], urinary Na<sup>+</sup> [mmol/24-hour], urinary K<sup>+</sup> excretion [mmol/24-hour]; model 4 is model 3 without BMI, four models per metabolite applied for each of systolic and diastolic blood pressure [SBP and DBP]; in total eight models), urinary formate (in all eight models) and hippurate (in six models) were negatively associated with both SBP and DBP, and the positive association of alanine with blood pressure was observed in five models (Holmes et al., 2008).

Another study based on the INTERMAP population includes 1 880 individuals from the USA and 444 individuals from the UK as a replication cohort. Using partial correlation analyses, twenty-nine urinary metabolites identified by NMR were associated with the body mass index (BMI) in 3 models (model 1 is adjusted for sex, age, and sample; model 2 is adjusted for all factors in model 1 plus physical activity, cardiovascular disease history, prescribed lipid-lowering drugs, medication for hypertension, nonsteroidal anti-inflammatory drug [NSAID] use, special diet, dietary supplement use, smoking, total energy intake per day [kcal/day], and education; model 3 is adjusted for all factors in models 1 and 2 plus 24-hour urinary creatinine). Twenty-five of these species were also found in the UK validation cohort. As well as this, 16 in 22 amino acids identified by the targeted analysis using ion exchange chromatography (IEC) were associated with

BMI after the Bonferroni correction for multiple testing (models 1 and 2) (Elliott et al., 2015).

The HELIX study (Maitre et al., 2018), a joint effort spanning six established and longitudinal birth cohorts in Europe, involved the analysis of 24-hour urine samples from 1 192 children who were monitored from December 2013 to February 2016. Using NMR spectroscopy, the study investigated four urinary metabolites linked to BMI through linear regression models that were adjusted for variables including analytical batching, age, sex, dietary intake habits across 11 food groups, and the type of sampling (night only, morning only, or pooled sample). Urinary 4-deoxyerythronic acid (metabolite standard deviation [SD]/zBMI: 0.21, 95% confidence interval [CI] 0.16 to 0.26) and valine (metabolite SD/zBMI: 0.09, 95% CI 0.04 to 0.15) displayed positive correlations with BMI, while urinary p-cresol sulphate (metabolite SD/zBMI: -0.10, 95% CI -0.16 to -0.05) and pantothenate (metabolite SD/zBMI: -0.12, 95% CI -0.17 to -0.07) were inversely associated with BMI. In terms of metabolic associations with sex, urinary isoleucine (-0.24 SD lower, 95% CI -0.37 to -0.12) was found at lower concentrations while 5-oxoproline (0.23 SD higher, 95% CI 0.11 to 0.36) and tyrosine were elevated (0.43 SD higher, 95% CI 0.31 to 0.55) in males. Furthermore, a common trait observed across six distinct study cohorts was a positive correlation between urinary creatinine level (0.39 SD higher, 95% CI 0.26 to 0.53) and age was recognized as a common characteristic across six different study cohorts (Lau et al., 2018).

In the study conducted with participants from the Northern Finland Birth Cohort (NFBC) 1966 (Nordström et al., 2022), 4 549 morning spot urine samples were gathered from individuals at the age of 46. Using the quantitative high-throughput urine NMR metabolomics platform, 43 urine metabolites were quantified from 1004 participants within the NFBC1966 study. For each quantified urinary metabolite, a linear regression model, adjusted for sex, was applied with BMI as the independent variable. The findings indicated that BMI showed a negative association with urinary p-cresol sulphate and hippurate, and while it displayed a positive association with 2-hydroxyisobutyrate, as well as branched-chain amino acids such as isoleucine and valine. In addition, aromatic amino acids tryptophan and tyrosine exhibited a positive association with BMI (Tynkkynen et al., 2019).

In the aforementioned studies, we note multiple concordant associations for BMI: BMI was positive with urinary valine but negative with urinary p-cresol sulphate. Additionally, in the INTERMAP study and NFBC1966 study, BMI was negatively associated with urinary hippurate, formate and 2-hydroxyisobutyrate,

and positively associated with glucose, isoleucine, as well as the aromatic amino acids tryptophan and tyrosine. Consistent results across different studies suggest the potential of urinary metabolomic data to provide additional molecular insights for large-scale epidemiological studies.

In the Study of Health in Pomerania (SHIP-0) (Völzke et al., 2011), participants consisted of 212 157 adult German residents in north-eastern Germany, and baseline examinations were performed between 1997 and 2001. A total of 3 611 individuals in SHIP-0 were characterized by urine (spot urine, non-fasting samples) NMR spectra. A metabolic assessment of biological age was created through sex-stratified multivariable linear regressions, employing 59 urine metabolites as predictors and chronological age as an outcome. The metabolic age score was validated in two large independent population-based samples. In SHIP-0, the metabolic age score consistently demonstrated a strong correlation with the chronological age ( $r=0.74$ ,  $p<1.23\times 10^{-269}$  in men;  $r=0.78$ ,  $p<3.32\times 10^{-277}$  in women). In the SHIP-TREND ( $n=850$ ) study, the new sample from the same area was independently selected, comparable assessments to those conducted in SHIP-0 were carried out between 2008 and 2012. In both men and women, the metabolic age score demonstrated a notably high correlation with chronological age ( $r=0.53$ ,  $p<2.45\times 10^{-32}$  in men;  $r=0.61$ ,  $p<1.06\times 10^{-51}$  in women). Additionally, the metabolic age score has an independent predictive value for all-cause mortality for survival over approximately 13 years of follow-up of SHIP-0 (sex and age adjusted model: hazard ratio [HR]=1.24, 95% CI 1.14 to 1.35; fully adjusted model: HR=1.18, 95% CI 1.07 to 1.29). Furthermore, it has prognostic value for predicting weight loss in a sample of individuals who have undergone bariatric surgery (Hertel et al., 2016). However, recent longitudinal studies have questioned this type of ageing concept since with extensive longitudinal data, cross-sectionally calculated metabolic ageing was found not to be related to longitudinal metabolic changes (Ala-Korpela et al., 2023).

**Biomarkers:** There are five studies focused on the association between urinary metabolites and clinical biomarkers. Three of the studies are based on the SHIP-TREND study (Entenmann et al., 2017; Lange et al., 2018; Pietzner et al., 2017), and the other two studies are involved with the Inter99 study (Friedrich et al., 2017, 2018), which is a population-based non-pharmacological lifestyle intervention study. In total, 4 511 subjects participated in a 5-year follow-up examination (Jørgensen et al., 2003).

In SHIP-TREND studies, Entenmann et al. (2017) ( $n = 931$  individuals) investigated the interrelations between serum osteocalcin (OCN) and (spot) urine

metabolome as well as plasma metabolome by untargeted LC-MS/MS. OCN has been reported to have metabolic implications such as insulin sensitivity or steroid production (Gao et al., 2016; Mochida et al., 2011). Seventeen urine metabolites were associated with OCN by linear regression analyses controlling for age, sex, waist circumference, physical activity, eGFR, intake of oral contraceptive or hormone replacement therapy and smoking behaviours. Notably, prolylhydroxyproline ( $\beta=0.603$ , 95% CI 0.537 to 0.668) and glycyproline ( $\beta=0.508$ , 95% CI 0.424 to 0.591) as well as the unknown metabolite X-18927 ( $\beta=0.596$ , 95% CI 0.507 to 0.685) in urine showed a positive association with serum OCN.

With regard to studies based on SHIP-TREND, utilizing both MS and NMR technologies, Pietzner et al. (2017) (n=925 individuals) reported linear regression analyses between inflammatory markers such as high-sensitivity C-reactive protein (hsCRP), white blood cell count (WBC), and fibrinogen, and 587 urine as well as 613 plasma metabolites, controlling for age, sex, waist circumference, smoking behaviour, and physical activity. hsCRP, WBC, and fibrinogen showed notable associations with 22, 3, and 16 out of 587 urine metabolites, respectively. In addition, a subset of 10 metabolites achieved a moderately robust classification between healthy participants and an advanced inflammatory state by random forest analysis (median area under the receiver operating curve [AUC]=0.83). In particular, urinary 3'-sialyllactose was found to be a novel metabolite associated with inflammation. Based on the same analytical technique and biological samples (urine and plasma), Lange et al. (2018) (n=952 individuals) found the associations between free thyroxine (FT4) and urine (12 of 587) as well as plasma metabolites (106 of 613) using linear regression analyses controlling for sex, age, waist circumference, eGFR, physical activity, smoking behaviour, and serum alanine aminotransferase activities. However, it should be noted that the aforementioned biomarker studies all use cross-sectional epidemiological data set involving a limited number of individuals, resulting in a sample that is not representative of the population. Moreover, as the lack of independent replication makes it impossible to dissect potential cohort-related bias in the results, the results must be treated with caution.

In Inter99 studies, Friedrich et al. (2017) performed the cross-sectional and longitudinal analysis between urinary metabolites and the serum unbound fraction of FT4 as well as thyroid-stimulating hormone (TSH). Urinary metabolites of participants from Inter99 (baseline: n=6 184; 1year: n=2 482; 3year: n=2 496; and 5year: n=4 511 subjects) and Health2006/Health2008 (n=3 788 subjects),

replication cohort) populations were analysed by NMR spectroscopy. In cross-sectional analyses, linear regression models adjusted for sex, age, BMI, haemoglobin A1c (HbA1c), low-density lipoprotein cholesterol (LDL-C) and SBP revealed associations between serum FT4 levels with five urinary metabolites including alanine (Inter99:  $\beta=-0.267$ , false discovery rate [FDR]<0.01; Health2006/Health2008:  $\beta=-0.147$ , FDR<0.01), glycine (Inter99:  $\beta=-0.062$ , FDR=0.05; Health2006/Health2008:  $\beta=-0.090$ , FDR=0.01), N,N-dimethylglycine (DMG, Inter99:  $\beta=-0.098$ , FDR=0.04; Health2006/Health2008:  $\beta=-0.083$ , FDR=0.04), trigonelline (Inter99:  $\beta=-0.112$ , FDR<0.01; Health2006/Health2008:  $\beta=-0.065$ , FDR=0.02) and lactic acid (Inter99:  $\beta=-0.183$ , FDR<0.01; Health2006/Health2008:  $\beta=-0.140$ , FDR<0.01). And four urine metabolites including dimethylamine (DMA, Inter99:  $\beta=-0.079$ , FDR<0.01; Health2006/Health2008:  $\beta=-0.068$ , FDR=0.01), D-glucose-beta (Inter99:  $\beta=-0.028$ , FDR=0.01; Health2006/Health2008:  $\beta=-0.049$ , FDR<0.01), glycine (Inter99:  $\beta=0.023$ , FDR=0.01; Health2006/Health2008:  $\beta=-0.030$ , FDR=0.02) and lactic acid (Inter99:  $\beta=-0.035$ , FDR=0.01; Health2006/Health2008:  $\beta=-0.064$ , FDR<0.01) were inversely associated with TSH. In longitudinal analyses, mixed linear regression models controlling the same covariates based on Inter99 data revealed that lower levels of urinary alanine ( $\beta=0.223$ , FDR<0.01), DMA ( $\beta=0.201$ , FDR<0.01), glycine ( $\beta=0.073$ , FDR=0.01), lactic acid ( $\beta=0.121$ , FDR<0.01) and DMG ( $\beta=0.128$ , FDR<0.01) were related to a higher decline in FT4 levels over time. In addition, DMA was positively associated with TSH levels in mixed models ( $\beta=0.046$ , FDR=0.04). The detected associations underlined the importance of thyroid hormones in glucose homeostasis.

After one year, this research group reported another longitudinal study: NMR spectroscopy was utilized to analyse urinary metabolites in 3 986 participants from the Inter99 study at both baseline and during a 5-year follow-up. Linear regression analyses, adjusted for sex, age, BMI, SBP, and LDL-C, were employed to identify associations between urine metabolites and changes over the 5-year period in markers of glucose homeostasis, such as fasting glucose, HbA1c and a homeostasis model assessment of insulin resistance (HOMA-IR). Elevated baseline levels of urinary alanine, betaine, DMG, creatinine and trimethylamine were associated with an increase in HbA1c from baseline to follow-up. Conversely, formate and trigonelline levels were associated with a reduction in HbA1c over time. Similar findings were observed in the analyses of 5-year changes in fasting glucose and HOMA-IR index (Friedrich et al., 2018). Both studies involved longitudinal analyses (5 years of follow-up), and the former also performed cross-

sectional analyses and independent replication, enhancing the statistical power of the results.

**Nutrition:** There are three studies focused on nutritional science involved with urinary metabolome. Two different strategies were mainly utilized for nutritional research: (1) find specific or discriminatory molecular signals of individual nutrients; and (2) find discriminatory metabolic panel corresponding to a specific diet pattern.

Lau et al. (2018) summarized associations between urinary metabolites and intakes of 11 main food groups (cereal, meat, fish, lipids, dairy, vegetables, potatoes, fruits, bakery products, sweets, beverages). Multiple linear regression analysis followed by meta-analysis were performed on each metabolite—dietary factor pair after adjusting for analytical batching, sampling, sex, age, and zBMI score. The results showed that creatine was positively associated with meat intake (SD/times per week: 0.025; 95% CI 0.012 to 0.039). Hippurate showed a positive correlation with fruit (SD/times per week: 0.026; 95% CI 0.018 to 0.034) and vegetable intake (SD/times per week: 0.021; 95% CI 0.01 to 0.031). Proline betaine (SD/times per week: 0.017; 95% CI 0.009 to 0.025), N-methylnicotinic acid (SD/times per week: 0.017; 95% CI 0.009 to 0.025) and scyllo-inositol (SD/times per week: 0.014; 95% CI 0.007 to 0.021) exhibited positive associations with fruit intake, whereas glutamine (SD/times per week: -0.013; 95% CI -0.021 to -0.005), alanine (SD/times per week: -0.014; 95% CI -0.022 to -0.006), and leucine (SD/times per week: -0.014; 95% CI -0.022 to -0.006) demonstrated negative associations with fruit intake. Furthermore, pantothenic acid (SD/times per week: 0.009; 95% CI 0.004 to 0.015) was found to be positively associated with dairy intake, while and acetic acid (SD/times per week: 0.045; 95% CI 0.023 to 0.068) showed a positive association with potato consumption.

Gibson et al. (2020) used cross-sectional data from INTERMAP (n=4 680), where fish consumption was assessed by four 24-hour dietary recalls. Partial Pearson correlation coefficients adjusting for sex, age, and centre were calculated using the combined sample population from 4 countries (Japan, China, the UK, and USA). The fish consumption was associated with creatine ( $p < 1 \times 10^{-15}$ ), DMA ( $p = 1 \times 10^{-6}$ ), ethyl glucuronide ( $p < 1 \times 10^{-11}$ ), homarine ( $p < 1 \times 10^{-13}$ ), taurine ( $p < 1 \times 10^{-15}$ ), trimethylamine N-oxide (TMAO,  $p < 1 \times 10^{-23}$ ), and trimethyllysine ( $p < 1 \times 10^{-9}$ ). In particular, the strong association between homarine ( $p < 1 \times 10^{-20}$ ) and shellfish consumption was observed. ROC curves were employed to evaluate the predictive capability of the specific urinary metabolites in distinguishing between total fish intake and separately for shellfish only. In Japan, TMAO and taurine

revealed AUC values of 0.81 and 0.78 for high/low fish intake, respectively, and homarine demonstrated an AUC of 0.80 for high/non-shellfish intake. However, these optimistic models were based on only 3 input metabolites, and several analyses used OPLS-DA prediction scores to examine the predictive ability of these discriminating metabolites without replication, resulting in overfitting of the model. Therefore, the relevant results need to be supported by further study.

Although both the aforementioned studies have focused on discovering diet-specific biomarkers, the “single nutrient” approach may not be sufficient to consider the complex interactions between nutrients in free-living populations. The effects of a single nutrient may be too small to detect and may be confused by the effects of different dietary patterns. Therefore, the use of multiple metabolic profiles associated with the overall diet pattern is a potentially more practical and beneficial solution (Ala-Korpela, 2018; Bogl et al., 2013; Hu, 2002). However, there is a paucity of studies with a sufficient sample size in this field (De Filippis et al., 2016). The following is an atypical dietary pattern study where, instead of directly finding the dietary pattern-related metabolome, it observes correlations between metabolic profiles with the index of dietary pattern.

Stratakis et al. (2022) explored urinary metabolic biomarkers associated with diet quality in a multi-country examination of European children, involving 1 147 participants from six distinct cohorts within the HELIX project. Information on the children's usual habitual diets was gathered through a semi-quantitative food-frequency questionnaire, which included 15 specific inquiries aimed at assessing their adherence to a Mediterranean diet. To assess diet quality of children, the Mediterranean Diet Quality Index in children and adolescents (KIDMED) (Serra-Majem et al., 2004) and the proportion of ultra-processed foods (UPF) (Monteiro et al., 2019) in the overall diet (Monteiro et al., 2019) were applied. A panel of four metabolites such as hippurate, N-methylnicotinic acid, sucrose and urea were related to both diet quality indicators. This metabolite panel demonstrated AUC values of 0.75 and 0.7 for high/low KIDMED and high/low UPF, respectively. This finding indicates that urinary metabolic biomarkers possess the capacity to distinguish the quality of childhood diets and could offer mechanistic insights into the common biological pathways underlying both healthy and unhealthy dietary patterns.

### *Disease association studies*

Urinary biomarkers are of interest in extending our information on physiological processes and disease pathogenesis. A total of eight studies describe associations between urinary metabolites and diseases such as lung cancer (Mathé et al., 2014), inborn errors of metabolism (IEMs) (Aygen et al., 2014; Hampe et al., 2017), diabetes mellitus (Friedrich et al., 2015; Salihovic et al., 2020), and kidney diseases (McMahon et al., 2017; Mutter et al., 2022; Zhang et al., 2022).

**Cancer:** Mathe et al. (2014) collected urine samples from 469 patients with lung cancer and 536 controls. Urinary metabolites profiling was detected by untargeted high-resolution ultra-performance liquid chromatography coupled with electrospray ionisation/quadrupole-time-of-flight-mass spectrometry (UPLC-ESI-QTOFMS). Creatine nucleoside, cortisol sulfate, NANA and 561+ were the top four metabolites associated with diagnostic and prognostic of lung cancer. These biomarkers were quantified at two different time points and further validated in an independent sample set consisting of 80 cases and 78 controls. Creatine riboside ( $p < 0.00001$ ), and NANA ( $p < 0.00001$ ) were elevated in lung cancer cases for training and validation sets and they were also associated with the worse prognosis of lung cancer in all cases (HR=1.81 [ $p = 0.0002$ ], and 1.54 [ $p = 0.025$ ], respectively) as well as increased in tumour tissue ( $p < 0.00001$ ,  $p < 0.003$ , respectively). The present study revealed that creatine riboside and NANA might be robust urinary biomarkers associated with early detection of lung cancer diagnosis and indicative of a poorer prognosis.

**IEMs:** It has been reported that metabolomics has succeeded in turning biomarkers into clinical aspect of IEMs screening (Xia et al., 2013). MS/MS-based screening for IEMs in neonates is now routinely done in most industrialized countries (Chace, 2001). Hampe et al. (2017) screened 23 140 urine samples from Indian children along with relevant demographic and clinical data, subsequent computer-aided analysis of amino acids, organic acids, sugars, and fatty acids by GC-MS. In total, 323 (1.40%) positive cases were observed. An altered pattern of organic acid excretion characteristic of positive cases in the present study were consistent with IEMs (Kimura et al., 1999). The most common diseases are primary lactic acidemia and organic acidemia, followed by aminoacidopathies. Hampe et al. proposed that GC-MS is beneficial for the quick detection, accurate identification, and precise quantification of various urinary biomarkers, and for providing an effective platform for routine and high-risk screening of IEMs. In addition, Aygen et al. (2014) demonstrated that NMR may offer a rapid and versatile alternative. In

their study, they collected urine samples from 989 neonates in 14 clinical centres in Turkey, and 65 urinary metabolites were identified by untargeted and targeted NMR spectroscopy in two different laboratories. The primary objective of the present study was to investigate the range of variation in concentrations and chemical shifts of specific metabolites without clinically relevant findings detected in the urine of Turkish neonates. The secondary aim was to incorporate findings from these healthy neonates into the NMR database, facilitating routine and completely automated screening for IEMs.

While new-born screening technology offers a remarkable chance diagnose and treat these disorders with unprecedented effectiveness, it poses a challenge for the biochemical genetics laboratory. This technology identifies a disease state primarily through patterns of metabolic changes, often in the absence of typical clinical symptoms and physical indicators of the disorder. Due to technically demanding procedures with considerable interlaboratory variability, there is a lack of guidelines in practical applications (Dietzen et al., 2009).

**Diabetes mellitus:** Friedrich et al. (2015) assessed the association of urinary metabolites measured via NMR spectroscopy with incident type 2 diabetes (T2D) in 2 709 individuals (1 353 men and 1 356 women, respectively) from the SHIP-TREND study. Within 5 years, 87 men (6.4%) and 50 women (3.7%) developed T2D. Adjusted for age and waist circumference, logistic regression analysis revealed that 5 urinary metabolites were associated with incident T2D in men, while 16 urinary metabolites were associated with incident T2D in women. In men, these include glucose, glycine, lactate, valine and 4-hydroxyphenylacetate, while in women, these include acetate, alanine, carnitine, creatinine, ethanol, formate, glucose, glycine, glycolate, lactate, DMG, TMAO, tau-methylhistidine, trigonelline, urea and 3-hydroxyisovalerate. ROC analysis showed that compared with the standard model with classical risk factors including age, waist circumference, HbA1c and hsCRP, the model additionally adjusted for the urinary glucose, trigonelline and TMAO, and exhibited superior discriminatory ability between diabetic cases and non-diabetic cases in women (AUC=0.8744 and 0.9025, respectively,  $p=0.019$ ). However, in men, the additional adjustment of urinary metabolites did not improve discrimination between new-onset diabetes and non-diabetic cases. The present study revealed the potential of urine metabolomics as potential non-invasive diabetes screening approach and found pronounced sex differences.

Salihovic et al. (2020) analysed urine samples via UPLC-MS from 789 individuals of the PIVUS study (discovery, 108 prevalent cases of T2D) and 635

individuals of the ULSAM study (replication, 89 cases of prevalent T2D). In the replication sample ULSAM, 2 of 62 initially identified metabolites were correlated with prevalent T2D in logistic regression analyses, adjusted for sex, age (for PIVUS only) and urinary creatinine: 3-hydroxyundecanoyl-carnitine (odds ratio [OR]=0.61, 95% CI 0.47 to 0.79,  $p=1.56\times 10^{-4}$ ) and nonanoylcarnitine (OR=0.71, 95% CI 0.56 to 0.89,  $p=3.11\times 10^{-3}$ ). Moreover, with both cohorts combined without T2D at baseline (n=1 227), the sample was randomly split into a two-thirds training set (n=818) and one-third test set (n=409). Over a maximum of 12 years of follow-up, there were 36 and 10 T2D cases in the training and test sets, respectively. Six out of sixty-two metabolites chosen by least absolute shrinkage and selection operator regression in the training set enhanced the prediction of T2D. The present study proposed urinary 3-hydroxyundecanoyl-carnitine as a biomarker candidate.

**Kidney diseases:** McMahon et al. (2017) examined the correlation between urinary metabolites, genetic variants, and incident CKD in the Framingham Heart Study, involving 193 control subjects and 193 cases, with an average follow-up duration of 9.7 years.. A total of 154 urinary metabolites were measured by LC-MS. Urinary glycine (OR=0.59, 95% CI 0.43 to 0.80) and histidine (OR=0.65, 95% CI 0.50 to 0.85) were associated with the lower risk of CKD in the multivariable analysis adjusted for baseline eGFR and the presence of diabetes, hypertension, and dipstick proteinuria. The relationship between urinary glycine (HR=0.82, 95% CI 0.69 to 0.99) and incident CKD was replicated in the Atherosclerosis Risk in Communities (ARIC) cohort (n=998) via a Cox regression analysis adjusted for sex, age, centre, prevalent hypertension, baseline eGFR and albuminuria, and urine collection after oral glucose tolerance testing. Moreover, genomic association of urine metabolomics identified lysine and NG-monomethyl-L-arginine related to CKD.

Mutter et al. (2022) utilized a high-throughput NMR platform to analyse 51 metabolites in 24-hour urine samples collected from 2 670 patients with T1D from the Finnish Diabetic Nephropathy Study. These individuals were monitored for an average of  $9.0 \pm 5.0$  years until the first indication of progression of diabetic kidney disease (DKD), end-stage kidney disease (ESKD) or study end. Seven urinary metabolites were associated with overall progression of DKD in Cox models that adjusted for sex, age and diabetes duration, baseline glycaemic control, albuminuria and CKD stages: leucine (HR=1.47, 95% CI 1.30 to 1.66), valine (HR=1.38 95% CI 1.22 to 1.56), isoleucine (HR=1.33, 95% CI 1.18 to 1.50), pseudouridine (HR=1.25 95% CI 1.11 to 1.42), threonine (HR=1.27, 95% CI 1.11 to 1.46), and citrate (HR=0.84 95% CI 0.75 to 0.93). 2-Hydroxyisobutyrate was

correlated with overall progression (HR=1.30, 95% CI 1.16 to 1.45) and progression from normoalbuminuria (HR=1.56, 95% CI 1.25 to 1.95). Six amino acids and pyroglutamate were related to progression from macroalbuminuria. These findings revealed new potential urinary biomarkers correlated with progression of diabetic nephropathy and underscore the viability of conducting routine large-scale analyses of urinary metabolites, given the reliability of high-throughput urine NMR metabolomic technology.

Zhang et al. (2022) analysed urine samples (n=995) from Chronic Renal Insufficiency (CRIC) Study for relative metabolite abundance using untargeted flow-injection mass spectrometry, resulting in 698 annotated metabolite ions. Following adjustments for clinical variables, 6 metabolite ions were found to be correlated with eGFR slope, while 99 metabolite ions were associated with ESKD. Subsequently, they conducted a targeted quantitative analysis of 15 amino acid metabolism metabolites from the 99 ions and validated the prognostic value of 13 (of 15). However, whether the present study could help acknowledge risk stratification and treatment strategies for DKD needs to be validated in an independent cohort.

From the aforementioned published large-scale studies of urine metabolomics, it can be seen that these limited studies involve a wide range of research such as general characteristics of the population, biomarkers and diseases, demonstrating the potential of urine sample as a very useful biological resource for epidemiological application. However, most existing studies lack large-scale population-based, comprehensive metabolomics quantitative data and independently validated datasets. Therefore, urine samples contain a wealth of underutilized information for epidemiological and translational applications that deserves further exploration.



### **3 Aim of study**

This thesis focuses on the epidemiological applications of quantitative urinary NMR metabolomics to understand the characteristics of urinary metabolite distribution at the population level and to explore the underutilized epidemiological information in urine samples.

In particular, the aims of the present work are as follows:

1. Distribution of urinary metabolites at the population level (Study I and Study III).
2. Characteristics of normalization methods in quantitative urinary metabolomics (Study II).
3. Clinical and biochemical associations with urinary metabolites (Study I and Study III).



## 4 Methods

### 4.1 Epidemiological cohorts and study populations

In this thesis, three population-based cohorts from Finland were utilized. Informed written consent was obtained from all participants, and the studies were approved by the local ethics committees and conducted in accordance with the Declaration of Helsinki. Urine samples collected from all cohorts were quantified by the quantitative high-throughput urine NMR metabolomics platform. Next, a brief description of the cohorts is introduced.

**Northern Finland Birth Cohorts 1966 and 1986:** The NFBC studies encompass two longitudinal birth cohorts initiated to investigate elements influencing preterm birth and consequent morbidity in the two northernmost provinces of Finland, including Oulu and Lapland. NFBC1966 enrolled 12 058 children born alive into the cohort, representing 96% of all births in the region in 1966 (Nordström et al., 2022). In 2012, when cohort members were 46 years old, all those who were alive and residing in Finland with a known address (n=10 331) were invited to take part in an extensive health assessment that involved both questionnaires and clinical examination. Together, they included items about lifestyle (nutrition, smoking, sleep, and physical activity), social background, diagnosed diseases, medication, psychiatric symptoms, organ-specific symptoms, occupational health and workload, economy, functioning, personal traits, quality of life, family history of diseases and use of health services. Questionnaires were submitted by a total of 7 146 cohort members. As for basic clinical health assessments, they were extended to all cohort members via mail, and were carried out by three teams of research nurses in 36 different towns across Finland (n=5 832). Examinations comprised, for example, anthropometric parameters, brachial blood pressure, physical performance, heart rate variability tests, 12-lead electrocardiogram, tolerance tests and pressure pain threshold, spirometry, skin allergy tests (prick tests), oral glucose tolerance tests (0-, 30-, 60- and 120-min samples), cognitive tests and objective measures of sleep and physical activity. Biological samples (blood, urine, saliva, faeces, and hair) were obtained from all who attended the clinical examination. The serum (n=5 788) and urine (n=4 549) samples were taken after overnight fasting.

NFBC1986 comprised 9 432 children who were born alive between 1<sup>st</sup> July 1985 and 30<sup>th</sup> June 1986 and included in the cohort. The most recent data collection,

involving clinical examinations and questionnaires, took place when the cohort members reached the age of 33 years. This encompassed postal questionnaires for the entire cohort and a health examination for those residing in the city of Oulu and its surrounding areas (250 km from Oulu). These examinations were carried out from May 2019 to December 2020. In total, 3 587 cohort members returned their questionnaires, and basic clinical health examinations were conducted for 1 807 of them. More information on the cohort and the data collection can be found at <http://www.oulu.fi/nfbc>.

**Cardiovascular Risk in Young Finns Study (YFS):** The YFS (Raitakari et al., 2008) is among the largest longitudinal studies investigating cardiovascular risk spanning from childhood to adulthood. Initiated in 1980, the study involved 3 596 subjects participated. Since then, several follow-up studies have been carried out on this cohort, with the latest field study conducted from 2018 to 2020. This phase of the study extended invitations not only to the original YFS participants (G1), but also to their parents (G0) and children (G2). A total of 7 349 participants took part in these three generational field studies. The examinations encompassed comprehensive data collection through questionnaires, physical measurements, and blood tests. Primary study variables focused on serum lipoproteins, obesity indices, insulin, inflammation markers, blood pressure, homocysteine, diet, life-style factors, general health status, alcohol use, smoking status, psychological factors, physical activity, and socio-economic status. Participants (n=474) included in the present study consisted of 182 participants from G1, 161 from G0, and 131 from G2. More information on the cohort can be found at <https://youngfinnsstudy.utu.fi/index.html>.

The population data used in each studies are as follows: Study I explored quantified urinary glucose data from both NFBC1966 (n=4 482) and NFBC1986 (n=1 010), as well as 51 clinical biochemical measures available for corresponding participants from NFBC1966 (n=4 275). Study II utilized 44 quantified urinary metabolites and two representative clinical measures such as BMI and MAP in a random subset of 994 individuals from NFBC1966. Study III employed the 61 quantified urinary metabolites in a random subset of 994 individuals from the NFBC1966. The automated quantified data of 13 urinary metabolites and extensive clinical data from NFBC1966 (n=4 505), NFBC1986 (n=1 010) and YFS (n=474) were also investigated. The populations and study designs of studies I-III are summarized in Table 3, while the specific information about clinical and biochemical measures and the quantified urinary metabolites employed in each study will be illustrated in 4.1.1 and 4.1.2.

**Table 3. The study designs.**

Studies	Cohorts	Individuals and women (%)	Clinical and biochemical measures	Covariates	Quantified urinary metabolites	Normalization methods
<b>Study I:</b>						
There is always glucose in normal urine						
	NFBC1966	n=4 482 (57)	n=51	Sex	Urinary glucose	Creatinine
	NFBC1986	n=1 010 (58)	-	-		
<b>Study II:</b>						
Characteristics of normalization methods in quantitative urinary metabolomics						
	NFBC1966	n=994 (58)	BMI and MAP	Sex	n=44	ABS, IS-CREA, IS-GLUC, IS-UREA, IS-PSEURID, CS, PQN and DESeq2
<b>Study III:</b>						
Clinical and biochemical associations of urinary metabolites						
	NFBC1966	n=4 505 (57) and random subset n=994 (58)	n=49	Sex, age (YFS only), MAP, fasting glucose, fasting insulin, smoking history, total TG, CRP and eGFR	n=13 (including creatinine) for the whole NFBC1966 and n=61 (including creatinine) for the random subset of NFBC1966 (n=994)	Creatinine
	NFBC1986	n=1 010 (58)	n=29		n=13 (including creatinine)	
	YFS	n=474 (62)	n=39		n=13 (including creatinine)	

### 4.1.1 Applied clinical and biochemical measures

Study I performed the regression analysis between urinary glucose and serum glucose, eGFR as well as multiplication of serum glucose and eGFR (using CKD-EPI formula) (Inker et al., 2012) in participants (n=4 275) categorized by a serum fasting glucose concentration: serum fasting glucose<5.6mM (n=2 639) or serum fasting glucose≥5.6mM (n=1 636). In addition, it initially explored the correlation adjusted for sex between urinary glucose and 51 clinical and biochemical measures. The specific clinical and biochemical measures for the whole NFBC1966 and the two fasting glucose subgroups in Study I are illustrated in Table 4.

**Table 4. Characteristics of the NFBC1966 applied in Study I.**

Characteristics	NFBC1966	Fasting glucose <5.6mol/L	Fasting glucose ≥5.6mol/L
Number	4 275	2 639	1 636
Women (%)	57	67	39
BMI	26 (24–29)	25 (23–28)	28 (25–31)
Waist-to-hip ratio	0.91 (0.85–0.97)	0.88 (0.83–0.94)	0.96 (0.90–1.00)
Body fat mass	21 (16–28)	19 (15–25)	24 (18–31)
Body fat (%)	28 (22–35)	28 (22–34)	28 (22–36)
Total body water (L)	39 (34–47)	36 (33–44)	45 (37–50)
Systolic blood pressure (mmHg)	124 (114–135)	121 (112–131)	129 (119–140)
Diastolic blood pressure (mmHg)	84 (77–91)	82 (75–89)	87 (80–95)
Fitness score	74 (69–79)	75 (70–79)	74 (67–79)
Basal metabolic rate (calories)	1 513 (1 359–1 757)	1 439 (1 331–1 665)	1 680 (1 447–1 848)
Grip strength average (kg)	32 (26–45)	30 (25–41)	39 (28–48)
Alcohol intake (g/day)	4.4 (1.0–13.3)	3.8 (0.8–9.7)	6.5 (1.4–16.7)
Leucine (mmol/L)	0.08 (0.07–0.10)	0.08 (0.07–0.09)	0.09 (0.08–0.10)
Isoleucine (mmol/L)	0.05 (0.04–0.07)	0.05 (0.04–0.06)	0.06 (0.05–0.07)
Valine (mmol/L)	0.20 (0.18–0.23)	0.19 (0.17–0.22)	0.22 (0.19–0.25)
Alanine (mmol/L)	0.45 (0.41–0.50)	0.44 (0.40–0.49)	0.48 (0.44–0.53)
Glutamine (mmol/L)	0.57 (0.53–0.61)	0.57 (0.52–0.61)	0.57 (0.53–0.62)
Glycine (mmol/L)	0.29 (0.26–0.33)	0.29 (0.26–0.34)	0.28 (0.25–0.31)
Histidine (mmol/L)	0.07 (0.06–0.08)	0.07 (0.06–0.08)	0.07 (0.06–0.08)
Tyrosine (mmol/L)	0.06 (0.05–0.06)	0.05 (0.05–0.06)	0.06 (0.05–0.07)
Glycated haemoglobin (%)	5.5 (5.2–5.7)	5.4 (5.2–5.6)	5.6 (5.4–5.9)
Fasting insulin (IU/L)	7.8 (5.3–11.5)	6.6 (4.6–9.3)	10.5 (7.1–15)

Characteristics	NFBC1966	Fasting glucose	Fasting glucose
		<5.6mol/L	≥5.6mol/L
Fasting glucose (mmol/L)	5.4 (5.1–5.8)	5.2 (5.0–5.4)	5.9 (5.7–6.1)
Lactate (mmol/L)	1.42 (1.23–1.67)	1.35 (1.18–1.58)	1.52 (1.34–1.79)
Pyruvate (mmol/L)	0.09 (0.08–0.12)	0.09 (0.07–0.11)	0.10 (0.08–0.13)
Glycerol (mmol/L)	0.07 (0.06–0.09)	0.07 (0.06–0.09)	0.08 (0.07–0.10)
Apolipoprotein B (g/L)	1.02 (0.88–1.20)	0.98 (0.86–1.15)	1.09 (0.94–1.27)
Total cholesterol (mmol/L)	5.61 (4.97–6.33)	5.58 (4.97–6.26)	5.64 (4.98–6.40)
LDL cholesterol (mmol/L)	2.17 (1.81–2.61)	2.13 (1.78–2.55)	2.25 (1.86–2.69)
Total triglycerides (mmol/L)	1.11 (0.82–1.57)	0.99 (0.76–1.36)	1.36 (0.96–1.86)
Apolipoprotein A-I (g/L)	1.74 (1.60–1.89)	1.77 (1.63–1.91)	1.69 (1.55–1.84)
HDL cholesterol (mmol/L)	1.65 (1.37–1.93)	1.74 (1.47–1.99)	1.5 (1.26–1.77)
Acetoacetate (mmol/L)	0.04 (0.03–0.05)	0.04 (0.03–0.05)	0.03 (0.03–0.05)
C-reactive protein (mg/L)	0.82 (0.45–1.63)	0.71 (0.41–1.41)	1.00 (0.53–2.03)
Testosterone (nmol/L)	1.23 (0.77–15.21)	1.03 (0.73–12.97)	10.91 (0.92–16.7)
Sex hormone binding globulin (SHBG, nmol/L)	42 (29.2–60.6)	48.4 (34.5–68.6)	33.2 (25.15–46.45)
Leukocytes (× 10 <sup>9</sup> cells/L)	5.3 (4.5–6.4)	5.2 (4.4–6.2)	5.6 (4.7–6.7)
Platelets (× 10 <sup>9</sup> cells/L)	247 (215–285)	248 (216–287)	245 (213–283)
Erythrocytes (× 10 <sup>12</sup> cells/L)	4.7 (4.4–4.9)	4.6 (4.3–4.8)	4.8 (4.6–5.1)
Amylase (U/L)	53 (42–66)	54 (43–68)	50 (40–63)
Bilirubin (μmol/L)	11 (9–15)	12 (9–15)	11 (9–15)
Alkaline phosphatase (U/L)	61 (51–73)	58 (49–70)	66 (55–78)
Alanine aminotransferase (U/L)	25 (18–36)	22 (17–30)	31 (22–44)
Aspartate aminotransferase (U/L)	21 (17–26)	20 (17–25)	23 (19–29)
Gamma-glutamyltransferase (U/L)	23 (15–38)	19 (14–30)	30 (20–51)
Albumin (g/L)	45 (43–46)	45 (43–46)	46 (44–47)
Urea (mmol/L)	5.3 (4.5–6.2)	5.1 (4.4–6)	5.5 (4.7–6.5)
Uric acid (μmol/L)	298 (249–353)	278 (236–327)	333 (283–389)
Creatine kinase (U/L)	106 (77–153)	99 (74–144)	117 (85–167)
Creatinine (μmol/L)	0.07 (0.06–0.08)	0.07 (0.06–0.08)	0.07 (0.07–0.08)
eGFR (mL/min/1.73m <sup>2</sup> )	104 (95–107)	103 (95–107)	104 (96–108)
Thyroid-stimulating hormone (mIU/L)	1.42 (1.01–1.96)	1.41 (1–1.93)	1.44 (1.02–2.02)

Study II used two representative clinical measures including BMI (median=26, 95% CI 23 to 29) and MAP (median=97, 95% CI 89 to 105) in a random subset (n=994) of NFBC1966. Study III involved 49 (+smoking, numbers 3–52 in the Table 5)

clinical and biochemical measures from the NFBC1966 (n=4 505), of which 29 (+smoking, numbers 3–14, 23, 25, 31, 33, 36, 38, 40–52 in the Table 5) were available in the NFBC1986 (n=1 010) and 39 (+age and smoking, numbers 2–3, 8–10, 14–41, 44–45, 47–52 in the Table 5) were available in the YFS (n=474). The characteristics of the three independent population cohorts are given in Table 5.

A wide range of clinical and biochemical measures including clinical measures, serum NMR measurements as well as calculated/estimated values were used. For example, anthropometry measures (BMI, waist-hip-ratio, body fat percentage, visceral fat area, total body water), blood pressure (systolic blood pressure, diastolic blood pressure, pulse pressure), fitness indicators (fitness score, basal metabolic rate, grip strength), amino acids (leucine, isoleucine, valine, alanine, glutamine, glycine, phenylalanine and tyrosine quantified by serum NMR metabolomics), glycaemic traits and gluconeogenesis measures (clinical measures: HbA1c, fasting insulin; serum NMR measurements: fasting glucose, lactate, pyruvate, citrate, glycerol), lipids and lipoproteins (serum NMR measurements: Apolipoprotein B, Apolipoprotein A-I; clinical measures: total triglycerides, HDL cholesterol), miscellaneous blood biomarkers (serum NMR measurements: acetoacetate, beta-hydroxybutyrate, GlycA; clinical measures: C-reactive protein, haemoglobin, leukocytes, platelets, erythrocytes), liver and pancreas function biomarkers (clinical measures: bilirubin, alkaline phosphatase, alanine aminotransferase, gamma-glutamyl transferase), kidney function indicators [clinical measures: uric acid, creatinine, and eGFR estimated by CKD-EPI formula (Inker et al., 2012)], as well as several calculated the risk scores of disease [CVD FINRISK (Bhopal et al., 2005), CKD Nelson (Nelson et al., 2019), CKD O'Seaghdha (O'Seaghdha et al., 2012), CKD Chien (Chien et al., 2010)]. The specific formula used to calculate risk scores are shown in Table 6.

**Table 5. Characteristics of the three independent study populations. Table 5 is modified under CC BY License from Original publication III © 2023 Authors (Li et al., 2023).**

Characteristics	NFBC1966	NFBC1986	YFS
1. Number	4 505	1 010	474
2. Age (years)	46.7 (46.2–47.1)	33.7 (33.4–34.1)	50.1 (31.3–68.5)
3. BMI (kg/m <sup>2</sup> )	26 (23–29)	25 (23–28)	27 (24–30)
4. Waist-to-hip ratio	0.91 (0.85–0.97)	0.91 (0.87–0.97)	
5. Body fat (%)	28 (22–35)	26 (20–34)	
6. Visceral fat area (cm <sup>2</sup> )	99 (76–127)	81 (59–121)	
7. Total body water (L)	39 (34–47)	39 (33–47)	
8. Systolic blood pressure (mmHg)	124 (114–135)	111 (103–120)	128 (117–142)

Characteristics	NFBC1966	NFBC1986	YFS
9. Diastolic blood pressure (mmHg)	84 (77–92)	74 (68–80)	78 (71–85)
10. Pulse (bpm)	69 (62–77)	71 (64–79)	69 (62–77)
11. Fitness score	74 (69–79)	75 (70–80)	
12. Basal metabolic rate (calories)	1 509 (1 358–1 756)	1 520 (1 353–1 753)	
13. Grip strength average (kg)	32 (26–45)	35 (30–47)	
14. Smoking prevalence (%)	17.5	14	18.7
15. Leucine (mmol/L)	0.08 (0.07–0.10)		0.12 (0.10–0.14)
16. Isoleucine (mmol/L)	0.05 (0.05–0.07)		0.06 (0.05–0.07)
17. Valine (mmol/L)	0.20 (0.18–0.23)		0.23 (0.21–0.26)
18. Alanine (mmol/L)	0.45 (0.41–0.5)		0.37 (0.32–0.43)
19. Glutamine (mmol/L)	0.57 (0.53–0.61)		0.75 (0.70–0.80)
20. Glycine (mmol/L)	0.29 (0.26–0.33)		0.26 (0.23–0.32)
21. Phenylalanine (mmol/L)	0.08 (0.07–0.08)		0.06 (0.05–0.07)
22. Tyrosine (mmol/L)	0.06 (0.05–0.06)		0.06 (0.06–0.07)
23. Glycated haemoglobin (%)	5.5 (5.2–5.7)	5.2 (5.0–5.4)	5.5 (5.3–5.8)
24. Fasting insulin (IU/L)	7.9 (5.4–11.7)		9.2 (5.4–13.4)
25. Fasting glucose (mmol/L)	5.4 (5.1–5.8)	4.9 (4.7–5.2)	5.4 (5.1–5.8)
26. Lactate (mmol/L)	1.4 (1.2–1.7)		2.0 (1.7–2.4)
27. Pyruvate (mmol/L)	0.09 (0.08–0.12)		0.06 (0.05–0.08)
28. Citrate (mmol/L)	0.12 (0.11–0.13)		0.04 (0.04–0.05)
29. Glycerol (mmol/L)	0.07 (0.06–0.09)		0.12 (0.09–0.15)
30. Apolipoprotein B (g/L)	1.02 (0.88–1.2)		0.93 (0.79–1.11)
31. Total triglycerides (mmol/L)	1.03 (0.76–1.47)	0.78 (0.57–1.09)	1.13 (0.85–1.55)
32. Apolipoprotein A-I (g/L)	1.7 (1.6–1.9)		1.5 (1.4–1.7)
33. HDL cholesterol (mmol/L)	1.5 (1.3–1.8)	1.4 (1.2–1.7)	1.3 (1.1–1.6)
34. Acetoacetate (mmol/L)	0.04 (0.03–0.05)		0.02 (0.01–0.04)
35. Beta-hydroxybutyrate (mmol/L)	0.12 (0.10–0.16)		0.05 (0.02–0.11)
36. C-reactive protein (mg/L)	0.82 (0.45–1.65)	0.71 (0.36–1.55)	1.03 (0.53–2.37)
37. GlycA (mmol/L)	1.4 (1.3–1.5)		0.89 (0.81–0.96)
38. Haemoglobin (g/L)	141 (132–150)	136 (128–146)	144 (137–152)
39. Leukocytes ( $\times 10^9$ cells/L)	5.4 (4.5–6.4)		6.1 (5.2–7.1)
40. Platelets ( $\times 10^9$ cells/L)	247 (215–286)	239 (210–274)	254 (219–292)
41. Erythrocytes ( $\times 10^{12}$ cells/L)	4.7 (4.4–4.9)	4.6 (4.3–4.9)	4.8 (4.5–5.1)
42. Bilirubin ( $\mu$ mol/L)	11 (9–15)	12 (9–16)	
43. Alkaline phosphatase (U/L)	61 (51–73)	55 (45–66)	
44. Alanine aminotransferase (U/L)	25 (18–36)	21 (16–31)	22 (15–31)
45. Gamma-glutamyltransferase (U/L)	23 (15–38)	15 (11–24)	24 (17–38)
46. Uric acid ( $\mu$ mol/L)	297 (249–353)	303 (252–356)	
47. Creatinine ( $\mu$ mol/L)	67 (59–75)	65 (58–74)	76 (67–87)
48. eGFR (mL/min/1.73m <sup>2</sup> )	104 (95–107)	115 (107–118)	87 (75–100)
49. CVD FINRISK	0.57 (0.20–1.66)	0.10 (0.04–0.32)	1.19 (0.2–6.62)
50. CKD Nelson risk	1.4 (1.0–2.4)	0.29 (0.23–0.47)	6.6 (1.2–31.7)

Characteristics	NFBC1966	NFBC1986	YFS
51. CKD O'Seaghdha risk	0.76 (0.76–1.61)	0.24 (0.23–0.25)	3.65 (0.50–20.89)
52. CKD Chien risk	6.1 (4.7–8.4)	1.8 (1.4–2.3)	10.2 (2.7–40)

**Table 6. Calculation formula of disease risk score.**

Risk score	Formula	Approximation*
CVD FINRISK (Bhopal et al., 2005)	In men: $\log \text{ odds} = 10.2133 - 0.0985 \times \text{age in years} - 0.6072 \times \text{smoking (1 if a current smoker, 0 otherwise)} - 0.0110 \times \text{systolic blood pressure in mmHg} - 0.3421 \times \text{total cholesterol in mmol/L} + 1.2506 \times \text{HDL-cholesterol (mmol/l)} - 0.4888 \times \text{diabetes (1 if diagnosed diabetes and 0 otherwise)}$ ; The 10 year risk (per cent) was calculated as $1/(1 + \exp(\log \text{ odds})) \times 100$ .	/
CKD Nelson (Nelson et al., 2019)	In women: $\log \text{ odds} = 10.6882 - 0.0954 \times \text{age} - 0.6785 \times \text{smoking} - 0.0186 \times \text{systolic blood pressure} - 0.1969 \times \text{total cholesterol} + 1.9039 \times \text{HDL cholesterol} - 0.8799 \times \text{diabetes}$ . The 10 year risk (per cent) was calculated as $1/(1 + \exp(\log \text{ odds})) \times 100$ . In nondiabetic: $1 - \exp(-5^{\wedge}1.055408 \times \exp[-3.609661 + 0.2582196 \times (\text{age}/5 - 11) + 0.1821665 \times (\text{if female}) + 0.1808945 \times (\text{if black}) + 0.4581006 \times (15 - \min(\text{eGFR}, 90))/5] - 0.3159218 \times \max(0, \text{eGFR} - 90))/5 + 0.1953927 \times (\text{if has history of CVD}) + 0.1213741 \times (\text{if ever smoking}) + 0.3543645 \times (\text{if hypertensive}) + 0.0630538 \times (\text{BMI}/5 - 5.4) + 0.3519087 \times (\log_{10} \text{ ACR-expected log}_{10} \text{ACR})$ In diabetic: $1 - \exp(-5^{\wedge}0.9766551 \times \exp[-2.647004 + 0.1351572 \times (\text{age}/5 - 11) + 0.1381975 \times (\text{if female}) + 0.0920208 \times (\text{if black}) + 0.3546697 \times (15 - \min(\text{eGFR}, 90))/5] - 0.1525133 \times \max(0, \text{eGFR} - 90))/5 + 0.1870637 \times (\text{if has history of CVD}) + 0.0619679 \times (\text{HbA1c} - 7) + 0.1078296 \times (\text{if insulin use}) - 0.150944 \times (\text{if no DM medication use}) + 0.023959 \times (\text{HbA1c} - 7) \times (\text{if insulin use}) + 0.0398424 \times (\text{HbA1c} - 7) \times (\text{if no DM medication use}) - 0.00084 \times (\text{if ever smoking}) + 0.3653268 \times (\text{if hypertensive}) + 0.050306 \times (\text{BMI}/5 - 5.4) + 0.3737905 \times (\log_{10} \text{ACR} - 1)$	if insulin in use: for diabetes are positive (only 22 T2D cases and 2 T1D cases), others are negative. if no DM medication in use: For diabetes are negative and normal people are positive. if hypertensive: hypertension was defined as blood pressure more than 140/90 mm Hg. Due to the good health and youth in most participants, some approximations were used in these calculations: History of CVD: no cases; urine ACR: 0 in all individuals.
CKD O'Seaghdha (O'Seaghdha et al., 2012)	$1 - (1 - 0.092)^{\wedge}0.5 / ((1 - 0.092)^{\wedge}0.5 + (1 - (1 - 0.092)^{\wedge}0.5) \times \exp(-6.235 + 0.095 \times \text{age} + 0.476 \times (\text{if diabetics}) + 0.761 \times (\text{if hypertensive}) + 0.779 \times (\text{if } 75 \leq \text{eGFR} < 90) + 1.558 \times (\text{if } 60 \leq \text{eGFR} < 75) + 0.300 \times (\text{if ACR} \geq 30 \text{ or dipstick} \geq \text{trace}))$	ACR: 0 in all individuals.
CKD Chien (Chien et al., 2010)	$1 - 0.9632^{\wedge} \exp(-6.8 + 0.077 \times \text{age} + 0.366 \times (\text{if diabetics}) + 1.24 \times (\text{if history of stroke}) + 0.059 \times \text{BMI} + 0.018 \times \text{systolic blood pressure})$	History of stroke: no cases.

\*: Approximation was used for the unavailable variables of formula  
Abbreviations: DM, diabetes mellitus; ACR, Albumin-creatinine ratio.

#### **4.1.2 Applied quantified urinary metabolomics data**

Metabolite quantification: Urinary metabolites are quantified from the NMR spectra with a semi-automated methodology utilising sophisticated constrained total line shape (CTLS) fitting analysis tools developed for high-precision quantitative NMR spectroscopy (Mierisová & Ala-Korpela, 2001; Soinen et al., 2005). In CTLS fitting, the characteristic peak groupings and appropriate mathematical constraints are considered the molecular features of the individual metabolites. Based on this approach, Tynkkynen et al. published quantitative data for 43 (+creatinine) urinary metabolites of 1004 urine samples from NFBC1966 in 2019 (Tynkkynen et al., 2019). In the latest work, the semi-quantitative data were expanded to 60 (+creatinine) urinary metabolites. Although the CTLS analyses represent the state-of-the-art and most robust methods in NMR spectral analyses, they typically require a considerable amount of manual control of the analysis parameters and call for a visual assessment of individual fitting results. Thus, at best, they can be operated in a semi-automated manner with different spectral regions, i.e. one or a few metabolites, analysed at a time. Unfortunately, all this makes the overall spectral analyses tedious and time consuming.

The aforementioned challenges are the very reason why an automated regression analysis approach for urine NMR metabolomics is developed. Such an approach has proven superior in the case of quantitative serum NMR metabolomics with current data available for a plethora of various epidemiological and genetic applications and spanning over 1.8 million samples and counting (Ala-Korpela et al., 2022; Julkunen et al., 2021, 2023; Locke et al., 2019; Würtz et al., 2017). The robust CTLS analyses form the essential base for eventually automating the quantitative metabolite analyses via metabolite-specific regression models. The process of constructing automated quantification of urine metabolites from NMR spectra includes two parts: an alignment and pre-processing of spectral data and the training and validation of linear regression models for metabolite concentrations. Alignment and pre-processing include the identification and selection of metabolite resonance(s) and fine-tuning the optimal spectral data for quantification. Training and validation of the model utilize the fine-tuned spectral data and generalize the implementation of the automated quantification model. The automated quantitative analyses use R to develop Bayesian regression-based models using metabolite signal areas determined from ca. Only the results of two automated regression models (glucose and creatinine) have been published so far (Li, Ihanus, et al., 2022;

Tynkkynen et al., 2019). In the most recent work, 11 automated quantification models were added, namely 2-hydroxyisobutyrate, valine, alanine, pseudouridine, dimethylamine, glycine, citrate, urea, formate, trigonelline, and hippurate.

There are many reasons for the immensely complex development of robust automated regression models for urinary metabolites such as drastic chemical shift variations, extra (random) signals, signal overlap, and complicated varying baseline. Since the expected time for completing the development of automated quantitative models for approximately 50 urinary metabolites is still a number of years, we consider it of high value to provide the concentrations and population distributions of 61 metabolites in a random subset of the 994 participants in NFBC1966 (supplementary Figures 2–4, in III). Importantly, the concentration and distribution data for the random subset of NFBC1966 match very well to the corresponding data for the entire cohort of 4505 participants in NFBC1966 (supplementary Figures 5–8, in III). This type of quantitative population information constitutes substantial epidemiological novelty at this point in this area of research.

Since the relaxation time used in the measurement is rather short, it affects the area of the metabolites signal, as well as the reference concentration TSP (3-(trimethylsilyl) propionic-2,2,3,3 acid D4), in different ways depending on the relaxation time of the associated proton. In the present study, the correction coefficient for each quantified metabolite was used to obtain the true absolute metabolite concentrations. The correction coefficients were obtained by measuring a set of NFBC1966 urine samples also with long relaxation delays and comparing the resulting concentrations to those obtained via the high-throughput parameters.

Quantitative urinary metabolite data applied in each study are as follows: Study I utilized urinary glucose concentration (referenced to creatinine) quantified by the automated regression model in participants from NFBC1966 (n=4 482) and NFBC1986 (n=1 010) with a median relative concentration of 24.2 (95% CI 20.5 to 28.6) and 22.5 (95% CI 19.1 to 26.4) mM per 1 mM creatinine in NFBC1966 and NFBC1986, respectively.

Study II involved 43 urinary metabolites (+creatinine) previously semi-quantified with NMR spectroscopy in morning spot urine samples from 994 individuals in NFBC1966 (Tynkkynen et al., 2019). Urinary metabolite concentrations (referenced to creatinine) in men and women of the present study are illustrated in Table 7 (metabolites without \*).

Study III analysed 12 urine metabolites (+creatinine) that were simultaneously quantified in NFBC1966 (n=4 505), NFBC1986 (n=1 010), and YFS (n=474) by automated regression models. The concentrations of these metabolites referenced

to creatinine are summarized in Table 8. In addition, another 17 metabolites (metabolites with \*) were semi-quantified in a random subset of NFBC1966 (n=994). In total, 60 urinary metabolites referenced concentrations including the aforementioned 43 metabolites are summarized in Table 7.

**Table 7. Sixty-one urinary metabolite concentrations in men and women from the random subset (n=994) of NFBC1966 (metabolites labelled with \* only quantified in Study III). Table 7 is reprinted under CC BY License from Original publication III © 2023 Authors (Li et al., 2023).**

Metabolites	Absolute concentrations (µmol/L)		Creatinine-referenced concentrations (µM/mM creatinine)	
	Men (n=419)	Women (n=575)	Men (n=419)	Women (n=575)
<b>Amino acids</b>				
Alanine	246 (172–369)	191 (131–311)	18.90 (14.19–26.47)	20.78 (15.96–28.68)
Glycine	932 (647–1437)	1101 (724–1861)	73.60 (50.74–103.6)	124 (82.6–185)
Histidine	976 (667–1421)	652 (406–1067)	76.43 (56.42–107.6)	68.98 (47.15–99.54)
Threonine	107 (68–155)	89 (57–139)	8.29 (5.61–12.26)	9.83 (6.73–13.15)
Isoleucine	10 (7–15)	9 (7–14)	0.82 (0.65–1.08)	0.96 (0.74–1.24)
Leucine*	26 (19–36)	20 (14–30)	2.09 (1.64–2.64)	2.09 (1.65–2.63)
Valine	48 (26–50)	30 (21–43)	2.92 (2.34–3.65)	3.10 (2.55–3.88)
Tryptophan	76 (51–115)	61 (39–95)	6.09 (4.64–8.19)	6.23 (4.47–8.56)
Tyrosine	138 (97–196)	92 (60–144)	10.91 (8.21–14.38)	9.87 (7.21–13.21)
<b>Metabolism of amino acids</b>				
Betaine*	69 (41–119)	58 (37–96)	5.18 (3.36–8.62)	6.47 (4.08–10.32)
Creatine	116 (70–223)	158 (77–591)	8.21 (6.14–14.08)	14.52 (7.90–55.39)
3-Hydroxyisobutyrate	128 (80–187)	108 (70–164)	9.33 (7.10–12.63)	10.87 (8.20–14.70)
Hippurate	4 774 (2 724–7 843)	4 688 (2 600–7 881)	380 (229–613)	502 (289–762)
Urea	296 129 (246165–346324)	257 252 (212231–308767)	23 110 (18 371–29 362)	27 650 (20 610–34 328)
Pyroglutamate*	263 (202–330)	238 (181–302)	20.69 (17.56–23.84)	24.55 (20.36–28.99)
<b>Carbohydrate metabolism</b>				
Glucose	295 (217–390)	241 (178–343)	22.62 (19.54–26.42)	25.33 (21.66–30.41)
Lactate	76 (48–116)	74 (46–123)	5.52 (3.94–7.79)	7.40 (5.38–11.14)

Metabolites	Absolute concentrations (µmol/L)		Creatinine-referenced concentrations (µM/mM creatinine)		P
	Men (n=419)	Women (n=575)	Men (n=419)	Women (n=575)	
Citrate	1 716 (1 090–2 612)	2 366 (1 506–3 470)	140 (91–195)	251 (177–332)	2.9E-56
Cis-aconitate	221 (154–308)	190 (134–273)	17.27 (13.64–20.90)	19.77 (16.34–24.41)	1E-13
Fumarate*	6 (4–10)	7 (4–11)	0.40 (0.27–0.65)	0.65 (0.45–0.99)	2.6E-22
Succinate*	28 (19–44)	42 (25–68)	2.21 (1.57–3.24)	4.45 (3.05–6.28)	4.7E-58
Mannitol*	448 (244–917)	437 (233–878)	34.32(19.07–69.43)	49.27 (25.83–91.47)	0.49
Glucuronate	277 (200–356)	244 (178–342)	20.64 (17.55–24.93)	24.57 (20.89–30.04)	3.4E-20
<b>Nucleotide</b>					
<b>metabolism</b>					
Hypoxanthine	41 (25–67)	33 (19–54)	3.15 (2.24–4.76)	3.28 (2.29–4.96)	0.22
Pseudouridine	343 (264–441)	295 (212–414)	26.55 (24.89–28.28)	30.65 (28.21–32.92)	1.6E-63
Uracil	58 (39–84)	60 (41–92)	4.59 (3.28–6.77)	6.47 (4.75–8.74)	7.9E-18
<b>Nicotinate and</b>					
<b>nicotinamide</b>					
<b>metabolism</b>					
2-PY	167 (110–247)	133 (83–203)	12.47 (9.53–17.76)	13.48 (9.03–19.02)	0.23
N1-	56 (37–85)	47 (31–69)	4.37 (3.12–6.35)	4.78 (3.42–6.73)	0.01
Methylnicotinamide					
Trigonelline	921 (463–1 508)	802 (396–1 373)	70.70 (38.71–120.8)	86.44 (46.20–139)	0.0027
<b>Caffeine</b>					
<b>metabolism</b>					
3-Methylxanthine*	91 (59–140)	94 (53–156)	7.10 (4.93–10.19)	9.71 (6.18–14.58)	1.1E-12
<b>Microbial</b>					
<b>metabolism</b>					
3-Hydroxyhippurate	592 (272–1 112)	449 (192–875)	51.26 (23.75–90.10)	47.54 (21.45–89.35)	0.93
4-Hydroxyhippurate	94 (64–144)	96 (59–157)	7.11 (5.10–10.63)	10.22 (6.92–14.52)	1.5E-17
4-	179 (126–249)	151 (99–219)	13.26 (10.40–18.38)	14.85 (11.95–19.94)	0.00013
Hydroxyphenylacetat					

Metabolites	Absolute concentrations (µmol/L)			Creatinine-referenced concentrations (µM/mM creatinine)		
	Men (n=419)	Women (n=575)	P	Men (n=419)	Women (n=575)	P
Acetate	48 (34–72)	47 (32–74)	0.78	3.78 (2.67–5.60)	4.86 (3.22–7.91)	1.1E-10
Dimethylamine	402 (297–507)	329 (233–457)	3.3E-09	29.61 (27.31–32.24)	32.91 (30.60–36.45)	3.2E-34
Formate	265 (179–360)	216 (148–314)	3.1E-06	19.92 (14.54–30.23)	22.92 (15.88–32.73)	0.0013
HPHPA	327 (158–564)	323 (160–575)	0.69	27.31 (14.32–45.34)	33.97 (17.44–58.68)	0.031
Indoxyl sulfate	332 (227–476)	330 (210–510)	0.82	25.67 (19.85–33.91)	33.90 (25.25–45.12)	6.9E-20
Methanol*	31 (23–43)	30 (23–42)	0.5	2.44 (1.72–3.66)	3.21 (2.05–5.25)	2.4E-09
Methylamine	32 (20–48)	28 (17–45)	0.0083	2.45 (1.80–3.20)	2.78 (1.98–3.80)	0.000027
p-Cresol sulfate	299 (165–483)	372 (205–619)	0.000063	24.08 (13.70–37.70)	38.45 (23.23–59.16)	2E-21
Phenylacetylglutamin e*	717 (452–1 030)	706 (442–1 150)	0.84	56.50 (38.35–76.29)	75.40 (52.53–101)	1.1E-16
TMAO	515 (329–746)	418 (265–650)	0.00003	36.82 (27.37–54.29)	39.98 (29.54–64.38)	0.0063
Modification of histones						
2-Hydroxyisobutyrate	55 (40–78)	45 (30–65)	6.4E-11	4.30 (3.59–5.20)	4.59 (3.66–5.68)	0.0088
<b>Dietary metabolites</b>						
1-Methylhistidine*	383 (287–496)	277 (203–369)	1.3E-25	29.08 (24.49–34.48)	27.94 (22.63–34.06)	0.0027
2-Furoylglycine	57 (32–112)	48 (27–83)	0.00026	4.60 (2.75–7.96)	4.55 (2.69–8.35)	0.015
3-Methylhistidine	506 (214–1 180)	416 (179–1 036)	0.018	38.05 (18.40–87.78)	45.40 (18.58–102)	0.31
Arabinose	117 (82–172)	103 (70–159)	0.0025	9.05 (6.99–11.68)	10.95 (8.53–14.15)	1.3E-12
Levoglucosan*	94 (61–162)	86 (57–170)	0.5	7.41 (6.64–12.15)	9.20 (5.66–16.53)	0.92
Proline-betaine*	200 (70–455)	173 (51–493)	0.31	15.76 (5.68–36.69)	17.94 (5.21–51.71)	0.021
Quinate*	301 (208–437)	283 (191–428)	0.071	24.25 (18.04–34.58)	30.37 (21.21–41.10)	0.00053
Scyllitol*	66 (44–99)	69 (45–108)	0.28	5.06 (3.73–7.16)	7.06 (5.09–9.87)	4.7E-23
Sucrose	40 (21–87)	39 (21–88)	0.95	3.20 (1.64–6.36)	4.17 (2.27–8.22)	0.00005
Trans-ferulate*	24 (18–32)	23 (16–32)	0.35	1.69 (1.35–2.11)	2.09 (1.59–2.78)	0.000032
Xylose	71 (49–117)	71 (46–118)	0.68	5.33 (3.77–8.71)	6.99 (5.06–11.94)	1.8E-11
<b>Miscellaneous</b>						
3-Hydroxyisovalerate	70 (48–98)	57 (38–84)	4E-07	5.33 (4.24–6.91)	5.81 (4.65–7.45)	0.00036

Metabolites	Absolute concentrations (µmol/L)		Creatinine-referenced concentrations (µM/mM creatinine)		P
	Men (n=419)	Women (n=575)	Men (n=419)	Women (n=575)	
4-Deoxyerythronate*	125 (86–163)	95 (64–138)	9.15 (7.36–11.68)	9.94 (7.54–12.71)	0.0043
4-Deoxythreonate*	281 (195–422)	188 (131–263)	21.62 (17.67–28.62)	18.53 (15.20–23.70)	6.3E-12
Sumiki's acid	38 (18–74)	29 (15–59)	2.94 (1.45–5.69)	2.94 (1.64–5.69)	0.39
Trans-aconitate	57 (38–79)	45 (31–68)	4.43 (3.57–5.31)	4.57 (3.75–5.91)	0.0049
Creatinine (mmol/L)	13 (10–16)	10 (7–14)	1000	1000	/

**Table 8. Concentrations of 12 urinary metabolites (+creatinine) in NFBC1966, NFBC1986 and YFS. Table 8 is reprinted under CC BY License from Original publication III © 2023 Authors (Li et al., 2023).**

Metabolites ( $\mu\text{M}/\text{mM}$ creatinine)	NFBC1966 (n=4 505)	NFBC1986 (n=1 010)	YFS (n=474)
2-Hydroxyisobutyrate	4.5 (3.64–5.52)	4.27 (3.51–5.39)	5.75 (4.59–7.23)
Valine	3.09 (2.49–3.86)	2.82 (2.28–3.55)	3.04 (2.44–3.86)
Alanine	20.71 (15.42–28.28)	17.61 (13.68–24.52)	24.84 (18.93–35.15)
Pseudouridine	29.44 (26.16–33.46)	28.52 (25.06–32.92)	31.73 (27.85–36.01)
Glucose	24.17 (20.6–28.83)	22.38 (19–26.58)	29.37 (23.91–41.52)
Dimethylamine	31.85 (29.02–35.39)	31.93 (28.88–35.69)	34.2 (30.52–39.36)
Glycine	97.85 (66.16–152.45)	90.77 (60.81–140.47)	99.63 (66.78–150.9)
Citrate	201.62 (134.64–289.16)	153.78 (95.85–232.79)	263.72 (175.51–390.45)
Urea	26 118 (19 805–33 823)	25 356 (19 130–33 108)	29 186 (21 519–39 244)
Formate	22.39 (15.58–32.06)	21.54 (15.46–30.22)	23.45 (16.34–34.51)
Trigonelline	82.89 (42.59–133.82)	58.48 (25.26–101.88)	47.28 (15.95–102.71)
Hippurate	466.84 (267.05–705.55)	364.13 (212.24–579.65)	323.41 (175.92–559.13)
Creatinine (mmol/L)	10.79 (7.5–14.83)	9.24 (6.37–13.28)	6.47 (2.98–11.9)

## 4.2 Applied normalization methods

Since urine serves as a waste biological fluid with uncontrolled chemical properties, urine sample volumes and metabolite concentrations can be largely varied even within the same individual. Thus, an additional normalization step is required during data analysis (Li, Tynkkynen, et al., 2022). In Study I and III, the absolute concentration of each urinary metabolite was divided by the concentration of urinary creatinine. In Study II, seven normalization protocols were applied, for example four internal standard (IS) normalization methods, a constant sum (CS) normalization (Torgrip et al., 2008), PQN (Dieterle et al., 2006), and the DESeq2 method (Love et al., 2014). Table 9 lists the methods' biological rationale, as well as their advantages and disadvantages for epidemiological studies. The basic features of these methods are summarized below.

IS-CREA, IS-GLUC, IS-UREA and IS-PSEURID: The practice involves dividing the absolute concentration of each urine metabolite is divided by the concentration of the internal reference metabolite. It's common to utilize creatinine concentration as the internal standard (IS-CREA, Study I and III) (Li, Tynkkynen, et al., 2022). In addition, normalization to glucose (IS-GLUC), urea (IS-UREA) and pseudouridine (IS-PSEURID) were also applied in Study II.

CS: The method entails dividing the absolute concentration of each metabolite by the total concentration of all quantified metabolites.

PQN: This approach operates under the assumption that biologically relevant concentration changes impact only a restricted set of metabolites, whereas the sample dilution influences all metabolite concentrations (Dieterle et al., 2006). The approach commences with CS normalization, followed by the computation of median relative metabolite concentrations of all samples, which serve as reference values. Subsequently, the quotient between each metabolite and its corresponding reference value is computed for every sample. Afterward, the median quotient of all metabolites is calculated for each sample. Ultimately, this median quotient is employed as the scaling factor, correcting for the concentration of each metabolite.

DESeq2: This method shares a similar principle with PQN, but it is designed for RNA-seq analysis (Love et al., 2014). In DESeq2, the reference concentration for each metabolite is determined as the geometric mean of the absolute metabolite concentrations across all samples, as opposed to the median used in PQN. Following this, the subsequent steps in DESeq2 are the same as those in PQN.

**Table 9. The characteristics of employed normalization methods. Table 9 is reprinted under CC BY License from Original publication II © 2022 Authors (Li, Tynkkynen, et al., 2022).**

Method	Abbr.	Description	Pros for Epidemiology	Cons for Epidemiology	Ref.
Absolute concentrations	ABS	Using the original quantified concentrations of the urinary metabolites from the <sup>1</sup> H NMR spectra, i.e., no normalization method applied.	Original data are preserved; the concentration values are straightforward to interpret. Useful if urine volume is noteworthy, e.g. fluid balance. Good for timed collections (e.g. 24-hour urine) where exact amount excreted per time unit can be calculated. Morning spot urine accurate enough to detect large effect sizes in epidemiological studies.	Urinary volume, and thus absolute metabolite concentrations, varies greatly day-to-day and person-to-person. Random spot urine samples are thus likely to be too confounded to use without normalization.	(Vart & Grams, 2016)
Normalization to an internal metabolite standard	IS-CREA	The concentration of each metabolite is divided by the concentration of an internal standard. Here, creatinine, glucose, urea and pseudouridine were used.	Creatinine comes from non-enzymatic breakdown of creatine phosphate in muscles, it is typically produced at a constant rate, and it is stable and inert in plasma. Creatinine is freely filtered by the kidneys and not reabsorbed. The most applied reference, allows straightforward comparisons between studies and meta-analyses.	The stable excretion of creatinine may not hold in the presence of external stimuli or pathophysiological conditions. Renal filtration and excretion of creatinine depend on circulating creatinine concentrations that are dependent on muscle mass – thus the urinary creatinine concentrations can be biased, e.g. for elderly and between men and women. A small amount of creatinine is secreted by the proximal tubule, resulting in a potentially study-dependent bias.	(Eaton & Pooler, 2018; Forni Ogna et al., 2015; Herrera & Rodríguez-Iturbe, 1998; Saccanti, 2017; Waikar et al., 2010)

Method	Abbr.	Description	Pros for Epidemiology	Cons for Epidemiology	Ref.
	IS-GLUC	<p>Glucose is freely filtered by the kidneys and mostly reabsorbed. The mechanisms of glucose reabsorption are well understood. Under normoglycemia, the plasma glucose level can be considered stable, and there is always a detectable amount of glucose in normal urine.</p>	<p>At high plasma glucose concentrations (&gt;10 mmol/L) the tubular reabsorption saturates, triggering a pronounced part of filtered glucose to be excreted into the urine.</p> <p>Also in normoglycemia the urinary glucose is dependent on the circulating glucose which is widely variable at the population level and also affected by the fasting/non-fasting status.</p> <p>Urinary glucose concentration is dependent on the glomerular filtration rate; this can cause bias in cohorts with large variation in kidney function.</p>	<p>Plasma urea concentrations vary widely depending on protein intake, changes in tissue catabolism, and in various pathological conditions.</p> <p>Urea is a waste product, and its excretion is under partial hormonal regulation resulting in variably large amounts to be excreted into the urine.</p>	<p>(Li, Ihanus, et al., 2022)</p>
	IS-UREA		<p>Serum urea is applied as a marker of renal function for routine clinical analysis.</p> <p>Urea is freely filtered by the kidneys and about 50% of it is reabsorbed.</p>	<p>Plasma urea concentrations vary widely depending on protein intake, changes in tissue catabolism, and in various pathological conditions.</p> <p>Urea is a waste product, and its excretion is under partial hormonal regulation resulting in variably large amounts to be excreted into the urine.</p>	<p>(Eaton &amp; Pooler, 2018; Gowda et al., 2010)</p>

Method	Abbr.	Description	Pros for Epidemiology	Cons for Epidemiology	Ref.
	IS- PSEURID		<p>Pseudouridine is a low-molecular-mass, water-soluble compound with no significant protein binding in serum.</p> <p>Pseudouridine is freely filtered by the kidneys and not reabsorbed. It is not re-utilized or metabolized in the body. Pseudouridine concentrations reflect the whole-body turn-over of RNA and its excretion appears constant.</p> <p>Pseudouridine concentrations appear independent of muscle mass.</p>	<p>Pseudouridine concentrations appear related to kidney function with potential associated bias.</p> <p>It is the most common RNA modification with links to and potential variation in multiple metabolic diseases.</p>	(Bernert et al., 1988; Niewczasz et al., 2017; Sekula et al., 2016)
Constant sum	CS	<p>The concentration of each metabolite is divided by the total sum of all metabolite concentrations.</p> <p>Water dilution ideally affects all metabolites equally, thus a linear sum over all metabolites should capture volume factor accurately despite (random) variation in any specific metabolite.</p>	<p>A generic algorithm that can be applied to any metabolomics platform without requirements for a specific set of metabolites.</p> <p>Water dilution ideally affects all metabolites equally, thus a linear sum over all metabolites should capture volume factor accurately despite (random) variation in any specific metabolite.</p>	<p>Abundance of urinary metabolites resembles the Pareto distribution; a few abundant molecules (e.g. urea) typically dominate the concentration sum. The distribution of a metabolite is usually fat-tailed, thus extreme values may reduce normalization accuracy. Both issues mean that the benefit of averaging across many metabolites may be lost in real data.</p>	(Craig et al., 2006; Torgrip et al., 2008)

Method	Abbr.	Description	Pros for Epidemiology	Cons for Epidemiology	Ref.
Probabilistic quotient normalization	PQN	A robust version of the CS principle that addresses the Pareto issue of CS (by standardized abundances) and the outlier issue (by the median estimator).	A generic algorithm that can be applied to any metabolomics platform. While water dilution ideally affects all metabolites equally, it is plausible that only a small subset will be affected by the biological phenomenon under study. The median estimator is not much affected by a few biologically driven or random metabolites, thus PQN can capture the volume factor accurately in most situations.	All normalized concentrations are inter-dependent (i.e. if some metabolites go up, then others must go down to maintain balance). This means that undesired correlated variation/confounding across many metabolites may cause normalization artefacts. Since abundances are standardized, including low-abundance metabolites near the detection limit may amplify the impact of measurement noise.	(Dieterle et al., 2006; Torgrip et al., 2008)
A method for differential gene expression analysis based on the negative binomial distribution	DESeq2	The DESeq2 is a variant of PQN developed for RNA-seq data. Uses geometric mean instead of the median to standardize abundances.	Same benefits as PQN. Geometric mean is better suited for concentrations with limited numerical precision.	Same downsides as in PQN.	(Love et al., 2014)

### 4.3 Statistical analyses

All statistical analyses were performed using the R software.

#### *Study I. There is always glucose in normal urine.*

Correlation and regression analyses of urinary glucose were conducted on urinary glucose data from the NFBC1966 cohort. The participants with urinary glucose levels exceeding the third quartile plus eight times of the interquartile range (n=75) and those with levels below the detection limit (n=36) as well as those without available serum fasting glucose (n=96), were excluded from the subsequent analysis. The analyses were performed separately for normoglycemic participants (serum fasting glucose <5.6 mmol/L; n=2 639) and for those with serum fasting glucose concentrations were  $\geq 5.6$  mmol/L (n=1 636). The correlation analyses examined the associations between urinary glucose and serum glucose, as well as the multiplication of serum glucose and eGFR for all participants (n=4 275). Associations between the urinary glucose (log-transformed) and serum glucose, eGFR or the product of serum glucose and eGFR were assessed in the previously mentioned categories. These analyses were performed by linear regression analyses, either without covariate, with adjustment for sex, or with further adjustment for sex and BMI. To facilitate comparisons across various measures, association magnitudes are presented in units of standard deviation (SD). Partial rank correlations adjusted for sex were employed to demonstrate the correlations between urinary glucose and the 51 customary clinical and biochemical measures. The heat map is structured using column dimensional hierarchical clustering. The multiple comparison corrected p-value threshold of 0.001 (0.05/43 via the Bonferroni method, where 43 is the number of principal components needed to explain >99% of variation in the metabolic information) was utilized to indicate evidence supporting an association. Those associations for which the p-value fulfils this threshold are called robust, and this statement applies to the corresponding p-value in the following studies.

### *Study II. Characteristics of normalization methods in quantitative urinary metabolomics.*

In Study II, quantitative metabolomics data for the 44 urinary metabolites together with the clinical measures (sex, BMI, and MAP) available for 994 individuals from NFBC1966 were utilized. In all analyses, the absolute urinary metabolite concentrations and seven kinds of normalized concentrations were applied.

The associations between urinary metabolites were determined using Spearman's rank correlations adjusted for sex. The results were visually represented in heat maps. One of these heat maps, which utilized urine creatinine normalization, was structured using two-dimensional hierarchical clustering. The remaining heat maps followed the identical order of metabolites and clusters for straightforward visual comparison. In addition, creatinine was included as the final row in all heat maps. In these corresponding heat maps, the correlations of reference metabolites were left empty. Furthermore, the rank correlation analyses were reiterated without any adjustment, both for absolute and all referenced concentrations.

BMI and MAP were used as exemplars of clinical measurements to observe the potential effects and differences of different normalization strategies on the epidemiological association of urinary metabolites. A linear regression model without adjustment or adjusted for sex was used between 44 urinary metabolites and BMI as well as MAP, respectively. Prior to conducting the regression analysis, extreme metabolite concentrations were truncated to the third quartile plus eight times of the interquartile range. Subsequently, the metabolite concentrations underwent a logarithmic transformation. Association magnitudes are expressed in SD units; the p-value threshold was set at 0.001 (0.05/39 via the Bonferroni method).

### *Study III. Clinical and biochemical associations of urinary metabolites.*

Sixty urinary metabolite concentrations (normalized to urinary creatinine) for the random subset of NFBC1966 (n=994) as well as 12 urinary metabolites (referenced concentrations) for the NFBC1966 (n=4 505), NFBC1986 (n=1 010), and YFS (n=474) were used in Study III.

To explore the associations between 12 quantified urinary metabolites with 49 clinical measurements, partial rank correlations adjusted for sex were used for the whole NFBC1966 as well as for men (n=1 950) and women (n=2 555) without

adjustment. Meta-analyses of the partial correlations (adjusted for sex in NFBC1966 and 1986, in addition to age for YFS) between the 12 automatically quantified urinary metabolites (referenced to urinary creatinine) and 49 customary clinical and biochemical measures were applied for all the available data (n up to 5 989) as well as in NFBC1986 and YFS (n up to 1 484). The results are presented in color-coded heat maps. The colour key on the top of the figure represents the availability of clinical and biochemical measures in the three cohorts. There were 20 measures available in all the 3 cohorts (green), 19 measures available only in NFBC1966 and YFS (pink), and 10 measures available only in NFBC1966 and NFBC1986 (blue). The heat map which indicated the correlation between 12 quantified urinary metabolites with 49 clinical measures in the NFBC1966 was fully organized by via two-dimensional hierarchical clustering. For easy visual comparison, the other heat maps were organized in the same order and clusters of metabolites and clinical measurements. Fifty-six principal components explained over 99% of variation in the 60 creatinine-referenced urinary metabolite concentrations and the 49 clinical and biochemical measures. Thus, a multiple comparison corrected p-value threshold of 0.0009 (0.05/56 via the Bonferroni method) was used to suggest evidence supporting an association.

Linear regression analyses adjusted for sex (in addition to age in YFS) were used to analyse the associations between 12 quantified urinary metabolites and BMI as well as smoking history (current smokers vs non-smokers) in all cohorts. Extreme metabolite concentrations were truncated to the third quartile adding eight times of interquartile range and the metabolite concentrations were log-transformed. All the models were further individually adjusted for MAP, fasting glucose, fasting insulin, smoking history (only for BMI analysis), total triglycerides, CRP, eGFR and BMI (only for smoking analysis). Association magnitudes are reported in SD units to ease the comparison across multiple measures with different initial units and scales. All analyses were applied for each cohort and then meta-analysed.

Additionally, linear regression analyses for each of the three urinary amino acid concentrations including alanine, glycine and valine available for both NFBC1966 and YFS were also performed (as in our previous work for glucose (Li, Ihanus, et al., 2022)), namely with their corresponding serum concentrations, eGFR and the multiplication of serum concentration and eGFR. The models were further adjusted for BMI. The analyses were applied for each cohort and then meta-analysed.



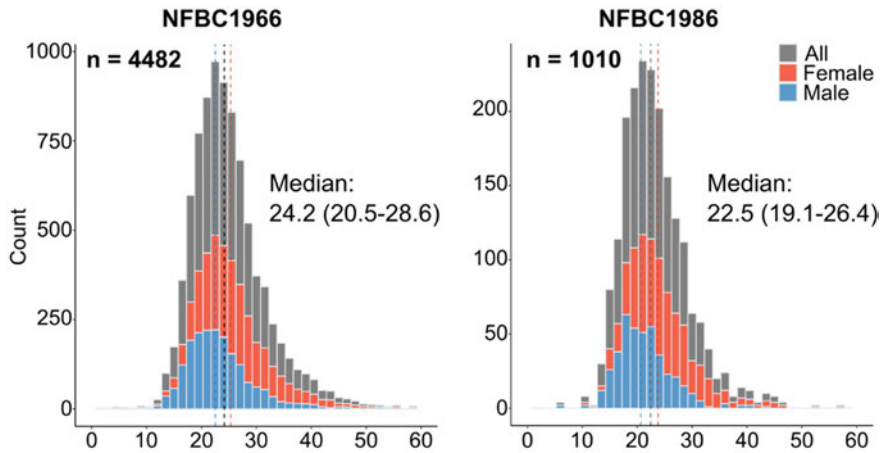
## **5 Results and discussion**

### **5.1 Characteristics of urinary glucose at the population level (Study I)**

The analysis of urine glucose has faced persistent methodological challenges. Issues primarily stem from the use of non-specific methods (FINE, 1965; Lund & Wolf, 1925) and the prevalent reliance on glucose test strips, which often have detection limits as high as 5.6 mmol/L. As a result, this has contributed to a widely held perception, and prevalent in the contemporary renal physiology textbooks (Eaton & Pooler, 2018), that the urine of healthy individuals typically does not contain glucose. This is an apparent misunderstanding as we summarize here from experiments performed in two independent populations by modern NMR spectroscopy (Tynkkynen et al., 2019). These new data are also placed in an epidemiological context with a variety of conventional clinical and biochemical measures.

#### ***5.1.1 Distribution of urinary glucose concentrations at the population level***

Urinary glucose was measured in 99.1% of the 5 492 urine samples. The concentration distributions were nearly identical, with a slight positive skew. Specifically, the median relative concentrations were 24.2 and 22.5 ( $\mu\text{M}/\text{mM}$  creatinine) in NFBC1966 ( $n=4\ 482$ ) and NFBC1986 ( $n=1\ 010$ ), respectively (Fig. 4). The concentrations seemed slightly higher in women, which could potentially be attributed to their lower muscle mass and subsequently lower circulating creatinine concentrations, resulting in lower creatinine excretion into the urine.



**Fig. 4. Distribution of urinary glucose concentration in NFBC1966 (n=4 482) and NFBC1986 (n=1 010). Values of >60  $\mu\text{M}/\text{mM}$  creatinine are not drawn for clarity (n=106 in NFBC1966 and n=7 in NFBC1986). Fig.4 is modified under CC BY License from Original publication I © 2022 Authors (Li, Ihanus, et al., 2022).**

### **5.1.2 Correlation and regression analyses of urinary glucose in NFBC1966**

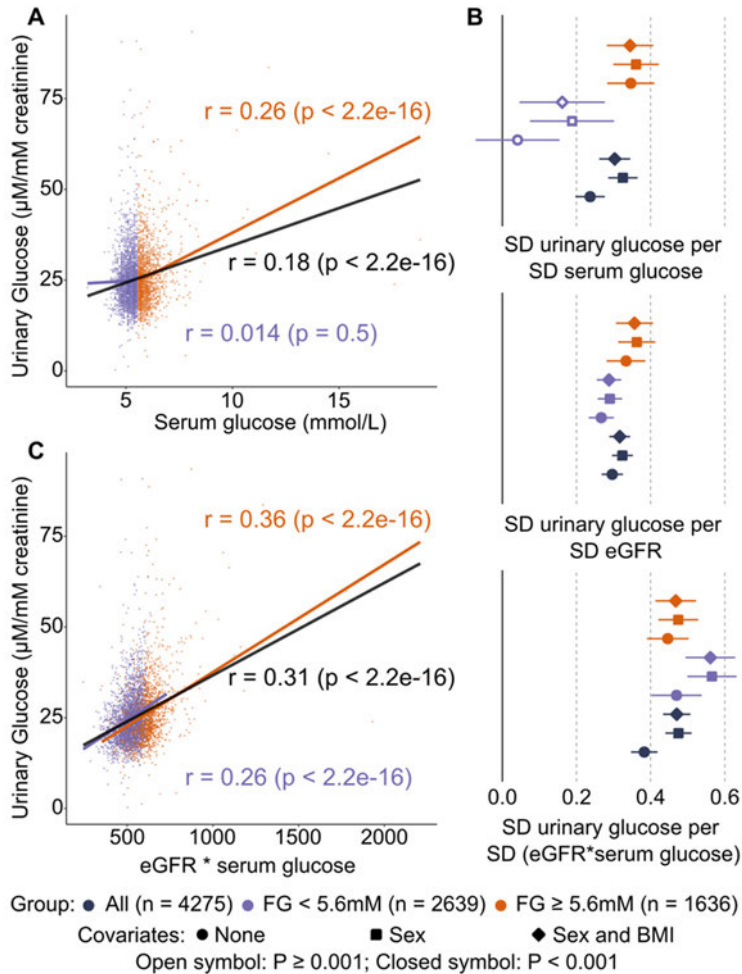
It is reasonable to expect that, under normoglycaemia, the kidneys would reabsorb all the filtered glucose. This reabsorption primarily occurs through two main processes: the sodium-glucose cotransporter 2 (SGLT2) in the early proximal tubule reabsorbs 97% of filtered glucose, and the remaining 3% is reabsorbed by sodium-glucose cotransporter 1 (SGLT1) in the late proximal tubule (Vallon, 2020). However, at elevated plasma glucose concentrations (>10 mmol/L), the capacity for tubular glucose reabsorption becomes saturated, leading to excretion of majority of filtered glucose into the urine.

The majority of the participants in NFBC1966 exhibited normoglycaemia (serum fasting glucose <5.6 mmol/L), indicating that their kidneys operating well below the saturation points for glucose reabsorption. As a result, no association was observed between serum and urinary glucose levels. However, the correlation appeared when the serum fasting glucose  $\geq 5.6$  mmol/L (Fig. 5A). This could be attributed to the early stages of glucose reabsorption saturation. Nevertheless, the consistently low concentration of glucose in the urine during normoglycaemia requires another explanation.

These findings could be interpreted as minor defects the coordinated operation of SGLT2 and SGLT1. Given that the kidneys filter the entire volume of plasma approximately 60 times per day, it is not surprising that a few glucose molecules might evade reabsorption by SGLT2 and SGLT1, although the process is recognized for its high efficiency (Vallon, 2020). This results in the non-specific excretion of glucose into the urine, even in instances of normoglycaemia.

The results of three regression analyses of urinary glucose with serum glucose, eGFR, and the multiplication of serum glucose and eGFR are shown in Fig. 5B. In all cases, positive associations were observed, indicating that urinary glucose concentrations are influenced by both circulating glucose and eGFR. Among normoglycemic individuals, the association between urine glucose and serum glucose was notably weak, aligning with the pattern seen in Fig. 5A. However, the relationship between eGFR and urinary glucose remained relatively consistent across all serum glucose concentrations. The most robust association was found between urinary glucose and the multiplication of serum glucose and eGFR, with consistent trends observed across all serum glucose levels, as shown in the correlation plot in Fig. 5C.

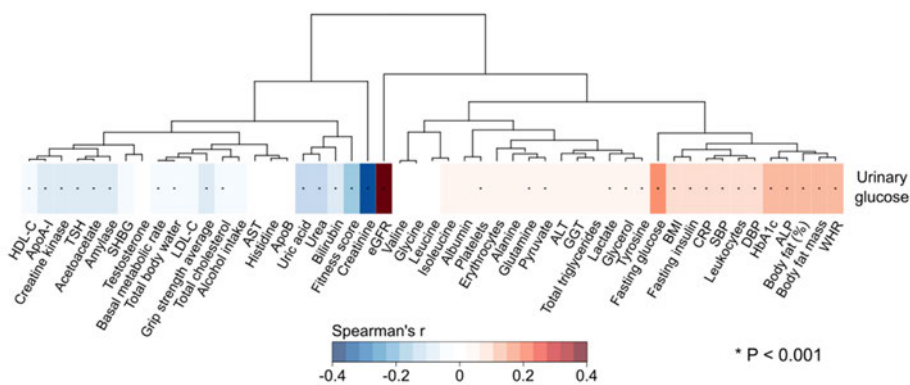
The interpretation of eGFR provides a clear indication of the connection between circulating glucose levels and urinary glucose concentrations. This discovery implies that quantitative measurements of urine glucose may offer valuable insights into renal function and the risk of disease at the population level. However, epidemiological data cannot directly assess how gluconeogenesis or other molecular mechanisms in the kidney impact urinary glucose levels. Additionally, based on the current data, potential changes in renal function associated with long-term conditions (i.e. diabetes) or immediate systemic conditions (i.e. meal-induced hyperglycaemia) cannot be speculated. Nonetheless, the direct quantitative molecular data obtained by NMR spectroscopy unequivocally demonstrate the consistent presence of glucose in the urine of normoglycemic individuals. Indeed, it is crucial for elementary teaching materials to acknowledge and rectify this fundamental issue.



**Fig. 5. Correlation and regression analyses of urinary glucose in NFB1966. The correlation analyses of urinary glucose with (A) serum glucose and with (C) the multiplication of serum glucose and eGFR are depicted for all participants, for normoglycemic participants (serum fasting glucose  $< 5.6$  mmol/L;  $n = 2639$ ) and for those whose serum fasting glucose concentrations were  $\geq 5.6$  mmol/L ( $n = 1636$ ). (B) Results from linear regression models for urinary glucose with serum glucose, eGFR or the multiplication of serum glucose and eGFR as the explanatory variable. The results for all participants, those who were normoglycemic and those whose serum fasting glucose concentrations were  $\geq 5.6$  mmol/L are colour coded in (A)-(C) (black, purple, and orange, respectively). Fig.5 is modified under CC BY License from Original publication I © 2022 Authors (Li, Ihanus, et al., 2022).**

### 5.1.3 Novel associations of urinary glucose with customary clinical and biochemical measures

Spearman's rank correlation, adjusted for sex between urinary glucose with a diverse array of customary clinical and biochemical measures at the population level, is shown in Fig. 6. For example, positive associations with urinary glucose were observed with body mass and composition (such as BMI, waist-to-hip-ratio, body fat mass, body fat [%]), diabetes risk (fasting insulin and HbA1c) and blood pressure. Urinary glucose was negatively associated with fitness indicators (grip strength, fitness score and body water content), with a few metabolic enzymes (amylase and creatine kinase) and with TSH. The negative associations were also seen with HDL-C and apolipoprotein A-I (ApoA-I), but no associations were seen with LDL-C, triglycerides, or apolipoprotein B (ApoB). In addition, a positive association was observed with alkaline phosphatase, but no association was shown with other commonly measured liver enzymes (gamma-glutamyl transferase, alanine, and aspartate aminotransferase). The observed associations of urinary glucose concentrations with a range of clinical and biochemical measures at the population level were first revealed, which suggested the non-specific release of glucose into the urine may have epidemiological and translational implications.



**Fig. 6. Partial rank correlations adjusted for sex to illustrate the associations between urinary glucose and 51 customary clinical and biochemical measures. The heat map is organized via column dimensional hierarchical clustering. Abbreviations: ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, gamma-glutamyl transferase; WHR, waist-to-hip ratio. Fig.6 is modified under CC BY License from Original publication | © 2022 Authors (Li, Ihanus, et al., 2022).**

### **5.1.4 Summary of Study I**

The present study demonstrates that the currently introduced method, the quantitative high-throughput urine NMR metabolomics platform, is feasible in large-scale epidemiological and clinical studies. Quantitative urinary glucose concentrations in two independent populations confirm the consistent presence of glucose in normal urine. Associations of urinary glucose with serum glucose, eGFR, and the multiplication of serum glucose and eGFR suggest that urinary glucose concentrations are influenced by circulating glucose and eGFR. Quantitative urine glucose data were also placed in an epidemiological context and in employing multiple routine clinical and biochemical measures, opening novel possibilities for exploring new insights into renal function and/or the risk of disease at the population level.

## **5.2 Characteristics of normalization methods in quantitative urinary metabolomics (Study II)**

Urine acts as the metabolic sewage of the human body and therefore its chemical properties are uncontrolled, contradicting the strict physiological regulation of blood plasma. This leads to a critical problem in urine metabolite analysis that volumes and metabolite concentrations vary widely, even within the same individual (Tynkkynen et al., 2019). Therefore, urinary metabolomics requires a normalization process to account for such variations. While utilizing 24-hour urine collections would greatly reduce dilution effect of urine samples, and morning spot urine samples partially address this issue (Vart & Grams, 2016). It is worth noting that these methods of urine collection are often impractical.

Because individual plasma creatinine concentrations are nearly constant over a wide range of GFR, and creatinine is freely filtered and not reabsorbed in the kidneys, the gold standard protocol in this field is to standardize to the concentration of urine creatinine (Kwan et al., 2020; Mathé et al., 2014; Mutter et al., 2022; Tynkkynen et al., 2019). GFR represents the volume of fluid filtered from the blood by the kidneys before water is reabsorbed. Consequently, GFR is not affected by occasional changes in urine output. While normalizing metabolite levels to urinary creatinine could exclude confounding factors of urine volume regulation (Eaton & Pooler, 2018), the concentrations of circulating creatinine are influenced by muscle mass and slight amounts of creatinine secreted by the proximal tubules, leading to a potential case-dependent bias (Eaton & Pooler, 2018;

Forni Ognà et al., 2015). Therefore, a number of other standardized methods, such as constant sum normalization (Emwas et al., 2016) and PQN (Dieterle et al., 2006) have been employed. The present study is the first to systematically compare various normalized methods for quantitative urinary metabolomics in epidemiological applications. Our main goals in this study were (1) to compare various normalization schemes for quantitative urine metabolomics data with conventional normalization using urinary creatinine, and (2) to determine the rationale for selecting appropriate normalization strategies for analysing urinary metabolite concentrations in epidemiology studies.

### **5.2.1 Intra-urinary metabolite associations and metabolic cluster analysis**

Urinary metabolite-metabolite associations in seven different normalization methods (Table 9) are compared in Figures 7–10. Fig. 7 illustrates the intra-urinary metabolite associations that have been normalized by IS-CREA (top-right), which is the prevailing approach for normalizing concentrations of urine metabolite. In addition, it displays associations using quantified absolute metabolite concentrations (down-left). To provide a clearer overview of the metabolite associations, a two-dimensional hierarchical clustering was applied to the associations based on IS-CREA. This clustering was then utilized in all visual representations for the other normalization methods, facilitating direct visual comparisons.

The general patterns of association among urinary metabolites are summarized by the eight distinct metabolic clusters, denoted by numbers on the left side of Fig. 7. Cluster 6, the largest cluster, demonstrates strong associations between all quantified amino acids, 3-hydroxyisobutyrate, 3-hydroxyisovalerate, and lactate. It exhibits abundant positive connections to Clusters 1–5, with some sparse negative links to Clusters 7 and 8. Cluster 1, the second largest cluster, encompasses various metabolites related to diet (arabinose, xylose, sucrose, and 2-furoylglycine), carbohydrate metabolism (cis-aconitate and glucuronate), microbial metabolism (4-hydroxyphenylacetate and 4-hydroxyhippurate), trans-aconitate and Sumiki's acid. It shows positive links to all other clusters but characterized by numerous negative correlations with Cluster 8. Cluster 7 is comprised of microbial metabolites 3-hydroxyhippurate and 3-(3-hydroxyphenyl)-3-hydroxypropanoate (HPHPA), along with hippurate and trigonelline. Cluster 2 combines metabolites related to carbohydrate metabolism (glucose and citrate) and microbial metabolism

(acetate and formate). Cluster 4 combines creatine and urea, both associated with amino acids metabolism, while Cluster 8 includes N1-Methyl-2-pyridone-5-carboxamide (2-PY) and N1-methylnicotinamide from nicotinate and nicotinamide (NAM) metabolism, along with uracil and hypoxanthine from nucleotide metabolism, and 3-methylhistidine related to diet. The other Clusters 3 and 5 are relatively small and consist of a diverse mixture of metabolites. However, all the metabolites in Cluster 5 could be identified as microbial metabolites.

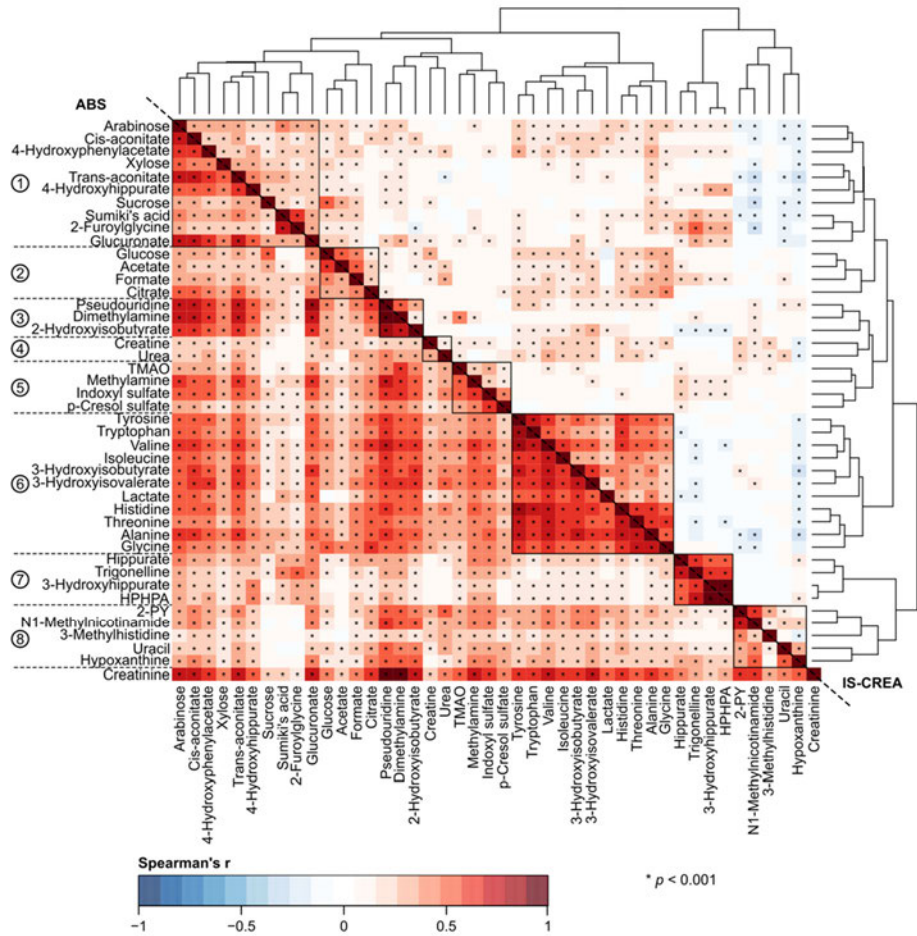
Notable similarities and dissimilarities of all eight different metabolite association patterns were revealed in Figures 7–10 (note the mirroring by the diagonal when comparing). The heat maps normalized using IS-CREA (Fig. 7 top-right) and IS-PSEURID (Fig. 9 down-left) display globally similar association patterns. This similarity is also observed between absolute concentrations (Fig. 7 down-left) and IS-UREA (Fig. 8 top-right), as well as the normalization by PQN (Fig. 10 down-left) and DESeq2 (Fig. 10 top-right). However, IS-GLUC (Fig. 8 down-left) and the CS normalization (Fig. 9 top-right) did not align well with any overall association structure in the other methods.

In terms of the absolute concentrations and the normalizations via IS-GLUC, IS-UREA, IS-PSEURID, and CS, the associations within the metabolite clusters are predominantly strongly positive, mirroring those observed with IS-CREA normalization, which was used in the cluster formation. However, the overall association structure for PQN and DESeq2 appear somewhat weaker compared to the other normalization methods, particularly in terms of many metabolite associations within clusters. While the CS normalization aligns quite closely with PQN and DESeq2 in the case of the overall associations for Clusters 7 and 8, it differs in its associations between Cluster 6 and Clusters 1–4, which are generally positive. Moreover, the positive links between Clusters 1 and 2 as well as Cluster 1 and 3 for the CS normalization contrast with those observed for the PQN and DESeq2 normalizations. As for these association patterns, the CS normalization closely resembles that of IS-CREA. The associations in the PQN and DESeq2 normalized data do not exhibit much similarity to those of the other normalization methods.

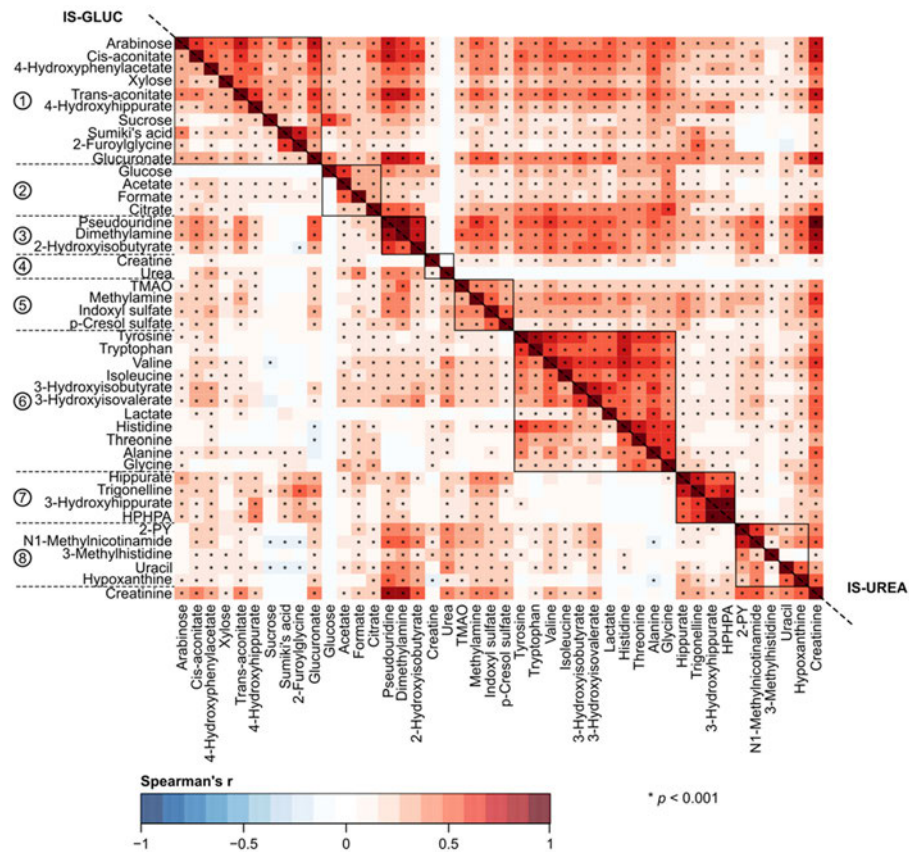
The associations between creatinine with the other urinary metabolites for all the normalization methods (except IS-CREA) are also depicted in Figures 7–10. The patterns are highly consistent for PQN and DESeq2 normalized data, exhibiting more negative associations than positive ones. In contrast, the associations vary somewhat among the other normalization methods, where nearly

all creatinine associations are positive. The heat maps without sex adjustment for Figures 7–10 closely resemble those in the sex-adjusted analysis.

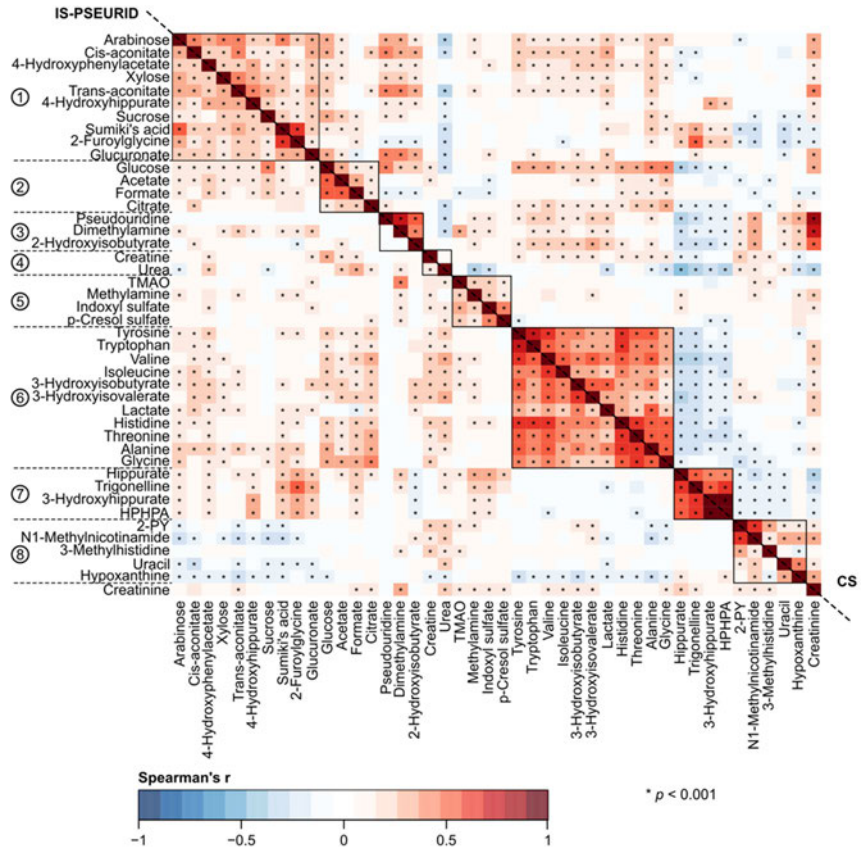
As outlined in Table 9, both urinary glucose (Li, Ihanus, et al., 2022) and urea (Eaton & Pooler, 2018; Gowda et al., 2010) have physiologically relevant limitations in being the optimal individual metabolite reference. Comparing the associations between metabolites in the normalized data using IS-GLUC and IS-UREA (Fig. 8) with those using absolute metabolite concentrations (Fig. 7) reveals very similar and predominantly positive correlations. This indicates that metabolite concentrations normalized via IS-GLUC and IS-UREA still retain a considerable component related to the total volume and dilution of the urine sample. Furthermore, urinary creatinine (Eaton & Pooler, 2018; Forni Ognà et al., 2015; Herrera & Rodríguez-Iturbe, 1998; Saccenti, 2017; Waikar et al., 2010) and pseudouridine (Bernert et al., 1988; Niewczas et al., 2017; Sekula et al., 2016) have fundamentally distinct biological backgrounds and limitations as reference molecules (Table 9). However, their correlation heat maps being almost identical (top-right in Fig. 7 and bottom-left in Fig. 9, respectively). Therefore, IS-PSEURID serve as a valuable normalization method to validate the consistency of results obtained using IS-CREA under different biological assumptions.



**Fig. 7.** The urinary metabolite-metabolite associations as indicated by Spearman's rank correlations adjusted for sex. The down-left triangle shows results for the absolute urinary metabolite concentrations (i.e. no normalization applied) and the top-right triangle for the creatinine normalization (IS-CREA). To make detailed comparisons of the metabolite-metabolite associations feasible between the different normalization schemes, two-dimensional hierarchical clustering was applied to organize the IS-CREA heat map. All heat maps are presented in the same order of metabolites with creatinine added to the last row. The reference metabolite correlations are left blank in their corresponding heat maps. As a by-product of the hierarchical clustering, eight urinary metabolite clusters were identified as numbered on the left and detailed in the Results section. Fig.7 is reprinted under CC BY License from Original publication II © 2022 Authors (Li, Tynkkynen, et al., 2022).



**Fig. 8.** The urinary metabolite-metabolite associations as indicated by Spearman's rank correlations adjusted for sex. The down-left triangle shows results for the glucose normalization (IS-GLUC) and the top-right triangle for the urea normalization (IS-UREA). The order of metabolites (with creatinine added to the last row) is the same as in Fig. 7 and based on the two-dimensional hierarchical clustering of the IS-CREA heat map. The reference metabolite correlations are left blank in their corresponding heat maps. Fig.8 is reprinted under CC BY License from Original publication II © 2022 Authors (Li, Tynkkyinen, et al., 2022).



**Fig. 9.** The urinary metabolite-metabolite associations as indicated by Spearman's rank correlations adjusted for sex. The down-left triangle shows results for the pseudouridine normalization (IS-PSEURID) and the top-right triangle for the constant sum normalization (CS). The order of metabolites (with creatinine added to the last row) is the same as in Fig. 7 and based on the two-dimensional hierarchical clustering of the IS-CREA heat map. The reference metabolite correlations are left blank in their corresponding heat maps. Fig.9 is reprinted under CC BY License from Original publication II © 2022 Authors (Li, Tynkkynen, et al., 2022).

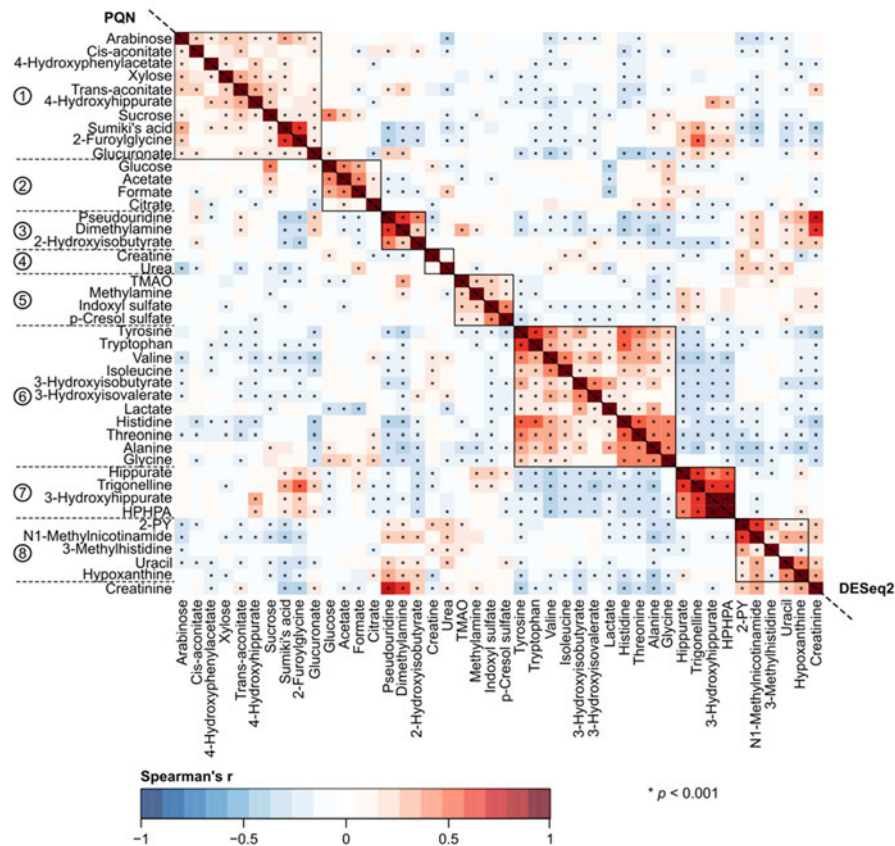


Fig. 10. The urinary metabolite-metabolite associations as indicated by Spearman's rank correlations adjusted for sex. The down-left triangle shows results for the probabilistic quotient normalization (PQN) and the top-right triangle for the DESeq2 normalization. The order of metabolites (with creatinine added to the last row) is the same as in Fig. 7 and based on the two-dimensional hierarchical clustering of the IS-CREA heat map. The reference metabolite correlations are left blank in their corresponding heat maps. Fig.10 is reprinted under CC BY License from Original publication II © 2022 Authors (Li, Tynkkyinen, et al., 2022).

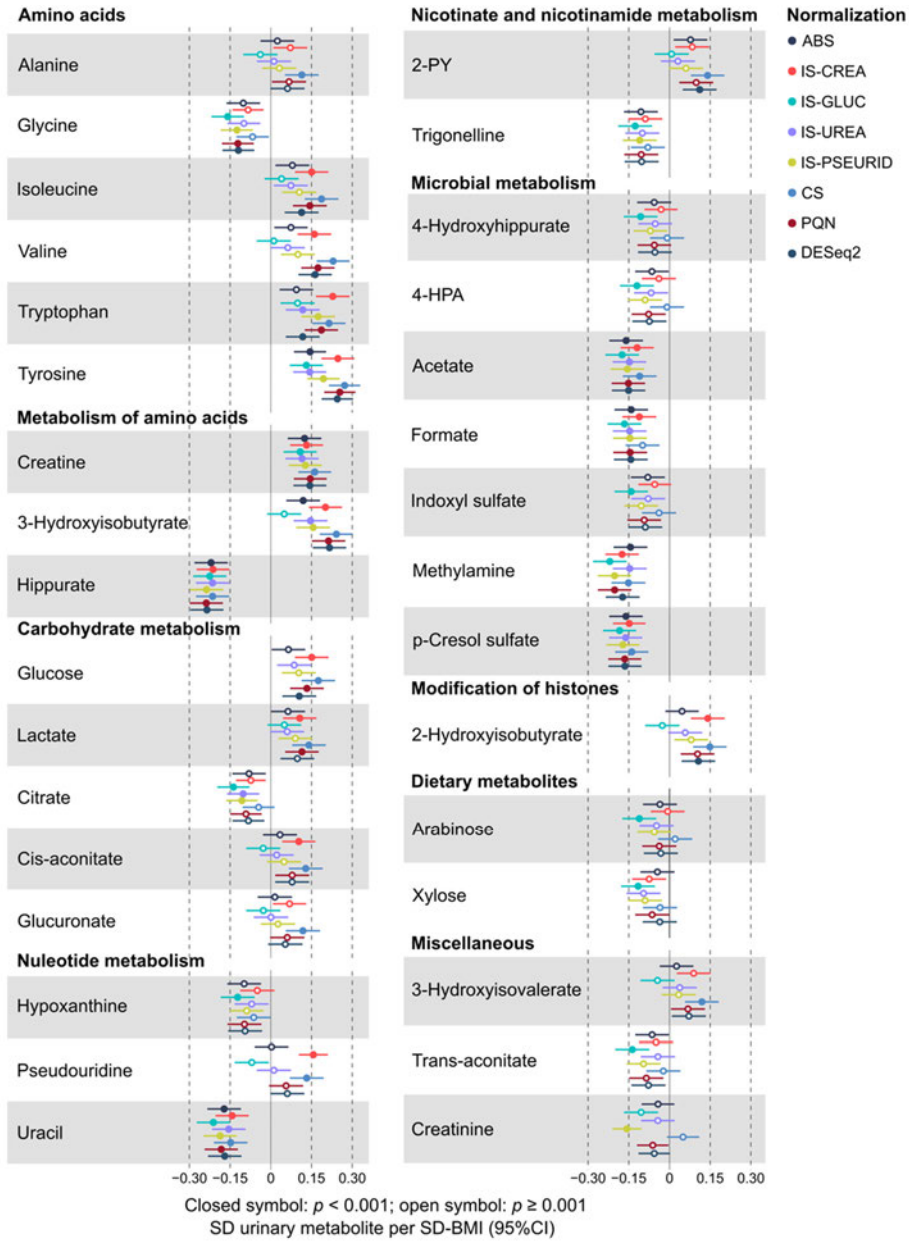
### 5.2.2 Epidemiological Exemplars: BMI and MAP

The linear regression analyses utilized urinary metabolite concentrations from seven normalization protocols, along with their absolute concentrations. Taking BMI and MAP as examples, robust associations were found for sex-adjusted BMI

with 18 urinary metabolites (Fig. 11) and MAP with 18 urinary metabolites (Fig. 12). Overall, for these epidemiological associations, results from all normalized data, as well as absolute concentrations of urinary metabolites, exhibit consistent outcomes. However, a more nuanced examination shows that the results for the concentrations normalized via IS-GLUC tend to differ slightly from other approaches. This is evident, for example, in various amino acid related metabolites and 2-hydroxyisobutyrate for both BMI and MAP. Apart from these biases for IS-GLUC, and some occasional deviations for IS-PSEURID, IS-UREA, and CS normalization approaches, the results of these epidemiological regression analyses are surprisingly consistent. In particular, the results of PQN and DESeq2 closely align with those of IS-CREA and CS normalization.

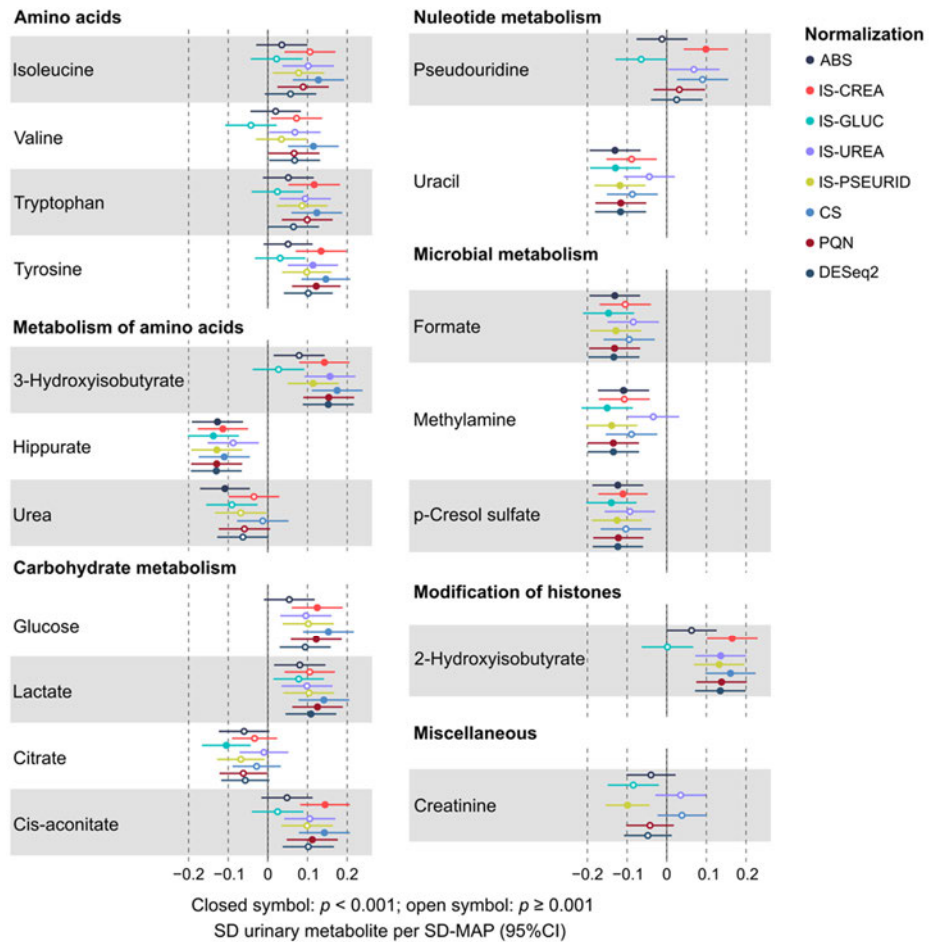
Among seven normalization protocols, three multi-metabolite normalization schemes were employed, which were based on the idea that the sum of multiple metabolites help stabilize individual metabolite variations and enhance the ability to correct for the variable dilution of urine samples (Craig et al., 2006; Torgrip et al., 2008). In present study, three such schemes—the CS normalization, PQN, and DESeq2—yielded intricate patterns of similarities and differences in metabolite-metabolite associations but exhibited remarkable consistency in epidemiological regression analyses concerning the two clinical measures. In these analyses, their estimates closely aligned with those of the IS-CREA. For epidemiological studies, these findings indicated that when a wealth of individual quantified metabolites are available, employing these multi-metabolite normalization protocol could examine the consistency of IS-CREA results and evaluate whether they could strengthen some associations due to potentially better accounting for the biologically relevant concentration changes of the urinary metabolites.

In addition, urine is the primary biological material used in the study of renal diseases that often alter creatinine secretion and excretion, which could compromise the reliability of the IS-CREA method. On the contrary, the quantile based DESeq2 method has become the standard choice for RNA analysis, particularly in situations where internal standards such as creatinine are not available. Hence, we suggest considering and perhaps placing greater reliance on the PQN method in cases where the accuracy of creatinine measurement may be compromised or influenced by pathophysiological factors.



**Fig. 11. The robust associations ( $p < 0.001$ ) of the urinary metabolite concentrations with BMI for the various normalization schemes. Abbreviations: 4-HPA, 4-**

hydroxyphenylacetate. Fig.11 is reprinted under CC BY License from Original publication II © 2022 Authors (Li, Tynkkynen, et al., 2022).



**Fig. 12. The robust associations ( $p < 0.001$ ) of the urinary metabolite concentrations with MAP for the various normalization schemes. Fig.12 is reprinted under CC BY License from Original publication II © 2022 Authors (Li, Tynkkynen, et al., 2022).**

### 5.2.3 Summary of Study II

The present study is the first to compare the various normalization protocols in the context of quantitative urinary metabolomics in an epidemiological framework.

Forty-four urinary metabolites (Table 7) were quantified in population-based morning urine samples from 994 individuals in NFBC1966 via NMR spectroscopy. Seven distinct normalization protocols were used to analyse metabolite-metabolite associations and the association of metabolite concentrations with two representative clinical measures (BMI and MAP). Four urinary metabolite concentrations were used as internal standards, such as creatinine (routine normalization method), glucose, urea, and pseudouridine. In addition, constant sum normalization, together with its two more complex versions (PQN and DESeq2) were tested. Since the study sample was morning spot urine, we also performed a comparison of the raw absolute concentrations of urine metabolites in the samples. Table 9 summarizes the biological rationale for each normalization scheme as well as their advantages and disadvantages for epidemiological studies. The present study focused on advancing the urine NMR metabolomics platform for large-scale epidemiology and genetics. The aim was to propose a high-throughput and cost-effective protocol for tens of thousands of samples in different laboratories and consistent with epidemiological and clinical sample collection routines (Tynkkynen et al., 2019). Therefore, by definition, a number of normalization approaches, such as urine osmolality or dry mass, are not considered feasible in this case (Cook et al., 2020; Wu & Li, 2016).

The results in this work can be summarized that although the impact and differences of various data normalization methods for urinary metabolite-metabolite associations (Figures 7–10) are pronounced, the epidemiological association patterns for clinical measures, e.g. BMI and MAP (Figures 11 and 12) are consistent. Therefore, especially for epidemiological research on the role of urinary metabolites, the custom of referencing to urinary creatinine seems sensible. Referencing to urinary pseudouridine may be helpful in checking the consistency of results with reference to creatinine under different biological assumptions. If considerable individual quantified metabolites are available, it might be useful to apply multi-metabolite normalization methods (e.g. CS and PQN) to check the consistency of results with reference to creatinine, and to assess whether they could enhance certain associations induced by biologically relevant concentration changes of urinary metabolites. The application of urinary metabolite concentrations for epidemiological analysis requires the comparison of different normalization protocols, considering physiology, metabolism, and renal function, which may help to explain intra-urinary metabolite associations in detail, as well as to nuancedly elucidate the epidemiological associations.

It is worth noting that the present study is based on a cross-sectional epidemiological dataset involving a limited number of individuals. For the aims of the study, the data were sufficient, as evidenced by the robustness of the multiple testing correction results across all the analyses. However, as is customary in epidemiological studies, potential biases related to the cohort cannot be fully disentangled, and caution should be exercised when generalizing the findings. The majority of associations between urinary metabolites with clinical measures have been previously observed in other cohorts and datasets. This suggests that the key findings related to the impact of different normalization methods would be broadly valid. With the availability of independent data, replication of these findings would, of course, be preferable. Furthermore, it is important to bear in mind that the participants involved in the present study were selected from a birth cohort with a average age of 46 years, representing a population of rather healthy individuals with only relatively minor variations in kidney and other organ functions. Therefore, the participants represent a population of rather healthy individuals with only relatively minor variations in kidney and other organ and metabolic functions. Consequently, these results should not be extrapolated to individuals or patients with clear organ or metabolic dysfunction.

### **5.3 Clinical and biochemical associations of urinary metabolites (Study III)**

The number of urinary biomarkers widely utilized as diagnostic aids in renal disease (albumin and creatinine) and diabetes mellitus (glucose) is limited. The rationale for these measurements is primarily to pinpoint high values that cross pre-set diagnostic limits, such as the standard urine glucose test strip with a detection limit of up to 5.6 mmol/L (Li, Ihanus, et al., 2022). Quantitative urine metabolic approaches involving a large enough number of individuals for appropriate epidemiological studies, aiming at improving the understanding of urinary metabolites in health, and as potential biomarkers of the risk of disease are almost non-existent (Elliott et al., 2015; Emwas et al., 2016; Holmes et al., 2008; Mutter et al., 2022; Nicholson et al., 2011; Tynkkynen et al., 2019; van Duynhoven & Jacobs, 2016). This is in notable contrast to the situation with the various metabolomics and lipidomics approaches already widely used in epidemiology and genetics for large-scale quantitative studies of systemic blood biomarkers (Ala-Korpela et al., 2022; Ala-Korpela & Davey Smith, 2016; Deelen et al., 2019;

Ekholm et al., 2020; Huynh et al., 2019; Julkunen et al., 2023; Locke et al., 2019; Würtz et al., 2017).

Nonetheless, it has been recognized that urinary metabolites have potentiality in epidemiology and translational medicine (Bouatra et al., 2013; Elliott et al., 2015; Emwas et al., 2016; Holmes et al., 2008; Nicholson et al., 2011; Tynkkynen et al., 2019). Urinary metabolites are physiologically related to the renal glomerular filtration and molecular reabsorption processes and reflect several key biochemical pathways associated with cardiometabolic status, gut microbial metabolic activities, and dietary features. Thus, detailed quantitative data on urinary metabolites could provide direct molecular probes to evaluate renal function and its inference under various metabolic conditions. To achieve this far-reaching goal, we have recently developed the basis for an open-access methodology for quantitative high-throughput urinary NMR metabolomics. (Li, Ihanus, et al., 2022; Li, Tynkkynen, et al., 2022; Tynkkynen et al., 2019).

The present study aims to expand population-level quantitative urine metabolite data to 61 species and provide their sex-specific reference concentrations and distributions in a population sample of 994 individuals. The first coherent set of automated quantification models for 12 urinary metabolites (+creatinine) were also presented, and the concentrations of these metabolites as well as their associations with extensive clinical measures from three independent population cohorts (up to 5 989 participants) were investigated. The epidemiological scale of urine metabolomics is extended to a new level in the present study, which involves independent replication, and provides a wealth of new metabolic findings related to renal function with potential translational relevance.

### ***5.3.1 Quantitative urinary metabolite abundance at the population level***

In a population sample of 994 individuals, the set of 61 urinary metabolites representing all the most abundant signals in urine NMR spectra have been quantified. The absolute and urinary creatinine normalized concentrations of 61 quantified urinary metabolites are illustrated in Fig. 13. By far, the most abundant metabolite in urine is urea with a median absolute concentration of more than 200 mmol/L. Also abundant is creatinine, with median absolute concentrations of more than 10 mmol/L, while the median absolute concentrations of hippurate and citrate are more than 1 mmol/L. Quantified urinary metabolites have a substantial

population variation, and the distributions of many metabolites are positively skewed (Table 10 and supplementary Figures 3 & 4, in III). In creatinine referenced concentrations, many metabolites are slightly more abundant in women than in men, but the differences in concentrations are small (Table 7 and supplementary Figures 2–4, in III). On average, women have lower muscle mass, resulting in lower circulating creatinine concentrations and therefore lower creatinine excretion into the urine. In random urine samples, the absolute metabolite concentrations would not be relevant. However, for the morning spot urine samples, as shown in Fig. 13, biological variability was somewhat reduced because of the corresponding time and conditions of the urine generation, i.e. overnight in mostly a fasted physiological state (Manjunath et al., 2003).

The methodological intra-assay coefficients of metabolite variations in percent (CV%)s for the new quantified 17 metabolites was similar to that of the earlier quantified 43 metabolites (Tynkkynen et al., 2019), and their intra- and inter-individual population variations for 30 consecutive days follow the same overall pattern as previously reported (Table 10). Intra-assay CV% refer to analysis of one urine sample as 10 replicates, reflecting the entire quantification process, such as all sample preparation steps, NMR experiments, and quantitation protocols. Most intra-assay CV% are less than 5%, illustrating high robustness and accuracy of the urine NMR spectroscopy and the whole quantification process. Intra-individual metabolite variation was measured by taking the average of 3 different volunteers over 30 consecutive days of urine collection. Since urine is a waste product, urinary metabolites vary widely within individuals (CV% for most metabolites are more than 20% over 30 days). However, inter-individual variation calculated based on 1003 different individuals from the NFBC1966 was even higher (CV% for most metabolites in 1003 individuals exceed 40%). Therefore, although intra-individual variation in urinary metabolites is high, the even larger population-based inter-individual variation provides a good base for epidemiological and genetic applications.

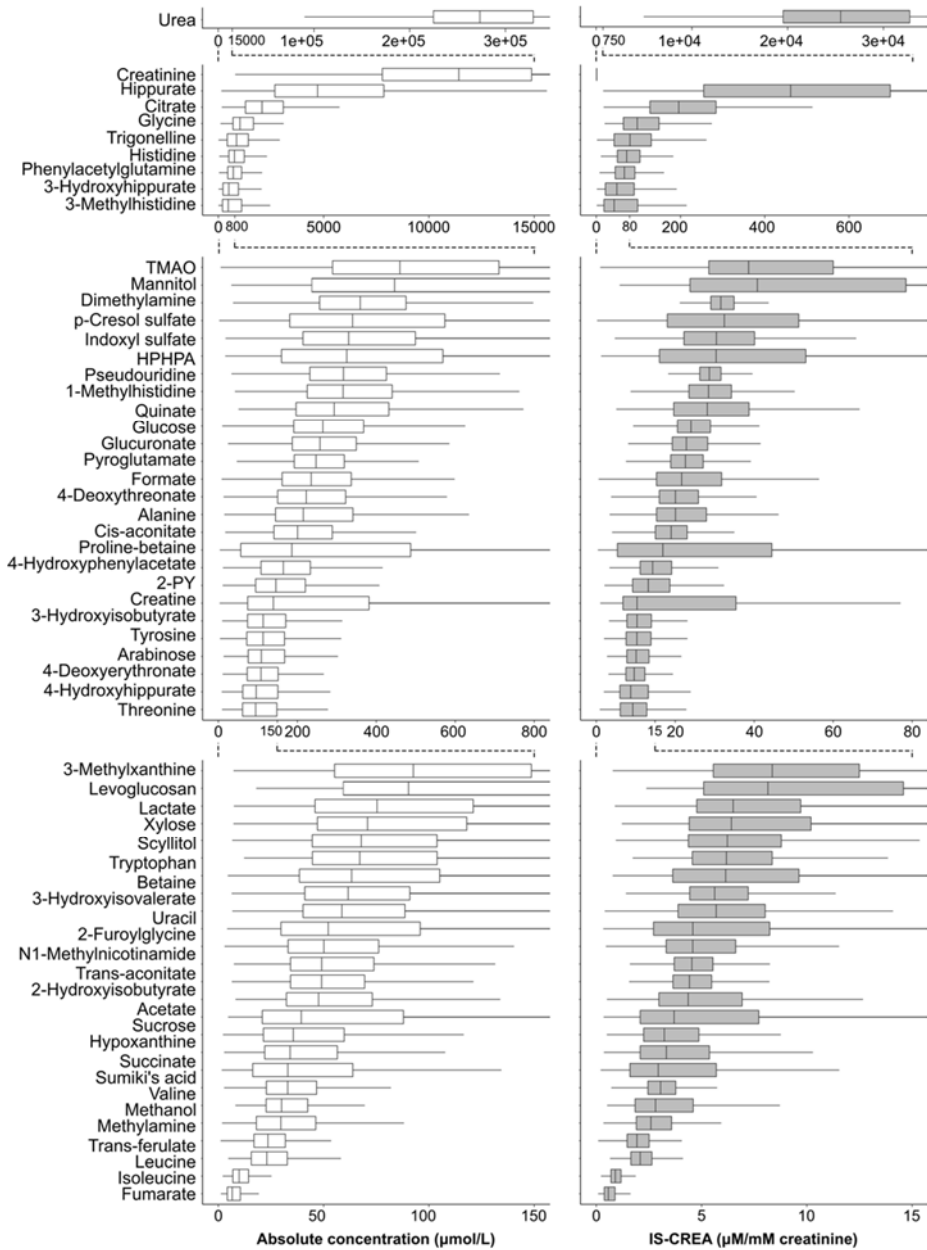


Fig. 13. Absolute (left) and creatinine-referenced (right) concentrations of 61 quantified urinary metabolites in a random subset ( $n=994$ ) of morning spot urine samples in the NFBC1966. The metabolites are presented in the descending order of median absolute

concentrations. Several different scales are used for the x-axes to provide a clear visualization for the large concentration ranges. Fig.13 is reprinted under CC BY License from Original publication III © 2023 Authors (Li et al., 2023).

**Table 10. Intra-assay, intra-individual and inter-individual variation of 60 quantified urine metabolites. Table 10 is modified under CC BY License from Original publication III © 2023 Authors (Li et al., 2023).**

Metabolite	Intra-assay CV (%) <sup>a,b</sup>	Intra-individual CV (%) <sup>a,c</sup>	Inter-individual CV (%) <sup>a,d</sup>
<b>Amino acids</b>			
Alanine	1.16	28.69	49.85
Glycine	2.21	34.71	73.99
Histidine	1.10	30.25	48.75
Threonine	4.57	38.58	75.55
Isoleucine	6.68	23.27	53.56
Leucine	6.34	26.48	58.25
Valine	4.72	20.28	39.22
Tryptophan	3.34	33.76	49.66
Tyrosine	3.35	32.09	45.76
<b>Metabolism of amino acids</b>			
Betaine	3.23	43.96	176.13
Creatine	4.12	126.19	240.36
3-Hydroxyisobutyrate	2.67	34.18	61.42
Hippurate	1.15	58.45	69.29
Phenylacetylglutamine	3.12	31.90	54.00
Urea	1.46	32.98	39.41
Pyroglutamate	4.40	17.64	27.64
<b>Carbohydrate metabolism</b>			
Glucose	2.91	13.76	1294.34
Lactate	4.28	44.26	477.64
Citrate	1.51	27.92	53.37
Cis-aconitate	0.85	22.28	39.64
Fumarate	- <sup>e</sup>	176.23	218.72
Succinate	13.36	32.08	179.91
Mannitol	- <sup>e</sup>	164.85	222.25
Glucuronate	4.07	18.31	66.41
Hypoxanthine	3.53	38.80	73.70
Pseudouridine	2.15	6.32	14.28
Uracil	4.29	37.71	148.66
<b>Nicotinate and NAM</b>			
2-PY	2.14	35.29	60.78
N1-Methylnicotinamide	1.32	28.24	52.30

Metabolite	Intra-assay CV (%) <sup>a,b</sup>	Intra-individual CV (%) <sup>a,c</sup>	Inter-individual CV (%) <sup>a,d</sup>
Trigonelline	0.79	68.71	74.78
<b>Caffeine metabolism</b>			
3-Methylxanthine	5.02	174.27	84.08
<b>Microbial metabolism</b>			
3-Hydroxyhippurate	2.56	51.81	100.10
4-Hydroxyhippurate	3.43	34.85	71.88
4-Hydroxyphenylacetate	2.24	28.73	52.08
Acetate	14.17	62.87	394.40
Dimethylamine	0.74	9.79	30.50
Formate	8.71	41.32	587.34
HPPHA	4.30	67.68	99.97
Methanol	1.91	60.33	114.21
Methylamine	3.17	32.07	49.79
p-Cresol sulphate	1.53	35.65	71.32
Trimethylamine <i>N</i> -oxide	1.63	80.89	127.18
<b>Modification of histones</b>			
2-Hydroxyisobutyrate	1.15	16.25	35.39
<b>Dietary metabolites</b>			
1-Methylhistidine	2.04	21.15	31.26
2-Furoylglycine	5.46	225.45	212.89
3-Methylhistidine	1.56	95.44	117.02
Arabinose	3.58	35.50	59.28
Indoxyl sulphate	1.59	32.24	46.79
Levoglucofan	1.52	304.04	190.10
Proline-betaine	2.71	132.03	139.93
Quinate	3.51	262.56	81.67
Scyllitol	1.19	22.05	57.91
Sucrose	4.45	194.15	555.87
Trans-ferulate	4.71	31.12	101.66
Xylose	3.38	99.60	112.05
<b>Miscellaneous</b>			
3-Hydroxyisovalerate	4.84	66.55	46.13
4-Deoxyerythronate	1.59	18.15	38.46
4-Deoxythreonate	1.67	30.42	38.60
Sumiki's acid	2.36	35.23	133.83
Trans-aconitate	4.42	50.71	59.02

<sup>a</sup>Concentrations are scaled to the concentration of creatinine; CV% = (standard deviation / average) × 100%

<sup>b</sup>One urine sample prepared and analysed as 10 replicates; reflects the entire quantitative process.

<sup>c</sup>A 30-day consecutive urine collection, averaged over 3 different volunteers.

<sup>d</sup>One thousand and three different individuals from the NFBC1966.

<sup>e</sup>Concentration of the metabolite below the detection limit in this urine sample.

### **5.3.2 Association clusters of urinary metabolites**

The associations between 12 urinary metabolites (referenced to urinary creatinine, Table 8) and 49 conventional clinical and biochemical measures (details are provided in 4.1.1 and Tables 5 and 6) are presented in a color-coded heat map (Fig. 14). Because quantitative urinary metabolites data at an epidemiological scale are scarce, the described associations are mostly novel. Although these associations are overall rather weak, 362 robust associations were detected (denoted with \* at  $p < 0.0009$ ). The heat map based on sex-adjusted associations across the entire NFBC1966 cohort ( $n = 4\ 505$ ) was organized via two-dimensional hierarchical clustering. Four 3-metabolite clusters were rendered, reflecting the clinical and biochemical associations of 12 urinary metabolites.

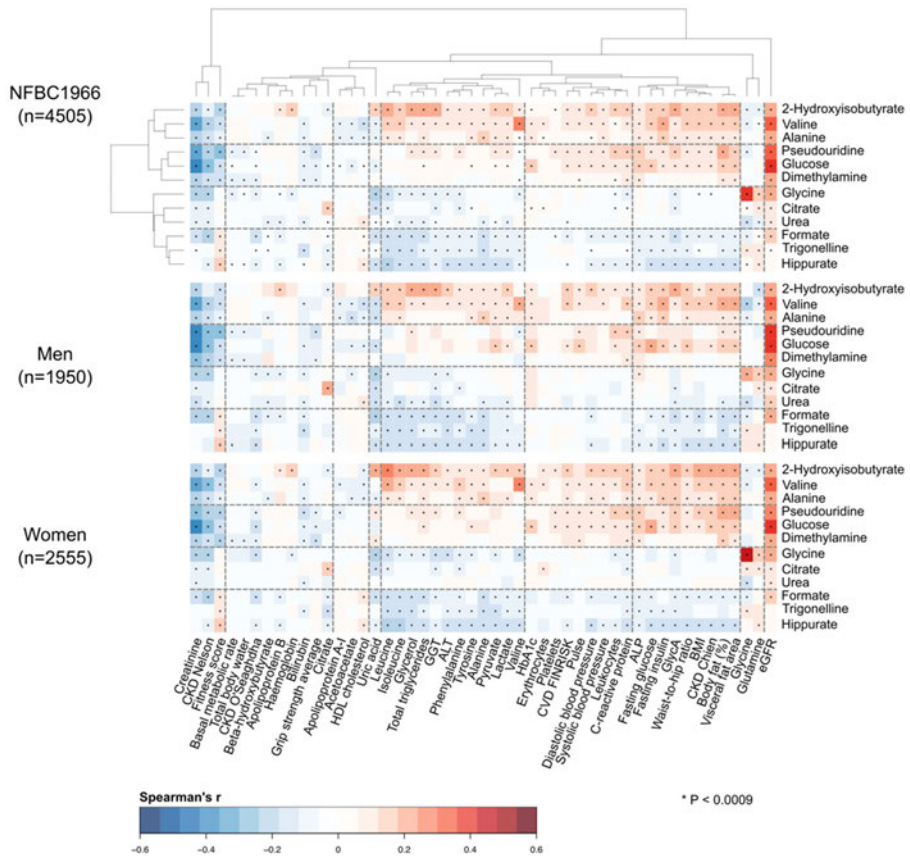
The strongest association is between urinary and serum glycine, and the association between urinary and serum valine was also quite strong. Positive correlations are observed between eGFR and all the urine metabolite clusters, with the strongest correlations for glucose, pseudouridine and valine. As expected, serum creatinine (used to estimate GFR) is negatively associated with these clusters. The two uppermost metabolite clusters in Fig. 14 (the first consisting of 2-hydroxyisobutyrate, valine, and alanine and the second consisting of pseudouridine, glucose, and dimethylamine) in terms of associations are generally similar. They were overall positive with over 60% of clinical and biochemical measures, including eGFR, various serum amino acids, serum triglycerides, glycaemic traits, lactate, pyruvate, inflammation (CRP and glycoprotein acetylation [GlycA]), liver function markers (ALP, ALT and GGT), multiple obesity indicators, blood pressure, the FINRISK CVD, and the Chien CKD risk scores.

In Fig. 14, the associations of the two lowermost metabolite clusters (the first consisting of glycine, citrate, and urea and the second consisting of formate, trigonelline, and hippurate) also represent roughly similar, although the former has the fewest associations with clinical and biochemical measures. For most of the aforementioned positive associations of the two uppermost metabolite clusters, their associations are usually negative. However, the correlations for serum glycine and glutamine are somewhat opposite, showing weak negative correlations with the two uppermost clusters and mixed negative and positive correlations with the two lowermost clusters. In addition to serum creatinine, serum citrate, bilirubin, and the CKD Nelson and O'Seaghda risk scores tend to have negative associations with all the urinary metabolite clusters. ApoB is positively associated with the uppermost cluster and negatively associated with the lowermost cluster. HDL

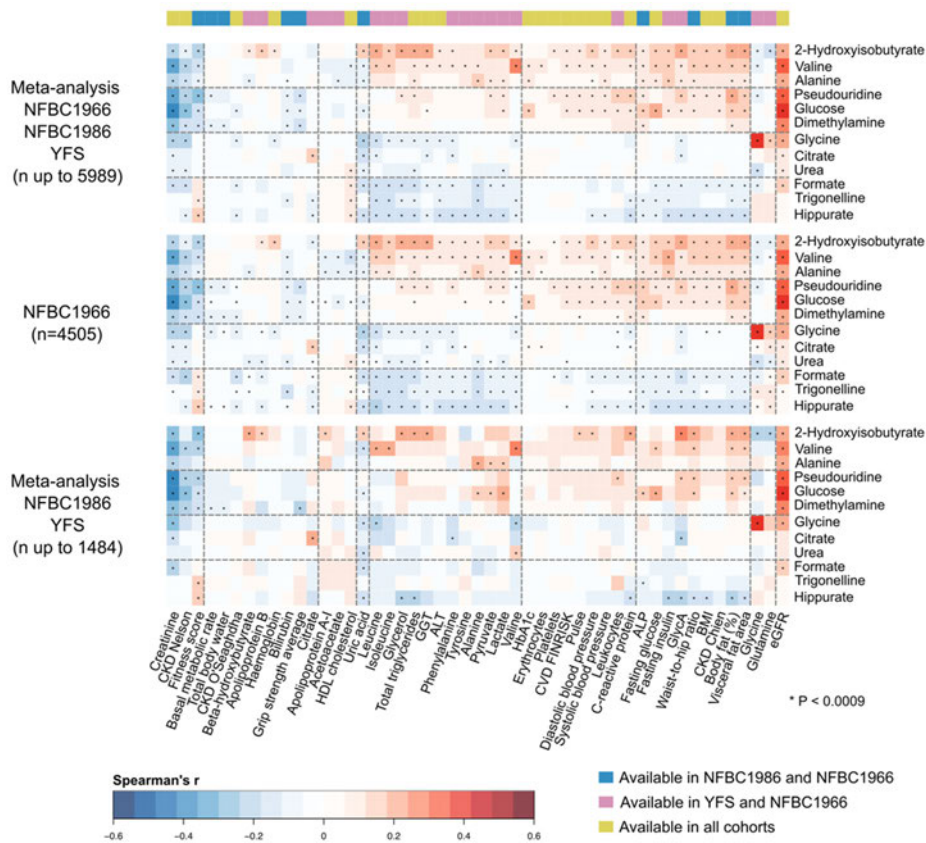
cholesterol has opposite associations to those of ApoB. And the fitness score is negatively correlated with the two uppermost clusters and positively correlated with the lowermost cluster.

In men and women, the correlations of urinary metabolite clusters and individual urinary metabolites with clinical and biochemical indicators are similar (Fig. 14). Furthermore, the correlations among the three independent cohorts appear to be consistent as shown in Fig. 15. Fewer urine samples were available for NFBC1986 (n=1 010) and YFS (n=474) than NFBC1966 (n=4 505). Thus, only the most prominent associations reach the multiple comparison corrected P-value threshold of 0.0009. Nonetheless, the overall association pattern closely matches that of NFBC1966.

As population-level biomarkers of health and disease, the various correlations between urinary metabolites and multiple clinical and biochemical indicators suggest that urinary metabolites are likely to have general value. In addition to confirmatory data on systemic metabolic results, urinary metabolites may be able to bring additional information, directly reflecting renal function if they are added to the systemic metabolic risk assessment. This is supported by previous findings which show that urine and serum metabolites are generally weakly correlated (Tynkkynen et al., 2019).



**Fig. 14.** The associations between the 12 automatically quantified urinary metabolites (referenced to urinary creatinine) and 49 customary clinical and biochemical measures as indicated by Spearman's rank correlations (adjusted for sex) for the entire NFBC1966 (n=4 505) as well as for men (n=1 950) and women (n=2 555). The two-dimensional hierarchical clustering is based on the results for the entire cohort and the resulting ordering is preserved in all the following heat maps. Four 3-metabolite clusters were rendered that reflect the clinical and biochemical associations of the urinary metabolites. P-value<0.0009 marked with \* in the map to indicate a robust, multiple testing corrected association. Fig.14 is reprinted under CC BY License from Original publication III © 2023 Authors (Li et al., 2023).



**Fig. 15.** Meta-analyses of the associations (Spearman's rank correlations adjusted for sex) between the 12 automatically quantified urinary metabolites (referenced to urinary creatinine) and 49 customary clinical and biochemical measures to illustrate the replication of the findings in all the three independent population cohorts. The uppermost heat map shows the full meta-analyses for all the available data (n up to 5 989). The heat map in the middle is for the entire NFBC1966 (the same heat map as in Fig. 14 to facilitate visual comparison). The lowermost heat map shows the meta-analysis for NFBC1986 and YFS (n up to 1 484). The heat maps are presented in the same order of metabolites and clusters as in Fig. 14. The colour key on the top of the figure represents the availability of clinical and biochemical measures in the three cohorts. There were 20 measures available in all the 3 cohorts (green), 19 measures available only in NFBC1966 and YFS (pink), and 10 measures available only in NFBC1966 and NFBC1986 (blue). P-value<0.0009 marked with \* in the map to indicate a robust, multiple testing corrected association. Fig.15 is reprinted under CC BY License from Original publication III © 2023 Authors (Li et al., 2023).

### **5.3.3 Metabolite associations with BMI**

Fig. 16A presents meta-analysis of associations of 12 urinary metabolites (referenced to creatinine) with BMI for the three independent cohorts (n up to 5 989). The results for each cohort are shown in supplementary Figure 9 in III. Only urea and dimethylamine were not notably associated with BMI. The base model was adjusted for sex (and age in YFS), then seven different adjustments (MAP, fasting glucose, fasting insulin, smoking history, total triglycerides, CRP and eGFR) were separately added to evaluate their potential confounding role. Fasting insulin strongly affected the association of multiple urinary metabolites with BMI. The strongest effects were for valine and alanine (amino acids), trigonelline (nicotinate and nicotinamide metabolism), and hippurate (amino acid metabolism). In addition to the effect of fasting insulin, the other adjustments in the regression model had very little, if any, effect on the overall associations. However, the adjustment of fasting glucose had similar but less pronounced effect on valine and alanine as fasting insulin. Adjusting for fasting glucose also had an expected strong effect on weakening the association of urinary glucose with BMI. The association of 2-hydroxyisobutyrate was strongly influenced by adjustments for MAP, fasting glucose, fasting insulin, total triglycerides, and CRP. These findings are consistent with recent longitudinal findings on trends in systemic metabolic ageing and obesity (Mäkinen et al., 2023; Miccheli et al., 2015).

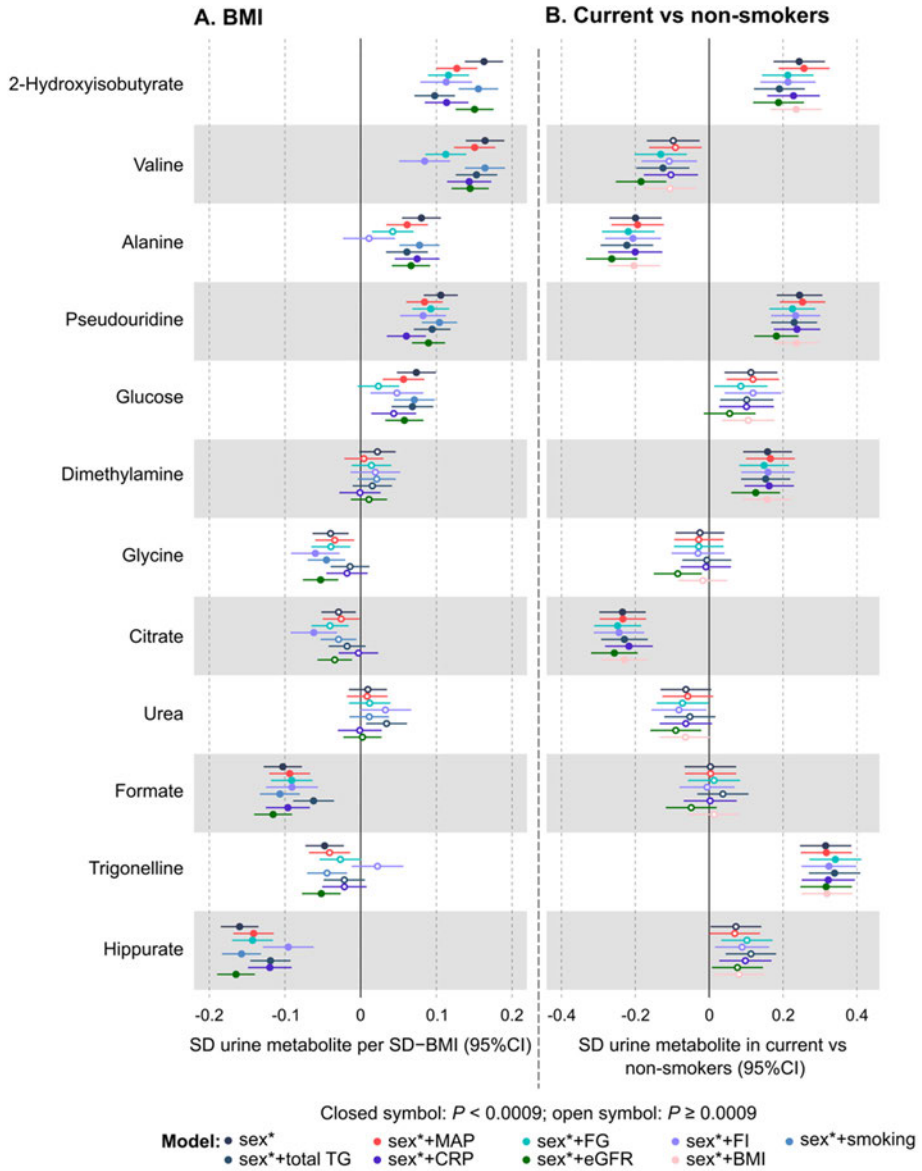
### **5.3.4 Metabolite associations with smoking history**

Figure 16B presents meta-analysis of association of 12 urinary metabolites (referenced to creatinine) with a smoking history (current smokers vs. non-smokers) for the three independent cohorts (n up to 5 989). The results for each cohort are shown in supplementary Figure 10 in III. The base model was adjusted for sex (and age in YFS), then seven different adjustments (MAP, fasting glucose, fasting insulin, total triglycerides, CRP, eGFR, and BMI) were separately added to evaluate their potential confounding role. Seven metabolites had robust associations with smoking history at the multiple testing corrected p-value threshold  $< 0.0009$ , namely 2-hydroxyisobutyrate, valine, alanine, pseudouridine, dimethylamine, citrate, and trigonelline. The different adjustments had very little, if any, impact on the associations of urinary metabolites with smoking history. These findings may be valuable as questionnaire-independent assessments of smoking status in large-scale epidemiological studies.

### **5.3.5 Unspecific renal excretion of amino acids**

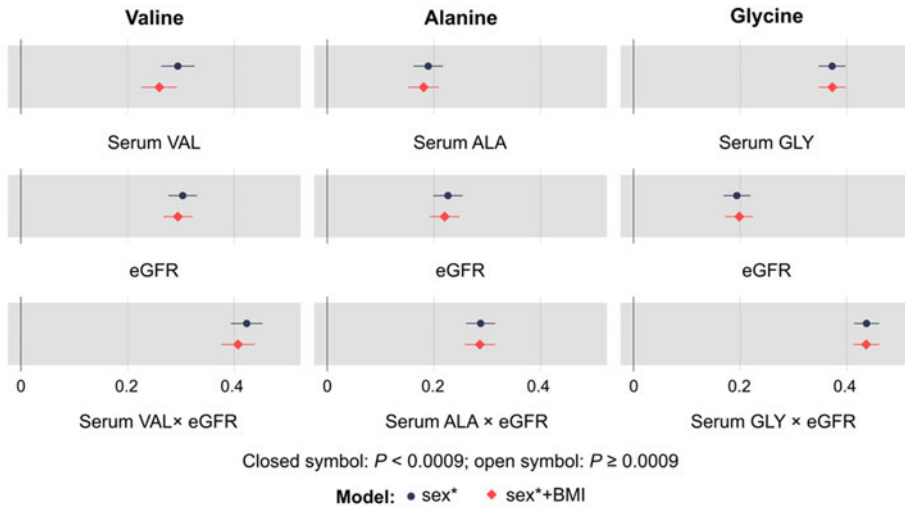
Three amino acids: valine, alanine, and glycine were included in the 12 quantified urinary metabolites. The positive associations are observed between these urinary amino acid and corresponding serum concentrations as well as eGFR, and the associations are strengthened for the multiplication of serum concentration and eGFR. The meta-analysis of the above associations for NFBC1966 and YFS available for serum amino acid data are shown in Fig. 17, while the results for individual cohort are presented in supplementary Figure 11 in III.

These associations might reflect the unspecific renal excretion of amino acids, similar to glucose (Li, Ihanus, et al., 2022). Amino acids, along with glucose, are indispensable for human metabolism, essentially all the amino acids filtered by the kidneys are also reabsorbed into circulation (or utilised in the kidneys) via a specific set of amino acid transporters (Mäkinen et al., 2022; Verrey et al., 2009). Similar to glucose, amino acids are present in the urine of all individuals. Despite the high efficiency of amino acid transporters, the large volume of filtered plasma would result in some nonspecific leakage of amino acids into the urine (Li et al., 2020; Verrey et al., 2009). Due to the paucity of quantitative metabolic studies of urine samples at the population level, we know very little about these kinds of nonspecific molecular processes in the kidney and their potential role as population-level biomarkers of renal function and/or the risk of disease (Li, Ihanus, et al., 2022).



**Fig. 16. Meta-analyses of the regression models for BMI (A) and smoking (B) with the 12 automatically quantified urinary metabolites (referenced to creatinine). The effects of sex (black), sex + MAP (red), sex + fasting glucose (cyan), sex + fasting insulin (lila), sex + smoking (light blue, applied to the BMI models only), sex + total triglycerides (blue), sex + CRP (violet), sex + eGFR (green), and sex + BMI (blush, applied to the**

smoking models only) were examined; \* indicates that for YFS also age was adjusted. The smoking data for the cohorts are: NFBC1966, 750 current and 3 544 non-smokers; NFBC1986, 115 current and 706 non-smokers; and YFS, 85 current and 370 non-smokers. Fig.16 is reprinted under CC BY License from Original publication III © 2023 Authors (Li et al., 2023).



**Fig. 17. Meta-analyses of the regression models for the three automatically quantified urinary amino acid (valine, alanine, and glycine) concentrations (referenced to creatinine) and their corresponding serum concentrations, eGFR, and the multiplication of the serum concentration and eGFR in NFBC1966 (n=4 505) and YFS (n=474). The effects of sex (black) and sex + BMI (red) were examined; \* indicates that for YFS also age was adjusted. Fig.17 is reprinted under CC BY License from Original publication III © 2023 Authors (Li et al., 2023).**

### 5.3.6 Summary of Study III

The present study presents novel quantitative data on 60 urinary metabolites (+creatinine) in a random subset of 994 individuals from the NFBC1966 cohort. This random subset represents the entire NFBC1966 cohort (n=4 505, supplementary Figures 5–8 in III), thus the quantitative data provide valuable reference concentrations for these urinary metabolites at the population level in both men and women (Fig. 13, Table 7 and supplementary Figures 2–4, in III). Furthermore, it also presents the first coherent automated quantification models for 12 urinary metabolites (+creatinine, supplementary Figure 1, III). The application

of these models made it possible to analyse metabolite concentrations in urine samples from nearly 6 000 individuals in three independent populations and to analyse their associations with 49 clinical and biochemical measures. The most of the reported associations are novel, as no epidemiological studies have yet combined a comprehensive quantitative urine metabolomics data with extensive additional clinical and biochemical data. Because the correlations between urinary metabolites and systemic metabolic markers are generally weak (Tynkkynen et al., 2019), urine samples might contain unique metabolic information not available in routine blood samples.

It is well known that deterioration of GFR is associated with ageing and cardiometabolic diseases (Sarnak et al., 2019; Vallon, 2020). The eGFR is a common clinical method to estimate individual renal function, which is usually estimated by age, sex, and serum creatinine (approximately not metabolised or reabsorbed by the kidneys) (Inker et al., 2012). However, the kidneys actively secrete a small amount of creatinine through the proximal tubules, and variations in systemic creatinine concentrations due to factors such as diet and muscle mass, can also affect the interpretation of eGFR-related results. Furthermore, metabolic processes in the kidney are multiple, with molecularly specific combinations of filtration and reabsorption determining the amount of a certain metabolite excreted into the urine. Thus, detailed quantitative data on specific urinary metabolites have the potential to provide direct molecular approaches to assess renal function and its metabolic inferences to multiple clinical and biochemical measures. In addition, in regression models of BMI and smoking history with urinary metabolites, adjusting for eGFR had little or no effect on relevant metabolites (Fig. 16 and supplementary Fig. 9 & 10, III), suggesting that the metabolite data are more valuable as they can bring in eGFR-unrelated information.

Urine metabolomics has the capacity to provide substantial scientific novelty. Taking 2-hydroxyisobutyrate as an example, the positive association of urinary 2-hydroxyisobutyrate with BMI appeared to be most extensively influenced by adjustments for MAP, fasting glucose, fasting insulin, total triglycerides, and CRP. Though it was also positively associated with smoking, no adjustment had a clear effect on this association. Generally, the correlations of 2-hydroxyisobutyrate were similar to those of valine (Figures 14 and 15), except that urinary valine had a negative association with smoking (Fig. 16B). 2-Hydroxyisobutyrate is a tertiary alcohol that may originate from gut microbial valine degradation (Elliott et al., 2015; Schifano et al., 2022; Wishart et al., 2018). Epidemiological data for this metabolite are almost completely lacking (Lotta et al., 2017), except that its urine

concentration was previously reported to be also strongly associated with BMI (Elliott et al., 2015). Furthermore, a recent study in patients with type 1 diabetes found a positive association between urinary 2-hydroxyisobutyrate and the progression of diabetic nephropathy (Mutter et al., 2022). All these findings indicate that urinary 2-hydroxyisobutyrate is related to insulin resistance (Calvani et al., 2010; Wang et al., 2017). A number of other associations have also been reported, such as that 2-hydroxyisobutyrate is part of a "specific obesity urinary metabolite" (Calvani et al., 2010) and is associated with various pregnancy problems (Diaz et al., 2011; Gil et al., 2018). However, all of these have applied OPLS-DA analysis, which is known to result in spurious results, especially when a large number of spectral data points are used as the basis for classification in very small datasets (Ala-Korpela, 2008, 2016; Ala-Korpela et al., 2012; Ala-Korpela & Davey Smith, 2016; Bevilacqua & Bro, 2020; Kjeldahl & Bro, 2010; Madsen et al., 2010).

The discussion above shows that there are few comprehensive quantitative data on urine metabolites in the current large-scale epidemiological studies, and then urine as a waste biofluid – closely related to kidney function – provides a metabolic view that is interdependent and complementary with systemic metabolism. Additionally, it is essential to avoid common limitations of multivariate metabolomics applications (often using OPLS-DA), which lead to false results induced by overfitting classification models with a large number of variables with very small numbers of participants (Ala-Korpela, 2008; Ala-Korpela et al., 2012; Ala-Korpela & Davey Smith, 2016; Bevilacqua & Bro, 2020; Smith et al., 2007; Soinen et al., 2015). In addition, quantitative metabolite data also provide an easy way to adjust confounding applied in this case and in the replication of results in this work, done in up to three independent cohorts. These are fundamental elements of triangulation (Lawlor et al., 2016) and scientific robustness (Ioannidis, 2014).

It is important to keep in mind with these observational epidemiological findings that cannot understand any mechanisms. However, this work is the first to demonstrate various association patterns of urinary metabolites with multiple descriptors of the renal-cardiometabolic system. Many associations of urinary metabolites with clinical indicators seem to be independent of fundamental systemic metabolic regulators, indicating that quantitative urine metabolomics may reflect specific insights about renal tubular filtration and reabsorption as well as kidney-specific molecular interactions. The consistent pattern of metabolic association of urine as a waste product in three independent cohorts is an intriguing result, demonstrating the high analytical robustness of the new quantitative method

and suggesting that urine metabolites may well have general value as population level health and disease biomarkers. For extending quantitative urinary metabolomics to large-scale epidemiological studies, the presented extensive data and results provide a reliable rationale, thereby gaining new insights into renal function and related metabolic disease biomarkers. As all the three populations studied in this work are from Finland, replication of these findings in other ethnicities and geographical locations would be valuable and also of high scientific interest due to the potential social and environmental impacts on urinary metabolite profiles and their clinical associations.

## 6 Conclusions

The aim of this thesis is to demonstrate the distribution characteristics of urinary metabolites at the population level and to find a rationale for appropriate normalization protocol for urinary metabolite concentrations for epidemiological application as well as explore clinical and biochemical associations with urinary metabolites at the population level.

Based on the quantitative high-throughput urine NMR metabolomics platform, 17 new metabolites in a random subset (n=994) of NFBC1966 were quantified by a semi-automated method using CTLS fitting analysis tool. The number of quantified urinary metabolites increases to 61 with relatively small methodological intra-assay metabolite CV%s (mostly less than 5%), illustrating the high robustness and accuracy of urine NMR spectroscopy. Additionally, although the intra-individual variation (mostly more than 20% over 30 days) is large, the even larger population-based inter-individual variation (mostly more than 40%) provide a good base for epidemiological and genetic applications. However, large variations of urine volume and metabolites concentrations of urine samples make it necessary to perform a normalization step during data analyses. Although quantitative data of urine metabolites provide sex-specific referenced concentrations and distributions for key urinary metabolites at the population level, women have slightly higher creatinine referenced concentrations of many metabolites than men, since women have a lower circulating creatinine level due to lower muscle mass and thus excrete lower amounts of creatinine in urine. In addition, 13 urine metabolites were quantified by the first coherent set of automated quantification models in three independent cohorts (n=5 989) in which the concentration and distribution data of the random subset of NFBC1966 match very well with corresponding data in the whole cohort, suggesting that this random subset is representative of the entire NFBC1966 cohort. Therefore, its quantitative data provide valuable reference concentrations for key urinary metabolites at the population level. Moreover, this automated quantification approach shows excellent quantitative performance and robustness, making it feasible for routine applications at large scale. As previously discussed, the development of both the semi-automated method and automated regression models is challenging, thus the population-level quantitative urine data included in the present study represent a considerable epidemiological novelty in this field of research.

Furthermore, this work compared various normalization strategies for quantitative urine metabolomics data as well as the absolute urinary metabolite

concentrations with customary normalization to urinary creatinine by analysing metabolite-metabolite associations and the association of metabolite concentrations with two representative clinical measures (BMI and MAP). The observed pronounced differences for urinary metabolite-metabolite associations suggest that comparing different normalization methods would help elucidate intra-urine metabolite associations. At the same time, overall consistent association patterns of the epidemiological regression analysis with clinical measurements illustrate the convention of using urinary creatinine as a reference, which seems judicious for epidemiological studies involving urinary metabolites. Moreover, normalization to urinary pseudouridine could serve as a valuable means to verify the consistency of creatinine reference results under different biological assumptions. If many individual quantified metabolites are available, multi-metabolite normalization methods may also be worth applying to validate the consistency of creatinine reference results. In conclusion, the comparison of different normalization protocols, considering physiology, metabolism, and renal function, facilitates the nuanced elucidation of associations of urinary metabolites in epidemiological studies.

In the present study, we noted that glucose in urine is a normal phenomenon (misleading claims that normal human urine does not contain glucose should be corrected in a number of renal physiology textbooks) (Eaton & Pooler, 2018) and the typical absolute urinary glucose concentrations are between 0.1 and 0.5 mmol/L. This phenomenon might reflect unspecific renal excretion of glucose also at low concentration ranges of circulating glucose (Li, Ihanus, et al., 2022). Amino acids, similar to glucose (despite the high efficiency of the amino acid transporters), and the large volume of plasma filtered would result in some unspecific leakage of amino acids into the urine (Li et al., 2020; Vallon, 2020). Moreover, the positive correlations between urine concentrations and serum concentrations as well as eGFR are both observed in glucose and amino acids, indicating that urinary metabolites are likely able to provide more information to directly reflect renal function.

Most importantly, the present study comprehensively explored the correlations between 12 urine metabolite concentrations and 49 conventional clinical measures, including kidney function indicators, various serum amino acids, lipids and lipoproteins, glycaemic traits and gluconeogenesis, inflammation, liver function markers, multiple obesity indicators, blood pressure, CVD and CKD risk scores, and these correlations appeared to be consistent among three independent cohorts. Moreover, the regression analyses of 12 urinary metabolites with BMI as well as

smoking history also showed similar association patterns among the three cohorts. The findings above suggest that urinary metabolites may have general value as population-level biomarkers of health and disease.

The work in this thesis presents a comprehensive quantitative urine metabolomics study at an epidemiological scale with independent replication. The novel extensive associations between quantitative urinary metabolites and clinical measures have the potential to obtain new insights into renal function and related disease biomarkers.



## List of references

- Ala-Korpela, M. (2007). Potential role of body fluid <sup>1</sup>H NMR metabonomics as a prognostic and diagnostic tool. *Expert Review of Molecular Diagnostics*, 7(6), 761–773. <https://doi.org/10.1586/14737159.7.6.761>
- Ala-Korpela, M. (2008). Critical evaluation of <sup>1</sup>H NMR metabonomics of serum as a methodology for disease risk assessment and diagnostics. *Clinical Chemistry and Laboratory Medicine*, 46(1), 27–42. <https://doi.org/10.1515/CCLM.2008.006>
- Ala-Korpela, M. (2016). Metabolomics in cardiovascular medicine: Not personalised, not diagnostic. *European Journal of Preventive Cardiology*, 23(17), 1821–1822. <https://doi.org/10.1177/2047487316664443>
- Ala-Korpela, M. (2018). Objective Metabolomics Research. *Clinical Chemistry*, 64(1), 30–33. <https://doi.org/10.1373/clinchem.2017.274852>
- Ala-Korpela, M., & Davey Smith, G. (2016). Metabolic profiling-multitude of technologies with great research potential, but (when) will translation emerge? *International Journal of Epidemiology*, 45(5), 1311–1318. <https://doi.org/10.1093/ije/dyw305>
- Ala-Korpela, M., & Holmes, M. V. (2020). Polygenic risk scores and the prediction of common diseases. *International Journal of Epidemiology*, 49(1), 1–3. <https://doi.org/10.1093/ije/dyz254>
- Ala-Korpela, M., Kangas, A. J., & Soininen, P. (2012). Quantitative high-throughput metabolomics: a new era in epidemiology and genetics. *Genome Medicine*, 4(4), 36. <https://doi.org/10.1186/gm335>
- Ala-Korpela, M., Lehtimäki, T., Kähönen, M., Viikari, J., Perola, M., Salomaa, V., Kettunen, J., Raitakari, O. T., & Mäkinen, V.-P. (2023). Cross-sectionally Calculated Metabolic Aging Does Not Relate to Longitudinal Metabolic Changes-Support for Stratified Aging Models. *The Journal of Clinical Endocrinology and Metabolism*, 108(8), 2099–2104. <https://doi.org/10.1210/clinem/dgad032>
- Ala-Korpela, M., Zhao, S., Järvelin, M.-R., Mäkinen, V.-P., & Ohukainen, P. (2022). Apt interpretation of comprehensive lipoprotein data in large-scale epidemiology: disclosure of fundamental structural and metabolic relationships. *International Journal of Epidemiology*, 51(3), 996–1011. <https://doi.org/10.1093/ije/dyab156>
- Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature*, 567(7748), 305–307. <https://doi.org/10.1038/d41586-019-00857-9>
- Antoniewicz, M. R., Stephanopoulos, G., & Kelleher, J. K. (2006). Evaluation of regression models in metabolic physiology: predicting fluxes from isotopic data without knowledge of the pathway. *Metabolomics : Official Journal of the Metabolomic Society*, 2(1), 41–52. <https://doi.org/10.1007/s11306-006-0018-2>
- Aygen, S., Dürr, U., Hegele, P., Kunig, J., Spraul, M., Schäfer, H., Krings, D., Cannet, C., Fang, F., Schütz, B., Bülbül, S. F. H., Aydin, H. I., Sarıcı, S. U., Yalaz, M., Ors, R., Atalan, R., & Tuncer, O. (2014). NMR-Based Screening for Inborn Errors of Metabolism: Initial Results from a Study on Turkish Neonates. *JIMD Reports*, 16, 101–111. [https://doi.org/10.1007/8904\\_2014\\_326](https://doi.org/10.1007/8904_2014_326)

- Babushok, V. I., Linstrom, P. J., Reed, J. J., Zenkevich, I. G., Brown, R. L., Mallard, W. G., & Stein, S. E. (2007). Development of a database of gas chromatographic retention properties of organic compounds. *Journal of Chromatography A*, 1157(1–2), 414–421. <https://doi.org/10.1016/j.chroma.2007.05.044>
- Baxmann, A. C., Ahmed, M. S., Marques, N. C., Menon, V. B., Pereira, A. B., Kirsztajn, G. M., & Heilberg, I. P. (2008). Influence of muscle mass and physical activity on serum and urinary creatinine and serum cystatin C. *Clinical Journal of the American Society of Nephrology : CJASN*, 3(2), 348–354. <https://doi.org/10.2215/CJN.02870707>
- Beckonert, O., Keun, H. C., Ebbels, T. M. D., Bundy, J., Holmes, E., Lindon, J. C., & Nicholson, J. K. (2007). Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nature Protocols*, 2(11), 2692–2703. <https://doi.org/10.1038/nprot.2007.376>
- Bermingham, K. M., Brennan, L., Segurado, R., Barron, R. E., Gibney, E. R., Ryan, M. F., Gibney, M. J., & O’Sullivan, A. M. (2021). Genetic and Environmental Contributions to Variation in the Stable Urinary NMR Metabolome over Time: A Classic Twin Study. *Journal of Proteome Research*, 20(8), 3992–4000. <https://doi.org/10.1021/acs.jproteome.1c00319>
- Bernert, J. T., Bell, C. J., Guntupalli, J., & Hannon, W. H. (1988). Pseudouridine is unsuitable as an endogenous renal clearance marker. *Clinical Chemistry*, 34(6), 1011–1017. <http://www.ncbi.nlm.nih.gov/pubmed/3378317>
- Bevilacqua, M., & Bro, R. (2020). Can We Trust Score Plots? *Metabolites*, 10(7). <https://doi.org/10.3390/metabo10070278>
- Bhopal, R., Fischbacher, C., Vartiainen, E., Unwin, N., White, M., & Alberti, G. (2005). Predicted and observed cardiovascular disease in South Asians: application of FINRISK, Framingham and SCORE models to Newcastle Heart Project data. *Journal of Public Health (Oxford, England)*, 27(1), 93–100. <https://doi.org/10.1093/pubmed/fdh202>
- Bogl, L. H., Pietiläinen, K. H., Rissanen, A., Kangas, A. J., Soininen, P., Rose, R. J., Ala-Korpela, M., & Kaprio, J. (2013). Association between habitual dietary intake and lipoprotein subclass profile in healthy young adults. *Nutrition, Metabolism, and Cardiovascular Diseases : NMCD*, 23(11), 1071–1078. <https://doi.org/10.1016/j.numecd.2012.11.007>
- Bouatra, S., Aziat, F., Mandal, R., Guo, A. C., Wilson, M. R., Knox, C., Bjorndahl, T. C., Krishnamurthy, R., Saleem, F., Liu, P., Dame, Z. T., Poelzer, J., Huynh, J., Yallou, F. S., Psychogios, N., Dong, E., Bogumil, R., Roehring, C., & Wishart, D. S. (2013). The Human Urine Metabolome. *PLoS ONE*, 8(9), e73076. <https://doi.org/10.1371/journal.pone.0073076>
- Cai, Q., Dekker, L. H., Bakker, S. J. L., de Borst, M. H., & Navis, G. J. (2020). Dietary Patterns Based on Estimated Glomerular Filtration Rate and Kidney Function Decline in the General Population: The Lifelines Cohort Study. *Nutrients*, 12(4). <https://doi.org/10.3390/nu12041099>

- Calvani, R., Miccheli, A., Capuani, G., Tomassini Miccheli, A., Puccetti, C., Delfini, M., Iaconelli, A., Nanni, G., & Mingrone, G. (2010). Gut microbiome-derived metabolites characterize a peculiar obese urinary metabotype. *International Journal of Obesity*, 34(6), 1095–1098. <https://doi.org/10.1038/ijo.2010.44>
- Chace, D. H. (2001). Mass spectrometry in the clinical laboratory. *Chemical Reviews*, 101(2), 445–477. <https://doi.org/10.1021/cr990077+>
- Chan, E. C. Y., Pasikanti, K. K., & Nicholson, J. K. (2011). Global urinary metabolic profiling procedures using gas chromatography–mass spectrometry. *Nature Protocols*, 6(10), 1483–1499. <https://doi.org/10.1038/nprot.2011.375>
- Chetwynd, A. J., Dunn, W. B., & Rodriguez-Blanco, G. (2017). Collection and Preparation of Clinical Samples for Metabolomics. *Advances in Experimental Medicine and Biology*, 965, 19–44. [https://doi.org/10.1007/978-3-319-47656-8\\_2](https://doi.org/10.1007/978-3-319-47656-8_2)
- 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. *The American Journal of Medicine*, 123(9), 836–846.e2. <https://doi.org/10.1016/j.amjmed.2010.05.010>
- Cook, T., Ma, Y., & Gamagedara, S. (2020). Evaluation of statistical techniques to normalize mass spectrometry-based urinary metabolomics data. *Journal of Pharmaceutical and Biomedical Analysis*, 177, 112854. <https://doi.org/10.1016/j.jpba.2019.112854>
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>
- Craig, A., Cloarec, O., Holmes, E., Nicholson, J. K., & Lindon, J. C. (2006). Scaling and normalization effects in NMR spectroscopic metabonomic data sets. *Analytical Chemistry*, 78(7), 2262–2267. <https://doi.org/10.1021/ac0519312>
- De Filippis, F., Pellegrini, N., Vannini, L., Jeffery, I. B., La Storia, A., Laghi, L., Serrazanetti, D. I., Di Cagno, R., Ferrocino, I., Lazzi, C., Turrone, S., Cocolin, L., Brigidi, P., Neviani, E., Gobetti, M., O'Toole, P. W., & Ercolini, D. (2016). High-level adherence to a Mediterranean diet beneficially impacts the gut microbiota and associated metabolome. *Gut*, 65(11), 1812–1821. <https://doi.org/10.1136/gutjnl-2015-309957>
- Deelen, J., Kettunen, J., Fischer, K., van der Spek, A., Trompet, S., Kastenmüller, G., Boyd, A., Zierer, J., van den Akker, E. B., Ala-Korpela, M., Amin, N., Demirkan, A., Ghanbari, M., van Heemst, D., Ikram, M. A., van Klinken, J. B., Mooijaart, S. P., Peters, A., Salomaa, V., ... Slagboom, P. E. (2019). A metabolic profile of all-cause mortality risk identified in an observational study of 44,168 individuals. *Nature Communications*, 10(1), 3346. <https://doi.org/10.1038/s41467-019-11311-9>
- Díaz, S. O., Pinto, J., Graça, G., Duarte, I. F., Barros, A. S., Galhano, E., Pita, C., Almeida, M. do C., Goodfellow, B. J., Carreira, I. M., & Gil, A. M. (2011). Metabolic biomarkers of prenatal disorders: an exploratory NMR metabonomics study of second trimester maternal urine and blood plasma. *Journal of Proteome Research*, 10(8), 3732–3742. <https://doi.org/10.1021/pr200352m>
- Dieterle, F., Ross, A., Schlotterbeck, G., & Senn, H. (2006). Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. *Analytical Chemistry*, 78(13), 4281–4290. <https://doi.org/10.1021/ac051632c>

- Dietzen, D. J., Rinaldo, P., Whitley, R. J., Rhead, W. J., Hannon, W. H., Garg, U. C., Lo, S. F., & Bennett, M. J. (2009). National academy of clinical biochemistry laboratory medicine practice guidelines: follow-up testing for metabolic disease identified by expanded newborn screening using tandem mass spectrometry; executive summary. *Clinical Chemistry*, *55*(9), 1615–1626. <https://doi.org/10.1373/clinchem.2009.131300>
- Dubin, R. F., & Rhee, E. P. (2020). Proteomics and Metabolomics in Kidney Disease, including Insights into Etiology, Treatment, and Prevention. *Clinical Journal of the American Society of Nephrology*, *15*(3), 404–411. <https://doi.org/10.2215/CJN.07420619>
- Dunn, W. B., Broadhurst, D., Begley, P., Zelena, E., Francis-McIntyre, S., Anderson, N., Brown, M., Knowles, J. D., Halsall, A., Haselden, J. N., Nicholls, A. W., Wilson, I. D., Kell, D. B., & Goodacre, R. (2011). Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nature Protocols*, *6*(7), 1060–1083. <https://doi.org/10.1038/nprot.2011.335>
- Dunn, W. B., Broadhurst, D. I., Atherton, H. J., Goodacre, R., & Griffin, J. L. (2011). Systems level studies of mammalian metabolomes: The roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chemical Society Reviews*, *40*(1), 387–426. <https://doi.org/10.1039/b906712b>
- Eaton, D. C. and Pooler, J. P. (2018). *Vander's Renal Physiology* (9th edn.). Hill, McGraw.
- Eklholm, J., Ohukainen, P., Kangas, A. J., Kettunen, J., Wang, Q., Karsikas, M., Khan, A. A., Kingwell, B. A., Kähönen, M., Lehtimäki, T., Raitakari, O. T., Järvelin, M.-R., Meikle, P. J., & Ala-Korpela, M. (2020). EpiMetal: an open-source graphical web browser tool for easy statistical analyses in epidemiology and metabolomics. *International Journal of Epidemiology*, *49*(4), 1075–1081. <https://doi.org/10.1093/ije/dyz244>
- Elbaz, A. M., Gani, A., Hourani, N., Emwas, A.-H., Sarathy, S. M., & Roberts, W. L. (2015). TG/DTG, FT-ICR Mass Spectrometry, and NMR Spectroscopy Study of Heavy Fuel Oil. *Energy & Fuels*, *29*(12), 7825–7835. <https://doi.org/10.1021/acs.energyfuels.5b01739>
- Elliott, P., Posma, J. M., Chan, Q., Garcia-Perez, I., Wijeyesekera, A., Bictash, M., Ebbels, T. M. D., Ueshima, H., Zhao, L., Van Horn, L., Daviglius, M., Stamler, J., Holmes, E., & Nicholson, J. K. (2015). Urinary metabolic signatures of human adiposity. *Science Translational Medicine*, *7*(285). <https://doi.org/10.1126/scitranslmed.aaa5680>
- Emwas, A.-H. M. (2015). The Strengths and Weaknesses of NMR Spectroscopy and Mass Spectrometry with Particular Focus on Metabolomics Research. In *Methods in Molecular Biology* (Vol. 1277, pp. 161–193). [https://doi.org/10.1007/978-1-4939-2377-9\\_13](https://doi.org/10.1007/978-1-4939-2377-9_13)
- Emwas, A.-H. M., Salek, R. M., Griffin, J. L., & Merzaban, J. (2013). NMR-based metabolomics in human disease diagnosis: applications, limitations, and recommendations. *Metabolomics*, *9*(5), 1048–1072. <https://doi.org/10.1007/s11306-013-0524-y>

- Emwas, A.-H., Roy, R., McKay, R. T., Tenori, L., Saccenti, E., Gowda, G. A. N., Raftery, D., Alahmari, F., Jaremko, L., Jaremko, M., & Wishart, D. S. (2019). NMR Spectroscopy for Metabolomics Research. *Metabolites*, *9*(7), 123. <https://doi.org/10.3390/metabo9070123>
- Emwas, A. H., Roy, R., McKay, R. T., Ryan, D., Brennan, L., Tenori, L., Luchinat, C., Gao, X., Zeri, A. C., Gowda, G. A. N., Raftery, D., Steinbeck, C., Salek, R. M., & Wishart, D. S. (2016). Recommendations and Standardization of Biomarker Quantification Using NMR-Based Metabolomics with Particular Focus on Urinary Analysis. *Journal of Proteome Research*, *15*(2), 360–373. <https://doi.org/10.1021/acs.jproteome.5b00885>
- Entenmann, L., Pietzner, M., Artati, A., Hannemann, A., Henning, A.-K., Kastenmüller, G., Völzke, H., Nauck, M., Adamski, J., Wallaschofski, H., & Friedrich, N. (2017). Comprehensive metabolic characterization of serum osteocalcin action in a large non-diabetic sample. *PloS One*, *12*(9), e0184721. <https://doi.org/10.1371/journal.pone.0184721>
- Fearnley, L. G., & Inouye, M. (2016). Metabolomics in epidemiology: from metabolite concentrations to integrative reaction networks. *International Journal of Epidemiology*, *45*(5), 1319–1328. <https://doi.org/10.1093/ije/dyw046>
- FINE, J. (1965). GLUCOSE CONTENT OF NORMAL URINE. *British Medical Journal*, *1*(5444), 1209–1214. <https://doi.org/10.1136/bmj.1.5444.1209>
- Forni Ognà, V., Ognà, A., Vuistiner, P., Pruijm, M., Ponte, B., Ackermann, D., Gabutti, L., Vakilzadeh, N., Mohaupt, M., Martin, P.-Y., Guessous, I., Péchère-Bertschi, A., Paccaud, F., Bochud, M., Burnier, M., & Swiss Survey on Salt Group. (2015). New anthropometry-based age- and sex-specific reference values for urinary 24-hour creatinine excretion based on the adult Swiss population. *BMC Medicine*, *13*, 40. <https://doi.org/10.1186/s12916-015-0275-x>
- Frérot, M., Lefebvre, A., Aho, S., Callier, P., Astruc, K., & Aho Glélé, L. S. (2018). What is epidemiology? Changing definitions of epidemiology 1978-2017. *PLOS ONE*, *13*(12), e0208442. <https://doi.org/10.1371/journal.pone.0208442>
- Friedrich, N., Budde, K., Suhre, K., Völker, U., John, U., Felix, S. B., Kroemer, H. K., Grabe, H. J., Völzke, H., Nauck, M., & Wallaschofski, H. (2015). Sex differences in urine metabolites related with risk of diabetes using NMR spectroscopy: results of the study of health in pomerania. *Metabolomics*, *11*(5), 1405–1415. <https://doi.org/10.1007/s11306-015-0795-6>
- Friedrich, N., Pietzner, M., Cannet, C., Thuesen, B. H., Hansen, T., Wallaschofski, H., Grarup, N., Skaaby, T., Budde, K., Pedersen, O., Nauck, M., & Linneberg, A. (2017). Urinary metabolomics reveals glycemic and coffee associated signatures of thyroid function in two population-based cohorts. *PloS One*, *12*(3), e0173078. <https://doi.org/10.1371/journal.pone.0173078>
- Friedrich, N., Skaaby, T., Pietzner, M., Budde, K., Thuesen, B. H., Nauck, M., & Linneberg, A. (2018). Identification of urine metabolites associated with 5-year changes in biomarkers of glucose homeostasis. *Diabetes & Metabolism*, *44*(3), 261–268. <https://doi.org/10.1016/j.diabet.2017.05.007>

- Galloway, J. W., Keijsers, B. J. F., & Williams, D. M. (2016). Saliva in studies of epidemiology of human disease: the UK Biobank project. *Periodontology* 2000, 70(1), 184–195. <https://doi.org/10.1111/prd.12108>
- Gao, J., Zhong, X., Ding, Y., Bai, T., Wang, H., Wu, H., Liu, Y., Yang, J., & Zhang, Y. (2016). Inhibition of voltage-gated potassium channels mediates uncarboxylated osteocalcin-regulated insulin secretion in rat pancreatic  $\beta$  cells. *European Journal of Pharmacology*, 777, 41–48. <https://doi.org/10.1016/j.ejphar.2016.02.060>
- Gao, S., Mutter, S., Casey, A., & Mäkinen, V.-P. (2019). Numero: a statistical framework to define multivariable subgroups in complex population-based datasets. *International Journal of Epidemiology*, 48(2), 369–374. <https://doi.org/10.1093/ije/dyy113>
- Gerstman, B. B. (2013). Epidemiology Past and Present. In A John Wiley & Sons (Ed.), *Epidemiology Kept Simple: An Introduction to Classic and Modern Epidemiology* (3rd edn., p. 2). Wiley.
- Gibson, R., Lau, C. H. E., Loo, R. L., Ebbels, T. M. D., Chekmeneva, E., Dyer, A. R., Miura, K., Ueshima, H., Zhao, L., Daviglius, M. L., Stamler, J., Van Horn, L., Elliott, P., Holmes, E., & Chan, Q. (2020). The association of fish consumption and its urinary metabolites with cardiovascular risk factors: The International Study of Macro-/Micronutrients and Blood Pressure (INTERMAP). *American Journal of Clinical Nutrition*, 111(2), 280–290. <https://doi.org/10.1093/ajcn/nqz293>
- Gil, A. M., Duarte, D., Pinto, J., & Barros, A. S. (2018). Assessing Exposome Effects on Pregnancy through Urine Metabolomics of a Portuguese (Estarreja) Cohort. *Journal of Proteome Research*, 17(3), 1278–1289. <https://doi.org/10.1021/acs.jproteome.7b00878>
- Gordis, L. (2014). *Epidemiology* (5th edn.). Elsevier.
- Gowda, S., Desai, P. B., Kulkarni, S. S., Hull, V. V., Math, A. A. K., & Vernekar, S. N. (2010). Markers of renal function tests. *North American Journal of Medical Sciences*, 2(4), 170–173. <http://www.ncbi.nlm.nih.gov/pubmed/22624135>
- Hall, J. E., Hall, M. E. (2020). *Guyton and Hall Textbook of Medical Physiology* (14th ed.). Elsevier Health Sciences.
- Hampe, M. H., Panaskar, S. N., Yadav, A. A., & Ingale, P. W. (2017). Gas chromatography/mass spectrometry-based urine metabolome study in children for inborn errors of metabolism: An Indian experience. *Clinical Biochemistry*, 50(3), 121–126. <https://doi.org/10.1016/j.clinbiochem.2016.10.015>
- Haq, K., & Patel, D. M. (2023). Urinalysis: Interpretation and Clinical Correlations. *The Medical Clinics of North America*, 107(4), 659–679. <https://doi.org/10.1016/j.mcna.2023.03.002>
- Herrera, J., & Rodríguez-Iturbe, B. (1998). Stimulation of tubular secretion of creatinine in health and in conditions associated with reduced nephron mass. Evidence for a tubular functional reserve. *Nephrology, Dialysis, Transplantation: Official Publication of the European Dialysis and Transplant Association - European Renal Association*, 13(3), 623–629. <https://doi.org/10.1093/ndt/13.3.623>

- Hertel, J., Friedrich, N., Wittfeld, K., Pietzner, M., Budde, K., Van der Auwera, S., Lohmann, T., Teumer, A., Völzke, H., Nauck, M., & Grabe, H. J. (2016). Measuring Biological Age via Metabonomics: The Metabolic Age Score. *Journal of Proteome Research*, *15*(2), 400–410. <https://doi.org/10.1021/acs.jproteome.5b00561>
- Holmes, E., Loo, R. L., Stamler, J., Bictash, M., Yap, I. K. S., 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. (2008). Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature*, *453*(7193), 396–400. <https://doi.org/10.1038/nature06882>
- Holmes, E., Tsang, T. M., Huang, J. T.-J., Leweke, F. M., Koethe, D., Gerth, C. W., Nolden, B. M., Gross, S., Schreiber, D., Nicholson, J. K., & Bahn, S. (2006). Metabolic Profiling of CSF: Evidence That Early Intervention May Impact on Disease Progression and Outcome in Schizophrenia. *PLoS Medicine*, *3*(8), e327. <https://doi.org/10.1371/journal.pmed.0030327>
- Holmes, M. V., & Davey Smith, G. (2019). Can Mendelian Randomization Shift into Reverse Gear? *Clinical Chemistry*, *65*(3), 363–366. <https://doi.org/10.1373/clinchem.2018.296806>
- Hu, F. B. (2002). Dietary pattern analysis: a new direction in nutritional epidemiology. *Current Opinion in Lipidology*, *13*(1), 3–9. <https://doi.org/10.1097/00041433-200202000-00002>
- Huynh, K., Barlow, C. K., Jayawardana, K. S., Weir, J. M., Mellett, N. A., Cinel, M., Magliano, D. J., Shaw, J. E., Drew, B. G., & Meikle, P. J. (2019). High-Throughput Plasma Lipidomics: Detailed Mapping of the Associations with Cardiometabolic Risk Factors. *Cell Chemical Biology*, *26*(1), 71-84.e4. <https://doi.org/10.1016/j.chembiol.2018.10.008>
- Inker, L. A., Schmid, C. H., Tighiouart, H., Eckfeldt, J. H., Feldman, H. I., Greene, T., Kusek, J. W., Manzi, J., Van Lente, F., Zhang, Y. L., Coresh, J., Levey, A. S., & CKD-EPI Investigators. (2012). Estimating glomerular filtration rate from serum creatinine and cystatin C. *The New England Journal of Medicine*, *367*(1), 20–29. <https://doi.org/10.1056/NEJMoa1114248>
- Ioannidis, J. P. A. (2014). How to make more published research true. *PLoS Medicine*, *11*(10), e1001747. <https://doi.org/10.1371/journal.pmed.1001747>
- 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 International*, *96*(5), 1048–1050. <https://doi.org/10.1016/j.kint.2019.07.012>
- Johnson, C. H., Ivanisevic, J., & Siuzdak, G. (2016). Metabolomics: beyond biomarkers and towards mechanisms. *Nature Reviews. Molecular Cell Biology*, *17*(7), 451–459. <https://doi.org/10.1038/nrm.2016.25>
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, *374*(2065), 20150202. <https://doi.org/10.1098/rsta.2015.0202>

- Jørgensen, T., Borch-Johnsen, K., Thomsen, T. F., Ibsen, H., Glümer, C., & Pisinger, C. (2003). A randomized non-pharmacological intervention study for prevention of ischaemic heart disease: baseline results Inter99. *European Journal of Cardiovascular Prevention and Rehabilitation : Official Journal of the European Society of Cardiology, Working Groups on Epidemiology & Prevention and Cardiac Rehabilitation and Exercise Physiology*, 10(5), 377–386. <https://doi.org/10.1097/01.hjr.0000096541.30533.82>
- Julkunen, H., Cichońska, A., Slagboom, P. E., Würtz, P., & Nightingale Health UK Biobank Initiative. (2021). Metabolic biomarker profiling for identification of susceptibility to severe pneumonia and COVID-19 in the general population. *ELife*, 10. <https://doi.org/10.7554/eLife.63033>
- Julkunen, H., Cichońska, A., Tiainen, M., Koskela, H., Nybo, K., Mäkelä, V., Nokso-Koivisto, J., Kristiansson, K., Perola, M., Salomaa, V., Jousilahti, P., Lundqvist, A., Kangas, A. J., Soininen, P., Barrett, J. C., & Würtz, P. (2023). Atlas of plasma NMR biomarkers for health and disease in 118,461 individuals from the UK Biobank. *Nature Communications*, 14(1), 604. <https://doi.org/10.1038/s41467-023-36231-7>
- Kaplan, O., van Zijl, P. C. M., & Cohen, J. S. (1990). Information from combined <sup>1</sup>H and <sup>31</sup>P NMR studies of cell extracts: Differences in metabolism between drug-sensitive and drug-resistant MCF-7 human breast cancer cells. *Biochemical and Biophysical Research Communications*, 169(2), 383–390. [https://doi.org/10.1016/0006-291X\(90\)90343-L](https://doi.org/10.1016/0006-291X(90)90343-L)
- Khamis, M. M., Adamko, D. J., & El-Aneed, A. (2017). Mass spectrometric based approaches in urine metabolomics and biomarker discovery. *Mass Spectrometry Reviews*, 36(2), 115–134. <https://doi.org/10.1002/mas.21455>
- Kimura, M., Yamamoto, T., & Yamaguchi, S. (1999). Automated metabolic profiling and interpretation of GC/MS data for organic acidemia screening: a personal computer-based system. *The Tohoku Journal of Experimental Medicine*, 188(4), 317–334. <https://doi.org/10.1620/tjem.188.317>
- Kind, T., Wohlgemuth, G., Lee, D. Y., Lu, Y., Palazoglu, M., Shahbaz, S., & Fiehn, O. (2009). FiehnLib : Mass Spectral and Retention Index Libraries for Metabolomics Based on Quadrupole and Time-of-Flight Gas Chromatography / Mass Spectrometry. *Analytical Chemistry*, 81(24), 10038–10048. <https://doi.org/10.1021/ac9019522>
- Kirwan, J. A., Brennan, L., Broadhurst, D., Fiehn, O., Cascante, M., Dunn, W. B., Schmidt, M. A., & Velagapudi, V. (2018). Preanalytical Processing and Biobanking Procedures of Biological Samples for Metabolomics Research: A White Paper, Community Perspective (for “Precision Medicine and Pharmacometabolomics Task Group”-The Metabolomics Society Initiative). *Clinical Chemistry*, 64(8), 1158–1182. <https://doi.org/10.1373/clinchem.2018.287045>
- Kjeldahl, K., & Bro, R. (2010). Some common misunderstandings in chemometrics. *Journal of Chemometrics*, 24(7–8), 558–564. <https://doi.org/10.1002/cem.1346>
- Kohler, I., Verhoeven, A., Derks, R. J., & Giera, M. (2016). Analytical pitfalls and challenges in clinical metabolomics. *Bioanalysis*, 8(14), 1509–1532. <https://doi.org/10.4155/bio-2016-0090>

- Kovesdy, C. P. (2022). Epidemiology of chronic kidney disease: an update 2022. *Kidney International Supplements*, 12(1), 7–11. <https://doi.org/10.1016/j.kisu.2021.11.003>
- Kurbatova, N., Garg, M., Whiley, L., Chekmeneva, E., Jiménez, B., Gómez-Romero, M., Pearce, J., Kimhofer, T., D'Hondt, E., Soininen, H., Kłoszewska, I., Mecocci, P., Tsolaki, M., Vellas, B., Aarsland, D., Nevado-Holgado, A., Liu, B., Snowden, S., Proitsi, P., ... Lovestone, S. (2020). Urinary metabolic phenotyping for Alzheimer's disease. *Scientific Reports*, 10(1), 21745. <https://doi.org/10.1038/s41598-020-78031-9>
- 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., & CRIC Study Investigators. (2020). Metabolomic Markers of Kidney Function Decline in Patients With Diabetes: Evidence From the Chronic Renal Insufficiency Cohort (CRIC) Study. *American Journal of Kidney Diseases : The Official Journal of the National Kidney Foundation*, 76(4), 511–520. <https://doi.org/10.1053/j.ajkd.2020.01.019>
- Lains, I., Gantner, M., Murinello, S., Lasky-Su, J. A., Miller, J. W., Friedlander, M., & Husain, D. (2019). Metabolomics in the study of retinal health and disease. *Progress in Retinal and Eye Research*, 69(March 2018), 57–79. <https://doi.org/10.1016/j.preteyeres.2018.11.002>
- Lange, T., Budde, K., Homuth, G., Kastenmüller, G., Artati, A., Krumsiek, J., Völzke, H., Adamski, J., Petersmann, A., Völker, U., Nauck, M., Friedrich, N., & Pietzner, M. (2018). Comprehensive Metabolic Profiling Reveals a Lipid-Rich Fingerprint of Free Thyroxine Far Beyond Classic Parameters. *The Journal of Clinical Endocrinology and Metabolism*, 103(5), 2050–2060. <https://doi.org/10.1210/jc.2018-00183>
- Lasky-Su, J., Kelly, R. S., Wheelock, C. E., & Broadhurst, D. (2021). A strategy for advancing for population-based scientific discovery using the metabolome: the establishment of the Metabolomics Society Metabolomic Epidemiology Task Group. *Metabolomics*, 17(5), 45. <https://doi.org/10.1007/s11306-021-01789-0>
- Lau, C. H. E., Siskos, A. P., Maitre, L., Robinson, O., Athersuch, T. J., Want, E. J., Urquiza, J., Casas, M., Vafeiadi, M., Roumeliotaki, T., McEachan, R. R. C., Azad, R., Haug, L. S., Meltzer, H. M., Andrusaityte, S., Petravičienė, I., Grazulevičienė, R., Thomsen, C., Wright, J., ... Coen, M. (2018). Determinants of the urinary and serum metabolome in children from six European populations. *BMC Medicine*, 16(1), 1–19. <https://doi.org/10.1186/s12916-018-1190-8>
- Lawlor, D. A., Tilling, K., & Davey Smith, G. (2016). Triangulation in aetiological epidemiology. *International Journal of Epidemiology*, 45(6), 1866–1886. <https://doi.org/10.1093/ije/dyw314>
- Levey, A. S., & Coresh, J. (2012). Chronic kidney disease. *The Lancet*, 379(9811), 165–180. [https://doi.org/10.1016/S0140-6736\(11\)60178-5](https://doi.org/10.1016/S0140-6736(11)60178-5)
- Lewis, G. D., Asnani, A., & Gerszten, R. E. (2008). Application of metabolomics to cardiovascular biomarker and pathway discovery. *Journal of the American College of Cardiology*, 52(2), 117–123. <https://doi.org/10.1016/j.jacc.2008.03.043>

- Li, J., Guasch-Ferré, M., Chung, W., Ruiz-Canela, M., Toledo, E., Corella, D., Bhupathiraju, S. N., Tobias, D. K., Tabung, F. K., Hu, J., Zhao, T., Turman, C., Feng, Y.-C. A., Clish, C. B., Mucci, L., Eliassen, A. H., Costenbader, K. H., Karlson, E. W., Wolpin, B. M., ... Liang, L. (2020). The Mediterranean diet, plasma metabolome, and cardiovascular disease risk. *European Heart Journal*, *41*(28), 2645–2656. <https://doi.org/10.1093/eurheartj/ehaa209>
- Li, T., Ihanus, A., Ohukainen, P., Järvelin, M.-R., Kähönen, M., Kettunen, J., Raitakari, O. T., Lehtimäki, T., Mäkinen, V.-P., Tynkkynen, T., & Ala-Korpela, M. (2023). Clinical and biochemical associations of urinary metabolites: quantitative epidemiological approach on renal-cardiometabolic biomarkers. *International Journal of Epidemiology*. <https://doi.org/10.1093/ije/dyad162>
- Li, T., Ihanus, A., Ohukainen, P., Järvelin, M.-R., Kettunen, J., Mäkinen, V.-P., Tynkkynen, T., & Ala-Korpela, M. (2022). There is always glucose in normal urine: unspecific excretion associated with serum glucose and glomerular filtration rate. *International Journal of Epidemiology*, *51*(6), 2022–2025. <https://doi.org/10.1093/ije/dyad060>
- Li, T., Tynkkynen, T., Ihanus, A., Zhao, S., Mäkinen, V.-P., & Ala-Korpela, M. (2022). Characteristics of Normalization Methods in Quantitative Urinary Metabolomics-Implications for Epidemiological Applications and Interpretations. *Biomolecules*, *12*(7). <https://doi.org/10.3390/biom12070903>
- Li, X., Zheng, S., & Wu, G. (2020). Amino Acid Metabolism in the Kidneys: Nutritional and Physiological Significance. *Advances in Experimental Medicine and Biology*, *1265*, 71–95. [https://doi.org/10.1007/978-3-030-45328-2\\_5](https://doi.org/10.1007/978-3-030-45328-2_5)
- Lind, M. V., Savolainen, O. I., & Ross, A. B. (2016). The use of mass spectrometry for analysing metabolite biomarkers in epidemiology: methodological and statistical considerations for application to large numbers of biological samples. *European Journal of Epidemiology*, *31*(8), 717–733. <https://doi.org/10.1007/s10654-016-0166-2>
- Lindon, J. C., & Nicholson, J. K. (2008). Analytical technologies for metabonomics and metabolomics, and multi-omic information recovery. *TrAC Trends in Analytical Chemistry*, *27*(3), 194–204. <https://doi.org/10.1016/j.trac.2007.08.009>
- Liu, X., Yin, P., Shao, Y., Wang, Z., Wang, B., Lehmann, R., & Xu, G. (2020). Which is the urine sample material of choice for metabolomics-driven biomarker studies? *Analytica Chimica Acta*, *1105*, 120–127. <https://doi.org/10.1016/j.aca.2020.01.028>
- Locke, A. E., Steinberg, K. M., Chiang, C. W. K., Service, S. K., Havulinna, A. S., Stell, L., Pirinen, M., Abel, H. J., Chiang, C. C., Fulton, R. S., Jackson, A. U., Kang, C. J., Kanachi, K. L., Koboldt, D. C., Larson, D. E., Nelson, J., Nicholas, T. J., Pietilä, A., Ramensky, V., ... Freimer, N. B. (2019). Exome sequencing of Finnish isolates enhances rare-variant association power. *Nature*, *572*(7769), 323–328. <https://doi.org/10.1038/s41586-019-1457-z>

- Lotta, L. A., Gulati, P., Day, F. R., Payne, F., Ongen, H., van de Bunt, M., Gaulton, K. J., Eicher, J. D., Sharp, S. J., Luan, J., De Lucia Rolfe, E., Stewart, I. D., Wheeler, E., Willems, S. M., Adams, C., Yaghootkar, H., EPIC-InterAct Consortium, Cambridge FPLD1 Consortium, Forouhi, N. G., ... Scott, R. A. (2017). Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance. *Nature Genetics*, *49*(1), 17–26. <https://doi.org/10.1038/ng.3714>
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. <https://doi.org/10.1186/s13059-014-0550-8>
- Lund, G. S., & Wolf, C. G. (1925). The Glucose Content of Normal Urine. *The Biochemical Journal*, *19*(4), 538–540. <https://doi.org/10.1042/bj0190538>
- Madsen, R., Lundstedt, T., & Trygg, J. (2010). Chemometrics in metabolomics--a review in human disease diagnosis. *Analytica Chimica Acta*, *659*(1–2), 23–33. <https://doi.org/10.1016/j.aca.2009.11.042>
- Maitre, L., de Bont, J., Casas, M., Robinson, O., Aasvang, G. M., Agier, L., Andrušaitytė, S., Ballester, F., Basagaña, X., Borràs, E., Brochot, C., Bustamante, M., Carracedo, A., de Castro, M., Dedele, A., Donaire-Gonzalez, D., Estivill, X., Evandt, J., Fossati, S., ... Vrijheid, M. (2018). Human Early Life Exposome (HELIX) study: a European population-based exposome cohort. *BMJ Open*, *8*(9), e021311. <https://doi.org/10.1136/bmjopen-2017-021311>
- Mäkinen, V.-P., Karsikas, M., Kettunen, J., Lehtimäki, T., Kähönen, M., Viikari, J., Perola, M., Salomaa, V., Järvelin, M.-R., Raitakari, O. T., & Ala-Korpela, M. (2022). Longitudinal profiling of metabolic ageing trends in two population cohorts of young adults. *International Journal of Epidemiology*, *51*(6), 1970–1983. <https://doi.org/10.1093/ije/dyac062>
- Mäkinen, V.-P., Kettunen, J., Lehtimäki, T., Kähönen, M., Viikari, J., Perola, M., Salomaa, V., Järvelin, M.-R., Raitakari, O. T., & Ala-Korpela, M. (2023). Longitudinal metabolomics of increasing body-mass index and waist-hip ratio reveals two dynamic patterns of obesity pandemic. *International Journal of Obesity (2005)*, *47*(6), 453–462. <https://doi.org/10.1038/s41366-023-01281-w>
- Manjunath, G., Tighiouart, H., Ibrahim, H., MacLeod, B., Salem, D. N., Griffith, J. L., Coresh, J., Levey, A. S., & Sarnak, M. J. (2003). Level of kidney function as a risk factor for atherosclerotic cardiovascular outcomes in the community. *Journal of the American College of Cardiology*, *41*(1), 47–55. [https://doi.org/10.1016/s0735-1097\(02\)02663-3](https://doi.org/10.1016/s0735-1097(02)02663-3)
- Marshall, D. D., & Powers, R. (2017). Beyond the paradigm: Combining mass spectrometry and nuclear magnetic resonance for metabolomics. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *100*, 1–16. <https://doi.org/10.1016/j.pnmrs.2017.01.001>

- Mathé, E. A., Patterson, A. D., Haznadar, M., Manna, S. K., Krausz, K. W., Bowman, E. D., Shields, P. G., Idle, J. R., Smith, P. B., Anami, K., Kazandjian, D. G., Hatzakis, E., Gonzalez, F. J., & Harris, C. C. (2014). Noninvasive urinary metabolomic profiling identifies diagnostic and prognostic markers in lung cancer. *Cancer Research*, *74*(12), 3259–3270. <https://doi.org/10.1158/0008-5472.CAN-14-0109>
- McCarthy, M., & Birney, E. (2021). Personalized profiles for disease risk must capture all facets of health. *Nature*, *597*(7875), 175–177. <https://doi.org/10.1038/d41586-021-02401-0>
- 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. (2017). Urinary metabolites along with common and rare genetic variations are associated with incident chronic kidney disease. *Kidney International*, *91*(6), 1426–1435. <https://doi.org/10.1016/j.kint.2017.01.007>
- Miccheli, A., Capuani, G., Marini, F., Tomassini, A., Praticò, G., Ceccarelli, S., Gnani, D., Baviera, G., Alisi, A., Putignani, L., & Nobili, V. (2015). Urinary (1)H-NMR-based metabolic profiling of children with NAFLD undergoing VSL#3 treatment. *International Journal of Obesity (2005)*, *39*(7), 1118–1125. <https://doi.org/10.1038/ijo.2015.40>
- Mierisová, S., & Ala-Korpela, M. (2001). MR spectroscopy quantitation: a review of frequency domain methods. *NMR in Biomedicine*, *14*(4), 247–259. <https://doi.org/10.1002/nbm.697>
- Mochida, T., Tanaka, T., Shiraki, Y., Tajiri, H., Matsumoto, S., Shimbo, K., Ando, T., Nakamura, K., Okamoto, M., & Endo, F. (2011). Time-dependent changes in the plasma amino acid concentration in diabetes mellitus. *Molecular Genetics and Metabolism*, *103*(4), 406–409. <https://doi.org/10.1016/j.ymgme.2011.05.002>
- Monteiro, C. A., Cannon, G., Levy, R. B., Moubarac, J.-C., Louzada, M. L., Rauber, F., Khandpur, N., Cediel, G., Neri, D., Martinez-Steele, E., Baraldi, L. G., & Jaime, P. C. (2019). Ultra-processed foods: what they are and how to identify them. *Public Health Nutrition*, *22*(5), 936–941. <https://doi.org/10.1017/S1368980018003762>
- Mutter, S., Valo, E., Aittomäki, V., Nybo, K., Raivonen, L., Thorn, L. M., Forsblom, C., Sandholm, N., Würtz, 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. <https://doi.org/10.1007/s00125-021-05584-3>
- Naghavi, M., Wang, H., Lozano, R., Davis, A., Liang, X., Zhou, M., Vollset, S. E., Abbasoglu Ozgoren, A., Abdalla, S., Abd-Allah, F., Abdel Aziz, M. I., Abera, S. F., Aboyans, V., Abraham, B., Abraham, J. P., Abuabara, K. E., Abubakar, I., Abu-Raddad, L. J., Abu-Rmeileh, N. M. E., ... Temesgen, A. M. (2015). Global, regional, and national age–sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet*, *385*(9963), 117–171. [https://doi.org/10.1016/S0140-6736\(14\)61682-2](https://doi.org/10.1016/S0140-6736(14)61682-2)
- Naylor, M., Brown, K. J., & Vasan, R. S. (2021). The Molecular Basis of Predicting Atherosclerotic Cardiovascular Disease Risk. *Circulation Research*, *128*(2), 287–303. <https://doi.org/10.1161/CIRCRESAHA.120.315890>

- 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., Köttgen, A., Naimark, D. M. J., Ohkubo, T., Prescott, G. J., Rebholz, C. M., Sabanayagam, C., ... CKD Prognosis Consortium. (2019). Development of Risk Prediction Equations for Incident Chronic Kidney Disease. *JAMA*, *322*(21), 2104–2114. <https://doi.org/10.1001/jama.2019.17379>
- Nicholson, G., Rantalainen, M., Maher, A. D., Li, J. V., Malmodin, D., Ahmadi, K. R., Faber, J. H., Hallgrímsson, I. B., Barrett, A., Toft, H., Krestyaninova, M., Viksna, J., Neogi, S. G., Dumas, M.-E., Sarkans, U., Silverman, B. W., Donnelly, P., Nicholson, J. K., Allen, M., ... Holmes, C. C. (2011). Human metabolic profiles are stably controlled by genetic and environmental variation. *Molecular Systems Biology*, *7*(1), 525. <https://doi.org/10.1038/msb.2011.57>
- Nielsen, F. (2016). Hierarchical Clustering. In *Introduction to HPC with MPI for Data Science* (pp. 195–211). Springer.
- 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. <https://doi.org/10.2337/dc16-0173>
- Niiranen, T. J., & Vasani, R. S. (2016). Epidemiology of cardiovascular disease: recent novel outlooks on risk factors and clinical approaches. *Expert Review of Cardiovascular Therapy*, *14*(7), 855–869. <https://doi.org/10.1080/14779072.2016.1176528>
- Nordström, T., Miettunen, J., Auvinen, J., Ala-Mursula, L., Keinänen-Kiukkaanniemi, S., Veijola, J., Järvelin, M.-R., Sebert, S., & Männikkö, M. (2022). Cohort Profile: 46 years of follow-up of the Northern Finland Birth Cohort 1966 (NFBC1966). *International Journal of Epidemiology*, *50*(6), 1786-1787j. <https://doi.org/10.1093/ije/dyab109>
- O’Seaghdha, C. M., Lyass, A., Massaro, J. M., Meigs, J. B., Coresh, J., D’Agostino, R. B., Astor, B. C., & Fox, C. S. (2012). A risk score for chronic kidney disease in the general population. *The American Journal of Medicine*, *125*(3), 270–277. <https://doi.org/10.1016/j.amjmed.2011.09.009>
- Peregrín-Alvarez, J. M., Sanford, C., & Parkinson, J. (2009). The conservation and evolutionary modularity of metabolism. *Genome Biology*, *10*(6), R63. <https://doi.org/10.1186/gb-2009-10-6-r63>
- Pietzner, M., Kaul, A., Henning, A.-K., Kastenmüller, G., Artati, A., Lerch, M. M., Adamski, J., Nauck, M., & Friedrich, N. (2017). Comprehensive metabolic profiling of chronic low-grade inflammation among generally healthy individuals. *BMC Medicine*, *15*(1), 210. <https://doi.org/10.1186/s12916-017-0974-6>
- Plomin, R., Haworth, C. M. A., & Davis, O. S. P. (2009). Common disorders are quantitative traits. *Nature Reviews. Genetics*, *10*(12), 872–878. <https://doi.org/10.1038/nrg2670>
- Puska, P., & Jains, P. (2020). The North Karelia Project: Prevention of Cardiovascular Disease in Finland Through Population-Based Lifestyle Interventions. *American Journal of Lifestyle Medicine*, *14*(5), 495–499. <https://doi.org/10.1177/1559827620910981>

- Rainville, P. D., Wilson, I. D., Nicholson, J. K., Isaac, G., Mullin, L., Langridge, J. I., & Plumb, R. S. (2017). Ion mobility spectrometry combined with ultra performance liquid chromatography/mass spectrometry for metabolic phenotyping of urine: Effects of column length, gradient duration and ion mobility spectrometry on metabolite detection. *Analytica Chimica Acta*, *982*, 1–8. <https://doi.org/10.1016/j.aca.2017.06.020>
- Raitakari, O. T., Juonala, M., Ronnema, T., Keltikangas-Jarvinen, L., Rasanen, L., Pietikainen, M., Hutri-Kahonen, N., Taittonen, L., Jokinen, E., Marniemi, J., Jula, A., Telama, R., Kahonen, M., Lehtimaki, T., Akerblom, H. K., & Viikari, J. S. (2008). Cohort Profile: The Cardiovascular Risk in Young Finns Study. *International Journal of Epidemiology*, *37*(6), 1220–1226. <https://doi.org/10.1093/ije/dym225>
- Rappaport, S. M., Barupal, D. K., Wishart, D., Vineis, P., & Scalbert, A. (2014). The Blood Exposome and Its Role in Discovering Causes of Disease. *Environmental Health Perspectives*, *122*(8), 769–774. <https://doi.org/10.1289/ehp.1308015>
- Rhee, C. M., & Kovesdy, C. P. (2015). Spotlight on CKD deaths—increasing mortality worldwide. *Nature Reviews Nephrology*, *11*(4), 199–200. <https://doi.org/10.1038/nrneph.2015.25>
- Rose, G. (1985). Sick individuals and sick populations. *International Journal of Epidemiology*, *14*(1), 32–38. <https://doi.org/10.1093/ije/14.1.32>
- Saccenti, E. (2017). Correlation Patterns in Experimental Data Are Affected by Normalization Procedures: Consequences for Data Analysis and Network Inference. *Journal of Proteome Research*, *16*(2), 619–634. <https://doi.org/10.1021/acs.jproteome.6b00704>
- Salihovic, S., Broeckling, C. D., Ganna, A., Prenni, J. E., Sundström, J., Berne, C., Lind, L., Ingelsson, E., Fall, T., Ärnlöv, J., & Nowak, C. (2020). Non-targeted urine metabolomics and associations with prevalent and incident type 2 diabetes. *Scientific Reports*, *10*(1), 16474. <https://doi.org/10.1038/s41598-020-72456-y>
- Sarnak, M. J., Amann, K., Bangalore, S., Cavalcante, J. L., Charytan, D. M., Craig, J. C., Gill, J. S., Hlatky, M. A., Jardine, A. G., Landmesser, U., Newby, L. K., Herzog, C. A., Cheung, M., Wheeler, D. C., Winkelmayer, W. C., Marwick, T. H., & Conference Participants. (2019). Chronic Kidney Disease and Coronary Artery Disease: JACC State-of-the-Art Review. *Journal of the American College of Cardiology*, *74*(14), 1823–1838. <https://doi.org/10.1016/j.jacc.2019.08.1017>
- Schifano, E., Conta, G., Preziosi, A., Ferrante, C., Batignani, G., Mancini, P., Tomassini, A., Sciubba, F., Scopigno, T., Uccelletti, D., & Miccheli, A. (2022). 2-hydroxyisobutyric acid (2-HIBA) modulates ageing and fat deposition in *Caenorhabditis elegans*. *Frontiers in Molecular Biosciences*, *9*, 986022. <https://doi.org/10.3389/fmolb.2022.986022>
- Sekula, P., Goek, O.-N., Quayle, L., Barrios, C., Levey, A. S., Römisch-Margl, W., Menni, C., Yet, I., Gieger, C., Inker, L. A., Adamski, J., Gronwald, W., Illig, T., Dettmer, K., Krumsiek, J., Oefner, P. J., Valdes, A. M., Meisinger, C., Coresh, J., ... Köttgen, A. (2016). A Metabolome-Wide Association Study of Kidney Function and Disease in the General Population. *Journal of the American Society of Nephrology: JASN*, *27*(4), 1175–1188. <https://doi.org/10.1681/ASN.2014111099>

- Serra-Majem, L., Ribas, L., Ngo, J., Ortega, R. M., García, A., Pérez-Rodrigo, C., & Aranceta, J. (2004). Food, youth and the Mediterranean diet in Spain. Development of KIDMED, Mediterranean Diet Quality Index in children and adolescents. *Public Health Nutrition*, 7(7), 931–935. <https://doi.org/10.1079/phn2004556>
- Shao, Y., Li, T., Liu, Z., Wang, X., Xu, X., Li, S., Xu, G., & Le, W. (2021). Comprehensive metabolic profiling of Parkinson’s disease by liquid chromatography-mass spectrometry. *Molecular Neurodegeneration*, 16(1), 4. <https://doi.org/10.1186/s13024-021-00425-8>
- Slupsky, C. M., Rankin, K. N., Wagner, J., Fu, H., Chang, D., Weljie, A. M., Saude, E. J., Lix, B., Adamko, D. J., Shah, S., Greiner, R., Sykes, B. D., & Marrie, T. J. (2007). Investigations of the Effects of Gender, Diurnal Variation, and Age in Human Urinary Metabolomic Profiles. *Analytical Chemistry*, 79(18), 6995–7004. <https://doi.org/10.1021/ac0708588>
- Smith, G. D. (2011). Epidemiology, epigenetics and the “Gloomy Prospect”: embracing randomness in population health research and practice. *International Journal of Epidemiology*, 40(3), 537–562. <https://doi.org/10.1093/ije/dyr117>
- Smith, G. D., Lawlor, D. A., Harbord, R., Timpson, N., Day, I., & Ebrahim, S. (2007). Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. *PLoS Medicine*, 4(12), e352. <https://doi.org/10.1371/journal.pmed.0040352>
- Sniderman, A. D., Thanassoulis, G., Wilkins, J. T., Furberg, C. D., & Pencina, M. (2018). Sick Individuals and Sick Populations by Geoffrey Rose: Cardiovascular Prevention Updated. *Journal of the American Heart Association*, 7(19), e010049. <https://doi.org/10.1161/JAHA.118.010049>
- Soininen, P., Haarala, J., Vepsäläinen, J., Niemitz, M., & Laatikainen, R. (2005). Strategies for organic impurity quantification by 1H NMR spectroscopy: Constrained total-line-shape fitting. *Analytica Chimica Acta*, 542(2), 178–185. <https://doi.org/10.1016/j.aca.2005.03.060>
- Soininen, P., Kangas, A. J., Würtz, P., Suna, T., & Ala-Korpela, M. (2015). Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circulation. Cardiovascular Genetics*, 8(1), 192–206. <https://doi.org/10.1161/CIRCGENETICS.114.000216>
- Stamler, J., Elliott, P., Dennis, B., Dyer, A. R., Kesteloot, H., Liu, K., Ueshima, H., Zhou, B. F., & INTERMAP Research Group. (2003). INTERMAP: background, aims, design, methods, and descriptive statistics (nondietary). *Journal of Human Hypertension*, 17(9), 591–608. <https://doi.org/10.1038/sj.jhh.1001603>
- Sterne, J. A., & Davey Smith, G. (2001). Sifting the evidence-what’s wrong with significance tests? *BMJ (Clinical Research Ed.)*, 322(7280), 226–231. <https://doi.org/10.1136/bmj.322.7280.226>

- Stratakis, N., Siskos, A. P., Papadopoulou, E., Nguyen, A. N., Zhao, Y., Margetaki, K., Lau, C.-H. E., Coen, M., Maitre, L., Fernández-Barrés, S., Agier, L., Andrusaityte, S., Basagaña, X., Brantsaeter, A. L., Casas, M., Fossati, S., Grazuleviciene, R., Heude, B., McEachan, R. R., ... Chatzi, L. (2022). Urinary metabolic biomarkers of diet quality in European children are associated with metabolic health. *ELife*, *11*. <https://doi.org/10.7554/eLife.71332>
- Titan, S. M., Venturini, G., Padilha, K., Tavares, G., Zatz, R., Bensenor, I., Lotufo, P. A., Rhee, E. P., Thadhani, R. I., & Pereira, A. C. (2019). Metabolites related to eGFR: Evaluation of candidate molecules for GFR estimation using untargeted metabolomics. *Clinica Chimica Acta: International Journal of Clinical Chemistry*, *489*(April 2018), 242–248. <https://doi.org/10.1016/j.cca.2018.08.037>
- Torgrip, R. J. O., Åberg, K. M., Alm, E., Schuppe-Koistinen, I., & Lindberg, J. (2008). A note on normalization of biofluid 1D 1H-NMR data. *Metabolomics*, *4*(2), 114–121. <https://doi.org/10.1007/s11306-007-0102-2>
- Tynkkynen, T., Wang, Q., Ekholm, J., Anufrieva, O., Ohukainen, P., Vepsäläinen, J., Männikkö, M., Keinänen-Kiukaanniemi, S., Holmes, M. V., Goodwin, M., Ring, S., Chambers, J. C., Kooner, J., Järvelin, M.-R., Kettunen, J., Hill, M., Davey Smith, G., & Ala-Korpela, M. (2019). Proof of concept for quantitative urine NMR metabolomics pipeline for large-scale epidemiology and genetics. *International Journal of Epidemiology*, *48*(3), 978–993. <https://doi.org/10.1093/ije/dyy287>
- Vallon, V. (2020). Glucose transporters in the kidney in health and disease. *Pflugers Archiv : European Journal of Physiology*, *472*(9), 1345–1370. <https://doi.org/10.1007/s00424-020-02361-w>
- van den Berg, R. A., Hoefsloot, H. C. J., 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. <https://doi.org/10.1186/1471-2164-7-142>
- van Duynhoven, J. P. M., & Jacobs, D. M. (2016). Assessment of dietary exposure and effect in humans: The role of NMR. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *96*, 58–72. <https://doi.org/10.1016/j.pnmrs.2016.03.001>
- van Roekel, E. H., Lofffield, E., Kelly, R. S., Zeleznik, O. A., & Zanetti, K. A. (2019). Metabolomics in epidemiologic research: challenges and opportunities for early-career epidemiologists. *Metabolomics : Official Journal of the Metabolomic Society*, *15*(1), 9. <https://doi.org/10.1007/s11306-018-1468-z>
- Vart, P., & Grams, M. E. (2016). Measuring and Assessing Kidney Function. *Seminars in Nephrology*, *36*(4), 262–272. <https://doi.org/10.1016/j.semnephrol.2016.05.003>
- Verrey, F., Singer, D., Ramadan, T., Vuille-dit-Bille, R. N., Mariotta, L., & Camargo, S. M. R. (2009). Kidney amino acid transport. *Pflugers Archiv : European Journal of Physiology*, *458*(1), 53–60. <https://doi.org/10.1007/s00424-009-0638-2>
- Visscher, P. M., Yengo, L., Cox, N. J., & Wray, N. R. (2021). Discovery and implications of polygenicity of common diseases. *Science (New York, N.Y.)*, *373*(6562), 1468–1473. <https://doi.org/10.1126/science.abi8206>

- Völzke, H., Alte, D., Schmidt, C. O., Radke, D., Lorbeer, R., Friedrich, N., Aumann, N., Lau, K., Piontek, M., Born, G., Havemann, C., Ittermann, T., Schipf, S., Haring, R., Baumeister, S. E., Wallaschofski, H., Nauck, M., Frick, S., Arnold, A., ... Hoffmann, W. (2011). Cohort profile: the study of health in Pomerania. *International Journal of Epidemiology*, *40*(2), 294–307. <https://doi.org/10.1093/ije/dyp394>
- Waikar, S. S., Sabbiseti, V. S., & Bonventre, J. V. (2010). Normalization of urinary biomarkers to creatinine during changes in glomerular filtration rate. *Kidney International*, *78*(5), 486–494. <https://doi.org/10.1038/ki.2010.165>
- Wang, Q., Holmes, M. V., Davey Smith, G., & Ala-Korpela, M. (2017). Genetic Support for a Causal Role of Insulin Resistance on Circulating Branched-Chain Amino Acids and Inflammation. *Diabetes Care*, *40*(12), 1779–1786. <https://doi.org/10.2337/dc17-1642>
- Wang, Q., Würtz, P., Auro, K., Mäkinen, V.-P., Kangas, A. J., Soinen, P., Tiainen, M., Tynkkynen, T., Jokelainen, J., Santalahti, K., Salmi, M., Blankenberg, S., Zeller, T., Viikari, J., Kähönen, M., Lehtimäki, T., Salomaa, V., Perola, M., Jalkanen, S., ... Ala-Korpela, M. (2016). Metabolic profiling of pregnancy: cross-sectional and longitudinal evidence. *BMC Medicine*, *14*(1), 205. <https://doi.org/10.1186/s12916-016-0733-0>
- Ware, J. H. (2006). The Limitations of Risk Factors as Prognostic Tools. *New England Journal of Medicine*, *355*(25), 2615–2617. <https://doi.org/10.1056/NEJMp068249>
- Waxman, A., & World Health Assembly. (2004). WHO global strategy on diet, physical activity and health. *Food and Nutrition Bulletin*, *25*(3), 292–302. <https://doi.org/10.1177/156482650402500310>
- Wishart, D. S. (2019). Metabolomics for Investigating Physiological and Pathophysiological Processes. *Physiological Reviews*, *99*(4), 1819–1875. <https://doi.org/10.1152/physrev.00035.2018>
- Wishart, D. S., Feunang, Y. D., Marcu, A., Guo, A. C., Liang, K., Vázquez-Fresno, R., Sajed, T., Johnson, D., Li, C., Karu, N., Sayeeda, Z., Lo, E., Assempour, N., Berjanskii, M., Singhal, S., Arndt, D., Liang, Y., Badran, H., Grant, J., ... Scalbert, A. (2018). HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Research*, *46*(D1), D608–D617. <https://doi.org/10.1093/nar/gkx1089>
- 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., ... Gautam, V. (2022). HMDB 5.0: the Human Metabolome Database for 2022. *Nucleic Acids Research*, *50*(D1), D622–D631. <https://doi.org/10.1093/nar/gkab1062>
- Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, *58*(2), 109–130. [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1)
- Wu, Y., & Li, L. (2016). Sample normalization methods in quantitative metabolomics. *Journal of Chromatography A*, *1430*, 80–95. <https://doi.org/10.1016/j.chroma.2015.12.007>

- Würtz, 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. *American Journal of Epidemiology*, *186*(9), 1084–1096. <https://doi.org/10.1093/aje/kwx016>
- Würtz, P., Mäkinen, V.-P., Soininen, P., Kangas, A. J., Tukiainen, T., Kettunen, J., Savolainen, M. J., Tammelin, T., Viikari, J. S., Rönnemaa, T., Kähönen, M., Lehtimäki, T., Ripatti, S., Raitakari, O. T., Järvelin, M.-R., & Ala-Korpela, M. (2012). Metabolic signatures of insulin resistance in 7,098 young adults. *Diabetes*, *61*(6), 1372–1380. <https://doi.org/10.2337/db11-1355>
- Xia, J., Broadhurst, D. I., Wilson, M., & Wishart, D. S. (2013). Translational biomarker discovery in clinical metabolomics: an introductory tutorial. *Metabolomics: Official Journal of the Metabolomic Society*, *9*(2), 280–299. <https://doi.org/10.1007/s11306-012-0482-9>
- Ye, T., Mo, H., Shanaiah, N., Gowda, G. A. N., Zhang, S., & Raftery, D. (2009). Chemoselective <sup>15</sup>N tag for sensitive and high-resolution nuclear magnetic resonance profiling of the carboxyl-containing metabolome. *Analytical Chemistry*, *81*(12), 4882–4888. <https://doi.org/10.1021/ac900539y>
- Yousri, N. A., Kastenmüller, G., Gieger, C., Shin, S.-Y., Erte, I., Menni, C., Peters, A., Meisinger, C., Mohny, R. P., Illig, T., Adamski, J., Soranzo, N., Spector, T. D., & Suhre, K. (2014). Long term conservation of human metabolic phenotypes and link to heritability. *Metabolomics: Official Journal of the Metabolomic Society*, *10*(5), 1005–1017. <https://doi.org/10.1007/s11306-014-0629-y>
- 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. <https://doi.org/10.1159/000521940>

## Original publications

- I Li, T., Ihanus, A., Ohukainen, P., Järvelin, M. R., Kettunen, J., Mäkinen, V. P., Tynkkynen, T., & Ala-Korpela, M. (2022). There is always glucose in normal urine: unspecific excretion associated with serum glucose and glomerular filtration rate *International Journal of Epidemiology*, 51(6), 2022–2025, doi: 10.1093/ije/dyac060
- II Li, T., Tynkkynen, T., Ihanus, A., Zhao, S., Mäkinen, V. P., & Ala-Korpela, M. (2022). Characteristics of Normalization Methods in Quantitative Urinary Metabolomics-Implications for Epidemiological Applications and Interpretations. *Biomolecules*, 12(7), doi: 10.3390/biom12070903
- III Li, T.<sup>†</sup>, Ihanus, A.<sup>†</sup>, Ohukainen, P., Järvelin, M. R., Kähönen, M., Kettunen, J., Raitakari, O.T., Lehtimäki, Terho., Mäkinen, V. P., Tynkkynen, T.<sup>‡</sup>, & Ala-Korpela, M.<sup>\* ‡</sup>. (2023) Clinical and biochemical associations of urinary metabolites: Quantitative epidemiological approach on renal-cardiometabolic biomarkers. *International Journal of Epidemiology*, doi: 10.1093/ije/dyad162

Reprinted with permission under Creative Commons CC BY 4.0 license (Publications I, II, III © 2022, 2023 Authors)<sup>1</sup>.

Original publications are not included in the electronic version of the dissertation.

---

<sup>1</sup> <https://creativecommons.org/licenses/by/4.0/>



1757. Paukkonen, Leila (2023) Patient participation, adherence, and activation for self-management among adult patients with multimorbidity in primary healthcare settings
1758. Närhi, Anne Linnea (2023) Malocclusions and their associations with oral health-related quality of life in the Northern Finland Birth Cohort 1966
1759. Elseragy, Amr (2023) Predictive features and potential treatment targets for oral tongue cancer
1760. Keinänen, Anna-Leena (2024) Koulutusinterventioiden vaikuttavuus suun terveydenhuollon ammattilaisten opiskelijaohjausosaamiseen kliinisessä harjoittelussa
1761. Tikkanen, Ville (2024) Adolescent social functioning and later psychiatric morbidity in genetic high- and low-risk adoptees. The Finnish Adoptive Family Study of Schizophrenia
1762. Mäkikyrö, Elina (2024) Asthma subtypes formed by epidemiological and clinical methods
1763. Tervahauta, Elisa (2024) Occlusal characteristics, temporomandibular disorders, and oral health-related quality of life in the Northern Finland Birth Cohort 1966
1764. Mosorin, Maria-Elina (2024) Metabolic effects of hormonal contraception in women with and without polycystic ovary syndrome
1765. Puhakka, Soile (2024) Natural residential environment and physical activity in adolescence and midlife : the role of greenness, landscape diversity, and nature relatedness
1766. Pöyry, Hilla (2024) Pediatric emergency department visits and parental assessment of acutely ill children
1767. Kerimaa, Heli (2024) Leikki-ikäisten lasten ja heidän vanhempinsa päiväkirurgiaan valmistaminen
1768. Valinen, Yrtti (2024) ADAM12 and first trimester trisomy screening
1769. Hanhivaara, Jaakko (2024) Lumbosacral transitional vertebrae : prevalence, associated degenerative changes, and association with low back pain in Finnish population
1770. Lehtovirta, Sami (2024) Utility of quantitative magnetic resonance imaging for articular cartilage assessment in clinical scenarios

S E R I E S E D I T O R S

**A**  
**SCIENTIAE RERUM NATURALIUM**  
*University Lecturer Mahmoud Filali*

**B**  
**HUMANIORA**  
*University Lecturer Santeri Palviainen*

**C**  
**TECHNICA**  
*Senior Research Fellow Antti Kajjalainen*

**D**  
**MEDICA**  
*University Lecturer Pirjo Kaakinen*

**E**  
**SCIENTIAE RERUM SOCIALIUM**  
*University Lecturer Henri Pettersson*

**E**  
**SCRIPTA ACADEMICA**  
*Strategy Officer Mari Katvala*

**G**  
**OECONOMICA**  
*University Researcher Marko Korhonen*

**H**  
**ARCHITECTONICA**  
*Associate Professor Anu Soikkeli*

**EDITOR IN CHIEF**  
*University Lecturer Santeri Palviainen*

**PUBLICATIONS EDITOR**

ISBN 978-952-62-3986-6 (Paperback)  
ISBN 978-952-62-3987-3 (PDF)  
ISSN 0355-3221 (Print)  
ISSN 1796-2234 (Online)