Aus dem Institut für Epidemiologie
Helmholtz Zentrum München

Direktor: Prof. Dr. Annette Peters

Prediction and progression of type 2 diabetes using molecular characterized signatures

Dissertation
zum Erwerb des Doktorgrades der Naturwissenschaften
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität München

vorgelegt von
Qiuling Dong

aus
Hubei, China

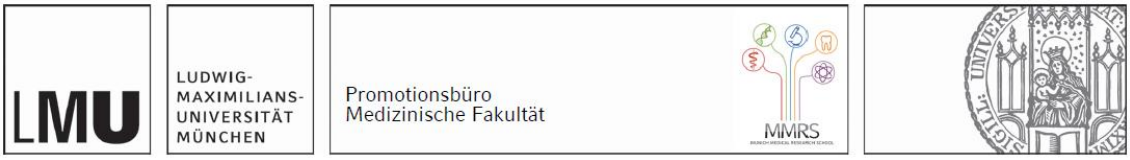
Jahr
2024

Mit Genehmigung der Medizinischen Fakultät
der Ludwig-Maximilians-Universität München

Betreuerin:	Prof. Dr. Annette Peters
Zweitgutachter:	Prof. Dr. Christian Unger
Dekan:	Prof. Dr. med. Thomas Gudermann

Tag der mündlichen Prüfung: 13.10.2025

Affidavit



Affidavit

Qiuling Dong

Surname, first name

Ingolstädter Landstraße 1

Street

D-85764 Neuherberg, Munich, Germany

Zip code, town, country

I hereby declare, that the submitted thesis entitled:

Prediction and progression of Type 2 Diabetes using molecular characterized signatures .....

is my own work. I have only used the sources indicated and have not made unauthorised use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.

I further declare that the dissertation presented here has not been submitted in the same or similar form to any other institution for the purpose of obtaining an academic degree.

Munich, 18.11.2024 — Qiuling Dong

place, date — Signature doctoral candidate

Table of Contents

List of abbreviations

Type 2 diabetes	T2D
Cooperative Health Research in the Region of Augsburg	KORA
Branched-chain amino acids	BCAA
Sphingomyelin	SM
Diacylphosphatidylcholine	PC aa
Normal glucose regulation	NGR
Impaired glucose regulation	IGR
Lysophosphatidylcholine acyl	LysoPC a
High-sensitivity C-reactive protein	hsCRP
Polygenic risk scores	PRS
Diabetes mellitus	DM
Maturity-onset diabetes of the young	MODY
World health organization	WHO
Fasting plasma glucose	FPG
2-h plasma glucose	2h-PG
Free fatty acid	FFA
Oral glucose tolerance test	OGTT
American diabetes association	ADA
Impaired fasting glucose	IFG
Impaired glucose tolerance	IGT
Hemoglobin a1c	HbA1c
European Association for the Study of Diabetes	EASD
Body mass index	BMI
Dipeptidyl peptidase 4 inhibitors	DPP4i
Genome-wide association studies	GWAS
Severe autoimmune diabetes	SAID
Estimated glomerular filtration rate	eGFR
Chronic kidney disease epidemiology collaboration	CKD-EPI
Chronic kidney disease	CKD
Nnovation Medicines Initiative - Diabetes Research on Patient Stratification	IMI-DIRECT
Quality control	QC
Haplotype reference consortium	HRC

Minor allele frequency	MAF
Two sample mendelian randomization	2SMR
Genotype-tissue expression	GTEx
Mouse genome informatics	MGI
Inverse-variance-weighted	IVW
Polygenic score	PGS
Phosphatidylcholines	PC
Fatty acid	FA
Polyunsaturated fatty acids	PUFAs
Lysophosphatidylcholine	LPC
Severe insulin-deficient diabetes	SIDD
Severe insulin resistant diabetes	SIRD
Mild obesity-related diabetes	MORD
Mild age-related diabetes	MARD

List of publications

This thesis consists of the following papers:

Qiuling D, Sidra S, Christian G, Rui W, et al. Metabolic Signatures Elucidate the Effect of Body Mass Index on Type 2 Diabetes. *Metabolites*. 2023 Feb 3;13(2):227.

Sapna S*, **Qiuling D***, Mark H*, et al. Role of human blood plasma metabolites in prediabetes and Type 2 Diabetes from IMI-DIRECT study. (*Joint first author) Diabetologia. 2024 Sep 30. doi: 10.1007/s00125-024-06282-6. Online ahead of print.

Qiuling D, Yue X, Stefan B, Markéta F, et al. Subphenotypes of adult-onset diabetes: Data-driven clustering in the population-based KORA cohort. Diabetes, Obesity and Metabolism. First published: 28 October 2024. https://doi.org/10.1111/dom.16022

Contribution to papers

1.1 Contribution to paper I

The study “Metabolic Signatures Elucidate the Effect of Body Mass Index on Type 2 Diabetes” was published in *Metabolites* in 2023. The research was conceptualized by Qiuling Dong in consultation with Prof. Annette Peters, Dr. Harald Grallert and Dr. Sapna Sharma. Qiuling Dong analyzed the data, evaluated and visualized the results, drafted the manuscript, and managed the publication process as first author. All co-authors critically reviewed and approved the manuscript.

1.2 Contribution to paper II

The study “Role of human blood plasma metabolites in prediabetes and type 2 diabetes from DIRECT study” was published in *Diabetologia* in 2024. The research was conceptualized and designed by Dr. Sapna Sharma and Qiuling Dong in consultation with Dr. Harald Grallert. Qiuling Dong conducted data analysis, evaluated and visualized the results, drafted the manuscript, and managed the publication process as shared first author with Dr. Sapna Sharma. All co-authors critically reviewed and approved the manuscript.

1.3 Contribution to paper III

The study “Subphenotypes of adult-onset diabetes: Data-driven clustering in the population-based KORA cohort” was published in *Diabetes, Obesity and Metabolism* in 2024. The research was designed by Qiuling Dong in consultation with Prof. Annette Peters and Dr. Harald Grallert. Qiuling Dong conducted the data analysis, evaluated and visualized the results, drafted the manuscript, and managed the publication process as first author. All co-authors critically reviewed and approved the manuscript.

2. Abstract (English):

Type 2 diabetes (T2D) is a prevalent and intricate metabolic condition stemming from the body's inefficient utilization of insulin. It is distinguished by elevated blood sugar levels attributed to compromised insulin secretion and resistance, with the majority of cases linked to overweight or obesity. T2D is a complex chronic disease with no apparent early symptoms or vary widely from individual to individual but untreated high blood sugar can damage multiple organs. So early detection of preclinical conditions and comprehension of the underlying mechanisms within subphenotypes of T2D are crucial, followed by the adoption of preventive and therapeutic approaches.

Precision medicine in diabetes entails optimizing the diagnosis, prediction, prevention, or treatment of diabetes through integration of multidimensional data while considering individual variations. It employs muti-omics technologies such as genotyping, transcriptomics, metabolomics combined with clinical phenotype to illustrate the systematic biology. There is growing interest in applying metabolic profiling to identify disease molecular signatures, as it offers a powerful approach for unraveling the complex relationships between obesity, metabolism, and diabetes progression. This thesis aims to advance precision health strategies for T2D by improving early detection with identified candidate biomarkers and capturing disease heterogeneity using data-driven classification approaches, that integrates metabolomics, genetics, and clinical assessment data.

This thesis firstly utilized 146 targeted metabolomic profiles obtained from the Cooperative Health Research in the Region of Augsburg (KORA) FF4 cohort comprising 1715 participants and correlating them with obesity and T2D. 42 and 3 metabolites were significantly correlated with body mass index (BMI) and T2D adjusted for multiple covariates, respectively, and were also replicated in the previous studies. Those metabolites included branched-chain amino acids (BCAA) and lipids. Sobel mediation test implied that lipids including sphingomyelin (SM) C16:1, SM C18:1 and diacyl phosphatidylcholine (PC aa) C38:3 mediated the impact of BMI on T2D. Additionally, mendelian randomization indicated a causal link where BMI influenced changes in SM C16:1 and PC aa C38:3, and alterations in SM C16:1, SM C18:1, and PC aa C38:3 contributed to T2D incident. Biological pathway analysis, alongside genetic studies, and experiments with mice, revealed that dysregulation of sphingolipid and phosphatidylcholine metabolism were pivotal factors in the early stages of T2D pathophysiology. Our findings highlight that these three identified metabolites play a mediating role in connecting BMI with T2D, shedding light on their significance in T2D development.

To further elucidate the role of metabolites in the glycemic deterioration, data from 3000 individuals enrolled in the Innovation Medicines Initiative - Diabetes Research on Patient Stratification (IMI-DIRECT) consortium were analyzed, with measurements available for 911 metabolites (132 targeted-metabolomics, 779 untargeted-metabolomics). In the targeted (and untargeted) metabolomics measurements, we observed 4 (15) and 34 (99) metabolites had significant variation in normal glucose regulation (NGR) group compared to those with impaired glucose regulation (IGR) and T2D groups respectively. Besides, for pre-diabetic group, 50 (108) metabolites were identified to be significantly distinct from T2D group. Metabolites identified through targeted metabolomics, such as lysophosphatidylcholine acyl (lysoPC a) C17:0 and the sum of hexoses and untargeted metabolomics including N-lactoylvaline, N-lactoylleucine, formiminoglutamate, carbohydrate lactate, and an unknown compound (X-24295) were significantly associated with HbA1c progression rate and predictive of incident prediabetes and diabetes. In the causal mediation test, we also

observed that these metabolites were significant mediators of glycemic deterioration from baseline to 18- and 48-month follow-ups. In mendelian randomization, we observed T2D exhibited a causal influence on the concentrations of three metabolites (hexose, glutamate and caproate (FA 6:0)), while four phosphatidylcholines such as PC aa C36:2 as well as the two omega-3 fatty acids stearidonate (FA18:4) and docosapentaenoate (n3 DPA; FA22:5) potentially played a causal role in the onset of T2D. Our findings suggest metabolites lysoPC a C17:0, N-lactoylvaline, N-lactoylleucine, formiminoglutamate, as well as lactate, and an unknown metabolite (X-24295) are linked with glycemic deterioration and are the mediators for developing of IGR or T2D, which help improve the early detection and understanding the progression of the disease.

In above sections we identified the biomarkers playing a role in the early-stage T2D progression, which could contribute to implementation of preventive and therapeutic strategies. Individual trajectories of hyperglycemia vary widely, necessitating a thorough comprehension of its mechanisms and the implementation of precision treatment for T2D patients with this condition. Therefore, we used data of 301 T2D individuals from KORA FF4 study for cluster analysis. We firstly replicated original cluster from the study by Ahlqvist et al. 2018 by forcing k=4 with same variables but three different scaling parameters and centroids combinations. We found original clusters were not effectively replicated, as evidenced by significantly different assignment frequencies and cluster characteristics between the ANDIS and KORA samples. New clusters were derived through open k-means analysis and the stability of new clusters was evaluated based on the assignment consistency across various sets of variables and Jaccard indices. K=3 clusters was used in the new clustering and additionally including high-sensitivity C-reactive protein (hsCRP) in the variable set yielded notable cluster stability with all Jaccard indices exceeding 0.75. Polygenic risk scores (PRS) and diabetes complications were delineated in the new three clusters as manifestations of inherent heterogeneity. The three de-novo derived clusters (n= 96, 172, 33, respectively) effectively captured heterogeneity within the sample and exhibited distinct distributions of PRS and diabetes complications, i.e. Cluster 1 was characterized by insulin resistance with high neuropathy prevalence, Cluster 2 was defined as age-related diabetes with higher prevalence of stroke and CKD, and Cluster 3 showed the highest genetic predisposition and risk of obesity-related diabetes. Our findings demonstrate that subphenotyping T2D based on unique clinical characteristics of the samples yields stable categorization and effectively captures T2D heterogeneity, thereby supporting the advancement of personalized treatment strategies.

In conclusion, this thesis shows that metabolic profiles can support the early detection of diabetes and deepen our understanding of the pathological mechanisms underlying T2D progression. Additionally, subtyping T2D aids elucidate its inherent heterogeneity and paves the way for personalized treatment approaches. Together, these insights offer valuable contributions to the advancement of precision health.

3. Zusammenfassung

Typ-2-Diabetes (T2D) ist eine weit verbreitete und komplexe Stoffwechselerkrankung, die auf eine ineffiziente Glukoseregulation zurückzuführen ist. Sie zeichnet sich durch einen erhöhten Blutzuckerspiegel aus, der durch eine beeinträchtigte Insulinsekretion und -resistenz bedingt ist, wobei die meisten Fälle auf Übergewicht oder Fettleibigkeit zurückzuführen sind. T2D ist eine chronische Stoffwechselerkrankung, deren Symptome im Frühstadium fehlen oder von Person zu Person stark variieren können. Unbehandelt kann ein hoher Blutzucker jedoch mehrere Organe schädigen. Daher ist die frühzeitige Erkennung präklinischer Erkrankungen und das Verständnis der zugrunde liegenden Mechanismen von entscheidender Bedeutung, um präventive und therapeutische Ansätze zu entwickeln.

Präzisionsmedizin bei Diabetes bedeutet die Optimierung der Diagnose, Vorhersage, Prävention oder Behandlung von Diabetes durch die Integration mehrdimensionaler Daten unter Berücksichtigung individueller Variationen. Es nutzt Muti-Omics-Technologien wie Genotypisierung, Transkriptomik und Metabolomik in Kombination mit klinischem Phänotyp, um die zugrundeliegende systemische Biologie zu veranschaulichen. Besonders zunehmendes Interesse besteht an der Anwendung von Stoffwechselprofilen zur Identifizierung von Krankheitsbiomarkern, da es sich um einen wirksamen Ansatz zur Aufdeckung des komplizierten Verlaufs zwischen Fettleibigkeit, Stoffwechsel und Diabetes handelt. Ziel dieser Arbeit ist es, zur Präzisionsmedizin von T2D beizutragen, indem die Früherkennung mit identifizierten Kandidaten-Biomarkern verbessert und Heterogenität mit datengesteuerter Klassifizierung auf der Grundlage von Metabolomik/Genetik und klinischen Daten erfasst wird.

In dieser Arbeit wurden zunächst 146 Metabolit-Profile, die aus der FF4-Kohorte der kooperativen Gesundheitsforschung in der Region Augsburg (KORA) mit 1715 Teilnehmern gewonnen wurden, mit Fettleibigkeit und T2D korreliert. 45 Metaboliten waren signifikant mit dem um mehrere Kovariaten adjustierten BMI und T2D assoziiert, die alle bereits bekannt waren. Bei diesen Metaboliten handelte es sich um verzweigtkettige Aminosäuren, Sphingolipide, Acylcarnitine, Lysophospholipide oder Phosphatidylcholine. Der Sobel-Mediationstest legt nahe, dass der Einfluss des BMI auf T2D über Lipide wie Sphingomyelin (SM) C16:1, SM C18:1 und Diacylphosphatidylcholin (PC aa) C38:3 vermittelt werden könnte. Darüber hinaus weist die Mendelsche Randomisierung auf einen Kausalzusammenhang hin, bei dem der BMI die Veränderungen in SM C16:1 und PC aa C38:3 beeinflusste und Veränderungen in SM C16:1, SM C18:1 und PC aa C38:3 zu T2D beitrugen. Die Analyse biologischer Signalwege sowie genetischer Studien und Experimente mit Mäusen zeigen, dass eine Fehlregulation des Sphingolipid- und Phosphatidylcholin-Stoffwechsels entscheidende Faktoren in den frühen Stadien der T2D-Pathophysiologie sind. Die Ergebnisse dieser Arbeit zeigen, dass diese drei identifizierten Metaboliten eine vermittelnde Rolle bei der Verbindung von BMI mit T2D spielen, was Aufschluss über ihre Bedeutung für die T2D-Pathologie gibt.

Um die Rolle von Metaboliten bei glykämischer Veränderungen weiter aufzuklären, wurden 3000 Personen des Konsortiums Innovation Medicines Initiative – Diabetes Research on Patient Stratification (IMI-DIRECT) mit 911 gemessenen Metaboliten (132 gezielte Metabolomik, 779 ungezielte Metabolomik) analysiert. Bei den gerichtete (und ungerichtete) Metabolomics-Messungen beobachteten wir, dass 4 (15) und 34 (99) Metaboliten signifikante Unterschiede in der Gruppe mit normaler Glukoseregulation (NGR) im Vergleich zu denen mit beeinträchtigter Glukoseregulation (IGR) bzw. T2D aufwiesen. Darüber hinaus wurden für die prädiabetische Gruppe 50 (108) Metaboliten identifiziert, die sich deutlich von der T2D-Gruppe unterschieden.

Wesentliche Metaboliten waren hauptsächlich verzweigtkettige Aminosäuren (BCAA), auch abgeleitete BCAA, Lipide, Xenobiotika und einige nicht annotierte Metabolite. Metaboliten wie LysoPC a C17:0, Summe der Hexosen (aus gerichtete Metabolomik), Aminosäuren aus dem BCAA-Metabolismus wie N-Lactoylvalin und N-Lactoylleucin, Formiminoglutamat sowie Laktat und ein unbekannter Metabolit (X-24295) waren mit der HbA1c-Progressionsrate korreliert und sagten das Auftreten von Prädiabetes/Diabetes voraus. Im kausalen Mediationstest beobachteten wir auch, dass diese Metaboliten signifikante Mediatoren der glykämischen Verschlechterung vom Ausgangswert bis zur Nachuntersuchung nach 18 und 48 Monaten waren. Die Mendelschen Randomisierung zeigte, dass T2D einen kausalen Einfluss auf die Konzentrationen von drei Metaboliten (Hexose, Glutamat und Caproat (FA 6:0)) hatte, während Lipide wie spezifische Phosphatidylcholine (PC aa C36:2) sowie die beiden Omega-3-Fettsäuren Stearidonat (FA18:4) und Docosapentaenoat (FA22:5) möglicherweise eine ursächliche Rolle bei der Entstehung von T2D spielen. Unsere Ergebnisse legen nahe, dass die Metaboliten LysoPC a C17:0, N-Lactoylvalin, N-Lactoylleucin, Formiminoglutamat sowie Laktat und ein unbekannter Metabolit (X-24295) mit einer glykämischen Verschlechterung assoziiert sind und die Mediatoren für die Entwicklung von IGR und T2D sind, die dazu beitragen könnten, die Früherkennung und das Verständnis der Pathologie der Krankheit zu verbessern.

In den obigen Abschnitten haben wir die Biomarker identifiziert, die bei der T2D-Progression im Frühstadium eine Rolle spielen und zur Umsetzung präventiver und therapeutischer Strategien beitragen könnten. Aufgrund er Komplexität und Heterogenität von T2D, ist ein grundlegendes Verständnis seiner Mechanismen und die Umsetzung einer präzisen Behandlung für T2D-Patienten mit dieser Erkrankung erforderlich. Deshalb wurden Clusteranalyse in Daten von n = 301 T2D-Personen aus der KORA FF4-Studie durchgeführt. Zunächst wurde der Originalcluster von Ahlqvist et al. Repliziert, k=4 unter Verwendung derselben Variablen wie Ahlqvist et al. erzwungen. Es wurden drei verschiedene Skalierungsparameter und Schwerpunktkombinationen untersucht. Dabei konnten die ursprünglichen Cluster nicht effektiv repliziert werden, da es deutlich unterschiedliche Zuordnungshäufigkeiten und Clustereigenschaften zwischen den ANDIS- und KORA-Proben gab. Daraufhin wurden neue Cluster mittels offener K-Means-Analyse abgeleitet und die Stabilität dieser Cluster auf der Grundlage der Zuweisungskonsistenz über verschiedene Variablensätze und Jaccard-Indizes bewertet. Bei der neuen Clusterbildung wurden 3 Cluster verwendet.Unter zusätzlicher Einbeziehung von hochsensiblen C-reaktivem Protein (hsCRP) in den Variablensatz ergab sich eine bemerkenswerte Clusterstabilität, wobei alle Jaccard-Indizes 0,75 überstiegen. Polygene Risikoscores (PRS) und Diabeteskomplikationen wurden in den neuen drei Clustern als Manifestationen inhärenter Heterogenität abgegrenzt. Die drei de-novo-abgeleiteten Cluster (n = 96, 172 bzw. 33) erfassten effektiv die Heterogenität innerhalb der Stichprobe und zeigten unterschiedliche Verteilungen von PRS- und Diabetes-Komplikationen, d. h. Cluster 1 war durch Insulinresistenz mit hoher Neuropathie-Prävalenz gekennzeichnet Cluster 2 wurde als altersbedingter Diabetes mit höherer Prävalenz von Schlaganfällen und CKD definiert, und Cluster 3 wies die höchste genetische Veranlagung und das höchste Risiko für Diabetes im Zusammenhang mit Fettleibigkeit auf. Unsere Ergebnisse deuten darauf hin, dass die T2D-Subphänotypisierung auf der Grundlage der einzigartigen klinischen Merkmale der Probe zu einer stabilen Kategorisierung führt und die T2D-Heterogenität effektiv erfasst, und dass dieser Ansatz die Weiterentwicklung der personalisierten Behandlung von Diabetes erleichtern könnte.

Zusammenfassend zeigt diese Arbeit, dass Metabolitprofile bei der Früherkennung von Diabetes helfen können, um die Pathologien der T2D-Progression zu verstehen, während die Subtypisierung von T2D dabei hilft, die zugrunde liegende Heterogenität aufzuklären und um

potenziell personalisierte Therapien zu entwickeln. Somit konnten wichtige Beiträge zur Präzisionsmedizin geleistet werden.

4. Introduction

4.1 Type 2 diabetes

World Health Organization (WHO) manifests that the global number of people with diabetes rose to 422 million in 2014, and it is estimated to impact approximately 700 million adults by 2050 [1]. Diabetes mellitus (DM) is a metabolic condition marked by abnormally high blood glucose levels due to the body's impaired ability to utilize insulin effectively. Over time, disruptions in insulin secretion or insulin function, and eventually both, can result in disturbances in carbohydrate, protein, and lipid metabolism. DM is a multifactorial condition encompassing type 1 diabetes (T1D), type 2 diabetes (T2D), gestational diabetes, maturity-onset diabetes of the young (MODY), and so on [2]. Here we mainly focus on prediabetes and T2D in our research.

According to WHO, T2D is diagnosed when fasting plasma glucose (FPG) $\geqslant$ 126 mg/dL (7.0 mmol/L), or when 2-h plasma glucose (2h-PG) $\geqslant$ 200 mg/dL (11.1 mmol/L). These could be measured by an oral glucose tolerance test (OGTT). American Diabetes Association (ADA) aligns with the WHO criteria but adds two more measurements, hemoglobin A1c (HbA1c) $\geqslant$ 6.5% (48 mmol/mol) and a random plasma glucose $\geqslant$ 200 mg/dL (11.1 mmol/L) can be also considered as a diagnostic tool [3, 4].

More than 95% of individuals with diabetes have T2D. It was initially observed primarily in adults but is now increasingly prevalent in children. T2D can cause progressive damage to the body, particularly affecting nerves and blood vessels. Prevention is often possible, with contributing factors including overweight, physical inactivity, and genetic predisposition. Timely diagnosis is essential to reduce the severe consequences of T2D. The most effective approach to identify diabetes is through routine examinations and blood tests performed by a medical professional. Symptoms of T2D can be subtle and may take several years to manifest, leading to delayed diagnosis and the onset of complications [5].

4.1.1 Risk factors

Several factors are considered to increase the risk of T2D, including obesity, fat distribution, physical inactivity, age, family history, prediabetes and race/ethnicity. Being overweight and prediabetes are key contributing factors [6, 7].

The occurrence of overweight and obesity is on a continuous rise among both adults and children. From 1975 to 2016, the global occurrence of overweight or obese children and adolescents rose from 4% to 18% [8]. Obesity significantly contributes to the onset of insulin resistance and diabetes [9], but the molecular pathway remains not fully understood. Prediabetes denotes a state where blood sugar levels exceed normal ranges but fall below the threshold for diabetes diagnosis. If not addressed, prediabetes frequently advances to type 2 diabetes [7] as shown in the KORA cohort which is a longitudinal population-based adult cohort study [10] and IMI-DIRECT cohort which is a longitudinal population-stratified adult cohort study [11].

4.1.1.1 Obesity and T2D

Excess body weight and obesity pose significant risks for the development of T2D [6]. T2D results from insulin resistance across multiple organs, coupled with insulin secretory decrease in β-cells [12]. The global surge in obesity prevalence is believed to contribute to the recent uptick in T2D

cases, as obesity exerts influence on both insulin action and β-cell function [6]. Elevated plasma free fatty acid (FFA) concentrations linked to obesity and T2D may adversely affect β-cells. Under basal conditions, circulating FFAs contribute to approximately 30% of insulin secretion, irrespective of diabetes status [13]. It is plausible that the interplay of elevated plasma FFAs, triglycerides, and glucose—referred to as "glucolipotoxicity" [14], combined with excessive plasma amino acids, can lead to β-cell dysfunction and death [15]. Research conducted in mouse models has revealed that adipose tissue generates proinflammatory cytokines, leading to insulin resistance [16]. Moreover, the observation that obesity in humans is characterized by increased infiltration of macrophages into adipose tissue [17] suggests that adipose inflammation is a key contributor to insulin resistance among individuals with obesity. Both animal and human studies provide evidence that alterations in adipose tissue metabolism and inflammation, induced by obesity, play a crucial role in regulating metabolic functions in other organs and contribute to β-cell dysfunction [17].

Lifestyle changes such as a modest weight loss (5-10% of baseline weight), combined with at least two and half hours of physical activity per week led to a decrease in diabetes incidence by over 50% [18]. This suggests that weight loss is an effective therapeutic strategy for managing T2D.

4.1.1.2 Prediabetes

Prediabetes is designated for an intermediate stage of hyperglycemia in which glycemic parameters are above normal but do not meet the threshold criteria for diabetes. Thus, predisposing the individuals to an elevated risk of T2D and cardiovascular disease onset [3, 19]. Prediabetes is commonly linked to conditions such as obesity, metabolic syndrome, dyslipidemia and hypertension [20].

WHO and ADA have different criteria for prediabetes. WHO defines impaired fasting glucose (IFG) as FPG of 110 to 125 mg/dL (6.1 to 6.9 mmol/L). Impaired glucose tolerance (IGT) is then defined as 2-hour plasma glucose level of 140-200 mg/dL (7.8 to 11.0 mmol/L) after 75 g of oral glucose. It can be also a combined approach of the two previously mentioned based on a 2 h OGTT [4]. Whereas ADA classifies IFG as 100-125 mg/dL (5.6 to 6.9 mmol/L) and IGT as 140-199 mg/dL (7.8 to 11.0 mmol/L). Furthermore, ADA also advocates for the inclusion of HbA1c in diabetes diagnostics and considers a level of 5.7% to 6.4% as a critical criterion. Employing HbA1c as a diagnostic factor, it tends to yield higher percentages of prediabetes compared to FPG [3, 21]. HbA1c reflects an average blood sugar level and was initially perceived as a more representative indicator of hyperglycemia [22, 23].

In the United States, approximately 10% of individuals with prediabetes progress to diabetes annually [7]. Rigorous lifestyle modifications, encompassing weight loss, enhanced physical activity, regular self-monitoring, led to a significant decrease in diabetes incidence over a 3-year timeframe [7]. Dietary patterns were suggested to be a very important factor in prevention of prediabetes and T2D [24].

4.1.2 Complications

Many essential organs, including the heart, nerves, eyes, and kidneys, are impacted by T2D [25]. Additionally, factors that elevate the risk of diabetes also predispose individuals to other severe ailments. Proper diabetes management and blood sugar control can lower the risk of complications and related health conditions.

The majority of morbidity associated with T2D stems from cardiovascular diseases, which may include coronary artery disease, heart failure, and stroke [26]. 20-30% of individuals experiencing acute coronary syndromes have T2D, with an additional 40% exhibiting impaired glucose tolerance [27]. Research indicates that mortality rates after acute myocardial infarction are roughly double for patients with diabetes compared to those without [27]. Diabetes is a well-established risk factor for stroke, primarily due to its ability to induce pathological changes in blood vessels throughout the body. When cerebral vessels are influenced, diabetes can directly lead to stroke. Besides, Diabetic stroke patients tend to have worse post-stroke recovery outcomes and elevated mortality rates [28]. Diabetic kidney disease is also prevalent. Approximately half of individuals diagnosed with T2D will eventually experience kidney disease. Poorly managed diabetes can harm the blood vessels responsible for filtering waste in the kidneys, resulting in kidney damage and contributing to elevated blood pressure [29]. Another common chronic complication is diabetic neuropathy, which manifests with sensory disturbances, muscle atrophy, difficulty walking, susceptibility to wounds, and severe pain in the lower limbs. Additionally, it can lead to symptoms such as pulmonary dysfunction, tachycardia, orthostatic hypotension, urinary incontinence, indigestion, nausea, and fluctuations between diarrhea and constipation [26, 30].

4.2 Precision medicine and treatment in diabetes

ADA and the European Association for the Study of Diabetes (EASD) have jointly issued consensus report based on expert opinion on precision medicine in diabetes [31]. "A strategy to enhance diabetes diagnosis, prediction, prevention, or treatment of diabetes by integrating multidimensional data, considering individual variations" is how the text defines precision diabetes medicine [31]. In diabetes precision medicine, the distinctive genetic profile of an individual, along with environmental or contextual information from clinical records and other 'omics data, is employed. This approach enables a comprehensive understanding of individual traits, variations, circumstances, and preferences [32].

Presently, the primary hurdle lies in the fact that all precision medicine approaches necessitate the generation, storage, and comprehension of extensive datasets, encompassing not only genetic information but also various OMICs levels. The challenge is to process and translate these datasets into clinically relevant applications [31]. Wang et al. 2012 used a metabolomic approach which identifed metabolites like glycine and lysophosphatidylcholine to be predictors for prediabetes and T2D [33]. Tulipani et al. 2016 showed metabolic traits such as glutamate, glycine and BCAA serve as biomarkers of obesity and are associated with an increased risk of diabetes development in people with prediabetes [34]. This study uncovered individualized molecular markers of early T2D onset. In the case of individuals with T2D, clinical phenotypes such as glutamate decarboxylase antibodies, age at diagnosis, body mass index (BMI), HbA1c, and homeostatic model assessment can be employed to classify patients into four T2D categories such as age or BMI related diabetes [35]. These identified clinical features hold the potential to predict patients who exhibit favorable responses to some glucose-lowering medications like dipeptidyl peptidase 4 inhibitors (DPP4i), as well as those who may respond less effectively or experience adverse outcomes [36, 37]. The reduced costs associated with genotyping panels, genome-wide association studies (GWAS), and polygenic risk scores offer novel insights into the prevention, treatment, and prospective clinical applicability of these findings [31, 38]. Additionally, intervention trials indicate that tailoring diet or increasing physical activity remains effective in preventing diabetes, irrespective of the underlying genetic risk [39, 40]. These studies suggest that multi-omics profiles combined with clinical features have capacity to inform precision medicine in diabetes.

4.3 Assessment on the molecular level in Diabetes and Obesity research

4.3.1 Metabolomics

Metabolic variations are a commonality in both obesity and T2D, and these conditions frequently coexist due to shared causes and interrelated factors [41, 42]. The field of metabolomics has been proven valuable in delineating human metabolism [43]. Numerous investigations have established connections between the plasma metabolome and obesity [44, 45], revealing potential metabolic dysregulation associated with obesity. These metabolites linked to obesity encompass amino acids and their catabolic products, lipids, and nucleotides. Several of these plasma metabolites have exhibited correlations with the incidence of type 2 diabetes [46-49] and cardiovascular disease [50-52] in various prospective studies. This suggests that these metabolites hold the potential to enhance the characterization of obesity beyond traditional anthropometric measures. Several inquiries have established connections between phenotype and metabolism, pinpointing serum or plasma metabolic markers independently linked to the development of both obesity and diabetes [53-56]. Numerous metabolites have been identified as associated with both diabetes and overweight or obesity. Within the domain of amino acid metabolism, increased levels are observed for BCAAs, cysteine, glutamine, phenylalanine, proline, while decreased levels are noted for asparagine, glycine and citrulline in both obesity and diabetes [57]. The diminished expression of mitochondrial branched-chain amino transferase is regarded as a contributing factor to the heightened concentrations of BCAAs in obesity, a condition linked with decreased serum insulin levels [58]. Metabolomics has been proven to be instrumental in clarifying how metabolites work in governing system, particularly in relation to obesity and diabetes.

Additionally, connecting metabolites with various omics, particularly genetics through genome-wide association (mGWAS), provides insights into the genetic impact on the metabolic compositions [59-61]. As mGWAS continues to expand its sample size and delve into more intricate metabolic traits, it facilitates a more holistic and systematic downstream analysis.

4.3.2 Genomics

Examining the genetic makeup enables the prediction of disease susceptibility. The elevated rates of obesity and the resulting clinical implications T2D, underscore the significant role played by environmental factors and their interplay with genetic variations in the development of diseases [62]. More than 900 independent Single Nucleotide Polymorphisms (SNPs) linked to BMI [63] and more than 230 loci impacting the risk of T2D are shown by GWAS [64].

Obesity is a multifaceted condition influenced by both genetic and environmental factors. Evidence from twin and family studies has highlighted the significant role of genetic elements in obesity, particularly in cases where there is a positive family history of obesity, leading to an elevated risk of childhood obesity. The concordance rate for obesity is notably higher in monozygotic twins compared to dizygotic twins. Twin studies estimate the heritability of obesity to be between 40% and 75%, underscoring the substantial genetic component contributing to obesity [65]. Recent progress in genetic testing has allowed for the discovery of genes associated with obesity. Genes such as LEP, LEPR, POMC, PCSK1, MC4R, SIM1, BDNF, and NTRK2 have been identified as causative factors for obesity. Next-generation sequencing (NGS) is becoming

increasingly prevalent and proving to be a valuable tool in clinical settings for identifying candidate genes linked to obesity [66].

T2D is recognized as an intricate, polygenic disorder. Similarly, individuals with one parent affected by type 2 diabetes face an elevated risk of developing the condition, reaching nearly 40%. Furthermore, if both parents are affected, the risk rises substantially, reaching up to 70% [67]. The exploration of the genetic foundation of T2D was constrained to linkage studies and candidate gene approaches, and through these investigations, rare familial types of T2D, along with genes linked to common forms of T2D (such as PPARG, KCNJ11, and TCF7L2), were identified [68]. Conventionally, genetic studies of these genes concentrate on genomic regions with substantial genetic impacts and already recognized disease pathways. GWAS facilitates significant advancements in genetic investigation of complex disorders by pinpointing novel genes implicated in the pathogenesis of diseases. This approach revealed a substantial subset of newly identified genes, including KLF14, ENPP1, ADIPOQ, IRS, GCKR, SREBF1, JAZF1, SCL30A8, TCF7L2, and others [67, 69]. Lately, there has been a surge in the global pursuit of understanding the genetic landscape of susceptibility and the etiological architecture of T2D through an increasing number of GWAS, meta-analysis studies, investigations into rare, structural, and protective variants, as well as sequencing in familial contexts [67].

4.3.3 Cluster analysis

Cluster analysis refers to a range of mathematical techniques that explore relationships among points in multi-dimensional space. These methods primarily rely on assessing similarities or dissimilarities (or proximities) between entities within the dataset [70-72]. As an example of biological application of cluster analysis we can point out is Ahlqvist et al. [35] adopted data-driven cluster analysis (k-means and hierarchical clustering) to identify subgroups of diabetes using six diabetes-related variables to categorize diabetic individuals into five clusters in which four of these clusters predominantly represent subgroups of T2D while one cluster is associated with severe autoimmune diabetes (SAID), primarily corresponding to T1D. This study presented a novel and precise diabetes classification and provided a crucial advancement towards precision medicine in diabetes.

The challenge in cluster analysis, also known as unsupervised classification, lies in determining the most effective way to divide data points. Deciding the optimal clustering method for a given dataset is crucial to attain precise and desired results, all while taking into consideration the inherent nature of the data [70]. K-means clustering is a data-driving method. It facilitates the classification of data by distributing or grouping data points into K clusters based on distinctive features. In data-based clusters, also known as center-based clusters, each data object within a cluster is closer to the center of that cluster than to the center of any others [73, 74]. Initially, the K-means algorithm establishes the initial K centroids, either by computing means from random subsets of the dataset or by selecting the first K elements. In each iteration, the algorithm assigns each data point to its nearest centroid and subsequently updates the centroids after forming the K clusters [72-74]. To assign a point to the closest centroid, a proximity measure is required to quantify the distance between data points. This involves an algorithm that consistently calculates the similarity of each point to every centroid [72].

Determining the optimal number of clusters is one of the most challenging and debated aspects when employing the K-means data clustering algorithm. The final selection or justification of a specific K value is notably dependent on the particular analysis or experiment. Moreover, K values

can be adjusted or modified to enhance and achieve more suitable results based on the characteristics of the data and the objectives of the analysis [75]. Numerous methods are available for selecting the optimal K value, including the Elbow method, Gap statistic algorithm, Silhouette coefficient algorithm, Canopy algorithm, NBClust, and others [75]. The Elbow method involves using the cluster centroid and the squared distance between entities within each cluster to generate a series of K values. The number of clusters (K) is then visualized by plotting the explained variation as a function [76]. On the other hand, the Silhouette method assesses how well an observation aligns with its own cluster, with a high value indicating a well-matched observation to its own cluster and a poor match to neighboring clusters [76]. This method is commonly used in many studies [35, 77, 78].

4.4 Inadequate early detection and systematic biological understanding of T2D

The transition from normal or impaired fasting glucose to T2D typically occurs gradually, and notably, its symptoms may go unnoticed for extended periods. Delays in diagnosis significantly contribute to inadequate management and an increased risk of complications [79]. Given the rising global prevalence and burden of diabetes, early identification of predisposition to T2D can significantly enhance opportunities for preventing and managing this condition effectively. Presently, elevated blood glucose and glycated hemoglobin are primarily employed for diabetes diagnosis. Nevertheless, it's vital to acknowledge that elevated blood sugar reflects a continuous and evolving process. These constraints could lead to misclassification and misdiagnosis [80]. One study [81] discovered that the rise in fasting glucose levels, along with higher BMI, blood pressure, lipids like triglycerides and lower HDL-cholesterol (HDL-C), are higher risk of developing diabetes.

On the other hand, personalized progression of hyperglycemia and subsequent diabetes complications vary widely because of various underlying biological causes [82]. Traditionally, diabetes is mainly classified into type 1 (T1D) and T2D, primarily determined by the presence (T1D) or absence (T2D) of autoantibodies. So current treatment guidelines addressing poor metabolic control once it has occurred but lack the ability to identify which patients would require more intensive care, limits their applicability.

These findings underscore the need for further research to establish optimal diagnostic criteria and diabetes subphenotyping strategies. Approaches such as deep molecular phenotyping hold promise for uncovering the underlying mechanisms of glucose dysregulation and informing more personalized, precise prevention and treatment strategies.

4.5 Aims of this thesis

The main objectives of this thesis are to enhance early detection and explore the role of molecular biomarkers and T2D heterogeneity using multidimensional data. This approach aligns with the concept of precision medicine in diabetes and allows for a comprehensive understanding of individual traits, variations, situations, and inclinations. Three studies were performed to answer the specific research questions:

To 1) use the targeted metabolomic profiles from the German KORA FF4 cohort to examine metabolite signatures associated with obesity and T2D involved in the development of obesity related T2D, to improve the early detection of incident of T2D. This corresponds to paper I [83].

2) Identify metabolites from both targeted metabolomics and untargeted metabolomics in IMI-DIRECT cohort linked with glycemic deterioration, from NGR to prediabetes and eventually to T2D, to further optimize the early diagnosis and understand the progression of T2D. This corresponds to paper II [84].

3) use a data-driven cluster analysis with clinical characteristics in KORA FF4 to classify T2D into different subphenotypes and provide a more accurate, clinically relevant stratification, marking a significant advancement toward precision treatment in diabetes. This corresponds to paper III [85].

5. Methods

5.1 Study population

KORA is a population-based study that comprises several deeply phenotyped epidemiological surveys in the Augsburg region of southern Germany. The study design, sampling method and data collection have been described elsewhere [86]. The baseline KORA S4 involved the examination of 4,261 individuals between 1999 and 2001. The first follow-up F4 comprised of 3,080 individuals between 2006 and 2008 and the second follow-up FF4 examined 2,269 participants between 2013 and 2014. **Figure 1** provides an overview of the KORA study. Written informed consent was received from all study participants. T2D was diagnosed based on OGTT according to WHO criteria or clinically diagnosed by physician. The estimated glomerular filtration rate (eGFR) calculation and chronic kidney disease (CKD) definition was described in detail here [83, 87]. Information regarding parental history of diabetes, myocardial infarction, stroke, neuropathy, and metformin intake or any antidiabetic medication intake was self-reported.

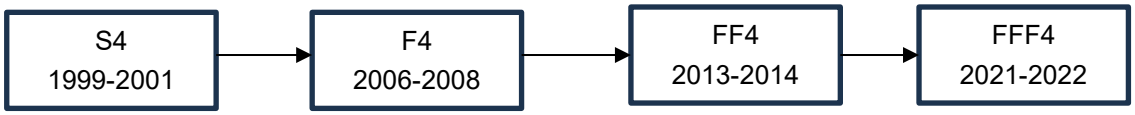


Figure 1. Overview of the timeline of KORA study for baseline surveys and follow-up examinations.

IMI-DIRECT is a population-stratified cohort study and multicenter, involved in diabetes research. Baseline clinical variables and omics were measured, follow-up studies are carried out at different intervals such as 18 month, 36 months and 48 months based on the feedback biomarkers to improve the study design. **Figure 2** provides an overview of the IMI-DIRECT study. Detailed characteristics of this cohort as well as inclusion/exclusion criteria were described somewhere else [11, 88]. NGR and IGR were defined based on ADA 2011 diagnostic criteria using HbA1c, fasting glucose and 2 h glucose. Prevalent T2D was identified through clinically diagnosis or the ADA 2011 criteria. Participants displayed NGR and prevalent T2D according to the glycaemic measures in cohort 1 were also included for further analyses (**Figure 2**).

Paper I and Paper III were conducted on the KORA FF4 cohort. Paper I was a cross sectional study analyzing 1715 participants with complete metabolomics data, consisting of 1276 non-obese participants and 439 obese ones based on BMI or 1415 non-diabetic individuals and 300 T2D ones based on OGTT or clinical diagnosis.

Paper III only included 301 participants with T2D for cluster analysis. Besides, PRS was calculated for all individuals with T2D (N=301) and all individuals without T2D but with available genetic data as the control group (N=1357) for comparison.

Paper II included a cross-sectional design for analyzing DIRECT data at baseline and a longitudinal design of follow-up 18 and 48 months. 3000 individuals with complete metabolomic measurements at baseline were included in our study and ones had complete follow-up information for further longitudinal analysis.

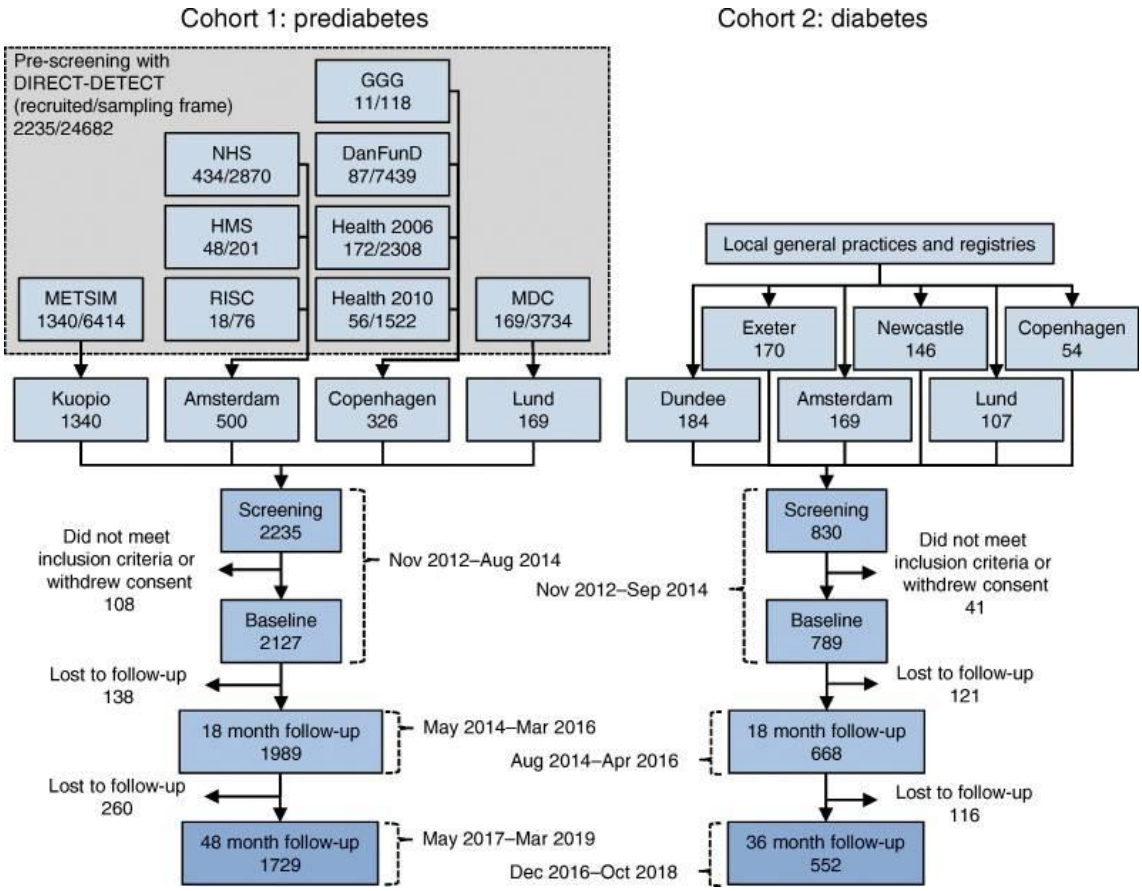


Figure 2. Overview of the baseline surveys and follow-up examinations of IMI-DIRECT study [11].

5.2 Metabolite quantification and normalization

Blood samples from KORA FF4 participants were measured with the AbsoluteIDQTM p180 Kit. The details were previously described here [89]. Finally, 146 metabolites were passed quality control (QC) for further analysis in this study in paper I.

Blood sample from IMI-DIRECT individuals at baseline were measured with the AbsoluteIDQTM p150 Kit (targeted metabolomics) and metabolon platform (untargeted metabolomics) respectively. 132 metabolites from targeted metabolomics and 779 ones from untargeted metabolomics were included in this study after QC in paper II. The measurement details were stated here [90].

The concentrations of all metabolites were natural log transformed and scaled (mean = 0, sd = 1) before statistical analysis to ensure comparability across the metabolites.

5.3 Genomics

Genotyping of KORA was performed by the Affymetrix Axiom Array. More than 500,000 autosomal SNPs thorough quality control were subsequently imputed based on the haplotype reference consortium (HRC) reference panel using Impute2 v2.3.2. Variants with certainty < 0.95, information metric < 0.7, low genotyping calls (geno 0.03), and Hardy–Weinberg equilibrium exact test p-value < 5x10-10 were excluded. Finally, 7,753,540 variants with a minor allele frequency (MAF) > 1% were kept for the analysis in the study.

5.4 Statistical analysis

Paper I

Firstly, I explored the candidate biomarkers associated with BMI and T2D through multivariable linear regression and logistic regression with different covariates. Sobel mediation test [91, 92] was performed to identify the metabolites which convey the effect of BMI on T2D. Lastly I performed two sample mendelian randomization (2SMR) to check the causal inference of BMI, identified metabolites and T2D.

SNPs and genes in humans were searched to find the association with three identified lipids (SM C16:1, SM C18:1, and PC aa C38:3) to comprehend the biological pathway. Genotype-Tissue Expression (GTEx) human database and the Mouse Genome Informatics (MGI) database were used to investigate the tissue-specific role of the related genes CERS4, PDXDC1 and FADS1-3. Mouse Genome Database [93] was used to identify gene expression correlations with relevant obesity and T2D traits in mice adipose, liver, muscle and brain tissue: an F2 cross of the inbred ApoE−/− C57BL/6J and C3H/HeJ strains [94].

Paper II

Estimates from multivariable logistic regression analysis were obtained using the concentration of each transformed metabolite as an independent variable and the glycemic status value (NGR vs. IGR, NGR vs. T2D, or IGR vs. T2D) as a dependent variable. A multivariable linear regression model was employed to examine metabolites associated with HbA1c progression rate.

Candidate biomarkers associated with prediabetes/diabetes as well as HbA1c progression rate were used for further analysis. Incident prediabetes or diabetes was carried out with the identified metabolites as exposure and glycemic status at follow-up timeline as outcome. A causal mediation test was performed to assess the mediation effect of the above identified metabolites. Here, baseline glycemic status was considered as an independent variable, the metabolites entered the model as a mediator whereas follow-up glycemic status was conducted as the dependent variable.

Causal inference used 2SMR methods and inverse-variance-weighted (IVW) method to explore the causal effect of T2D on metabolites and vice-versa the causal effect of metabolite on T2D by wald ratio.

Paper III

301 individuals with T2D from the KORA FF4 study were used for cluster analysis. For the replication of original clusters to KORA cohort, I used the same variables as ANDIS's cohort: age, BMI, HbA1c, homeostasis model assessment (HOMA) estimates of beta-cell function (HOMA2-B) and insulin resistance (HOMA2-IR) in KORA FF4. I employed k=4 since we did not include T1D in our study, with three different hyperparameter combinations: 1) scaling and centroids from the original ANDIS cohort, 2) scaling from KORA, but centroids from ANDIS, 3) both scaling and centroids from KORA. Each individual from KORA was assigned to one of the original clusters based on the smallest Euclidean distance to the respective cluster center. Transitions of individuals between clusters based on the different approaches were visualized by Sankey diagrams.

De-*novo* clusters were derived by a two-step approach, of which the optimal number of clusters k was determined based on silhouette width and elbow method, followed by k-means. The basic variable set: age at examination, BMI, HbA1C, HOMA2-B and HOMA2-IR. Additional variables hs-CRP, triglycerides (TG), HDL-cholesterol (HDL-C) and systolic blood pressure (SBP) were added individually to the basic set of variables to explore their impact on cluster performance.

The stability of de-*novo* clusters was assessed by assignment congruence over different variable sets and Jaccard indices. Sensitivity analysis was done for the final clusters by recalculating separately for men and women to assess potential sex differences.

PRS was calculated with driving the list of variants of the optimal T2D score from Polygenic Score (PGS) Catalog [95] as score number “PGS000014” and summing up the product of the dosage of risk allele multiplying with their respective weights in KORA. PRS was evaluated for both individuals with and without T2D in KORA FF4.

Differences of PRS in individuals with and without diabetes were quantified by logistic regression models with PRS as the exposure variable. One-way ANOVA and t-tests were employed to examine the disparities in PRS and frequency of risk alleles of the top significant 20 genetic variants across clusters. Differences in diabetes-related complications and parental history were assessed using Fisher’s exact test.

6. Results

Paper I

Obesity sets off a series of metabolic reactions that increase the risk of several accompanying conditions such as insulin resistance. In this study we addressed the first aim of this thesis and aimed to find the metabolites implicated in the progression from obesity to T2D.

We identified 42 metabolites associated with BMI independently, phosphatidylcholine diacyl (PC aa) C38:3, glutameta (Glu), sphingomyelin (SM) C16:1, SM C18:1, lysophosphatidylcholine acyl (lysoPC a) C17:0 and lysoPC a C18:2 were the six strongest ones. 3 metabolites were significantly linked to T2D, hydroxybutyrylcarnitine (C3-DC (C4-OH)), Alpha-Amino acid (alpha-AAA) and isoleucine (Ile). In the Sobel mediation test, three metabolites SM C16:1, SM C18:1, and PC aa C38:3 showed significant mediation effect of BMI on fasting glucose or HbA1c, suggesting these metabolites increased the risk of glycemic deterioration. The causality directions of BMI, three identified lipids and T2D were confirmed by 2sMR, we observed that BMI was a causal factor for the change of SM C16:1 and SM C18:1, while SM C16:1, SM C18:1, and PC aa C38:3 were causal factors for T2D incidence. These results suggest sphingomyelins and phosphatidylcholines could serve as a molecular mediator in the development of obesity related T2D.

Incorporating these three lipids with human genetics, we were informed that these three metabolites were linked to SNPs at the CERS4, PDXDC1 and FADS1-3 locus [96, 97]. CERS4 and FADS1-3 were found to impact the sphingolipids biosynthesis, impairing insulin sensitivity and pancreatic beta-cell function [98, 99]. PDXDC1 and FADS2 upregulate phosphatidylcholine, which suppresses key genes like IRS-2, disrupting insulin signaling [100].

Paper II

This paper addressed the second aim of this thesis and identified biomarkers assisting in categorizing individuals based on their glycemic deterioration, thereby contributing to aiding in the insight into the disease progression.

In the targeted assay we observed 4, 34 and 50 metabolites to be significantly different between NGR-IGR, NGR-T2D and IGR-T2D groups. While in the untargeted metabolomics panel, there were 15, 99 and 108 metabolites significantly having variations in each group. Significant metabolites were mainly BCAAs, also derived BCAAs, lipids, xenobiotics, and a few unknowns.

Metabolites from targeted metabolomics including lysoPC a C17:0 and sum of hexoses, and from untargeted metabolomics including N-lactoylvaline and N-lactoylleucine, formiminoglutamate, as well as carbohydrate lactate, and an unknown metabolite (X-24295) were linked to HbA1c progression rate as well as incidence of prediabetes/diabetes. In the causal mediation test, we also observed that these metabolites significantly mediated the glucose deterioration from baseline to follow-ups.

We utilized 2SMR to assess the causal directions between metabolites and T2D. Our analysis revealed that T2D causally affects the concentrations of three metabolites (hexose, glutamate and caproate (FA 6:0)). Additionally, four phosphatidylcholines such as PC aa C36:2 along with two omega-3 fatty acids stearidonate (FA18:4) and docosapentaenoate (n3 DPA; FA22:5) potentially contribute to the development of T2D.

Paper III

Data-driven clustering holds the potential to uncover the pathophysiology of glucose deterioration and the onset of comorbidities in individuals with T2D. This paper addressed the third aim of this thesis by a comprehensive statistical assessment of T2D subphenotyping in KORA FF4 cohort.

Participants in KORA were assigned to the corresponding clusters based on ANDIS's scaling and centroids parameters, it was observed that the relative cluster sizes in KORA differed from those observed in the ANDIS study. Severe insulin-deficient diabetes (SIDD) in KORA only collected 2% of the T2D cases compared to 17.5% in ANDIS. Over 80% of participants in KORA were classified into the MARD cluster, compared to approximately 40% in ANDIS. When clinical variables were scaled based on own scaling parameters, all these variables showed the same trend in KORA and ANDIS. The relative cluster sizes in KORA were similar to those observed in the ANDIS study, for example mild age-related diabetes (MARD) had the most participants for both KORA (46.8%) and ANDIS (39.1%), and 15.3% of individuals in KORA were allocated to SIDD which was similar to the ANDIS study (17.5%). Besides, 65% of participants were assigned to the same clusters when using ANDIS centroids, but either ANDIS scaling or KORA scaling. When we employed KORA's own scaling and centroids (using k-means, forced k=4), a novel distinct cluster was observed with the overall most modest metabolic impairments and low BMI. Only 45% of the participants had consistent cluster assignments between using ANDIS centroids and KORA centroids but same KORA scaling. Collectively, these findings imply that the original ANDIS clusters may not entirely capture the characteristics of the KORA sample.

De novo cluster analysis was derived in KORA, k=3 was determined to be the optimal number of clusters by silhouette width and the elbow plot. KORA participants were categorized into 3 groups by k-means with the basic variables. Cluster 1 (n=96, 31.9%) was characterized by hyperinsulinemia and insulin resistance, resembling the severe insulin resistant diabetes (SIRD) cluster identified in the ANDIS cohort; Cluster 2 (n=172, 57.1%) had older age, low BMI and low insulin resistance, akin to the mild age-related diabetes (MARD) cluster in the ANDIS cohort; Cluster 3 (n=33, 11.0%) showed insulin deficiency (low HOMA2-B), high BMI and poor glycemic control (high HbA1c), representing a distinct cluster not observed in the ANDIS cohort. When we added additional variables hsCRP, TG, HDL-C, or SBP, respectively, it was found 90%, 93%, 90% and 98% of participants were assigned to the same cluster compared to when using basic variables. To explore the influence of systemic inflammation distinguishing diabetes subtypes, we determined the clusters based on the variable set of age, BMI, HbA1c, HOMA2-B, and HOMA2-IR plus hsCRP as the final subphenotypes. The final clusters did not show substantial sex-specific effects as the majority of individuals (95%) were grouped to the same cluster for both men and women as in the initial data analysis. Moreover, the Jaccard indices of all final clusters were above 0.75, indicating reasonably high cluster stability.

The heterogeneity of the final clusters was assessed by PRS and diabetic complications. Cluster 1 had relatively lower genetic risk with no significant difference from the control group, but high prevalence of neuropathy. Cluster 2 had the middle genetic risk, relatively higher than the control group (but not significantly different from Cluster 1), and a higher prevalence of stroke and CKD. Cluster 3 showed a significantly higher PRS than both control group and Cluster 1, with more frequently a positive parental history of diabetes.

7. Discussion

7.1 Early progression of T2D

T2D is a chronic, metabolic disease, thus exploring the role of intermediate molecules may uncover new therapeutic targets for addressing early-stage T2D pathophysiology. Plasma metabolites circulate throughout the entire body and play a direct role in the molecular regulation of complex diseases, such as obesity, prediabetes, diabetes [57]. Assessing these metabolites offers molecular insights into their involvement in biological processes triggered by disease progression.

In Paper I, we analyzed KORA targeted metabolomics profiles to identify underlying links to metabolic pathways. T2D typically develops in the advanced stages of obesity, it is confirmed that SM C16:1, SM C18:1 and PC aa C38:3 were significantly correlated with obesity in our study. These metabolites were also reported to be strongly linked to T2D [101]. In the mediation test, these three lipids significantly conveyed the influence of BMI on fasting glucose/HbA1c. Integration with mendelian randomization indicates the direction of causality, suggesting that lipids such as SM C16:1 and PC aa C38:3 could serve as molecular mediators which contribute to the development of T2D. This metabolic process linking obesity and diabetes may be driven by modulation of inflammation through fatty acid (FA) and proinflammatory cytokines. Obese individuals are usually characterized with these two elevated elements which are known to activate sphingomyelinase (SMase) and converting sphingomyelins to ceramide, thus exerting an action of insulin resistance [102, 103]. High-fat diets, which lead to elevated FA and excessive of PC, contribute to obesity and diabetes in individuals [104]. Genetic factors may also modulate this process, offering potential avenues for intervention. FADS1-3 and PDXDC, associated with these three lipids, are revealed to be linked with polyunsaturated fatty acids (PUFAs) [105-108], which influence the biosynthesis of sphingolipid and phosphatidylcholines, modulating the risk of developing T2D [109, 110]. These findings suggest genetic predisposition and early variation in the metabolism of sphingolipids and phosphatidylcholines, which play a role in prediction the onset of T2D.

In Paper II, we reported candidate metabolites from both targeted metabolomics and untargeted metabolomics involved in the development of T2D in IMI-DIRECT study. 19 metabolites (4 from targeted and 15 from the global profiling) were significantly associated with prediabetes, which consisted of hexoses (H1), three phospholipids and five amino acids, five lipids, two carbohydrates, two unknown compounds, and one xenobiotic. Prediabetes is commonly linked to dyslipidemia, characterized by an altered lipid profile compared to individuals with normal glucose regulation [111]. In the further study we found that lysoPC a C17:0 and lactate, X-24295, formiminoglutamate, N-lactoylleucine, N-lactoylvaline were also liked with HbA1c progression and incident prediabetes/diabetes. Jenkins et al. [112] explored the source of circulating odd-chain fatty acids (C17:0, C15:0) by conducting both animal and human studies. Their findings suggested that dietary intake was associated with C15:0, whereas C17:0 was primarily synthesized by the body, indicating distinct origins and disparate roles in disease causation. One study also confirmed lysoPC a C17:0 was significantly associated with incident T2D [113], and the mechanism could be induced by FFA as it was reported inhibition of the transformation of FA palmitic acid to lysophosphatidylcholine (LPC), thereby preventing insulin resistance in mice models [114]. Causal mediation analysis further indicated the identified metabolites strongly mediates glycemic change from baseline status to follow-up. Elevated formiminoglutamate is a marker of folate deficiency, which is reported to be associated with destructed phospholipids homeostasis and a risk factor for diabetes mellitus [115]. It is also confirmed that formiminoglutamate was linked to an

increased risk of incident T2D in older Puerto Ricans [116]. N-lactoyl amino acids like N-lactoylleucine, N-lactoylvaline fused through a reaction between lactate and BCAAs are rarely reported in metabolomic datasets, a significant increase in all measured N-lactoyl amino acids was observed in T2D volunteers compared to those without T2D and metformin treatment increased their levels in TwinsUK cohort [117]. We could therefore hypothesis the mechanism maybe mediated by the medication, but the exact role in human body and pathways downstream are unclear and needs further research. A study of Swedish men proved that increased lactate concentrations in serum was independently associated with a higher incidence of T2D in a longitudinal study [118]. Lactate is generated in the cytoplasm through the glycolytic pathway under anaerobic conditions [119], other studies also showed that lactate production increases progressively during the early stages of T2D development [120, 121]. This was argued that augmented lactate levels are crucial in glucose transport and metabolism, profoundly contributing to insulin resistance [122]. Our study confirmed that lipids lysoPC a C17:0, amino acids formiminoglutamate along with novel N-lactoyl-amino acids N-lactoylleucine, N-lactoylvaline and carbohydrate lactate were significantly associated with glycemic deterioration and involvement in T2D progression.

7.2 Understanding of T2D heterogeneity

T2D is a complex, heterogeneous disease, so understanding the heterogeneity and enhancing the personalized treatment is required. The T2D classification system proposed by Ahlqvist et al. [35] has been validated in various populations and has proven to be a useful tool to further characterize potential pathophysiological pathways and diabetes progression. We firstly conducted an original cluster replication including a detailed overview of participant transitions between replicated clusters. We found that the characteristics of the KORA population were only partly reflected. Some of the inconsistencies in replicability may be due to variations in study design and participants' characteristics. We used age at examination for clustering in KORA which was significantly higher compared to the ANDIS cohort. Additionally, individuals in KORA showed better glycemic control and lower insulin resistance than those in the ANDIS sample [35], suggesting that KORA may have included a higher proportion of T2D cases with less severe disease. Besides, our HOMA models were calculated from fasting insulin rather than C-peptide which might induce differences in estimates. Our results thus highlight the value of subphenotyping by demonstrating the impact of specific study characteristics, and we add a potential new T2D subphenotype to the existing panel and we consider this finding important for personalized prevention.

De-*novo* cluster analysis was derived from KORA study, k=3 was found to be the best number of clusters rather than 4 based on silhouette and elbow plot. Mild obesity-related diabetes (MOD)-like cluster disappeared, which is consistent with the observation from Safai et al.[123]. We further assessed cluster performance and stability across different variable sets, basic variables with additionally hsCRP, HDL-C, TG, or SBP, respectively. We observed that these additional variables had minimal impact on the reassignment of individuals, with over 90% individuals remaining in their original clusters, suggesting high stability and robustness across different variable sets. One could thus hypothesize that the original variables likely encompass a significant portion of T2D heterogeneity and are sufficient for identifying clinically relevant T2D subphenotypes. In the current analysis we included hsCRP for clustering to consider the role of subclinical inflammation and assess potential differences within subphenotypes. Cluster 1 is most similar to SIRD with low hsCRP with a high proportion of newly diagnosed diabetes, Cluster 2 is most similar to MARD with the most favorable clinical characteristics. Cluster 3 is a distinct cluster, characterized by high BMI, high hsCRP and low HOMA2-B, closely resembling a typical patient encountered in

clinical practice. The mechanism could be the excess body weight triggers CRP stimulation and inflammation, which plays a role in the regulation of insulin action and insulin resistance [124, 125]. PRS and the prevalence of self-reported parental history of diabetes were both highest in Cluster 3. Besides, Cluster 3 had higher abundance of risk alleles which were mapped to locus TCF7L2, which is the most significant locus for T2D risk and the first to be consistently identified in genomic linkage studies [126]. TCF7L2 has also been reported to play essential developmental and metabolic roles in adipose tissue. The inactivation of adipocyte TCF7L2 in knockout mice promoted weight gain, and increased adipose tissue mass [127], and this phenotype was associated with adipose tissue inflammation [124, 128, 129]. So, Cluster 3 represents a T2D subphenotype associated with a greater genetic risk for both diabetes and obesity, and rigorous weight control could prove particularly beneficial to offset the higher genetic risk. Cluster 1 was characterized by a comparatively higher prevalence of neuropathy and this could be induced by insulin dysregulation which can lead to neuropathic changes in sensory neurons and the peripheral nervous system, which is significantly affected by diabetes [130]. Interestingly, Cluster 1 has significant lower HbA1c levels compared to Cluster 3 which also exhibits a relatively higher frequency of neuropathy, so it would be interesting to investigate glucose-lowering therapy in this cluster in the prevention of neuropathy. Additionally, lifestyle changes would be advantageous, including dietary adjustments to reduce calorie intake and limit high-glycemic-index carbohydrates, along with regular physical activity to boost calorie expenditure and improve insulin sensitivity in muscle tissue [131, 132]. Cluster 2 showed a relatively higher proportion of CKD cases and stroke, which could be due to the higher age in Cluster 2 as it is confirmed that age is a major risk factor for metabolic complications in T2D [133, 134]. Since the risk in Cluster 2 is primarily driven by aging—a non-modifiable factor—close monitoring of comorbidities is especially important for this group. Strict control of factors like blood pressure and renal function, potentially with medication, is recommended. Taken together, clustering based on age, BMI, HbA1c, HOMA2-B, HOMA2-ID with hsCRP provided the identification of a new distinct subphenotype with a potential genetic predisposition to obesity-induced inflammation. Therefore, this thesis demonstrates that to fully leverage the benefits of T2D subphenotyping, clustering approaches must be adapted and tailored to the respective sample, thereby enabling the development of more personalized and precise treatment strategies.

7.3 Limitation

We acknowledge several limitations in our studies. 1) Storing plasma samples for extended periods can lead to changes in metabolite concentrations [89], potentially affecting the associations. 2) Identifying metabolites, particularly in untargeted metabolomics, can be difficult due to gaps in database and the existence of unidentified or novel metabolites, often indicated by asterisk (*) in the metabolite names. and this may influence the associations; 3) The studies are mainly cross sectional or rely solely on baseline metabolites measurements, which limits their ability to distinguish between cause and consequence. Longitudinal analyses could be conducted to investigate the variation in metabolites concentrations across different stages to confirm our findings; 4) Regarding the T2D subphenotype study, the sample size of T2D is relatively small. It also had a limited sample of individuals with diabetes-related complications and family history information. Besides, we only investigated the prevalence of diabetic complications and incidence of complications need to be examined in future analysis.

8. Conclusion and Outlook

8.1 Conclusion

Amongst increasing global cases, effective adaptation programs are crucial to reducing the growing burden of diabetes-related hospitalizations and medical care. This thesis demonstrated several metabolic biomarkers like sphingolipids, phosphatidylcholines and amino acids that could contribute to prediction of T2D incident and involved in the T2D progression, which provide the clue for early detection of T2D in clinical. Moreover, this research work confirmed that T2D is heterogeneous, subphenotype classification based on the specific sample is necessary and assist in crafting personalized diabetes treatments. These research strategies outlined in this study align with the guidelines of precision medicine in diabetes are crucial to safeguard public health.

8.2 Outlook

Diabetes cannot be cured today but proactive measures can be taken for prevention and remission to alleviate the challenges of diabetes and improve the quality of life for individuals who have or are at risk of diabetes [135, 136].

Collectively, our findings provide insight into metabolic biomarkers participating in early T2D progression and disease heterogeneity based on data-driven classification, these align with the future goals of T2D prevention and treatment/remission. More studies with larger sample sizes are needed to corroborate our findings regarding the molecular biomarkers for early detection and essential classification accounting for individual differences.

Furthermore, metabolic profiles remain unclear among people with obesity who have or do not have T2D. In paper I, sphingolipids (SM C16:1, SM C18:1) and phosphatidylcholines (PC aa C38:3) were found to mediate the effect from obesity to T2D, and these three biomarkers were significantly higher in obese people but not significantly associated with T2D in the full model. From mediation test, they showed significant association with HbA1c and fasting glucose as these two are well-known clinical diagnotic factors for T2D. Hence, exploring whether molecular biomarkers can distinguish between obese individuals with and without T2D, while further exploring the role of metabolic biomarkers and the links between obesity and T2D, presents an intriguing avenue for more research.

Genetic is a well-established risk factor for T2D. A recent study demonstrated genetic contributes the heterogeneity of T2D pathophysiology by combining multi-ancestry genome-wide association study data with single-cell epigenomics [137]. Population-level genetic changes require many generations to take effect, this epidemic is primarily a result of recent environmental changes, suggesting that shifts in non-genetic factors have triggered the effects of pre-existing susceptibility genes [138]. It would be interesting to explore the potential role of the metabolites in groups with different genetic risk profiles and figure out the interaction with environment-metabolites-gene, shed more light on T2D mechanisms.

Results from a 5-year follow-up of the landmark Diabetes Remission Clinical Trial (DiRECT) study reveal that weight loss can potentially sustain remission of T2D for at least five years [139]. Remission refers to blood sugar levels can remain within a non-diabetic range long-term, without the

need for diabetes medication. In our paper I and paper II, we confirmed some metabolic biomarkers involved in T2D development, this contributes to T2D early detection and prevention. Thus, studies could target individuals already diagnosed with T2D, emphasizing the aspects of remission and treatment. A longitudinal study could be designed to investigate the correlation between BMI change and blood glucose levels. For instance, in Paper III, Cluster 3 was identified as obesity-related diabetes. It would be interesting to examine whether additional weight loss interventions could contribute to T2D remission within this group, but larger sample size would be required. On the other hand, personalized treatment is a prominent area of interest, demanding an understanding of individual trajectories of hyperglycemia and consequent diabetes complications stemming from diverse underlying biological factors. Therefore, employing effective classification methods based on multidimensional data is imperative. However, this process is not a one-time endeavor. In Paper III, we encountered challenges in replicating clusters from the ANDIS study due to differing cohort characteristics. Hence, deriving de *novo* clusters based on KORA's own clinical values proved to be more meaningful. This means another challenge of diabetes research would be the identification of suitable cohorts for translation taking into account race and ethnic differences. If this is considered, improved representation of global populations in subphenotyping studies would be covered [140].

9. References

[1] R. Franceschi, "Precision Medicine in Diabetes, Current Research and Future Perspectives," *J Pers Med,* vol. 12, no. 8, Jul 28 2022, doi: 10.3390/jpm12081233.

[2] A. Sapra and P. Bhandari, *Diabetes*. StatPearls [Internet], 2024.

[3] C. American Diabetes Association Professional Practice, "2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2022," *Diabetes Care,* vol. 45, no. Suppl 1, pp. S17-S38, Jan 1 2022, doi: 10.2337/dc22-S002.

[4] W. H. Organization, "Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia: report of a WHO/IDF consultation," 2006.

[5] "World Health Organization," *https://www.who.int/news-room/fact-sheets/detail/diabetes,* 5 April 2023 5 April 2023.

[6] S. Klein, A. Gastaldelli, H. Yki-Jarvinen, and P. E. Scherer, "Why does obesity cause diabetes?," *Cell Metab,* vol. 34, no. 1, pp. 11-20, Jan 4 2022, doi: 10.1016/j.cmet.2021.12.012.

[7] J. B. Echouffo-Tcheugui, L. Perreault, L. Ji, and S. Dagogo-Jack, "Diagnosis and Management of Prediabetes: A Review," *JAMA,* vol. 329, no. 14, pp. 1206-1216, Apr 11 2023, doi: 10.1001/jama.2023.4063.

[8] "World Health Organization," *https://www.who.int/health-topics/obesity#tab=tab_1*.

[9] D. M. Muoio and C. B. Newgard, "Obesity-related derangements in metabolic regulation," *Annu Rev Biochem,* vol. 75, pp. 367-401, 2006, doi: 10.1146/annurev.biochem.75.103004.142512.

[10] H. Luo *et al.*, "Associations of plasma proteomics with type 2 diabetes and related traits: results from the longitudinal KORA S4/F4/FF4 Study," *Diabetologia,* vol. 66, no. 9, pp. 1655-1668, Sep 2023, doi: 10.1007/s00125-023-05943-2.

[11] R. W. Koivula *et al.*, "Discovery of biomarkers for glycaemic deterioration before and after the onset of type 2 diabetes: descriptive characteristics of the epidemiological studies within the IMI DIRECT Consortium," *Diabetologia,* vol. 62, no. 9, pp. 1601-1615, Sep 2019, doi: 10.1007/s00125-019-4906-1.

[12] C. Bogardus and P. A. Tataranni, "Reduced early insulin secretion in the etiology of type 2 diabetes mellitus in Pima Indians," *Diabetes,* vol. 51 Suppl 1, pp. S262-4, Feb 2002, doi: 10.2337/diabetes.51.2007.s262.

[13] G. Boden, X. Chen, and N. Iqbal, "Acute lowering of plasma fatty acids lowers basal insulin secretion in diabetic and nondiabetic subjects," *Diabetes,* vol. 47, no. 10, pp. 1609-12, Oct 1998, doi: 10.2337/diabetes.47.10.1609.

[14] V. Poitout and R. P. Robertson, "Glucolipotoxicity: fuel excess and beta-cell dysfunction," *Endocr Rev,* vol. 29, no. 3, pp. 351-66, May 2008, doi: 10.1210/er.2007-0023.

[15] M. Prentki, M. L. Peyot, P. Masiello, and S. R. M. Madiraju, "Nutrient-Induced Metabolic Stress, Adaptation, Detoxification, and Toxicity in the Pancreatic beta-Cell," *Diabetes,* vol. 69, no. 3, pp. 279-290, Mar 2020, doi: 10.2337/dbi19-0014.

[16] G. S. Hotamisligil, N. S. Shargill, and B. M. Spiegelman, "Adipose expression of tumor necrosis factor-alpha: direct role in obesity-linked insulin resistance," *Science,* vol. 259, no. 5091, pp. 87-91, Jan 1 1993, doi: 10.1126/science.7678183.

[17] S. P. Weisberg, D. McCann, M. Desai, M. Rosenbaum, R. L. Leibel, and A. W. Ferrante, Jr., "Obesity is associated with macrophage accumulation in adipose tissue," *J Clin Invest,* vol. 112, no. 12, pp. 1796-808, Dec 2003, doi: 10.1172/JCI19246.

[18] W. C. Knowler *et al.*, "Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin," *N Engl J Med,* vol. 346, no. 6, pp. 393-403, Feb 7 2002, doi: 10.1056/NEJMoa012512.

[19] N. Bansal, "Prediabetes diagnosis and treatment: A review," *World J Diabetes,* vol. 6, no. 2, pp. 296-303, Mar 15 2015, doi: 10.4239/wjd.v6.i2.296.

[20] S. M. Grundy, "Pre-diabetes, metabolic syndrome, and cardiovascular risk," *J Am Coll Cardiol,* vol. 59, no. 7, pp. 635-43, Feb 14 2012, doi: 10.1016/j.jacc.2011.08.080.

[21] A. T. Kharroubi and H. M. Darwish, "Diabetes mellitus: The epidemic of the century," *World J Diabetes,* vol. 6, no. 6, pp. 850-67, Jun 25 2015, doi: 10.4239/wjd.v6.i6.850.

[22] A. R. Gosmanov and J. Wan, "Low positive predictive value of hemoglobin A1c for diagnosis of prediabetes in clinical practice," *Am J Med Sci,* vol. 348, no. 3, pp. 191-4, Sep 2014, doi: 10.1097/MAJ.0000000000000223.

[23] Z. T. Bloomgarden, S. E. Inzucchi, E. Karnieli, and D. Le Roith, "The proposed terminology 'A(1c)-derived average glucose' is inherently imprecise and should not be adopted," *Diabetologia,* vol. 51, no. 7, pp. 1111-4, Jul 2008, doi: 10.1007/s00125-008-1027-7.

[24] G. Pestoni *et al.*, "Association between dietary patterns and prediabetes, undetected diabetes or clinically diagnosed diabetes: results from the KORA FF4 study," *Eur J Nutr,* vol. 60, no. 5, pp. 2331-2341, Aug 2021, doi: 10.1007/s00394-020-02416-9.

[25] P. Farmaki, C. Damaskos, N. Garmpis, A. Garmpi, S. Savvanis, and E. Diamantis, "Complications of the Type 2 Diabetes Mellitus," *Curr Cardiol Rev,* vol. 16, no. 4, pp. 249-251, 2020, doi: 10.2174/1573403X1604201229115531.

[26] S. Vijan, "In the clinic. Type 2 diabetes," *Ann Intern Med,* vol. 152, no. 5, pp. ITC31-15; quiz ITC316, Mar 2 2010, doi: 10.7326/0003-4819-152-5-201003020-01003.

[27] R. Nesto, P. Libby, E. Brauwald, D. Zipes, and P. Libby, "Heart disease: A textbook of cardiovascular medicine," ed, 2001.

[28] R. Chen, B. Ovbiagele, and W. Feng, "Diabetes and Stroke: Epidemiology, Pathophysiology, Pharmaceuticals and Outcomes," *Am J Med Sci,* vol. 351, no. 4, pp. 380-6, Apr 2016, doi: 10.1016/j.amjms.2016.01.011.

[29] M. C. Thomas *et al.*, "Diabetic kidney disease," *Nat Rev Dis Primers,* vol. 1, p. 15018, Jul 30 2015, doi: 10.1038/nrdp.2015.18.

[30] V. Lagou *et al.*, "GWAS of random glucose in 476,326 individuals provide insights into diabetes pathophysiology, complications and treatment stratification," *Nat Genet,* vol. 55, no. 9, pp. 1448-1461, Sep 2023, doi: 10.1038/s41588-023-01462-3.

[31] W. K. Chung *et al.*, "Precision Medicine in Diabetes: A Consensus Report From the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD)," *Diabetes Care,* vol. 43, no. 7, pp. 1617-1635, Jul 2020, doi: 10.2337/dci20-0022.

[32] M. C. Riddle *et al.*, "Diabetes Care Editors' Expert Forum 2018: Managing Big Data for Diabetes Research and Care," *Diabetes Care,* vol. 42, no. 6, pp. 1136-1146, Jun 2019, doi: 10.2337/dci19-0020.

[33] R. Wang-Sattler *et al.*, "Novel biomarkers for pre-diabetes identified by metabolomics," *Mol Syst Biol,* vol. 8, p. 615, 2012, doi: 10.1038/msb.2012.43.

[34] S. Tulipani *et al.*, "Biomarkers of Morbid Obesity and Prediabetes by Metabolomic Profiling of Human Discordant Phenotypes," *Clin Chim Acta,* vol. 463, pp. 53-61, Dec 1 2016, doi: 10.1016/j.cca.2016.10.005.

[35] E. Ahlqvist *et al.*, "Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables," *Lancet Diabetes Endocrinol,* vol. 6, no. 5, pp. 361-369, May 2018, doi: 10.1016/S2213-8587(18)30051-2.

[36] J. M. Dennis, B. M. Shields, W. E. Henley, A. G. Jones, and A. T. Hattersley, "Disease progression and treatment response in data-driven subgroups of type 2 diabetes compared with models based on simple clinical features: an analysis using clinical trial data," *Lancet Diabetes Endocrinol,* vol. 7, no. 6, pp. 442-451, Jun 2019, doi: 10.1016/S2213-8587(19)30087-7.

[37] J. M. Dennis *et al.*, "Precision Medicine in Type 2 Diabetes: Clinical Markers of Insulin Resistance Are Associated With Altered Short- and Long-term Glycemic Response to DPP-4 Inhibitor Therapy," *Diabetes Care,* vol. 41, no. 4, pp. 705-712, Apr 2018, doi: 10.2337/dc17-1827.

[38] GoDarts *et al.*, "Common variants near ATM are associated with glycemic response to metformin in type 2 diabetes," *Nat Genet,* vol. 43, no. 2, pp. 117-20, Feb 2011, doi: 10.1038/ng.735.

[39] M. F. Hivert *et al.*, "Lifestyle and Metformin Ameliorate Insulin Sensitivity Independently of the Genetic Burden of Established Insulin Resistance Variants in Diabetes Prevention Program Participants," *Diabetes,* vol. 65, no. 2, pp. 520-6, Feb 2016, doi: 10.2337/db15-0950.

[40] M. F. Hivert *et al.*, "Updated genetic score based on 34 confirmed type 2 diabetes Loci is associated with diabetes incidence and regression to normoglycemia in the diabetes prevention program," *Diabetes,* vol. 60, no. 4, pp. 1340-8, Apr 2011, doi: 10.2337/db10-1119.

[41] G. A. Bray, "Medical consequences of obesity," *J Clin Endocrinol Metab,* vol. 89, no. 6, pp. 2583-9, Jun 2004, doi: 10.1210/jc.2004-0535.

[42] M. Stumvoll, B. J. Goldstein, and T. W. van Haeften, "Type 2 diabetes: principles of pathogenesis and therapy," *Lancet,* vol. 365, no. 9467, pp. 1333-46, Apr 9-15 2005, doi: 10.1016/S0140-6736(05)61032-X.

[43] G. J. Patti, O. Yanes, and G. Siuzdak, "Innovation: Metabolomics: the apogee of the omics trilogy," *Nat Rev Mol Cell Biol,* vol. 13, no. 4, pp. 263-9, Mar 22 2012, doi: 10.1038/nrm3314.

[44] J. E. Ho *et al.*, "Metabolomic Profiles of Body Mass Index in the Framingham Heart Study Reveal Distinct Cardiometabolic Phenotypes," *PLoS One,* vol. 11, no. 2, p. e0148361, 2016, doi: 10.1371/journal.pone.0148361.

[45] N. Kliemann *et al.*, "Metabolic signatures of greater body size and their associations with risk of colorectal and endometrial cancers in the European Prospective Investigation into Cancer and Nutrition," *BMC Med,* vol. 19, no. 1, p. 101, Apr 30 2021, doi: 10.1186/s12916-021-01970-1.

[46] C. Fernandez *et al.*, "Plasma Lipidome and Prediction of Type 2 Diabetes in the Population-Based Malmo Diet and Cancer Cohort," *Diabetes Care,* vol. 43, no. 2, pp. 366-373, Feb 2020, doi: 10.2337/dc19-1199.

[47] L. A. Lotta *et al.*, "Genetic Predisposition to an Impaired Metabolism of the Branched-Chain Amino Acids and Risk of Type 2 Diabetes: A Mendelian Randomisation Analysis," *PLoS Med,* vol. 13, no. 11, p. e1002179, Nov 2016, doi: 10.1371/journal.pmed.1002179.

[48] F. Ottosson, E. Smith, W. Gallo, C. Fernandez, and O. Melander, "Purine Metabolites and Carnitine Biosynthesis Intermediates Are Biomarkers for Incident Type 2 Diabetes," *J Clin Endocrinol Metab,* vol. 104, no. 10, pp. 4921-4930, Oct 1 2019, doi: 10.1210/jc.2019-00822.

[49] T. J. Wang *et al.*, "Metabolite profiles and the risk of developing diabetes," *Nat Med,* vol. 17, no. 4, pp. 448-53, Apr 2011, doi: 10.1038/nm.2307.

[50] F. Ottosson, P. Emami Khoonsari, M. J. Gerl, K. Simons, O. Melander, and C. Fernandez, "A plasma lipid signature predicts incident coronary artery disease," *Int J Cardiol,* vol. 331, pp. 249-254, May 15 2021, doi: 10.1016/j.ijcard.2021.01.059.

[51] P. Wurtz *et al.*, "Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts," *Circulation,* vol. 131, no. 9, pp. 774-85, Mar 3 2015, doi: 10.1161/CIRCULATIONAHA.114.013116.

[52] Y. Zheng *et al.*, "Metabolites of Glutamate Metabolism Are Associated With Incident Cardiovascular Events in the PREDIMED PREvencion con DIeta MEDiterranea (PREDIMED) Trial," *J Am Heart Assoc,* vol. 5, no. 9, Sep 15 2016, doi: 10.1161/JAHA.116.003755.

[53] Q. He *et al.*, "Comparison of serum metabolite compositions between obese and lean growing pigs using an NMR-based metabonomic approach," *J Nutr Biochem,* vol. 23, no. 2, pp. 133-9, Feb 2012, doi: 10.1016/j.jnutbio.2010.11.007.

[54] R. Williams *et al.*, "A multi-analytical platform approach to the metabonomic analysis of plasma from normal and Zucker (fa/fa) obese rats," *Mol Biosyst,* vol. 2, no. 3-4, pp. 174-83, Mar 2006, doi: 10.1039/b516356k.

[55] Y. Bao *et al.*, "Metabonomic variations in the drug-treated type 2 diabetes mellitus patients and healthy volunteers," *J Proteome Res,* vol. 8, no. 4, pp. 1623-30, Apr 2009, doi: 10.1021/pr800643w.

[56] J. Chen *et al.*, "Practical approach for the identification and isomer elucidation of biomarkers detected in a metabonomic study for the discovery of individuals at risk for diabetes by integrating the chromatographic and mass spectrometric information," *Anal Chem,* vol. 80, no. 4, pp. 1280-9, Feb 15 2008, doi: 10.1021/ac702089h.

[57] S. Park, K. C. Sadanala, and E. K. Kim, "A Metabolomic Approach to Understanding the Metabolic Link between Obesity and Diabetes," *Mol Cells,* vol. 38, no. 7, pp. 587-96, Jul 2015, doi: 10.14348/molcells.2015.0126.

[58] P. She *et al.*, "Disruption of BCATm in mice leads to increased energy expenditure associated with the activation of a futile protein turnover cycle," *Cell Metab,* vol. 6, no. 3, pp. 181-94, Sep 2007, doi: 10.1016/j.cmet.2007.08.003.

[59] D. Lanznaster, C. Veyrat-Durebex, P. Vourc'h, C. R. Andres, H. Blasco, and P. Corcia, "Metabolomics: A Tool to Understand the Impact of Genetic Mutations in Amyotrophic Lateral Sclerosis," *Genes (Basel),* vol. 11, no. 5, May 11 2020, doi: 10.3390/genes11050537.

[60] C. Gieger *et al.*, "Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum," *PLoS Genet,* vol. 4, no. 11, p. e1000282, Nov 2008, doi: 10.1371/journal.pgen.1000282.

[61] K. Suhre, J. Raffler, and G. Kastenmuller, "Biochemical insights from population studies with genetics and metabolomics," *Arch Biochem Biophys,* vol. 589, pp. 168-76, Jan 1 2016, doi: 10.1016/j.abb.2015.09.023.

[62] S. Wahl *et al.*, "Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity," (in English), *Nature,* vol. 541, no. 7635, pp. 81-+, Jan 5 2017, doi: 10.1038/nature20784.

[63] L. Yengo *et al.*, "Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry," *Hum Mol Genet,* vol. 27, no. 20, pp. 3641-3649, Oct 15 2018, doi: 10.1093/hmg/ddy271.

[64] A. Mahajan *et al.*, "Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation," *Nat Genet,* vol. 54, no. 5, pp. 560-572, May 2022, doi: 10.1038/s41588-022-01058-3.

[65] J. Wardle, S. Carnell, C. M. Haworth, and R. Plomin, "Evidence for a strong genetic influence on childhood adiposity despite the force of the obesogenic environment," *Am J Clin Nutr,* vol. 87, no. 2, pp. 398-404, Feb 2008, doi: 10.1093/ajcn/87.2.398.

[66] R. Mahmoud, V. Kimonis, and M. G. Butler, "Genetics of Obesity in Humans: A Clinical Review," *Int J Mol Sci,* vol. 23, no. 19, Sep 20 2022, doi: 10.3390/ijms231911005.

[67] R. B. Prasad and L. Groop, "Genetics of type 2 diabetes-pitfalls and possibilities," *Genes (Basel),* vol. 6, no. 1, pp. 87-123, Mar 12 2015, doi: 10.3390/genes6010087.

[68] E. Zeggini, "A new era for Type 2 diabetes genetics," *Diabet Med,* vol. 24, no. 11, pp. 1181-6, Nov 2007, doi: 10.1111/j.1464-5491.2007.02274.x.

[69] D. K. Sanghera and P. R. Blackett, "Type 2 Diabetes Genetics: Beyond GWAS," *J Diabetes Metab,* vol. 3, no. 198, Jun 23 2012, doi: 10.4172/2155-6156.1000198.

[70] M. Steinbach, L. Ertöz, and V. Kumar, "The challenges of clustering high dimensional data," in *New directions in statistical physics: econophysics, bioinformatics, and pattern recognition*: Springer, 2004, pp. 273-309.

[71] A. W. Edwards and L. L. Cavalli-Sforza, "A Method for Cluster Analysis," *Biometrics,* vol. 21, pp. 362-75, Jun 1965. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/14338671.

[72] P. Hansen and B. Jaumard, "Cluster analysis and mathematical programming," *Mathematical programming,* vol. 79, no. 1-3, pp. 191-215, 1997.

[73] D. Steinley, "K-means clustering: a half-century synthesis," *Br J Math Stat Psychol,* vol. 59, no. Pt 1, pp. 1-34, May 2006, doi: 10.1348/000711005X48266.

[74] G. F. Tzortzis and A. C. Likas, "The global kernel k-means algorithm for clustering in feature space," *IEEE Trans Neural Netw,* vol. 20, no. 7, pp. 1181-94, Jul 2009, doi: 10.1109/TNN.2009.2019722.

[75] C. Yuan and H. Yang, "Research on K-value selection method of K-means clustering algorithm," *J,* vol. 2, no. 2, pp. 226-235, 2019.

[76] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A k-means clustering algorithm," *Journal of the royal statistical society. series c (applied statistics),* vol. 28, no. 1, pp. 100-108, 1979.

[77] H. Tanabe *et al.*, "Factors Associated with Risk of Diabetic Complications in Novel Cluster-Based Diabetes Subgroups: A Japanese Retrospective Cohort Study," *J Clin Med,* vol. 9, no. 7, Jul 2 2020, doi: 10.3390/jcm9072083.

[78] M. Lugner *et al.*, "Comparison between data-driven clusters and models based on clinical features to predict outcomes in type 2 diabetes: nationwide observational study," *Diabetologia,* vol. 64, no. 9, pp. 1973-1981, 2021.

[79] D. Cavan, "Why screen for type 2 diabetes?," *Diabetes Res Clin Pract,* vol. 121, pp. 215-217, Nov 2016, doi: 10.1016/j.diabres.2016.11.004.

[80] J. Zhang, Z. Zhang, K. Zhang, X. Ge, R. Sun, and X. Zhai, "Early detection of type 2 diabetes risk: limitations of current diagnostic criteria," *Front Endocrinol (Lausanne),* vol. 14, p. 1260623, 2023, doi: 10.3389/fendo.2023.1260623.

[81] G. A. Nichols, T. A. Hillier, and J. B. Brown, "Progression from newly acquired impaired fasting glusose to type 2 diabetes," *Diabetes Care,* vol. 30, no. 2, pp. 228-33, Feb 2007, doi: 10.2337/dc06-1392.

[82] M. B. Davidson, "Diagnosing diabetes with glucose criteria: worshiping a false God," *Diabetes Care,* vol. 34, no. 2, pp. 524-6, Feb 2011, doi: 10.2337/dc10-1689.

[83] Q. Dong *et al.*, "Metabolic Signatures Elucidate the Effect of Body Mass Index on Type 2 Diabetes," *Metabolites,* vol. 13, no. 2, Feb 3 2023, doi: 10.3390/metabo13020227.

[84] S. Sharma *et al.*, "Role of human plasma metabolites in prediabetes and type 2 diabetes from the IMI-DIRECT study," *Diabetologia,* Sep 30 2024, doi: 10.1007/s00125-024-06282-6.

[85] Q. Dong *et al.*, "Subphenotypes of adult-onset diabetes: Data-driven clustering in the population-based KORA cohort," *Diabetes Obes Metab,* Oct 28 2024, doi: 10.1111/dom.16022.

[86] R. Holle, M. Happich, H. Lowel, H. E. Wichmann, and M. K. S. Group, "KORA--a research platform for population based health research," *Gesundheitswesen,* vol. 67 Suppl 1, pp. S19-25, Aug 2005, doi: 10.1055/s-2005-858235.

[87] L. A. Inker *et al.*, "Estimating glomerular filtration rate from serum creatinine and cystatin C," *N Engl J Med,* vol. 367, no. 1, pp. 20-9, Jul 5 2012, doi: 10.1056/NEJMoa1114248.

[88] R. W. Koivula *et al.*, "Discovery of biomarkers for glycaemic deterioration before and after the onset of type 2 diabetes: rationale and design of the epidemiological studies within the IMI DIRECT Consortium," *Diabetologia,* vol. 57, no. 6, pp. 1132-42, Jun 2014, doi: 10.1007/s00125-014-3216-x.

[89] M. Haid *et al.*, "Long-Term Stability of Human Plasma Metabolites during Storage at -80 degrees C," *J Proteome Res,* vol. 17, no. 1, pp. 203-211, Jan 5 2018, doi: 10.1021/acs.jproteome.7b00518.

[90] A. A. Brown *et al.*, "Genetic analysis of blood molecular phenotypes reveals common properties in the regulatory networks affecting complex traits," *Nat Commun,* vol. 14, no. 1, p. 5062, Aug 21 2023, doi: 10.1038/s41467-023-40569-3.

[91] K. J. Preacher and A. F. Hayes, "Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models," *Behav Res Methods,* vol. 40, no. 3, pp. 879-91, Aug 2008, doi: 10.3758/brm.40.3.879.

[92] M. E. Sobel, "Asymptotic confidence intervals for indirect effects in structural equation models," *Sociological methodology,* vol. 13, pp. 290-312, 1982.

[93] C. J. Bult, J. A. Blake, C. L. Smith, J. A. Kadin, J. E. Richardson, and G. Mouse Genome Database, "Mouse Genome Database (MGD) 2019," *Nucleic Acids Res,* vol. 47, no. D1, pp. D801-D806, Jan 8 2019, doi: 10.1093/nar/gky1056.

[94] X. Yang *et al.*, "Tissue-specific expression and regulation of sexually dimorphic genes in mice," *Genome Res,* vol. 16, no. 8, pp. 995-1004, Aug 2006, doi: 10.1101/gr.5217506.

[95] S. A. Lambert *et al.*, "The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation," *Nat Genet,* vol. 53, no. 4, pp. 420-425, Apr 2021, doi: 10.1038/s41588-021-00783-5.

[96] H. H. M. Draisma *et al.*, "Genome-wide association study identifies novel genetic variants contributing to variation in blood metabolite levels," *Nat Commun,* vol. 6, p. 7208, Jun 12 2015, doi: 10.1038/ncomms8208.

[97] A. A. Hicks *et al.*, "Genetic determinants of circulating sphingolipid concentrations in European populations," *PLoS Genet,* vol. 5, no. 10, p. e1000672, Oct 2009, doi: 10.1371/journal.pgen.1000672.

[98] S. R. Khan, Y. Manialawy, A. Obersterescu, B. J. Cox, E. P. Gunderson, and M. B. Wheeler, "Diminished sphingolipid metabolism, a hallmark of future type 2 diabetes pathogenesis, is linked to pancreatic β cell dysfunction," *IScience,* vol. 23, no. 10, 2020.

[99] A. Alexaki *et al.*, "De novo sphingolipid biosynthesis is required for adipocyte survival and metabolic homeostasis," *Journal of Biological Chemistry,* vol. 292, no. 9, pp. 3929-3939, 2017.

[100] A. Kumar *et al.*, "High-fat diet-induced upregulation of exosomal phosphatidylcholine contributes to insulin resistance," *Nat Commun,* vol. 12, no. 1, p. 213, Jan 11 2021, doi: 10.1038/s41467-020-20500-w.

[101] H.-S. Lee *et al.*, "Identification of putative biomarkers for type 2 diabetes using metabolomics in the Korea Association REsource (KARE) cohort," *Metabolomics,* vol. 12, pp. 1-12, 2016.

[102] W. L. Holland *et al.*, "Lipid-induced insulin resistance mediated by the proinflammatory receptor TLR4 requires saturated fatty acid-induced ceramide biosynthesis in mice," *J Clin Invest,* vol. 121, no. 5, pp. 1858-70, May 2011, doi: 10.1172/JCI43378.

[103] F. Samad, K. D. Hester, G. Yang, Y. A. Hannun, and J. Bielawski, "Altered adipose and plasma sphingolipid metabolism in obesity: a potential mechanism for cardiovascular and metabolic risk," *Diabetes,* vol. 55, no. 9, pp. 2579-87, Sep 2006, doi: 10.2337/db06-0330.

[104] X. Wei *et al.*, "Fatty acid synthesis configures the plasma membrane for inflammation in diabetes," *Nature,* vol. 539, no. 7628, pp. 294-298, Nov 10 2016, doi: 10.1038/nature20117.

[105] R. N. Lemaitre *et al.*, "Genetic loci associated with plasma phospholipid n-3 fatty acids: a meta-analysis of genome-wide association studies from the CHARGE Consortium," *PLoS Genet,* vol. 7, no. 7, p. e1002193, Jul 2011, doi: 10.1371/journal.pgen.1002193.

[106] W. Guan *et al.*, "Genome-wide association study of plasma N6 polyunsaturated fatty acids within the cohorts for heart and aging research in genomic epidemiology consortium," *Circ Cardiovasc Genet,* vol. 7, no. 3, pp. 321-331, Jun 2014, doi: 10.1161/CIRCGENETICS.113.000208.

[107] E. Lattka, T. Illig, B. Koletzko, and J. Heinrich, "Genetic variants of the FADS1 FADS2 gene cluster as related to essential fatty acid metabolism," *Curr Opin Lipidol,* vol. 21, no. 1, pp. 64-9, Feb 2010, doi: 10.1097/MOL.0b013e3283327ca8.

[108] H. T. Reardon *et al.*, "Dietary long-chain polyunsaturated fatty acids upregulate expression of FADS3 transcripts," *Prostaglandins Leukot Essent Fatty Acids,* vol. 88, no. 1, pp. 15-9, Jan 2013, doi: 10.1016/j.plefa.2012.02.003.

[109] D. Meierhofer, C. Weidner, and S. Sauer, "Integrative analysis of transcriptomics, proteomics, and metabolomics data of white adipose and liver tissue of high-fat diet and rosiglitazone-treated insulin-resistant mice identified pathway alterations and molecular hubs," *J Proteome Res,* vol. 13, no. 12, pp. 5592-602, Dec 5 2014, doi: 10.1021/pr5005828.

[110] B. Brayner, G. Kaur, M. A. Keske, and K. M. Livingstone, "FADS Polymorphism, Omega-3 Fatty Acids and Diabetes Risk: A Systematic Review," *Nutrients,* vol. 10, no. 6, Jun 13 2018, doi: 10.3390/nu10060758.

[111] S. B. Zaghlool *et al.*, "Revealing the role of the human blood plasma proteome in obesity using genetic drivers," *Nat Commun,* vol. 12, no. 1, p. 1279, Feb 24 2021, doi: 10.1038/s41467-021-21542-4.

[112] B. J. Jenkins *et al.*, "Odd Chain Fatty Acids; New Insights of the Relationship Between the Gut Microbiota, Dietary Intake, Biosynthesis and Glucose Intolerance," *Sci Rep,* vol. 7, p. 44845, Mar 23 2017, doi: 10.1038/srep44845.

[113] S. J. Yang, S. Y. Kwak, G. Jo, T. J. Song, and M. J. Shin, "Serum metabolite profile associated with incident type 2 diabetes in Koreans: findings from the Korean Genome and Epidemiology Study," *Sci Rep,* vol. 8, no. 1, p. 8207, May 29 2018, doi: 10.1038/s41598-018-26320-9.

[114] H. M. Han MyoungSook *et al.*, "Lysophosphatidylcholine as an effector of fatty acid-induced insulin resistance," 2011.

[115] L. Shao *et al.*, "Serum metabolomics-based heterogeneities and screening strategy for metabolic dysfunction-associated fatty liver disease (MAFLD)," *Clin Chim Acta,* vol. 538, pp. 203-210, Jan 1 2023, doi: 10.1016/j.cca.2022.12.014.

[116] S. Rivas-Tumanyan *et al.*, "Novel Plasma Metabolomic Markers Associated with Diabetes Progression in Older Puerto Ricans," *Metabolites,* vol. 12, no. 6, Jun 2 2022, doi: 10.3390/metabo12060513.

[117] B. Scott *et al.*, "Metformin and feeding increase levels of the appetite-suppressing metabolite Lac-Phe in humans," *Nat Metab,* Mar 18 2024, doi: 10.1038/s42255-024-01018-7.

[118] L. O. Ohlson *et al.*, "Risk factors for type 2 (non-insulin-dependent) diabetes mellitus. Thirteen and one-half years of follow-up of the participants in a study of Swedish men born in 1913," *Diabetologia,* vol. 31, no. 11, pp. 798-805, Nov 1988, doi: 10.1007/BF00277480.

[119] S. Romero-Garcia, M. M. Moreno-Altamirano, H. Prado-Garcia, and F. J. Sanchez-Garcia, "Lactate Contribution to the Tumor Microenvironment: Mechanisms, Effects on Immune Cells and Therapeutic Relevance," *Front Immunol,* vol. 7, p. 52, 2016, doi: 10.3389/fimmu.2016.00052.

[120] A. H. Barnett, A. J. Spiliopoulos, D. A. Pyke, W. A. Stubbs, J. Burrin, and K. G. Alberti, "Metabolic studies in unaffected co-twins of non-insulin-dependent diabetics," *Br Med J (Clin Res Ed),* vol. 282, no. 6277, pp. 1656-8, May 23 1981, doi: 10.1136/bmj.282.6277.1656.

[121] F. Berhane *et al.*, "Plasma Lactate Levels Increase during Hyperinsulinemic Euglycemic Clamp and Oral Glucose Tolerance Test," *J Diabetes Res,* vol. 2015, p. 102054, 2015, doi: 10.1155/2015/102054.

[122] Y. Wu, Y. Dong, M. Atefi, Y. Liu, Y. Elshimali, and J. V. Vadgama, "Lactate, a Neglected Factor for Diabetes and Cancer Interaction," *Mediators Inflamm,* vol. 2016, p. 6456018, 2016, doi: 10.1155/2016/6456018.

[123] N. Safai, A. Ali, P. Rossing, and M. Ridderstrale, "Stratification of type 2 diabetes based on routine clinical markers," *Diabetes Res Clin Pract,* vol. 141, pp. 275-283, Jul 2018, doi: 10.1016/j.diabres.2018.05.014.

[124] D. J. Freeman *et al.*, "C-reactive protein is an independent predictor of risk for the development of diabetes in the West of Scotland Coronary Prevention Study," *Diabetes,* vol. 51, no. 5, pp. 1596-600, May 2002, doi: 10.2337/diabetes.51.5.1596.

[125] N. R. Sproston and J. J. Ashworth, "Role of C-Reactive Protein at Sites of Inflammation and Infection," *Front Immunol,* vol. 9, p. 754, 2018, doi: 10.3389/fimmu.2018.00754.

[126] L. Del Bosque-Plata, E. Martinez-Martinez, M. A. Espinoza-Camacho, and C. Gragnoli, "The Role of TCF7L2 in Type 2 Diabetes," *Diabetes,* vol. 70, no. 6, pp. 1220-1228, Jun 2021, doi: 10.2337/db20-0573.

[127] G. Geoghegan *et al.*, "Targeted deletion of Tcf7l2 in adipocytes promotes adipocyte hypertrophy and impaired glucose metabolism," *Mol Metab,* vol. 24, pp. 44-63, Jun 2019, doi: 10.1016/j.molmet.2019.03.003.

[128] C. Gabay and I. Kushner, "Acute-phase proteins and other systemic responses to inflammation," *N Engl J Med,* vol. 340, no. 6, pp. 448-54, Feb 11 1999, doi: 10.1056/NEJM199902113400607.

[129] X. Chen *et al.*, "The Diabetes Gene and Wnt Pathway Effector TCF7L2 Regulates Adipocyte Development and Function," *Diabetes,* vol. 67, no. 4, pp. 554-568, Apr 2018, doi: 10.2337/db17-0318.

[130] C. W. Grote and D. E. Wright, "A Role for Insulin in Diabetic Neuropathy," *Front Neurosci,* vol. 10, p. 581, 2016, doi: 10.3389/fnins.2016.00581.

[131] O. Racz, M. Linkova, K. Jakubowski, R. Link, and D. Kuzmova, "[Barriers of the initiation of insulin treatment in type 2 diabetic patients - conquering the "psychological insulin resistance"]," *Orv Hetil,* vol. 160, no. 3, pp. 93-97, Jan 2019, doi: 10.1556/650.2019.31269. Az inzulinkezeles elkezdesenek gyakorlati akadalyai 2-es tipusu cukorbetegekben - a "pszichologiai inzulinrezisztencia" lekuzdese.

[132] H. Yaribeygi, S. L. Atkin, L. E. Simental-Mendia, and A. Sahebkar, "Molecular mechanisms by which aerobic exercise induces insulin sensitivity," *J Cell Physiol,* vol. 234, no. 8, pp. 12385-12392, Aug 2019, doi: 10.1002/jcp.28066.

[133] G. K. Bhatti, S. K. Bhadada, R. Vijayvergiya, S. S. Mastana, and J. S. Bhatti, "Metabolic syndrome and risk of major coronary events among the urban diabetic patients: North Indian Diabetes and Cardiovascular Disease Study-NIDCVD-2," *J Diabetes Complications,* vol. 30, no. 1, pp. 72-8, Jan-Feb 2016, doi: 10.1016/j.jdiacomp.2015.07.008.

[134] S. M. Shamshirgaran, A. Mamaghanian, A. Aliasgarzadeh, N. Aiminisani, M. Iranparvar-Alamdari, and J. Ataie, "Age differences in diabetes-related complications and glycemic control," *BMC Endocr Disord,* vol. 17, no. 1, p. 25, May 4 2017, doi: 10.1186/s12902-017-0175-5.

[135] S. Taheri, "Type 2 Diabetes Remission: A New Mission in Diabetes Care," *Diabetes Care,* vol. 47, no. 1, pp. 47-49, Jan 1 2024, doi: 10.2337/dci23-0062.

[136] E. W. Gregg *et al.*, "Improving health outcomes of people with diabetes: target setting for the WHO Global Diabetes Compact," *Lancet,* vol. 401, no. 10384, pp. 1302-1312, Apr 15 2023, doi: 10.1016/S0140-6736(23)00001-6.

[137] K. Suzuki *et al.*, "Genetic drivers of heterogeneity in type 2 diabetes pathophysiology," *Nature,* vol. 627, no. 8003, pp. 347-357, Mar 2024, doi: 10.1038/s41586-024-07019-6.

[138] P. W. Franks, E. Pearson, and J. C. Florez, "Gene-environment and gene-treatment interactions in type 2 diabetes: progress, pitfalls, and prospects," *Diabetes Care,* vol. 36, no. 5, pp. 1413-21, May 2013, doi: 10.2337/dc12-2211.

[139] M. E. Lean *et al.*, "5-year follow-up of the randomised Diabetes Remission Clinical Trial (DiRECT) of continued support for weight loss maintenance in the UK: an extension study," *Lancet Diabetes Endocrinol,* vol. 12, no. 4, pp. 233-246, Apr 2024, doi: 10.1016/S2213-8587(23)00385-6.

[140] J. S. Varghese, R. M. Carrillo-Larco, and K. V. Narayan, "Achieving replicable subphenotypes of adult-onset diabetes," *Lancet Diabetes Endocrinol,* vol. 11, no. 9, pp. 635-636, Sep 2023, doi: 10.1016/S2213-8587(23)00195-X.

Paper I

Title: Metabolic Signatures Elucidate the Effect of Body Mass Index on Type 2 Diabetes

Authors: Qiuling Dong, Sidra Sidra, Christian Gieger, Rui Wang-Sattler, Wolfgang Rathmann, Cornelia Prehn, Jerzy Adamski, Wolfgang Koenig, Annette Peters, Harald Grallert, Sapna Sharma

Journal: Metabolites

Status: Publishes

Volume:13

Page: 227

Year: 2023

doi: 10.3390/metabo13020227

Supplements: https://pmc.ncbi.nlm.nih.gov/articles/instance/9965667/bin/metabolites-13-00227-s001.zip

metabolites

MDPI

Article

Metabolic Signatures Elucidate the Effect of Body Mass Index on Type 2 Diabetes

Qiuling Dong [1,2,3], Sidra Sidra [4], Christian Gieger [1,2,5], Rui Wang-Sattler [6], Wolfgang Rathmann [7], Cornelia Prehn [8], Jerzy Adamski [9,10,11], Wolfgang Koenig [12,13,14], Annette Peters [2,5,15], Harald Grallert [1,2,5,*,†] and Sapna Sharma [1,2,16,*,†]

1 Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, 85764 Neuherberg, Germany
2 Institute of Epidemiology, Helmholtz Zentrum München, 85764 Neuherberg, Germany
3 Faculty of Medicine, Ludwig-Maximilians-University München, 81377 Munich, Germany
4 Institute for Medical Information Processing, Biometry and Epidemiology (IBE), Ludwig-Maximilians-Universität München, 81377 Munich, Germany
5 German Center for Diabetes Research (DZD), 85764 München-Neuherberg, Germany
6 Institute of Translational Genomics, Helmholtz Zentrum München, 85764 Neuherberg, Germany
7 Institute for Biometrics and Epidemiology, German Diabetes Center, Leibniz Center for Diabetes Research, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany
8 Metabolomics and Proteomics Core Facility, Helmholtz Zentrum München, 85764 Neuherberg, Germany
9 Institute of Experimental Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany
10 Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, 8 Medical Drive, Singapore 117597, Singapore
11 Institute of Biochemistry, Faculty of Medicine, University of Ljubljana, Vrazov trg 2, 1000 Ljubljana, Slovenia
12 German Research Center for Cardiovascular Disease (DZHK), Partner site Munich Heart Alliance, 81377 Munich, Germany
13 Deutsches Herzzentrum München, Technische Universität München, 81377 Munich, Germany
14 Institute of Epidemiology and Medical Biometry, University of Ulm, 89069 Ulm, Germany
15 Chair of Epidemiology, Faculty of Medicine, Ludwig-Maximilians-University München, 81377 Munich, Germany
16 Chair of Food Chemistry and Molecular Sensory Science, Technical University of Munich, 85354 Freising-Weihenstephan, Germany
* Correspondence: harald.grallert@helmholtz-muenchen.de (H.G.); sapna.sharma@tum.de (S.S.)
† These authors contributed equally to this work.

Abstract: Obesity plays an important role in the development of insulin resistance and diabetes, but the molecular mechanism that links obesity and diabetes is still not completely understood. Here, we used 146 targeted metabolomic profiles from the German KORA FF4 cohort consisting of 1715 participants and associated them with obesity and type 2 diabetes. In the basic model, 83 and 51 metabolites were significantly associated with body mass index (BMI) and T2D, respectively. Those metabolites are branched-chain amino acids, acylcarnitines, lysophospholipids, or phosphatidylcholines. In the full model, 42 and 3 metabolites were significantly associated with BMI and T2D, respectively, and replicate findings in the previous studies. Sobel mediation testing suggests that the effect of BMI on T2D might be mediated via lipids such as sphingomyelin (SM) C16:1, SM C18:1 and diacylphosphatidylcholine (PC aa) C38:3. Moreover, mendelian randomization suggests a causal relationship that BMI causes the change of SM C16:1 and PC aa C38:3, and the change of SM C16:1, SM C18:1, and PC aa C38:3 contribute to T2D incident. Biological pathway analysis in combination with genetics and mice experiments indicate that downregulation of sphingolipid or upregulation of phosphatidylcholine metabolism is a causal factor in early-stage T2D pathophysiology. Our findings indicate that metabolites like SM C16:1, SM C18:1, and PC aa C38:3 mediate the effect of BMI on T2D and elucidate their role in obesity related T2D pathologies.

Keywords: obesity; type 2 diabetes; metabolomics; mediation; mendelian randomization; type 2 diabetes pathology

Citation: Dong, Q.; Sidra, S.; Gieger, C.; Wang-Sattler, R.; Rathmann, W.; Prehn, C.; Adamski, J.; Koenig, W.; Peters, A.; Grallert, H.; et al. Metabolic Signatures Elucidate the Effect of Body Mass Index on Type 2 Diabetes. *Metabolites* **2023**, *13*, 227. https://doi.org/10.3390/metabo13020227

Academic Editors: Antonio Cittadini and Roberta D'Assante

Received: 1 December 2022
Revised: 26 January 2023
Accepted: 31 January 2023
Published: 3 February 2023

1. Introduction

According to the World health Organization (WHO), over 1 billion people worldwide are obese, including 650 million adults, 340 million adolescents and 39 million children, and this results in the degradation of health [1]. Obesity is a disease impacting most body systems and contributes to a range of noncommunicable diseases including cardiovascular disease, type 2 diabetes (T2D), and cancer [2–4]. It has been proven that being overweight or obese are the most critical conditions for risk of developing T2D and both are linked to metabolic syndrome [5]. Metabolic processes are regulated by various perturbations from its surrounding environment and several levels of enzymes [6]. The molecular mechanisms by which obesity affects T2D development include lipid metabolism, insulin sensitivity, and inflammation [7].

Increasing interest has been addressed in the application of metabolic profiling to the identification of disease biomarkers, as it is a potent approach to uncovering the convoluted progression between obesity, metabolism, and diabetes [8]. Stevens et al. outlined the metabolomic signature of human obesity and linked them to T2D parameters such as C-reactive protein (CRP) and HbA1c [9]. The study by Tulipani et al. shows metabolic traits [lyso]glycerophospholipids in particular lysophosphatidylcholines associated with morbid obesity and several amino acids glutamate, glycine and branch chain amino acids were biomarkers of risk of diabetes onset associated with obesity and prediabetes [10]. Lipidomics analysis has unraveled that several sphingomyelins, diacyl phosphatidylcholine, and lysophosphatidylcholine were associated with waist circumference whereas HOMA-IR was strongly related with specific lysophosphatidylcholines and diacyl phosphatidylcholines [11]. These studies provide support for the involvement of metabolites in progression of metabolic disease, but no emphasis was given to dissect the intermediate pathway between obesity and diabetes.

Small molecular lipids such as sphingolipids, glycerophospholipids, and fatty acids play vital roles in metabolic pathways related to health and disease. Sphingolipids are a class of lipids; simple sphingolipids include the sphingoid bases and ceramides. Ceramides are important bioactive lipids produced from three pathways: (i) the de novo pathway; (ii) the sphingomyelin pathway; and (iii) the salvage/recycling pathway [12]. Glycerophospholipids are a class of lipids that constitute a major component of cell membrane, which is generally composed of hydrophobic fatty acids and a hydrophilic phosphate group. The phosphate group is modified by different small molecules to form different kinds of glycerophospholipids, for example, by choline to form phosphatidylcholine [13]. Clinical studies have demonstrated that phospholipids including sphingolipids and glycerophospholipids are strongly associated with insulin sensitivity [14].

Genetic composition can be used to make predictions regarding disease susceptibility. The overgrown obesity rates and their clinical consequences (T2D) clearly indicate that non-genetic or environmental factors and their interaction with genetic variants are major players of disease development [15]. Genome-wide association studies show more than 900 genetic variants associated with BMI [16] and more than 230 loci influencing risk of T2D [17]. Furthermore, linking metabolites with other omics, especially genetics using genome-wide association (mGWAS), gives access to genetics' influence on the metabolic composition of key lipids, amino acids, and carbohydrates [18–20]. mGWAS, with a growing sample size and ascending complex metabolic traits, allows for a more comprehensive and systems-based downstream analysis.

In this work, we considered a targeted metabolomic analysis of 1715 participants enrolled in the KORA FF4 Cohort to investigate metabolite markers for obesity and T2D participate in development of obesity-related Type 2 diabetes. Metabolite profiles of 146 named serum metabolites were assessed and compared with publicly available studies. The metabolites mediation effect of BMI on T2D was investigated using a mediation test. Further, we used mendelian randomization (MR) to define metabolites that may be causally linked with BMI and T2D and vice versa using genetic variants. Finally, biological pathways and consequences were analyzed by incorporating genetics and mouse model data from

the literature, yielding the bioactive role of sphingolipids and glycerophospholipids in metabolic dysregulation and beta cell dysfunction.

2. Materials and Methods

2.1. Study Subjects and Sampling

The Cooperative Health Research in the Region of Augsburg (KORA) study is a population-based cohort study. The KORA FF4 study (2013–2014) is the second follow-up of KORA S4 (1999–2001). All samples included in the study were collected in the morning between 8:00 a.m. and 10:30 a.m. after at least 8 h of fasting. We examined 2216 individuals who had phenotype and metabolite measurements and excluded 501 participants in the analysis, including (1) underweight (BMI < 15 kg/m^2) or missing covariate values (n = 23), and (2) prediabetes (impaired fasting glycemia or impaired glucose tolerance, n = 390). It is reported that impaired fasting glucose and impaired glucose tolerance should be considered as different phenotypes from T2D, so we removed these participants [21]. Additionally excluded were (3) diagnosis for type 1 diabetes (n = 6) and (4) unclear type of diabetes mellitus (n = 82). The remaining dataset has 1715 participants, comprising 1276 non-obese participants (BMI < 30 kg/m^2) and 439 obese (BMI $\geq$ 30 kg/m^2), and 1415 non-diabetic participants and 300 individuals with type 2 diabetes. The incident T2D was defined based on an oral glucose tolerance test (OGTT) or a validated physician diagnosis. WHO diagnostic criteria were applied to the classification of KORA participants.

2.2. Metabolite Quantification and Normalization

Samples were collected and stored at −80 °C and profiling FF4 metabolomics were performed in February–October 2019. The stability was measured and validated [22]. Blood samples from KORA FF4 participants in the study were measured with the Absolute*IDQ*™ p180 Kit (BIOCRATES Life Sciences AG, Innsbruck, Austria). The assay procedures were previously described in detail [23]. Briefly, 10 µL serum samples were added to the 96-well kit plate with respective standards and dried under a nitrogen stream. Amino acids and biogenic amines were derivatized with 5% phenylisothiocyanate in ethanol/water/pyridine. After metabolite and standard extraction, using methanol containing 5 mM ammonium acetate, the eluate was diluted with water for LC MS/MS analysis and with the kits running solvent for FIA-MS/MS analysis. The analytical process was conducted by the MetIQ™ software package and a targeted profiling scheme was applied to quantitatively identify known metabolites. Metabolites that met any one of the three exclusion criteria were deleted: (1) coefficient of variance (CV) value of five reference samples was equal to or greater than 25%; (2) there were $\geq$ 50% of all measured sample concentrations lower than corresponding plate limit of detection (LOD), the plate LOD was defined as 3 times median of three zero samples in each plate; and (3) the non-detectable rate of all measured samples was equal to or greater than 50%. There were 146 metabolites that passed quality control (QC). Non-detectable values in sample data were randomly imputed ranging from 75% to 125% of the half of the lowest measured value of the metabolite in each plate. Afterwards, plate normalization factors (NFs) were taken into consideration and adjusted for metabolite concentrations to reduce the plate impact. The normalization process was described elsewhere [24]. Metabolite concentrations were natural-log transformed and scaled (mean = 0, sd = 1) to ensure comparability between the metabolites.2.3. Statistics

All statistical analyses were performed in R (version 4.1.0) and a two-sided p value < 0.05 was considered as statistically significant after the Bonferroni correction.

2.2.1. Multivariable Linear Regression and Logistic Regression

For BMI-metabolite associations, multivariable linear regression was employed with each metabolite as an independent variable and the BMI value as a dependent variable. This analysis was adjusted for covariates age, sex in basic model and including additional covariates like, physical activities, smoking status, systolic blood pressure, high-density lipoprotein cholesterol (HDL-C), triglyceride, fasting glucose levels in full model. In logistic

regression analysis for metabolite-T2D associations, odds ratios (ORs) for each metabolite between two groups were calculated. Logistic regression analysis was carried out with the diabetic status as a dependent variable and each metabolite as an independent variable. Same risk factors in the linear regression analyses with additional BMI were added as covariates in the logistic regression model and the same significance level was adopted.

2.2.2. Sobel Mediation Test

We performed Sobel tests [25,26] to assess whether metabolites carry the influence of BMI to T2D. All analyses were conducted in R by using the package 'bda' v15.2.5 and the functions mediation test. In order to adjust confounders, the residuals were obtained from a linear regression model that each metabolite was a dependent variable and covariates (age, sex, physical activity, smoking status, systolic blood pressure, HDL-C, and triglyceride) as independent variables. Afterwards, metabolite residual entered the Sobel test model as a mediator, and BMI as an independent variable, whereas fasting glucose or HbA1c was taken as the dependent variable. With these two approaches, we examined the mediation effect of metabolites. The p-value thresholds follow the Bonferroni-correction and metabolites with $p < 0.05$ were considered to have a significant mediation effect.

2.2.3. Mendelian Randomization

We checked for causal inference using two sample mendelian randomisation (2SMR) methods from the MRInstruments (0.3.2) and TwoSampleMR library (v0.5.6) [27]. 2SMR is a method to draw a causal relation using only summary statistics of genome wide association studies (GWAS) from two observational studies [27]. To assess the impact of BMI on metabolite levels, in a 2SMR test, BMI instruments were obtained from the GIANT-UK Biobank meta-analysis [16] and the corresponding SNP estimates on T2D were extracted from the mGWAS [28]. BMI instruments with genome-wide significance ($p < 1 \times 10^{-8}$) and an LD clumping threshold of 0.001 were considered. The exposure and outcome data were harmonized before performing the MR analysis by positioning the SNPs on the same effect allele. We used the IVW method to estimate the causal effect of BMI on metabolites. From the direction of metabolites to T2D, metabolite instruments were obtained from the metabolite-GWAS [28] and extracted the corresponding SNPs from the GWAS meta-analysis [29]. After LD clumping and harmonization, a Wald ratio method was selected in MR analysis to estimate the causal relationship due to the limited SNP instruments. For sensitivity analysis, we performed heterogeneity or horizontal pleiotropy based on the MR-Egger analysis.

3. Results

3.1. Associations of Metabolites with BMI and T2D

3.1.1. Characteristics of the KORA FF4 Participants

Among 1715 participants, 1276 individuals were non-obese (BMI < 30) and 439 were obese (BMI $\geq$ 30). As shown in Table 1, there was no significant difference in sex and alcohol consumption between obese and non-obese groups. Compared with the non-obese group, the blood pressure, triglycerides, and fasting glucose were significantly higher and HDL cholesterol was significantly lower in the obese group. Besides, for participants with BMI < 30, only 136 individuals (10.7%) developed T2D, whereas T2D was diagnosed more frequently in obese participants (37.6%).

Similarly, for alcohol consumption, no significant difference between healthy and T2D participants was observed. BMI, blood pressure, triglycerides, and fasting glucose were significantly higher and HDL cholesterol was significantly lower in the T2D group (Table 2). Compared with non-diabetic individuals, the cases of obesity in T2D groups (53.3%) were almost three times higher than in the normal participant's group (19.2%).

Table 1. Characteristics of the KORA FF4 participants based on their BMI. Mean and standard deviation are provided for quantitative variables. Count and percentage are provided for categorical variables. The significant difference of population characteristics between the individuals with obesity and the normal participants was calculated. Categorical variables were calculated via the chi square test. Student's *t* test was used for continuous variables. Abbreviations: HbA1C, glycated hemoglobin; HDL, high-density lipoprotein; LDL, low-density lipoprotein.

BMI	Overall	Non-Obese (BMI < 30 kg/m^2)	Obese (BMI $\geq$ 30 kg/m^2)	*p* Value
Sample size	1715	1276	439	
Age mean (SD)	59.0 (12.2)	58.1 (12.1)	62.0 (12.0)	<0.001
Sex woman (%)	904 (52.7)	683 (53.5)	221 (50.3)	0.268
Weight (kg) mean (SD)	78.6 (16.0)	72.7 (11.8)	95.7 (14.2)	<0.001
Height (cm) mean (SD)	169.1 (9.6)	169.5 (9.6)	167.7 (9.7)	<0.001
Alcohol (g/day) mean (SD)	14.2 (19.4)	14.5 (18.2)	13.5 (22.5)	0.392
Waist (cm) mean (SD)	95.6 (14.0)	90.2 (10.5)	111.6 (10.3)	<0.001
Waist-hip-ratio mean (SD)	0.9 (0.1)	0.88 (0.1)	0.96 (0.1)	<0.001
Fasting glucose (mmol/L) mean (SD)	5.6 (1.3)	5.4 (1.0)	6.3 (1.7)	<0.001
2 h post glucose (mmol/L) mean (SD)	5.8 (2.2)	5.5 (1.7)	6.9 (3.2)	<0.001
Systolic blood pressure (mmHg) mean (SD)	117.9 (17.2)	116.4 (16.6)	122.5 (18.1)	<0.001
Diastolic blood pressure (mmHg) mean (SD)	72.7 (9.5)	72.2 (9.1)	74.0 (10.3)	0.001
Smoking (%)				<0.001
Smoker	267 (15.6)	221 (17.3)	46 (10.5)	
Ex-smoker	658 (38.4)	461 (36.1)	197 (44.9)	
Never-smoker	790 (46.1)	594 (46.6)	196 (44.6)	
Physical activities inactive (%)	702 (40.9)	456 (35.7)	246 (56.0)	<0.001
HDL cholesterol (mmol/L) mean (SD)	1.7 (0.5)	1.8 (0.5)	1.5 (0.4)	<0.001
LDL cholesterol (mmol/L) mean (SD)	3.5 (0.9)	3.4 (0.9)	3.6 (0.9)	0.048
Triglycerides (mmol/L) mean (SD)	1.4 (0.8)	1.25 (0.8)	1.6 (0.9)	<0.001
HbA1c (%) mean (SD)	5.5 (0.7)	5.4 (0.6)	5.8 (0.9)	<0.001
Total cholesterol (mmol/L) mean (SD)	5.6 (1.00)	5.6 (1.0)	5.5 (1.0)	0.409
C-reactive protein (mg/L) mean (SD)	2.3 (4.4)	1.7 (3.8)	3.9 (5.5)	<0.001
Type 2 diabetesy (%)	300 (17.5)	136 (10.7)	164 (37.4)	<0.001

Table 2. Characteristics of the KORA FF4 participants based on their diabetic status. Mean and standard deviation is provided for quantitative variables. Count and percentage are provided for categorical variables. The significant difference of population characteristics between the diabetic patients and nondiabetic participants was tested, respectively. Categorical variables were calculated via chi square test. Student's *t* test was used for continuous variables.

Diabetes	Overall	T2D (No)	T2D (Yes)	*p* Value
Sample size	1715	1415	300	
Age mean (SD)	59.0 (12.2)	59.7 (12.2)	69.5 (10.0)	<0.001
Sex woman (%)	904 (52.7)	784 (55.4)	120 (40.0)	<0.001
Weight (kg) mean (SD)	78.6 (16.0)	76.8 (15.3)	87.2 (16.5)	<0.001
Height (cm) mean (SD)	169.1 (9.6)	169.4 (9.7)	167.2 (9.1)	<0.001
Alcohol (g/day) mean (SD)	14.2 (19.4)	13.9 (18.1)	15.8 (24.7)	0.115
Waist (cm) mean (SD)	95.6 (14.0)	93.1 (12.9)	107.8 (12.7)	<0.001
Waist-hip-ratio mean (SD)	0.9 (0.1)	0.9 (0.1)	1.0 (0.1)	<0.001
Fasting glucose (mmol/L) mean (SD)	5.6 (1.3)	5.2 (0.4)	7.6 (2.0)	<0.001
2 h post glucose (mmol/L) mean (SD)	5.8 (2.2)	5.4 (1.1)	12.6 (3.5)	<0.001
Systolic blood pressure (mmHg) mean (SD)	117.9 (17.2)	116.1 (16.2)	126.7 (18.8)	<0.001
Diastolic blood pressure (mmHg) mean (SD)	72.7 (9.5)	72.8 (9.1)	72.0 (11.1)	0.201
Smoking (%)				<0.001
Smoker	267 (15.6)	243 (17.2)	24 (8.0)	
Ex-smoker	658 (38.4)	524 (37.0)	134 (44.7)	
Never-smoker	790 (46.1)	648 (45.8)	142 (47.3)	
Physical activities inactive (%)	702 (40.9)	512 (36.2)	190 (63.3)	
HDL cholesterol (mmol/L) mean (SD)	1.72 (0.5)	1.76 (0.5)	1.48 (0.4)	<0.001
LDL cholesterol (mmol/L) mean (SD)	3.5 (0.9)	3.5 (0.9)	3.3 (0.9)	<0.001
Triglycerides (mmol/L) mean (SD)	1.4 (0.8)	1.3 (0.8)	1.8 (1.0)	<0.001
HbA1c (%) mean (SD)	5.5 (0.7)	5.3 (0.3)	6.5 (1.0)	<0.001
Total cholesterol (mmol/L) mean (SD)	5.6 (1.0)	5.6 (1.0)	5.3 (1.1)	<0.001
C-reactive protein (mg/L) mean (SD)	2.3 (4.4)	2.1 (4.3)	3.4 (4.6)	<0.001
BMI = Obese (%)	439 (25.6)	275 (19.4)	164 (54.7)	<0.001

3.1.2. Metabolites Associated with BMI and T2D

A linear regression model was used to investigate the BMI associated metabolites and a logistic regression model was employed for T2D associations. Model assumptions have been performed and reported in Supplemental Document S2. Only age and sex (adding BMI for T2D model) were added in the basic regression models. The numbers of significant metabolites were the highest, and 83 metabolites were significantly associated with BMI and 51 metabolites were significantly associated with T2D.

Next, we tested how covariates like lifestyle, lipids, and fasting glucose influenced the association between metabolites with BMI and T2D. When more covariates were included, the significant numbers decreased. In particular, the association between BMI and metabolites was affected mostly by lipids and blood pressure, which was indicated from the dramatically dropped number when lipids and blood pressure were added in the model. Fasting glucose influenced mostly the T2D association and the number of significant metabolites decreased from 41 to 3, which suggests many metabolites were associated with T2D mediated by fasting glucose (Figure 1).

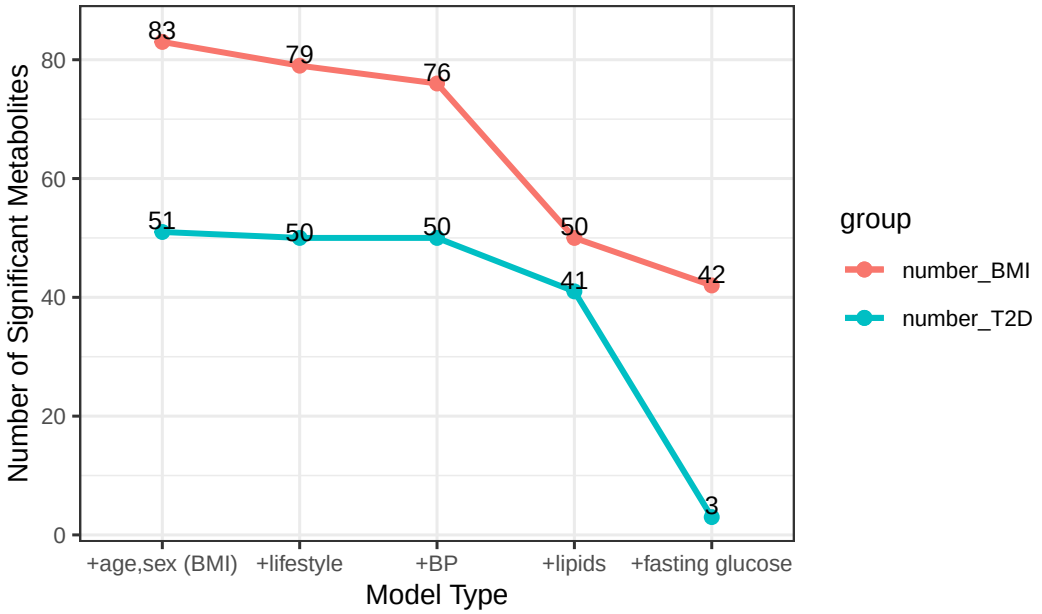


Figure 1. The number of metabolites significantly associated with BMI and T2D in different models after multiple testing correction. The first coordinate on *x*-axis shows basic model building upwards with including lifestyle, blood pressure, lipids, and fasting glucose parameters as covariates in the model. The *y*-axis depicts a number of significant metabolites resulting from each model as indicated on *x*-axis. Lifestyle includes smoking status and physical activities. BP: systolic blood pressure; lipids include HDL cholesterol (HDL-C) and triglycerides.

Obesity specific metabolites: Linear regression was used to execute a metabolite-wide association study in KORA FF4, and we identified 83 and 42 metabolites associations in the basic and full models after conservative Bonferroni correction for multiple testing. A volcano plot (Figure 2A,B) provides a quick visual identification of statistically significant metabolites with a larger effect size. The full summary statistics of different models are reported in the Supplemental Materials Tables S2 and S3. Table 3 shows only the metabolites significantly associated with BMI in the full model. Totally, 12 metabolites were negatively associated with BMI whereas 30 were positively associated in the full model. We confirmed the BMI metabolites associations using the published literature and almost all were replicated except for SM C20:2.

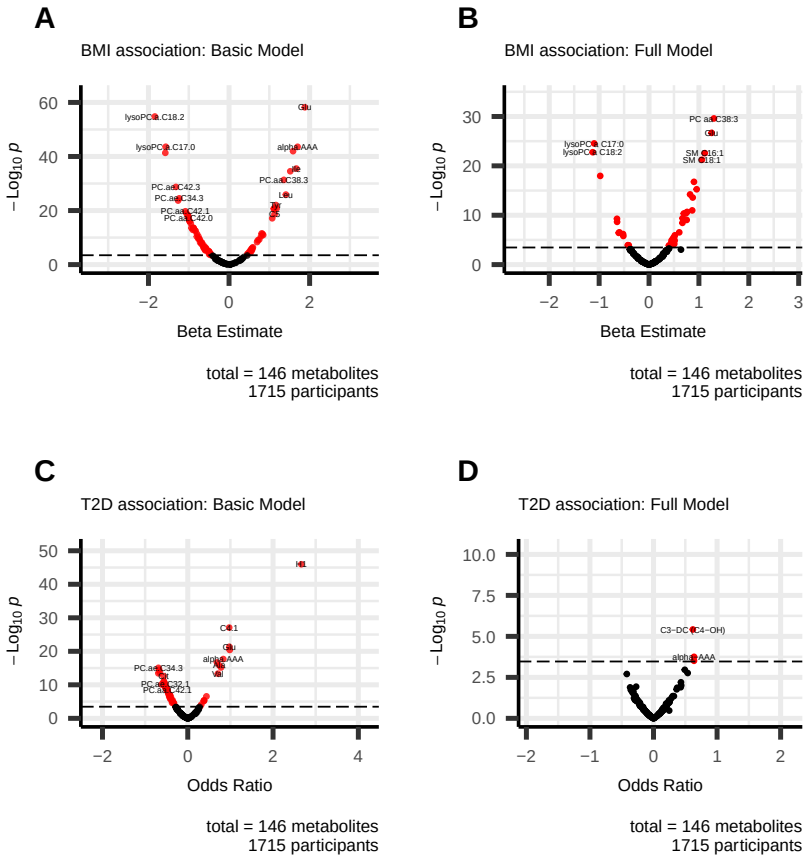


Figure 2. Volcano plots show the association of metabolites with BMI and T2D in the basic model (**A**,**C**) and the full model (**B**,**D**). Bonferroni correction *p*-value cut-off is 0.05/146 = 0.00034 was considered. Each dot represents a metabolite, and they are displayed based on the beta estimate or odds ratio (*x*-axis) and the negative logarithm (base 10) of the *p*-value (*y*-axis). The covariates for the basic model are age, sex, and (BMI); the covariates for the full model are age, sex, (BMI), smoking status, physical activities, HDL-C, blood pressure, triglycerides, and fasting glucose.

From this analysis we made the following four key observations.

(1) We have observed that all diacyl phosphatidylcholines (PC aa), acylcarnitines, biogenic amines, and sphingomyelins (SM) were positively associated with BMI. In particular, PC aa C38:3 was the strongest metabolite associated with BMI (1.301 [1.082–1.520], q-value = 3.65×10^{-28}. Glutamate (1.255 [1.032–1.478], q-value = 3.05×10^{-25}), SM C16:1 (1.118 [0.901–1.336], q-value = 3.87×10^{-21}), alpha-AAA (0.955 [0.726–1.184], q-value = 8.04×10^{-14}), and C0 (0.672 [0.462–0.882], q-value = 6.13×10^{-8}) were those with the strongest association in each category;

(2) Some amino acids were positively correlated with BMI. Among them, glutamate (1.255 [1.032–1.478], q-value = 3.05×10^{-25}) and Tyrosine (0.901 [0.695–1.106], q-value = 2.51×10^{-15}) have the strongest association. Others were inversely associated with BMI: Asparagine (−0.642 [−0.843–−0.44], q-value = 7.73×10^{-8}) and Glycine (−0.515 [−0.724–−0.305], q-value = 2.34×10^{-4});

(3) Three acylalkylphosphatidylcholine (PC ae) were positively associated with BMI, PC ae C36:5 (0.502 [0.29–0.713], q-value = 5.09×10^{-4}), PC ae C36:4 (0.457 [0.254–0.66], q-value = 1.56×10^{-3}), and PC ae C32:2 (0.506 [0.258–0.754], q-value = 9.52×10^{-3}); whereas others PC aes were negatively associated with BMI: PC ae C42:3 (−0.594 [−0.821–−0.368], q-value = 4.29×10^{-5}), PC ae C36:2 (−0.607 [−0.84–−0.373], q-value = 5.48×10^{-5}), PC ae C40:6 (−0.424 [−0.639–−0.209],

q-value = 1.66×10^{-2}), and PC ae C38:2 (−0.406 [−0.613–−0.199], q-value = 1.80×10^{-2});

(4) All lysophosphatidylcholines (lyso PC) were negatively associated with BMI. In particular, lysoPC a C17:0 (−1.1 [−1.305-−0.896], q-value = 4.20×10^{-23}) was the strongest.

Table 3. Metabolites significantly associated with BMI in the linear regression full model. The dependent variable was BMI, whereas the independent variables were the log transformed and standardized concentration of each metabolite, adjusted for age, sex, smoking status, physical activities, HDL-C, blood pressure, triglycerides, and fasting glucose. q-values were reported as *p* values adjusted for multiple testing by Bonferroni correction. Only metabolites with a *p*-value lower than 0.00034 (0.05/146) were included in this table.

Positively Associated				
Category	Metabolite	Beta Estimate (95% CI)	*p*-value	q-value
PC aa	PC aa C38:3	1.301 (1.082–1.520)	2.50×10^{-30}	3.65×10^{-28}
PC aa	PC aa C38:4	0.728 (0.514–0.943)	3.74×10^{-11}	5.47×10^{-9}
PC aa	PC aa C40:4	0.692 (0.471–0.913)	9.89×10^{-11}	1.44×10^{-7}
PC aa	PC aa C32:1	0.606 (0.375–0.837)	2.93×10^{-7}	4.28×10^{-5}
PC aa	PC aa C40:5	0.505 (0.279–0.730)	1.19×10^{-5}	1.74×10^{-3}
PC aa	PC aa C36:3	0.512 (0.281–0.742)	1.41×10^{-5}	2.06×10^{-3}
PC aa	PC aa C36:4	0.426 (0.207–0.644)	1.38×10^{-4}	2.01×10^{-2}
Amino Acids	Glutamate (Glu)	1.255 (1.032–1.478)	2.09×10^{-27}	3.05×10^{-25}
Amino Acids	Tyrosine (Tyr)	0.901 (0.695–1.106)	1.72×10^{-17}	2.51×10^{-15}
Amino Acids	Phenylalanine (Phe)	0.823 (0.618–1.027)	6.11×10^{-15}	8.92×10^{-13}
Amino Acids	Valine (Val)	0.876 (0.652–1.100)	2.60×10^{-14}	3.80×10^{-12}
Amino Acids	Isoleucine (Ile)	0.866 (0.618–1.114)	1.05×10^{-11}	1.53×10^{-9}
Amino Acids	Leucine (Leu)	0.755 (0.515–0.995)	9.02×10^{-10}	1.32×10^{-7}
Amino Acids	Alanine (Ala)	0.458 (0.242–0.673)	3.27×10^{-5}	4.78×10^{-3}
Amino Acids	Ornithine (Orn)	0.399 (0.195–0.603)	1.30×10^{-4}	1.90×10^{-2}
SM	SM C16:1	1.118 (0.901–1.336)	2.65×10^{-23}	3.87×10^{-21}
SM	SM C18:1	1.061 (0.848–1.273)	5.81×10^{-22}	8.48×10^{-20}
SM	SM C20:2	0.763 (0.541–0.985)	2.14×10^{-11}	3.12×10^{-9}
SM	SM C18:0	0.697 (0.490–0.903)	4.52×10^{-11}	6.60×10^{-9}
SM	SM C24:1	0.518 (0.310–0.726)	1.16×10^{-6}	1.69×10^{-4}
Biogenic Amines	Alpha-Amino acid (alpha-AAA)	0.955 (0.726–1.184)	5.51×10^{-16}	8.04×10^{-14}
Biogenic Amines	Kynurenine	0.743 (0.524–0.962)	3.81×10^{-11}	5.57×10^{-9}
Biogenic Amines	4-Hydroxyproline (t4-OH-Pro)	0.485 (0.279–0.691)	4.13×10^{-6}	6.02×10^{-4}
Acylcarnitines	Carnitine (C0)	0.672 (0.462 -0.882)	4.20×10^{-10}	6.13×10^{-8}
Acylcarnitines	Valerylcarnitine (C5)	0.700 (0.478–0.922)	7.96×10^{-10}	1.16×10^{-7}
Acylcarnitines	Propionylcarnitine (C3)	0.670 (0.449–0.891)	3.50×10^{-9}	5.11×10^{-7}
Acylcarnitines	Butyrylcarnitine (C4)	0.457 (0.247–0.667)	2.15×10^{-5}	3.14×10^{-3}
PC ae	PC ae C36:5	0.502 (0.290–0.713)	3.49×10^{-6}	5.09×10^{-4}
PC ae	PC ae C36:4	0.457 (0.254–0.660)	1.07×10^{-5}	1.56×10^{-3}
PC ae	PC ae C32:2	0.506 (0.258–0.754)	6.52×10^{-5}	9.52×10^{-3}
Negatively Associated				
Category	Metabolite	Beta Estimate (95% CI)	*p*-value	q-value
lysoPC	lysoPC a C17:0	−1.1 (−1.305–−0.896)	2.88×10^{-25}	4.20×10^{-23}
lysoPC	lysoPC a C18:2	−1.129 (−1.348–−0.911)	1.72×10^{-23}	2.51×10^{-21}
lysoPC	lysoPC a C18:1	−0.978 (−1.193–−0.763)	1.08×10^{-18}	8.72×10^{-15}
lysoPC	lysoPC a C16:0	−0.640 (−0.849–−0.432)	2.19×10^{-9}	3.20×10^{-7}
lysoPC	lysoPC a C18:0	−0.521 (−0.725–−0.316)	6.48×10^{-7}	9.46×10^{-5}
lysoPC	lysoPC a C20:4	−0.415 (−0.627–−0.203)	1.28×10^{-4}	1.86×10^{-2}
Amino Acids	Asparagine (Asn)	−0.642 (−0.843–−0.44)	5.30×10^{-10}	7.73×10^{-8}
Amino Acids	Glycine (Gly)	−0.515 (−0.724–−0.305)	1.60×10^{-6}	2.34×10^{-4}
PC ae	PC ae C42:3	−0.594 (−0.821–−0.368)	2.94×10^{-7}	4.29×10^{-5}
PC ae	PC ae C36:2	−0.607 (−0.840–−0.373)	3.75×10^{-7}	5.48×10^{-5}
PC ae	PC ae C40:6	−0.424 (−0.639–−0.209)	1.14×10^{-4}	1.66×10^{-2}
PC ae	PC ae C38:2	−0.406 (−0.613–−0.199)	1.23×10^{-4}	1.80×10^{-2}

To investigate the direction of effect across BMI class (normal, overweight, and obese), the six most significant metabolites from the full model were visualized by violin-box plots stratified by BMI in Figure 3. PC aa C38:3, glutameta (Glu), SM C16:1 and SM C18:1 showed synchronized direction with BMI, increasing concentrations with increased BMI, whereas lysoPC a C17:0 and lysoPC a C18:2 reversed, which is consistent with the result from the linear regression model.

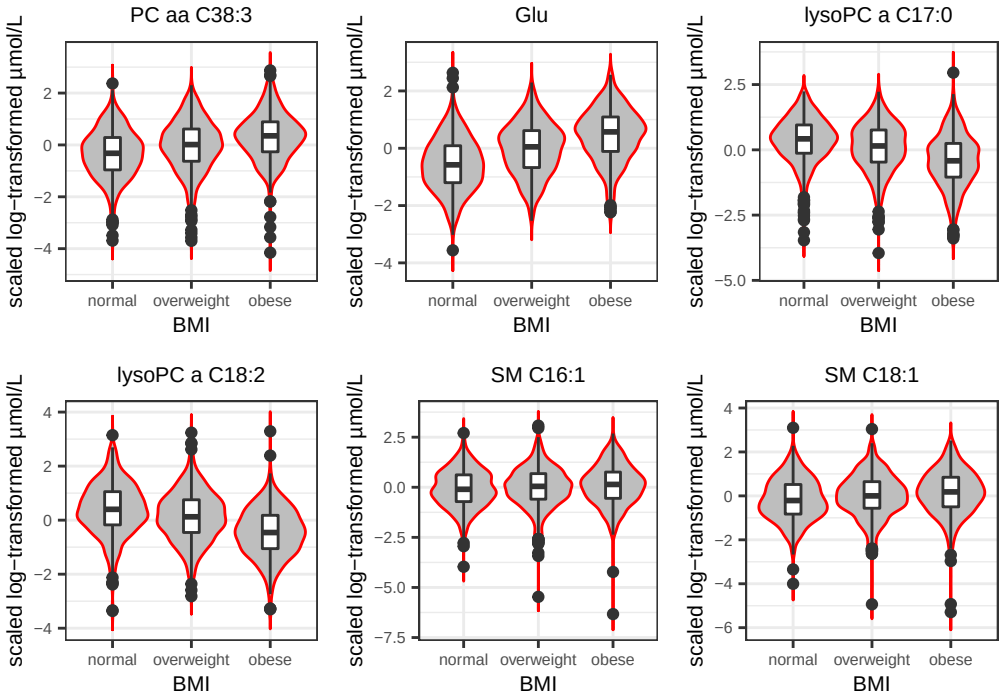


Figure 3. Violin-boxplots show the top six significant metabolite distributions of study subjects divided in three different classes of BMI, normal ($18.5 \leq$ BMI < 25), overweight ($25 \leq$ BMI < 30), and obese (BMI ≥ 30). The box contains 50% of the participants. The middle line stands for median dividing the box into two areas. The 25th and 75th percentile of the distribution are represented by upper and lower hinges.

T2D specific metabolites: multivariable logistic regression analysis was conducted with known diabetes-related variables as covariates to identify significant metabolites. Similarly, alcohol was not included in the model as a covariate because there was no significant difference between T2D and healthy individuals. A volcano plot (Figure 2C,D) represents the result of the logistic regression model. The full summary statistics of different models are reported in the Supplemental Materials Tables S4 and S5. Table 4 shows only the metabolites significantly associated with T2D in the full model. Three metabolites, C3-DC (C4-OH), alpha-AAA and isoleucine (Ile) were observed to have significant associations in the full model after conservative Bonferroni correction. All of them were positively correlated with T2D and replicated by the published literature (details in the Section 4).

Table 4. Metabolites significantly associated with T2D in the logistic regression full model. The dependent variable was T2D status, whereas the independent variables were the log transformed and standardized concentration of each metabolite, adjusted for age, sex, BMI, smoking status, physical activities, HDL-C, blood pressure, triglycerides, and fasting glucose. q-values were reported as *p*-values adjusted for multiple testing by Bonferroni correction. Only metabolites with *p*-value lower than 0.00034 (0.05/146) were included in this table.

Category	Metabolite	Odds Ratios (95% CI)	*p*-Value	q-Value
Acylcarnitines	Hydroxybutyrylcarnitine (C3-DC (C4-OH))	0.619 (0.363–0.888)	3.79×10^{-6}	5.54×10^{-4}
Biogenic Amines	Alpha-Amino acid (alpha-AAA)	0.638 (0.308–0.977)	1.77×10^{-4}	2.58×10^{-2}
Amino Acids	Isoleucine (Ile)	0.637 (0.293–0.987)	3.08×10^{-4}	4.50×10^{-2}

Figure 4 displays the violin-boxplots of the three significant metabolites in the T2D full model. The concentrations of C3-DC (C4-OH), alpha-AAA, and Ile increased among the group with T2D, which is consistent with the result from the logistic regression model.

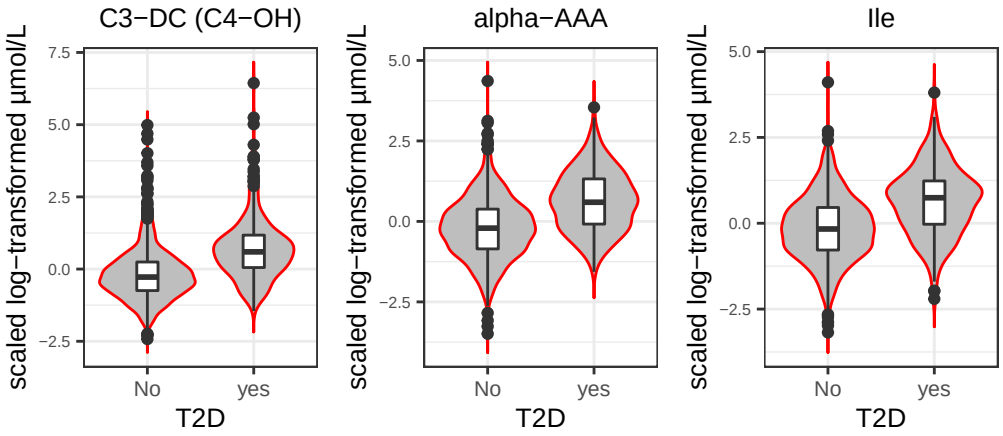


Figure 4. Violin-box plots show the distribution of three significant metabolites stratified by diabetic status. The box contains 50% of the observations. The middle line stands for median dividing the box into two areas. The 25th and 75th percentile of the distribution are represented by upper and lower hinges.

3.2. Sobel Mediation Test

A Sobel mediation test was conducted to investigate whether a mediator carries the effect of an independent variable on a dependent variable. In our research, we used fasting glucose or HbA1c as T2D indicators to test the metabolite mediation of the effect of BMI on T2D. In order to adjust the influence of the confounders, the metabolite residual, calculated from the linear regression model between each metabolite and covariates, was used as a mediator in the test. The significant mediators are shown in Table 5 and full statistics are shown in Supplementary Materials Table S6 and Table S7, respectively. The mediation of the associations between BMI and fasting glucose via the 12 metabolites were Bonferroni-corrected significant (q-value < 0.05) whereas nine metabolite mediations were significant between BMI and HbA1c. Among all these metabolites, sum of hexose, SM C16:1, glutamate, PC aa C38:3, alpha-AAA, isoleucine, lyso PC a C18:0, and leucine were significant in both tests, which suggests their robust mediation effects. The sum of hexose owned the strongest mediation in both studies, which was not very surprising as it mainly represents the glucose in human blood. A summarizing plot of the mediation analysis is shown in Figure 5.

Table 5. Results for mediation analysis with the BMI as independent variable, metabolite as potential mediator, fasting glucose or HbA1c as dependent variable. q-values were reported as *p*-value adjusted for multiple testing by Bonferroni correction.

Sobel Test (Metabolite, BMI, Fasting Glucose)			Sobel Test (Metabolite, BMI, HbA1c)		
Metabolite	***p*-Value**	**q-Value**	**Metabolite**	***p*-Value**	**q-Value**
Sum of hexoses (H1)	1.49×10^{-16}	2.18×10^{-14}	Sum of hexoses (H1)	1.14×10^{-15}	1.66×10^{-13}
SM C16:1	2.88×10^{-7}	4.20×10^{-5}	Isoleucine (Ile)	1.08×10^{-5}	1.58×10^{-3}
Glutamate (Glu)	1.27×10^{-6}	1.85×10^{-4}	SM C16:1	1.40×10^{-5}	2.04×10^{-3}
PC aa C38:3	2.62×10^{-6}	3.82×10^{-4}	lysoPC a C18:0	5.56×10^{-5}	8.11×10^{-3}
lysoPC a C17:0	1.31×10^{-5}	1.91×10^{-3}	Leucine (Leu)	1.05×10^{-4}	1.53×10^{-2}
Alpha-Amino acid (alpha-AAA)	1.58×10^{-5}	2.3×10^{-3}	Glutamate (Glu)	1.06×10^{-4}	1.55×10^{-2}
Isoleucine (Ile)	1.95×10^{-5}	2.84×10^{-3}	lysoPC a C16:0	1.12×10^{-4}	1.63×10^{-2}
lysoPC a C18:0	5.00×10^{-5}	7.30×10^{-3}	Alpha-Amino acid (alpha-AAA)	1.48×10^{-4}	2.16×10^{-2}
Alanine (Ala)	6.94×10^{-5}	1.01×10^{-2}	PC aa C38:3	3.14×10^{-4}	4.59×10^{-2}
SM C18:1	1.33×10^{-4}	1.94×10^{-2}			
Leucine (Leu)	1.48×10^{-4}	2.16×10^{-2}			
SM C20:2	2.91×10^{-4}	4.24×10^{-2}			

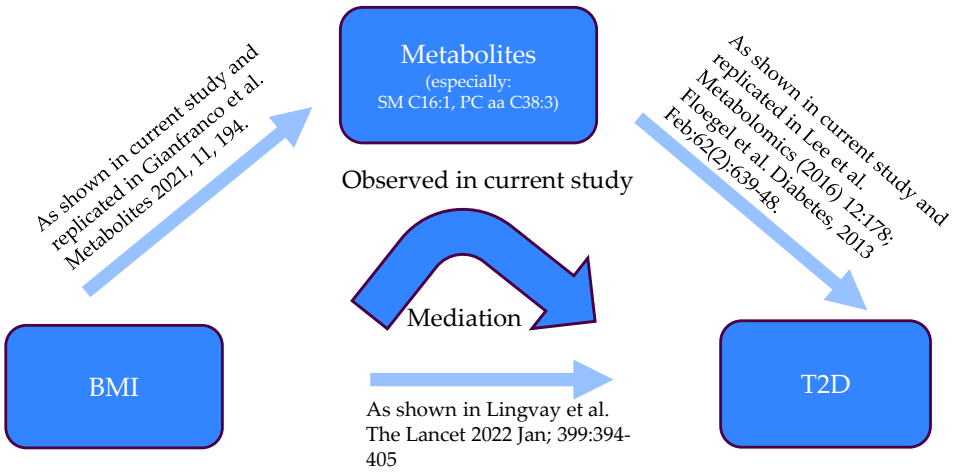
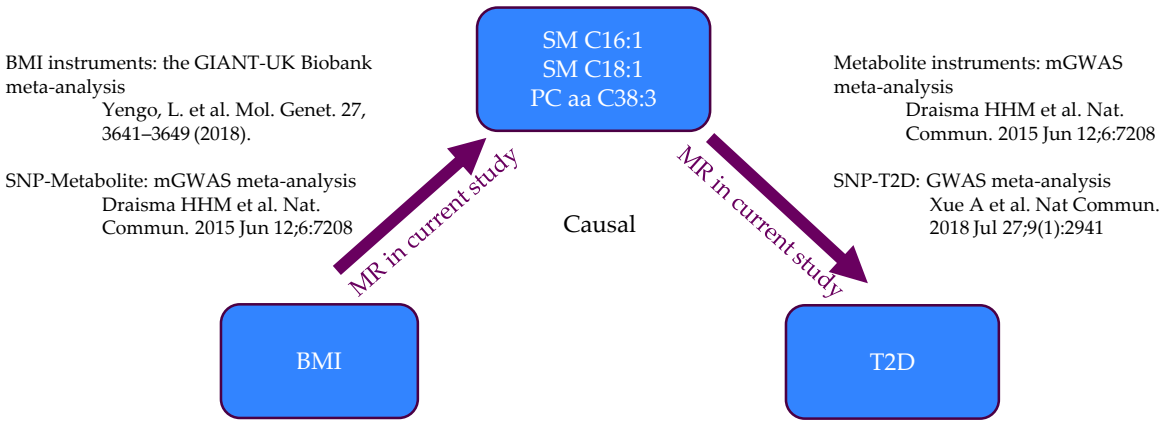


Figure 5. Schematic representation of the mediation analysis [2,30–32].

3.3. Mendelian Randomization

To assess the causality relationship between BMI, identified metabolites from mediation test and T2D, we employed two-sample mendelian randomization (MR) tests. We conducted a two-sample (2SMR) mendelian randomization analysis in two directions (BMI-to-metabolite, metabolite-to-T2D, Figure 6). BMI instruments were extracted from the GIANT-UK Biobank meta-analysis [16] and then the corresponding SNPs estimated on T2D were selected from the published metabolite-GWAS [28]. Metabolite instruments were obtained from the same metabolite-GWAS [28] and extracted the corresponding SNPs from the GWAS meta-analysis [29]. The 2SMR analysis results are presented using the Inverse Variance Weighted (IVW) method in BMI to metabolite direction and the Wald ratio method in metabolite to T2D direction. Only SM C16:1, SM C18:1, and PC aa C38:3 have available instruments in both directions, so we showed the MR results of these three metabolites in this study.



Metabolite	Beta Estimate	se	p-value
BMI to SM C16:1	0.0234	0.0117	4.6×10^{-2}
BMI to SM C18:1	0.0201	0.0142	0.156
BMI to PC aa C38:3	0.0434	0.0162	7.21×10^{-3}

Metabolite	Beta Estimate	se	p-value
SM C16:1 to T2D	0.896	0.288	1.89×10^{-3}
SM C18:1 to T2D	0.790	0.254	1.89×10^{-3}
PC aa C38:3 to T2D	0.462	0.191	1.56×10^{-2}

Figure 6. Schematic diagram is suggestive of relationships between BMI, metabolites and T2D. The studies we used for MR were listed in the figure. ß-estimate stands for beta coefficient, se stands for standard error [16,28,29].

Our results indicated that the change of BMI could cause the concentration change of SM C16:1 and PC aa C38:3. The change of SM C16:1, SM C18:1, and PC aa C38:3 contributes to the development of T2D, which suggests lipids like SM C16:1 and PC aa C38:3 are intermediate molecules involved in the progression from obesity to T2D. Sensitivity analysis was carried out to test if these results were robust from proof of heterogeneity or horizontal pleiotropy, which was supported by the MR-Egger analysis. For BMI to SM C16:1, Q statistic from the heterogeneity measure was not significant (p_Het 0.51 > 0.05), indicating there was no heterogeneity. For BMI to PC aa C38:3, the *p*-value (p_Het 0.03) was slightly lower than 0.05, showing heterogeneity between different instruments, and random effect was selected to report the result. The MR–Egger intercept test (p_Pleio > 0.05) suggested no directional pleiotropy for both metabolites. For the direction of metabolites to T2D, we did not perform the sensitivity analysis as only one SNP instrument was available for each metabolite.

3.4. The Biological Role of SM C16:1, SM C18:1, and PC aa C38:3 in Transition to T2D

In order to understand the biological pathway of these three lipids (SM C16:1, SM C18:1, and PC aa C38:3), we searched for the associated SNPs and genes in humans. The metabolite SM C18:1 was reported to be associated with SNP rs12610250-A, the locus CERS4 [28]. PC aa C38:3 was significantly correlated with rs7200543-A, locus PDXDC1 and rs968567-T, locus FADS2 [28]. Both SM C16:1 and SM C18:1 were associated with rs174547-C, rs174537-A, rs102275-G, rs174546-A, rs174556-A, rs1535-G, rs174449-G, rs1000778-A, the locus FADS1-3 [33]. CERS4 and FADS1-3 were identified to influence the biosynthesis of sphingolipids including sphingomyelins and ceramides [28,33], which could be produced from each other by hydrolysis and synthase [34]. It was reported sphingomyelins were essential for insulin secretion in rat beta cells [35] and beta cell viability [36]. Mice model and cell experiments demonstrated that inhibition of ceramide biosynthesis impaired insulin sensitivity and caused pancreatic beta-cell dysfunction [36,37]. This is consistent with the result of negative associations between SM C16:1, SM C18:1, and T2D in the current study (basic model). The specific variants of PDXDC1 and FADS2 were found to upregulate phosphatidylcholine [28]. Increased phosphatidylcholines bind to

and activate the aryl hydrocarbon receptor (AhR) expressed in hepatocytes and inhibition of the essential genes including IRS-2 for promotion of the insulin pathway [38]. We observed the consistent result that PC aa C38:3 was positively associated with T2D in a human study [32]. These observations support the particular sphingolipid and phosphatidylcholine dysmetabolism as a causal factor in early-stage T2D progression (shown in Figure 7).

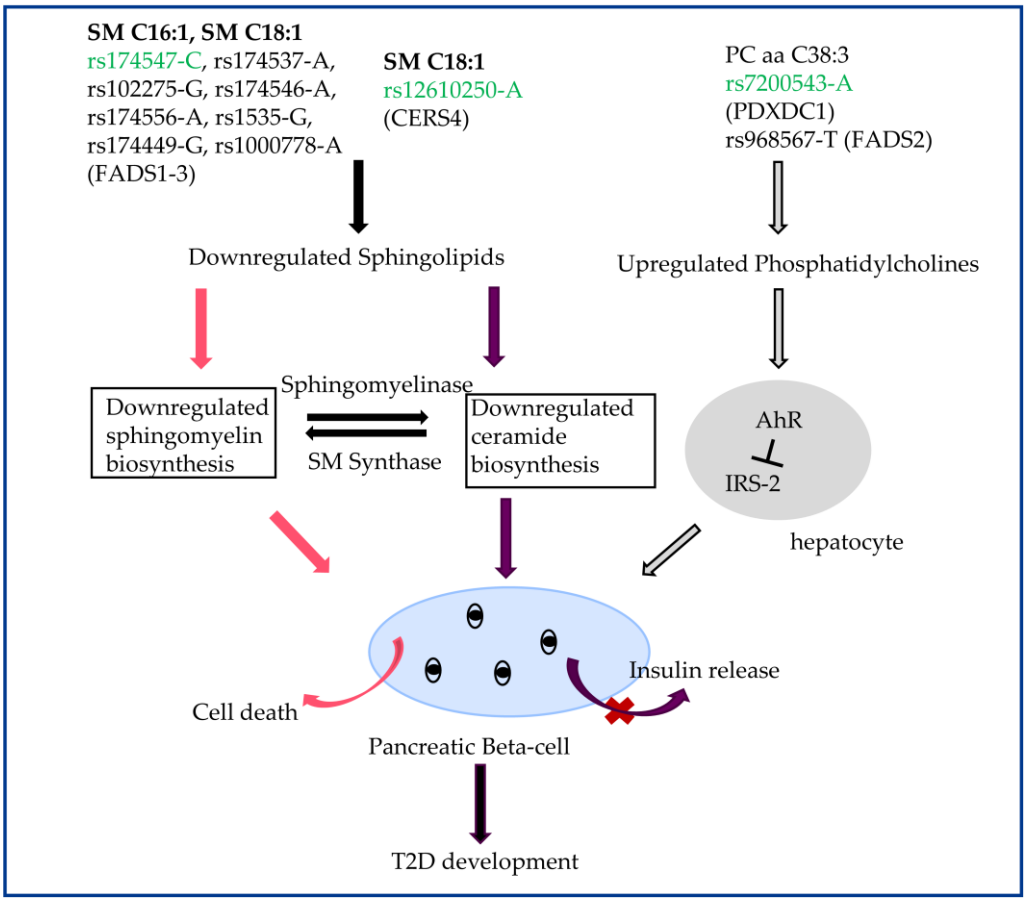


Figure 7. Schematic representation of the pathway analysis of diminished sphingolipid metabolism to a transition of T2D. The SNPs marked with green are the ones used in the MR test. The red pathway is generally involved in sphingomyelins (SM), the purple and gray pathways are for ceramides, and phosphatidylcholines (PC), respectively. All three kinds of metabolites influence insulin release.

4. Discussion

Obesity triggers a cascade of metabolic processes that raise the stake of various comorbidities including insulin resistance and glycemic deterioration causing T2D. Understanding the role of intermediate molecules involved in the process from obesity to T2D offers a therapeutic strategy to early-stage T2D pathophysiology. In our study, we assessed the functionally characterized targeted metabolite profiles of KORA FF4 participants for underlying metabolic pathway links. The major results of the present study are (1) identification of several metabolite changes among subjects with obesity and diabetic status, (2) metabolites such as SM C16:1, SM C18:1, and PC aa C38:3 show significant mediation effect of BMI on T2D, (3) the causality direction of BMI, three lipids (SM C16:1, SM C18:1, PC aa C38:3), and T2D, and (4) the biological consequences of the downregulated sphingolipids and upregulated phosphatidylcholine.

It is strongly suggested that in blood, elevated concentrations of branched-chain amino acids are associated with an increased risk of type 2 diabetes mellitus [39,40]. In our

study among metabolites associated with BMI, the branched chain amino acids (BCAAs), isoleucine (Ile), leucine (Leu), and valine (Val) were positively correlated and have been confirmed in several studies [30,41,42].In fact, isoleucine was positively associated with T2D in the full model and replicated in the literature [21]. Isoleucine (Ile) and leucine (Leu) also appear to be mediators between BMI and T2D. Other amino acids such as glutamate, alanine, tyrosine, and phenylalanine significantly changed among different BMIs and these also have been found in other studies [30,43,44]. Other studies speculate the reason could be that high concentration of BCAAs causes insulin resistance by activating the mammalian target of rapamycin (mTOR) signaling [45,46]. There might be a mechanism proposed for branched-chain-keto acid dehydrogenase (BCKD) inhibition and suppression of enzymatic catabolism of amino acids in individuals with obesity [47].

Acylcarnitines like carnitine (C0), valerylcarnitine (C5), propionylcarnitine (C3) increased in individuals with higher BMI, which is in line with other studies [30,44]. Hydroxybutyrylcarnitine (C3-DC (C4-OH)) was positively associated with T2D [48]. Several studies indicate an increase in plasma acyl carnitines in patients with T2D [30,31] and it is attributed to an incomplete long chain fatty acyl-CoA oxidation of fatty acids [43,49].

Biogenic amines were found to be related with obesity and T2D. Alpha-aminoadipic acid (alpha-AAA) and kynurenine were positively associated with BMI. Meantime, alpha-aminoadipic acid was also positively associated with T2D in the full model and showed significant mediation of BMI to T2D. Alpha-aminoadipic acid is an intermediate in the metabolism of lysine and rat studies indicate that aminoadipic acid is elevated in the pre-diabetic state and so it could be a predictive biomarker for the development of diabetes [50].

Considering glycerophospholipids, all diacylphosphatidylcholines (PC aa) increased with increased BMI, such as PC aa C38:3, PC aa C38:4, PC aa C40:4, PC aa C32:1, and especially PC aa C38:3, the strongest metabolite with the lowest *p*-value, which is in line with Frigerio et al. [30]. All lysophosphatidylcholines (lyso PCs) were observed to have negative association with BMI. lysoPC a C17:0, lysoPC C18:2, and lysoPC C18:1 were the strongest negatively correlated with BMI, consistent with several other studies [10,51]. Only a few acylalkylphosphatidylcholine (PC ae) increased with BMI (PC ae C36:5, PC ae C36:4, PC ae C32:2) whereas many decreased (PC ae C42:3, PC ae C36:2, PC ae C40:6, PC ae C38:2). Moreover, PC aa C38:3, LysoPC a C16:0, LysoPC a C17:0, and LysoPC a C18:0 were observed to mediate from BMI to T2D, and this is a novel finding in our study. Phospholipids such as phosphatidylcholines (PC) are the essential constituent of cellular membranes and are critical for cellular signal transduction [52]. The LysoPCs (16:0, 17:0, 18:0) negatively associated with T2D in the basic model in our cohort have been considered to be involved in pro-inflammatory and atherogenic [53], but their major role still needs to be elucidated. PC aa C38:3 is reported to be positively associated with incident T2D [32], and mediation analysis and mendelian randomization results indicate it could be the intermediate molecules involved in obesity-related T2D development. The mechanisms governing the PC-mediated association between obesity and T2D could be via fatty acid (FA) and insulin signaling pathways. High-fat diets, inducing overproduction of PC, result in obesity and diabetes in individuals [54,55]. It is stated that abnormally high PC lipids affect energy metabolism and insulin signaling [56,57]. Mice fed with high-fat diets show upregulation of exosomal phosphatidylcholine, which results in binding to the aryl hydrocarbon receptor (AhR) [38], a transcription factor expressed in hepatocytes to integrate dietary and metabolic processes, and thus inhibition of the insulin response.

The Frigerio et al. study [30] confirms that sphingomyelins (SM), SM C16:1, and SM C18:1 were significantly associated with BMI. In the mediation test, both SM C16:1 and SM C18:1 have significant mediation effects of BMI on fasting glucose. These two metabolites have been shown to be associated with BMI and T2D in other studies [30,31]. Integrating with mendelian randomization suggests the causality direction and sphingomyelins such as SM C16:1 could be the molecular mediators of obesity-to-T2D evolution. Sphingomyelins are one of the most abundant sphingolipids in bodily fluids and in tissues, which is a lipid class with both signaling and structural properties and was reported to be

related to the development of major metabolic and cardiovascular diseases [58–60]. The metabolic link between obesity and diabetes could be induced by modulating inflammation via FA and proinflammatory cytokines. Increased bioavailability of free fatty acid (FFA) and proinflammatory cytokines are characterized in obese subjects; sphingolipid metabolism is affected through both substrate supply and regulation of the enzymes [61,62]. Through the use in vivo and vitro mice models, it is confirmed that saturated FAs stimulate toll-like receptor 4 (TLR-4), activating sphingomyelinase (SMase) and converting sphingomyelins to ceramide, which reduces sphingomyelins content and exerts an action of insulin resistance [63]. SMase is also observed to be activated by proinflammatory cytokines tumor necrosis factor-alpha (TNF-α), resulting in an increased ceramide production from C57BL/6J mice with the intraperitoneal administration of TNF-α [64]. These events can lead to pancreatic β-cell dysfunction and T2D development in obese subjects. A study by Kelli M Sas et al. [65] investigates the role of perturbed ceramide metabolism in diabetic kidney disease (DKD). Ceramides were measured in the plasma and kidney cortex of a C57BLKS db/db mouse model of DKD which revealed long-chain ceramides (C14:0, C16:0, C18:0, C20:0) and a glucosylceramide (Glu-Cer C18:0) were increased in diabetic mouse plasma, whereas very-long-chain (C24:0, C24:1) ceramides and glucosylceramide (Glu-Cer C16:0) were decreased in diabetic mouse kidney tissue. However, circulating metabolites from the KORA study show exactly the opposite role of ceramide through SMase and genetics variants.

T2D usually occurs at the later stage of obesity, and we confirmed that lipids like SM C16:1, SM C18:1 and PC aa C38:3 could mediate the effect of BMI on T2D and also be a causal factor for T2D development. Therefore, we incorporated human genetics with mice model experiments to figure out the biological pathway. It was reported that FADS1-3 and CERS4 genetic variants with specific minor alleles (Figure 7) are associated with downregulated sphingolipids [28,33] whereas PDXDC1 and FADS2 upregulated phosphatidylcholine (Figure 7) [28], which contributes to promoting T2D pathophysiology [33,38]. CERS4 is the gene responsible for encoding ceramide synthases. Several knockout mice studies report that the inhibition of ceramide biosynthesis provokes both insulin resistance and the glucose homeostasis disruption [37,66,67]. This is contradictory with the above section which states increased ceramide causes insulin resistance. It may be attributed to that only general routes of metabolism are discussed, and specific sphingolipid species and sphingolipid metabolic pathways stay unintelligible. The function of the PDXDC1 protein, a vitamin B6-dependent decarboxylase, is not well known. It was observed in previous GWAS that PDXDC1 is linked with omega-3 (n-3) and omega-6 (n-6) polyunsaturated fatty acids (PUFAs) [68,69]. Insulin-resistance in mice induced by high-fat diets showed downregulation PDXDC1 in the liver [70]. These events suggest PDXDC1 plays a role in the fatty acids metabolism to influence phosphatidylcholine biosynthesis, regulating the risk of insulin resistance and T2D. The FADS1-3 genetic locus, which encodes FA desaturase enzymes, derive PUFAs via endogenous desaturation and elongation of fatty acids [71,72]. FADS1-3 are reported to share genome-wide significant associations with almost all cardiometabolic phenotypes such as dyslipidemia, fatty liver, obesity, and T2D [73–75]. The possible interpretation could be similar with PDXDC1—that the FADS genetic variants, which influence FA desaturase enzyme activity to affect sphingolipid and phosphatidylcholines biosynthesis, modulate the risking of developing T2D [76,77]. It has been observed that the FADS genes are associated with the differences in adipose tissue, body weight, and glucose homeostasis and these are regulated by PUFAs [78], which is consistent with our results that FADS1-3 have strong correlations with obesity and T2D traits in adipose, liver, and muscle tissues in ApoE−/− C57BL/6J and C3H/HeJ mice (Supplementary Figure S3). These data suggest genetic predisposition and early alterations in sphingolipids and phosphatidylcholines metabolism contribute to prediction of T2D incident.

This study has several advantages and limitations. A high number of participants were included in the study to investigate the metabolite signatures associated with obesity and T2D. We employed mediation testing to discover the novel metabolites which mediated

the effect from BMI on T2D. MR tests and mice model experiments from the literature were used to establish plausible biological pathways. The most important point from this study is that lipids SM C16:1, SM C18:1, and PC aa C38:3 could be biomarkers for early stage T2D diagnosis. However, there are still some limitations that could be investigated in further studies. It is reported that storage of plasma samples for up to five years results in altered concentrations of metabolites [22] and this may influence the associations. Sphingomyelins SM C16:1 and SM C18:1 were found to be positively associated with obesity but negatively with T2D (basic model) and this is also replicated in the literature [30,31]. This could be caused from SMase converting sphingomyelins to ceramide at the later stage of obesity [63,64] and could be the reason why sphingomyelins have a positive effect on incident T2D from MR results but were negatively associated with prevalent T2D in a cross-sectional study; however, the molecular mechanism was not confirmed. Longitudinal analyses could be performed to study how metabolite concentrations change at different stages and if they are able to predict the onset of obesity related T2D. In our study, we observed sphingolipids' metabolic pathway linked obesity and T2D but how specific metabolites SM C16:1, SM C18:1, and PC aa C38:3 work is still ambiguous and requires additional experiments to confirm more detailed molecular behavior. In the current study, metabolites were associated with BMI and T2D considering traditional covariates. Moreover, other complication factors like depressive symptoms or kidney disease or dietary intake might also have an influence on metabolic traits, which are not considered in this study.

5. Conclusions

This study assessed metabolic profiles from a targeted approach based on the KORA FF4 cohort. The cross-sectional analysis showed metabolic biomarkers related to obesity and T2D.For the first time, we show metabolites like SM C16:1, SM C18:1, and PC aa C38:3 performed significant mediation effects of BMI on T2D. MR analysis and mice model experiments provided new evidence in sphingolipid-driven alterations in insulin secretion and T2D development. This translates previous findings from mice models to the human metabolism. This study contributes to human validation of SM C16:1, SM C18:1, and PC aa C38:3 as biomarkers for obesity-related T2D pathophysiology that could be regarded as potential clinical targets for risk evaluation and disease monitoring. In conclusion, the findings reported here shed new light on new potential therapeutic strategies from the perspective of metabolic signatures.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/metabo13020227/s1, Supplemental Document S1: Figure S1: Human tissue-specific gene expression and regulation. Human RNA-seq data from GTEx showing the transcript per million (TPM) expression values for the genes encoding the proteins; Figure S2. Mice tissue-specific gene expression and regulation; Figure S3. The F2 dataset which is a cross of the inbred ApoE−/− C57BL/6J and C3H/HeJ strains fed a high fat + cholesterol diet; Table S1: Complete information about all the 146 considered metabolites, including the category, abbreviations and full name used; Table S2: Association of BMI with metabolites in the basic model; Table S3: Association of BMI with metabolites in the full model; Table S4, Association of T2D with metabolites in the basic model; Table S5, Association of T2D with metabolites in the full model. Table S6, mediation test of metabolites from BMI to T2D; Table S7, mediation test of metabolite residues from BMI to T2D. Table S8: Association of BMI with metabolites in the basic model for female individuals; Table S9: Association of BMI with metabolites in the basic model for male individuals. Table S10: Association of BMI with metabolites in the full model for female individuals. Table S11: Association of BMI with metabolites in the full model for male individuals. Table S12: Association of T2D with metabolites in the basic model for female individuals; Table S13: Association of T2D with metabolites in the basic model for male individuals. Table S14: Association of T2D with metabolites in the full model for female individuals. Table S15: Association of T2D with metabolites in the full model for male individuals. Supplemental Document S2: Testing of Assumptions for Multiple Linear Regression and Logistic Regres-

sion Model. Figures S4–S19: The linearity graphs for representative metabolites for each category. Tables S16–S31: VIF values for multicollinearity between predictors in the linear regression model and logistic regression model [79–84].

Author Contributions: Conceptualization, Q.D., S.S. (Sapna Sharma), H.G. and A.P.; methodology, Q.D. and S.S. (Sapna Sharma); software, Q.D. and S.S. (Sapna Sharma); validation, Q.D. and S.S. (Sapna Sharma); formal analysis, Q.D.; investigation, Q.D. and S.S. (Sapna Sharma); resources, C.G., R.W.-S., W.R., J.A., W.K., A.P. and H.G.; data curation, R.W.-S., W.R., J.A. and W.K.; writing—original draft preparation, Q.D.; writing—review and editing, Q.D., S.Sd. (Sidra Sidra), S.S. (Sapna Sharma), H.G., C.G., R.W.-S., C.P., W.R., J.A., W.K. and A.P.; visualization, Q.D.; supervision, S.S. (Sapna Sharma) and H.G.; project administration, H.G.; funding acquisition, H.G. and A.P. All authors have read and agreed to the published version of the manuscript.

Funding: The KORA study was initiated and financed by Helmholtz Munich—German research center for Environmental Health, which is financed by the German Federal Ministry of Education and Research and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC-Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ. This research was supported by China Scholarship Council (CSC) (No. 202008310176).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board (PV K003/22g).

Informed Consent Statement: Written informed consent has been obtained from the patients to publish this paper.

Data Availability Statement: The KORA FF4 datasets are not publicly available but can be accessed upon application through the KORA-PASST (Project application self-service tool, https://www.helmholtz-munich.de/epi/research/cohorts/kora-cohort/data-use-and-access-via-korapasst/index.html, accessed on 13 May 2022).

Acknowledgments: We express our appreciation to all KORA study participants for donating their blood and time. We thank the field staff in Augsburg conducting the KORA studies. We would like to thank the China Scholarship Council (CSC) for the financial support (No. 202008310176).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. World Health Organization. Available online: https://www.who.int/news/item/04-03-2022-world-obesity-day-2022-accelerating-action-to-stop-obesity (accessed on 4 March 2022).
2. Lingvay, I.; Sumithran, P.; Cohen, R.V.; le Roux, C.W. Obesity management as a primary treatment goal for type 2 diabetes: Time to reframe the conversation. *Lancet* **2022**, *399*, 394–405. [CrossRef] [PubMed]
3. Lauby-Secretan, B.; Scoccianti, C.; Loomis, D.; Grosse, Y.; Bianchini, F.; Straif, K.; Handbook, I.A.R.C. Body Fatness and Cancer—Viewpoint of the IARC Working Group. *N. Engl. J. Med.* **2016**, *375*, 794–798. [CrossRef]
4. Singh, G.M.; Danaei, G.; Farzadfar, F.; Stevens, G.A.; Woodward, M.; Wormser, D.; Kaptoge, S.; Whitlock, G.; Qiao, Q.; Lewington, S.; et al. The Age-Specific Quantitative Effects of Metabolic Risk Factors on Cardiovascular Diseases and Diabetes: A Pooled Analysis. *PLoS ONE* **2013**, *8*, e0065174. [CrossRef]
5. Okamura, T.; Hashimoto, Y.; Hamaguchi, M.; Obora, A.; Kojima, T.; Fukui, M. Ectopic fat obesity presents the greatest risk for incident type 2 diabetes: A population-based longitudinal study. *Int. J. Obesity* **2019**, *43*, 139–148. [CrossRef] [PubMed]
6. Varemo, L.; Nookaew, I.; Nielsen, J. Novel insights into obesity and diabetes through genome-scale metabolic modeling. *Front. Physiol.* **2013**, *4*, 92. [CrossRef] [PubMed]
7. Muoio, D.M.; Newgard, C.B. Obesity-related derangements in metabolic regulation. *Annu. Rev. Biochem.* **2006**, *75*, 367–401. [CrossRef]
8. Park, S.; Sadanala, K.C.; Kim, E.K. A Metabolomic Approach to Understanding the Metabolic Link between Obesity and Diabetes. *Mol. Cells* **2015**, *38*, 587–596. [CrossRef] [PubMed]
9. Stevens, V.L.; Carter, B.D.; McCullough, M.L.; Campbell, P.T.; Wang, Y. Metabolomic Profiles Associated with BMI, Waist Circumference, and Diabetes and Inflammation Biomarkers in Women. *Obesity* **2020**, *28*, 187–196. [CrossRef]
10. Tulipani, S.; Palau-Rodriguez, M.; Alonso, A.M.; Cardona, F.; Marco-Ramell, A.; Zonja, B.; de Alda, M.L.; Munoz-Garach, A.; Sanchez-Pla, A.; Tinahones, F.J.; et al. Biomarkers of Morbid Obesity and Prediabetes by Metabolomic Profiling of Human Discordant Phenotypes. *Clin. Chim. Acta* **2016**, *463*, 53–61. [CrossRef]
11. Rauschert, S.; Uhl, O.; Koletzko, B.; Kirchberg, F.; Mori, T.A.; Huang, R.C.; Beilin, L.J.; Hellmuth, C.; Oddy, W.H. Lipidomics Reveals Associations of Phospholipids With Obesity and Insulin Resistance in Young Adults. *J. Clin. Endocr. Metab.* **2016**, *101*, 871–879. [CrossRef]

12. Sokolowska, E.; Blachnio-Zabielska, A. The Role of Ceramides in Insulin Resistance. *Front. Endocrinol.* **2019**, *10*, 577. [CrossRef] [PubMed]
13. Chang, W.G.; Hatch, G.M.; Wang, Y.; Yu, F.; Wang, M. The relationship between phospholipids and insulin resistance: From clinical to experimental studies. *J. Cell Mol. Med.* **2019**, *23*, 702–710. [CrossRef]
14. Borkman, M.; Storlien, L.H.; Pan, D.A.; Jenkins, A.B.; Chisholm, D.J.; Campbell, L.V. The Relation between Insulin Sensitivity and the Fatty-Acid Composition of Skeletal-Muscle Phospholipids. *N. Engl. J. Med.* **1993**, *328*, 238–244. [CrossRef] [PubMed]
15. Wahl, S.; Drong, A.; Lehne, B.; Loh, M.; Scott, W.R.; Kunze, S.; Tsai, P.C.; Ried, J.S.; Zhang, W.H.; Yang, Y.W.; et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature* **2017**, *541*, 81. [CrossRef]
16. Yengo, L.; Sidorenko, J.; Kemper, K.E.; Zheng, Z.; Wood, A.R.; Weedon, M.N.; Frayling, T.M.; Hirschhorn, J.; Yang, J.; Visscher, P.M.; et al. Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. *Hum. Mol. Genet.* **2018**, *27*, 3641–3649. [CrossRef] [PubMed]
17. Mahajan, A.; Spracklen, C.N.; Zhang, W.; Ng, M.C.Y.; Petty, L.E.; Kitajima, H.; Yu, G.Z.; Rueger, S.; Speidel, L.; Kim, Y.J.; et al. Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *Nat. Genet.* **2022**, *54*, 560–572. [CrossRef] [PubMed]
18. Lanznaster, D.; Veyrat-Durebex, C.; Vourc'h, P.; Andres, C.R.; Blasco, H.; Corcia, P. Metabolomics: A Tool to Understand the Impact of Genetic Mutations in Amyotrophic Lateral Sclerosis. *Genes* **2020**, *11*, 537. [CrossRef]
19. Gieger, C.; Geistlinger, L.; Altmaier, E.; Hrabe de Angelis, M.; Kronenberg, F.; Meitinger, T.; Mewes, H.W.; Wichmann, H.E.; Weinberger, K.M.; Adamski, J.; et al. Genetics meets metabolomics: A genome-wide association study of metabolite profiles in human serum. *PLoS Genet.* **2008**, *4*, e1000282. [CrossRef]
20. Suhre, K.; Raffler, J.; Kastenmuller, G. Biochemical insights from population studies with genetics and metabolomics. *Arch. Biochem. Biophys.* **2016**, *589*, 168–176. [CrossRef]
21. Wang-Sattler, R.; Yu, Z.; Herder, C.; Messias, A.C.; Floegel, A.; He, Y.; Heim, K.; Campillos, M.; Holzapfel, C.; Thorand, B.; et al. Novel biomarkers for pre-diabetes identified by metabolomics. *Mol. Syst. Biol.* **2012**, *8*, 615. [CrossRef]
22. Haid, M.; Muschet, C.; Wahl, S.; Romisch-Margl, W.; Prehn, C.; Moller, G.; Adamski, J. Long-Term Stability of Human Plasma Metabolites during Storage at -80 degrees C. *J. Proteome Res.* **2018**, *17*, 203–211. [CrossRef] [PubMed]
23. Zukunft, S.; Prehn, C.; Rohring, C.; Moller, G.; Hrabe de Angelis, M.; Adamski, J.; Tokarz, J. High-throughput extraction and quantification method for targeted metabolomics in murine tissues. *Metabolomics* **2018**, *14*, 18. [CrossRef] [PubMed]
24. Huang, J.; Covic, M.; Huth, C.; Rommel, M.; Adam, J.; Zukunft, S.; Prehn, C.; Wang, L.; Nano, J.; Scheerer, M.F.; et al. Validation of Candidate Phospholipid Biomarkers of Chronic Kidney Disease in Hyperglycemic Individuals and Their Organ-Specific Exploration in Leptin Receptor-Deficient db/db Mouse. *Metabolites* **2021**, *11*, 89. [CrossRef]
25. Preacher, K.J.; Hayes, A.F. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behav. Res. Methods* **2008**, *40*, 879–891. [CrossRef]
26. Sobel, M.E. Asymptotic confidence intervals for indirect effects in structural equation models. *Sociol. Methodol.* **1982**, *13*, 290–312. [CrossRef]
27. Hemani, G.; Zhengn, J.; Elsworth, B.; Wade, K.H.; Haberland, V.; Baird, D.; Laurin, C.; Burgess, S.; Bowden, J.; Langdon, R.; et al. The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **2018**, *7*, e34408. [CrossRef]
28. Draisma, H.H.M.; Pool, R.; Kobl, M.; Jansen, R.; Petersen, A.K.; Vaarhorst, A.A.M.; Yet, I.; Haller, T.; Demirkan, A.; Esko, T.; et al. Genome-wide association study identifies novel genetic variants contributing to variation in blood metabolite levels. *Nat. Commun.* **2015**, *6*, 7208. [CrossRef] [PubMed]
29. Xue, A.; Wu, Y.; Zhu, Z.; Zhang, F.; Kemper, K.E.; Zheng, Z.; Yengo, L.; Lloyd-Jones, L.R.; Sidorenko, J.; Wu, Y.; et al. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat. Commun.* **2018**, *9*, 2941. [CrossRef]
30. Frigerio, G.; Favero, C.; Savino, D.; Mercadante, R.; Albetti, B.; Dioni, L.; Vigna, L.; Bollati, V.; Pesatori, A.C.; Fustinoni, S. Plasma Metabolomic Profiling in 1391 Subjects with Overweight and Obesity from the SPHERE Study. *Metabolites* **2021**, *11*, 194. [CrossRef]
31. Lee, H.-S.; Xu, T.; Lee, Y.; Kim, N.-H.; Kim, Y.-J.; Kim, J.-M.; Cho, S.Y.; Kim, K.-Y.; Nam, M.; Adamski, J.; et al. Identification of putative biomarkers for type 2 diabetes using metabolomics in the Korea Association REsource (KARE) cohort. *Metabolomics* **2016**, *12*, 178. [CrossRef]
32. Floegel, A.; Stefan, N.; Yu, Z.; Muhlenbruch, K.; Drogan, D.; Joost, H.G.; Fritsche, A.; Haring, H.U.; Hrabe de Angelis, M.; Peters, A.; et al. Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. *Diabetes* **2013**, *62*, 639–648. [CrossRef] [PubMed]
33. Hicks, A.A.; Pramstaller, P.P.; Johansson, A.; Vitart, V.; Rudan, I.; Ugocsai, P.; Aulchenko, Y.; Franklin, C.S.; Liebisch, G.; Erdmann, J.; et al. Genetic determinants of circulating sphingolipid concentrations in European populations. *PLoS Genet.* **2009**, *5*, e1000672. [CrossRef] [PubMed]
34. Straczkowski, M.; Kowalska, I.; Nikolajuk, A.; Dzienis-Straczkowska, S.; Kinalska, I.; Baranowski, M.; Zendzian-Piotrowska, M.; Brzezinska, Z.; Gorski, J. Relationship between insulin sensitivity and sphingomyelin signaling pathway in human skeletal muscle. *Diabetes* **2004**, *53*, 1215–1221. [CrossRef] [PubMed]
35. Subathra, M.; Qureshi, A.; Luberto, C. Sphingomyelin Synthases Regulate Protein Trafficking and Secretion. *PLoS ONE* **2011**, *6*, e23644. [CrossRef] [PubMed]

36. Khan, S.R.; Manialawy, Y.; Obersterescu, A.; Cox, B.J.; Gunderson, E.P.; Wheeler, M.B. Diminished Sphingolipid Metabolism, a Hallmark of Future Type 2 Diabetes Pathogenesis, Is Linked to Pancreatic beta Cell Dysfunction. *iScience* **2020**, *23*, 101566. [CrossRef]
37. Alexaki, A.; Clarke, B.A.; Gavrilova, O.; Ma, Y.; Zhu, H.; Ma, X.; Xu, L.; Tuymetova, G.; Larman, B.C.; Allende, M.L.; et al. De Novo Sphingolipid Biosynthesis Is Required for Adipocyte Survival and Metabolic Homeostasis. *J. Biol. Chem.* **2017**, *292*, 3929–3939. [CrossRef] [PubMed]
38. Kumar, A.; Sundaram, K.; Mu, J.Y.; Dryden, G.W.; Sriwastva, M.K.; Lei, C.; Zhang, L.F.; Qiu, X.L.; Xu, F.Y.; Yan, J.; et al. High-fat diet-induced upregulation of exosomal phosphatidylcholine contributes to insulin resistance. *Nat. Commun.* **2021**, *12*. [CrossRef] [PubMed]
39. Lynch, C.J.; Adams, S.H. Branched-chain amino acids in metabolic signalling and insulin resistance. *Nat. Rev. Endocrinol.* **2014**, *10*, 723–736. [CrossRef]
40. Yoon, M.S. The Emerging Role of Branched-Chain Amino Acids in Insulin Resistance and Metabolism. *Nutrients* **2016**, *8*, 405. [CrossRef]
41. Wang, S.M.; Yang, R.Y.; Wang, M.; Ji, F.S.; Li, H.X.; Tang, Y.M.; Chen, W.X.; Dong, J. Identification of serum metabolites associated with obesity and traditional risk factors for metabolic disease in Chinese adults. *Nutr. Metab. Cardiovas.* **2018**, *28*, 112–118. [CrossRef]
42. Bagheri, M.; Djazayery, A.; Farzadfar, F.; Qi, L.; Yekaninejad, M.S.; Aslibekyan, S.; Chamari, M.; Hassani, H.; Koletzko, B.; Uhl, O. Plasma metabolomic profiling of amino acids and polar lipids in Iranian obese adults. *Lipids Health Dis.* **2019**, *18*, 94. [CrossRef] [PubMed]
43. Rangel-Huerta, O.D.; Pastor-Villaescusa, B.; Gil, A. Are we close to defining a metabolomic signature of human obesity? A systematic review of metabolomics studies. *Metabolomics* **2019**, *15*, 93. [CrossRef]
44. Libert, D.M.; Nowacki, A.S.; Natowicz, M.R. Metabolomic analysis of obesity, metabolic syndrome, and type 2 diabetes: Amino acid and acylcarnitine levels change along a spectrum of metabolic wellness. *PeerJ* **2018**, *6*, e5410. [CrossRef] [PubMed]
45. Siddik, M.A.; Shin, A.C. Recent Progress on Branched-Chain Amino Acids in Obesity, Diabetes, and Beyond. *Endocrinol. Metab.* **2019**, *34*, 234–246. [CrossRef] [PubMed]
46. Zhao, X.; Han, Q.; Liu, Y.J.; Sun, C.L.; Gang, X.K.; Wang, G.X. The Relationship between Branched-Chain Amino Acid Related Metabolomic Signature and Insulin Resistance: A Systematic Review. *J. Diabetes Res.* **2016**, *2016*, 2794591. [CrossRef]
47. Adams, S.H. Emerging Perspectives on Essential Amino Acid Metabolism in Obesity and the Insulin-Resistant State. *Adv. Nutr.* **2011**, *2*, 445–456. [CrossRef]
48. Fikri, A.M.; Smyth, R.; Kumar, V.; Al-Abadla, Z.; Abusnana, S.; Munday, M.R. Pre-diagnostic biomarkers of type 2 diabetes identified in the UAE's obese national population using targeted metabolomics. *Sci. Rep.* **2020**, *10*, 17616. [CrossRef] [PubMed]
49. Adams, S.H.; Hoppel, C.L.; Lok, K.H.; Zhao, L.; Wong, S.W.; Minkler, P.E.; Hwang, D.H.; Newman, J.W.; Garvey, W.T. Plasma acylcarnitine profiles suggest incomplete long-chain fatty acid beta-oxidation and altered tricarboxylic acid cycle activity in type 2 diabetic African-American women. *J. Nutr.* **2009**, *139*, 1073–1081. [CrossRef]
50. Wijekoon, E.P.; Skinner, C.; Brosnan, M.E.; Brosnan, J.T. Amino acid metabolism in the Zucker diabetic fatty rat: Effects of insulin resistance and of type 2 diabetes. *Can. J. Physiol. Pharm.* **2004**, *82*, 506–514. [CrossRef]
51. Carayol, M.; Leitzmann, M.F.; Ferrari, P.; Zamora-Ros, R.; Achaintre, D.; Stepien, M.; Schmidt, J.A.; Travis, R.C.; Overvad, K.; Tjonneland, A.; et al. Blood Metabolic Signatures of Body Mass Index: A Targeted Metabolomics Study in the EPIC Cohort. *J. Proteome Res.* **2017**, *16*, 3137–3146. [CrossRef]
52. Cole, L.K.; Vance, J.E.; Vance, D.E. Phosphatidylcholine biosynthesis and lipoprotein metabolism. *BBA Mol. Cell. Biol. Lipids* **2012**, *1821*, 754–761. [CrossRef]
53. Matsumoto, T.; Kobayashi, T.; Kamata, K. Role of lysophosphatidylcholine (LPC) in atherosclerosis. *Curr. Med. Chem.* **2007**, *14*, 3209–3220. [CrossRef]
54. Pacana, T.; Cazanave, S.; Verdianelli, A.; Patel, V.; Min, H.K.; Mirshahi, F.; Quinlivan, E.; Sanyal, A.J. Dysregulated Hepatic Methionine Metabolism Drives Homocysteine Elevation in Diet-Induced Nonalcoholic Fatty Liver Disease. *PLoS ONE* **2015**, *10*, e0136822. [CrossRef] [PubMed]
55. Wei, X.C.; Song, H.W.; Yin, L.; Rizzo, M.G.; Sidhu, R.; Covey, D.F.; Ory, D.S.; Semenkovich, C.F. Fatty acid synthesis configures the plasma membrane for inflammation in diabetes. *Nature* **2016**, *539*, 294. [CrossRef] [PubMed]
56. van der Veen, J.N.; Lingrell, S.; McCloskey, N.; LeBlond, N.D.; Galleguillos, D.; Zhao, Y.Y.; Curtis, J.M.; Sipione, S.; Fullerton, M.D.; Vance, D.E.; et al. A role for phosphatidylcholine and phosphatidylethanolamine in hepatic insulin signaling. *FASEB J.* **2019**, *33*, 5045–5057. [CrossRef]
57. Kim, Y.C.; Seok, S.; Byun, S.; Kong, B.; Zhang, Y.; Guo, G.; Xie, W.; Ma, J.; Kemper, B.; Kemper, J.K. AhR and SHP regulate phosphatidylcholine and S-adenosylmethionine levels in the one-carbon cycle. *Nat. Commun.* **2018**, *9*, 540. [CrossRef]
58. Park, T.S.; Panek, R.L.; Mueller, S.B.; Hanselman, J.C.; Rosebury, W.S.; Robertson, A.W.; Kindt, E.K.; Homan, R.; Karathanasis, S.K.; Rekhter, M.D. Inhibition of sphingomyelin synthesis reduces atherogenesis in apolipoprotein E-knockout mice. *Circulation* **2004**, *110*, 3465–3471. [CrossRef]
59. Russo, S.B.; Ross, J.S.; Cowart, L.A. *Sphingolipids in Obesity, Type 2 Diabetes, and Metabolic Disease*; Springer: Vienna, Austria, 2013; pp. 373–401. [CrossRef]

60. Hammad, S.M.; Pierce, J.S.; Soodavar, F.; Smith, K.J.; Al Gadban, M.M.; Rembiesa, B.; Klein, R.L.; Hannun, Y.A.; Bielawski, J.; Bielawska, A. Blood sphingolipidomics in healthy humans: Impact of sample collection methodology. *J. Lipid Res.* **2010**, *51*, 3074–3087. [CrossRef]
61. Shimabukuro, M.; Zhou, Y.T.; Levi, M.; Unger, R.H. Fatty acid-induced beta cell apoptosis: A link between obesity and diabetes. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 2498–2502. [CrossRef] [PubMed]
62. Hu, W.; Ross, J.; Geng, T.Y.; Brice, S.E.; Cowart, L.A. Differential Regulation of Dihydroceramide Desaturase by Palmitate versus Monounsaturated Fatty Acids IMPLICATIONS FOR INSULIN RESISTANCE. *J. Biol. Chem.* **2011**, *286*, 16596–16605. [CrossRef]
63. Holland, W.L.; Bikman, B.T.; Wang, L.P.; Yuguang, G.; Sargent, K.M.; Bulchand, S.; Knotts, T.A.; Shui, G.H.; Clegg, D.J.; Wenk, M.R.; et al. Lipid-induced insulin resistance mediated by the proinflammatory receptor TLR4 requires saturated fatty acid-induced ceramide biosynthesis in mice. *J. Clin. Investig.* **2011**, *121*, 1858–1870. [CrossRef]
64. Samad, F.; Hester, K.D.; Yang, G.; Hannun, Y.A.; Bielawski, J. Altered adipose and plasma sphingolipid metabolism in obesity—A potential mechanism for cardiovascular and metabolic risk. *Diabetes* **2006**, *55*, 2579–2587. [CrossRef]
65. Sas, K.M.; Nair, V.; Byun, J.; Kayampilly, P.; Zhang, H.; Saha, J.; Brosius, F.C., 3rd; Kretzler, M.; Pennathur, S. Targeted Lipidomic and Transcriptomic Analysis Identifies Dysregulated Renal Ceramide Metabolism in a Mouse Model of Diabetic Kidney Disease. *J. Proteom. Bioinform.* **2015**, *2015* (Suppl. 14), 2. [CrossRef]
66. Lee, S.Y.; Lee, H.Y.; Song, J.H.; Kim, G.T.; Jeon, S.; Song, Y.J.; Lee, J.S.; Hur, J.H.; Oh, H.H.; Park, S.Y.; et al. Adipocyte-Specific Deficiency of De Novo Sphingolipid Biosynthesis Leads to Lipodystrophy and Insulin Resistance. *Diabetes* **2017**, *66*, 2596–2609. [CrossRef] [PubMed]
67. Park, J.W.; Park, W.J.; Kuperman, Y.; Boura-Halfon, S.; Pewzner-Jung, Y.; Futerman, A.H. Ablation of Very Long Acyl Chain Sphingolipids Causes Hepatic Insulin Resistance in Mice Due to Altered Detergent-Resistant Membranes. *Hepatology* **2013**, *57*, 525–532. [CrossRef] [PubMed]
68. Lemaitre, R.N.; Tanaka, T.; Tang, W.H.; Manichaikul, A.; Foy, M.; Kabagambe, E.K.; Nettleton, J.A.; King, I.B.; Weng, L.C.; Bhattacharya, S.; et al. Genetic Loci Associated with Plasma Phospholipid n-3 Fatty Acids: A Meta-Analysis of Genome-Wide Association Studies from the CHARGE Consortium. *PloS Genet.* **2011**, *7*, e1002193. [CrossRef] [PubMed]
69. Guan, W.H.; Steffen, B.T.; Lemaitre, R.N.; Wu, J.H.Y.; Tanaka, T.; Manichaikul, A.; Foy, M.; Rich, S.S.; Wang, L.; Nettleton, J.A.; et al. Genome-Wide Association Study of Plasma N6 Polyunsaturated Fatty Acids Within the Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium. *Circ. Cardiovasc. Genet.* **2014**, *7*, 321–331. [CrossRef]
70. Meierhofer, D.; Weidner, C.; Sauer, S. Integrative Analysis of Transcriptomics, Proteomics, and Metabolomics Data of White Adipose and Liver Tissue of High-Fat Diet and Rosiglitazone-Treated Insulin-Resistant Mice Identified Pathway Alterations and Molecular Hubs. *J. Proteome Res.* **2014**, *13*, 5592–5602. [CrossRef]
71. Lattka, E.; Illig, T.; Koletzko, B.; Heinrich, J. Genetic variants of the FADS1 FADS2 gene cluster as related to essential fatty acid metabolism. *Curr. Opin. Lipidol.* **2010**, *21*, 64–69. [CrossRef]
72. Reardon, H.T.; Hsieh, A.T.; Park, W.J.; Kothapalli, K.S.D.; Anthony, J.C.; Nathanielsz, P.W.; Brenna, J.T. Dietary long-chain polyunsaturated fatty acids upregulate expression of FADS3 transcripts. *Prostag. Leukotr. Ess.* **2013**, *88*, 15–19. [CrossRef]
73. Dupuis, J.; Langenberg, C.; Prokopenko, I.; Saxena, R.; Soranzo, N.; Jackson, A.U.; Wheeler, E.; Glazer, N.L.; Bouatia-Naji, N.; Gloyn, A.L.; et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.* **2010**, *42*, 464. [CrossRef]
74. Martinelli, N.; Girelli, D.; Malerba, G.; Guarini, P.; Illig, T.; Trabetti, E.; Sandri, M.; Friso, S.; Pizzolo, F.; Schaeffer, L.; et al. FADS genotypes and desaturase activity estimated by the ratio of arachidonic acid to linoleic acid are associated with inflammation and coronary artery disease. *Am. J. Clin. Nutr.* **2008**, *88*, 941–949. [CrossRef] [PubMed]
75. Wang, L.; Athinarayanan, S.; Jiang, G.; Chalasani, N.; Zhang, M.; Liu, W. Fatty acid desaturase 1 gene polymorphisms control human hepatic lipid composition. *Hepatology* **2015**, *61*, 119–128. [CrossRef] [PubMed]
76. Meldrum, S.J.; Li, Y.C.; Zhang, G.C.; Heaton, A.E.M.; D'Vaz, N.; Manz, J.; Reischl, E.; Koletzko, B.V.; Prescott, S.L.; Simmer, K. Can polymorphisms in the fatty acid desaturase (FADS) gene cluster alter the effects of fish oil supplementation on plasma and erythrocyte fatty acid profiles? An exploratory study. *Eur. J. Nutr.* **2018**, *57*, 2583–2594. [CrossRef] [PubMed]
77. Brayner, B.; Kaur, G.; Keske, M.A.; Livingstone, K.M. FADS Polymorphism, Omega-3 Fatty Acids and Diabetes Risk: A Systematic Review. *Nutrients* **2018**, *10*, 758. [CrossRef] [PubMed]
78. Ralston, J.C.; Matravadia, S.; Gaudio, N.; Holloway, G.P.; Mutch, D.M. Polyunsaturated Fatty Acid Regulation of Adipocyte FADS1 and FADS2 Expression and Function. *Obesity* **2015**, *23*, 725–728. [CrossRef]
79. Stancakova, A.; Paananen, J.; Soininen, P.; Kangas, A.J.; Bonnycastle, L.L.; Morken, M.A.; Collins, F.S.; Jackson, A.U.; Boehnke, M.L.; Kuusisto, J.; et al. Effects of 34 risk loci for type 2 diabetes or hyperglycemia on lipoprotein subclasses and their composition in 6,580 nondiabetic Finnish men. *Diabetes* **2011**, *60*, 1608–1616. [CrossRef]
80. Khamlaoui, W.; Mehri, S.; Hammami, S.; Hammouda, S.; Chraeif, I.; Elosua, R.; Hammami, M. Association Between Genetic Variants in FADS1-FADS2 and ELOVL2 and Obesity, Lipid Traits, and Fatty Acids in Tunisian Population. *Clin. Appl.Thromb.-Hem.* **2020**, *26*, 1076029620915286. [CrossRef] [PubMed]
81. Bult, C.J.; Blake, J.A.; Smith, C.L.; Kadin, J.A.; Richardson, J.E.; Anagnostopoulos, A.; Asabor, R.; Baldarelli, R.M.; Beal, J.S.; Bello, S.M.; et al. Mouse Genome Database (MGD) 2019. *Nucleic Acids Res.* **2019**, *47*, D801–D806. [CrossRef]
82. Yang, X.; Schadt, E.E.; Wang, S.; Wang, H.; Arnold, A.P.; Ingram-Drake, L.; Drake, T.A.; Lusis, A.J. Tissue-specific expression and regulation of sexually dimorphic genes in mice. *Genome Res.* **2006**, *16*, 995–1004. [CrossRef]

83. Langfelder, P.; Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **2008**, *9*, 559. [CrossRef] [PubMed]
84. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning: With Applications in R*, Uncorrected ed.; Springer: New York, NY, USA, 2013.

Paper II

Title: Role of human plasma metabolites in prediabetes and type 2 diabetes from the IMI-DIRECT study

Authors: Sapna Sharma*, Qiuling Dong*, Mark Haid*, Jonathan Adam, Roberto Bizzotto, Juan J. Fernandez-Tajes, Angus G. Jones, Andrea Tura, Anna Artati, Cornelia Prehn, Gabi Kastenmüller, Robert W. Koivula, Paul W. Franks, Mark Walker, Ian M. Forgie, Giuseppe Giordano, Imre Pavo, Hartmut Ruetten, Manolis Dermitzakis, Mark I. McCarthy, Oluf Pedersen, Jochen M. Schwenk, Konstantinos D. Tsirigos, Federico De Masi, Soren Brunak, Ana Viñuela, Andrea Mari, Timothy J. McDonald, Tarja Kokkola, Jerzy Adamski, Ewan R. Pearson & Harald Grallert (* joint first authors)

Journal: Diabetologia

Status: Published

Year: 2024

doi: 10.1007/s00125-024-06282-6

Supplements: https://link.springer.com/article/10.1007/s00125-024-06282-6#Sec17

Diabetologia (2024) 67:2804–2818
https://doi.org/10.1007/s00125-024-06282-6

ARTICLE

Role of human plasma metabolites in prediabetes and type 2 diabetes from the IMI-DIRECT study

Sapna Sharma[1] · Qiuling Dong[1,2] · Mark Haid[3] · Jonathan Adam[1,4] · Roberto Bizzotto[5] · Juan J. Fernandez-Tajes[6] · Angus G. Jones[7] · Andrea Tura[5] · Anna Artati[3] · Cornelia Prehn[3] · Gabi Kastenmüller[8] · Robert W. Koivula[9] · Paul W. Franks[10] · Mark Walker[11] · Ian M. Forgie[12] · Giuseppe Giordano[13] · Imre Pavo[14] · Hartmut Ruetten[15] · Manolis Dermitzakis[16,17,18] · Mark I. McCarthy[6] · Oluf Pedersen[19,20] · Jochen M. Schwenk[21] · Konstantinos D. Tsirigos[22] · Federico De Masi[22] · Soren Brunak[20,23] · Ana Viñuela[24] · Andrea Mari[5] · Timothy J. McDonald[25] · Tarja Kokkola[26] · Jerzy Adamski[27,28,29] · Ewan R. Pearson[12] · Harald Grallert[1,4]

Received: 28 February 2024 / Accepted: 29 July 2024 / Published online: 30 September 2024

Abstract

Aims/hypothesis Type 2 diabetes is a chronic condition that is caused by hyperglycaemia. Our aim was to characterise the metabolomics to find their association with the glycaemic spectrum and find a causal relationship between metabolites and type 2 diabetes.

Methods As part of the Innovative Medicines Initiative - Diabetes Research on Patient Stratification (IMI-DIRECT) consortium, 3000 plasma samples were measured with the Biocrates Absolute*IDQ* p150 Kit and Metabolon analytics. A total of 911 metabolites (132 targeted metabolomics, 779 untargeted metabolomics) passed the quality control. Multivariable linear and logistic regression analysis estimates were calculated from the concentration/peak areas of each metabolite as an explanatory variable and the glycaemic status as a dependent variable. This analysis was adjusted for age, sex, BMI, study centre in the basic model, and additionally for alcohol, smoking, BP, fasting HDL-cholesterol and fasting triacylglycerol in the full model. Statistical significance was Bonferroni corrected throughout. Beyond associations, we investigated the mediation effect and causal effects for which causal mediation test and two-sample Mendelian randomisation (2SMR) methods were used, respectively.

Results In the targeted metabolomics, we observed four (15), 34 (99) and 50 (108) metabolites (number of metabolites observed in untargeted metabolomics appear in parentheses) that were significantly different when comparing normal glucose regulation vs impaired glucose regulation/prediabetes, normal glucose regulation vs type 2 diabetes, and impaired glucose regulation vs type 2 diabetes, respectively. Significant metabolites were mainly branched-chain amino acids (BCAAs), with some derivatised BCAAs, lipids, xenobiotics and a few unknowns. Metabolites such as lysophosphatidylcholine a C17:0, sum of hexoses, amino acids from BCAA metabolism (including leucine, isoleucine, valine, *N*-lactoylvaline, *N*-lactoylleucine and formiminoglutamate) and lactate, as well as an unknown metabolite (X-24295), were associated with HbA_{1c} progression rate and were significant mediators of type 2 diabetes from baseline to 18 and 48 months of follow-up. 2SMR was used to estimate the causal effect of an exposure on an outcome using summary statistics from UK Biobank genome-wide association studies. We found that type 2 diabetes had a causal effect on the levels of three metabolites (hexose, glutamate and caproate [fatty acid (FA) 6:0]), whereas lipids such as specific phosphatidylcholines (PCs) (namely PC aa C36:2, PC aa C36:5, PC ae C36:3 and PC ae C34:3) as well as the two *n*-3 fatty acids stearidonate (18:4n3) and docosapentaenoate (22:5n3) potentially had a causal role in the development of type 2 diabetes.

Conclusions/interpretation Our findings identify known BCAAs and lipids, along with novel *N*-lactoyl-amino acid metabolites, significantly associated with prediabetes and diabetes, that mediate the effect of diabetes from baseline to follow-up (18 and 48 months). Causal inference using genetic variants shows the role of lipid metabolism and *n*-3 fatty acids as being causal for metabolite-to-type 2 diabetes whereas the sum of hexoses is causal for type 2 diabetes-to-metabolite. Identified metabolite markers are useful for stratifying individuals based on their risk progression and should enable targeted interventions.

Sapna Sharma, Qiuling Dong and Mark Haid contributed equally to this study.

Extended author information available on the last page of the article

Research in context

What is already known about this subject?

- Type 2 diabetes is a chronic condition that is characterised by hyperglycaemia
- The primary objectives of the DIRECT consortium include the identification of potential biomarkers that can assist in categorising individuals based on their glycaemic deterioration

What is the key question?

- What is the correlation between metabolomics and the glycaemic spectrum, and is there a causal relationship between any metabolites and type 2 diabetes?

What are the new findings?

- Novel metabolites such as picolinoylglycine, *N*-lactoylvaline, *N*-lactoylleucine and formiminoglutamate, as well as lactate and an unknown metabolite (X-24295, from untargeted metabolomics), were associated with prediabetes, type 2 diabetes and HbA_{1c} progression rate and were significant mediators of type 2 diabetes from baseline to 18 and 48 months of follow-up
- Causal inference using genetic variants shows that the *n*-3 fatty acids stearidonate (18:4n3) and docosapentaenoate (22:5n3) have a causal role in type 2 diabetes

How might this impact on clinical practice in the foreseeable future?

- Identified metabolite markers may be useful for stratifying individuals based on their risk of progression and may enable targeted interventions

Keywords Causality · Glycaemic traits · HbA_{1c} · IMI-DIRECT · Mediation · Metabolomics · *N*-lactoylaminoacids · Patient stratification · Targeted metabolomics · Type 2 diabetes · Untargeted metabolomics

Abbreviations

2SMR	Two-sample MR
GWAS	Genome-wide association study
H1	Hexoses
IGR	Impaired glucose regulation
IGT	Impaired glucose tolerance
IMI-DIRECT	Innovative Medicines Initiative - Diabetes Research on Patient Stratification
Lac-Phe	*N*-lactoyl-phenylalanine
LysoPC	Lysophosphatidylcholine
MOVE	Multi-omics variational autoencoders
MR	Mendelian randomisation
NA	Unidentified metabolite
NGR	Normal glucose regulation
PC	Phosphatidylcholine

Introduction

Type 2 diabetes is a complex and common metabolic disorder, resulting from the body's ineffective use of insulin. It can be characterised by hyperglycaemia (high blood sugar) due to impaired insulin secretion and insulin resistance, with most affected people being overweight or obese [1]. Impaired glucose tolerance (IGT) and impaired fasting glucose, together known as impaired glucose regulation (IGR) or prediabetes, characterise an intermediate condition before converging towards diabetes. Recent studies show that a complex interplay of genetic susceptibility, environmental factors, lifestyle (including diet, physical activity, smoking and alcohol consumption), clinical heterogeneity, drugs and gut microbiome orchestrates the development of type 2 diabetes [2]. Over time, individuals with type 2 diabetes are more likely to have a higher risk for heart attacks, strokes [3], neuropathy (nerve damage), retinopathy (causing blindness) and kidney failure as well as several infectious diseases including COVID-19, reducing life quality and causing social burden [4, 5].

Metabolomics profiles involve a set of low-molecular-weight biochemicals (metabolites) that includes sugars, amino acids, organic acids, nucleotides, lipids, xenobiotics and other compound classes. Identifying biochemical changes occurring between prediabetes and diabetes improves risk prediction for better-targeted prevention [6, 7]. In addition, genetic composition can be used to make predictions regarding disease susceptibility. Genome-wide association studies (GWAS) show that more than 400 loci influence the risk of type 2 diabetes [8] and that 900 genetic variants have been associated with BMI [9]. Therefore,

linking metabolites with genetics gives access to genetics' influence on the metabolic compositions [10–13], providing comprehensive molecular understanding of the disease.

In the Innovative Medicines Initiative - Diabetes Research on Patient Stratification (IMI-DIRECT), we characterised 132 metabolites from targeted measurements and 779 metabolites from untargeted measurements profiled in 3000 individuals at baseline. The study population was stratified by following ADA 2011 glycaemic categories as follows: 23.89% (n=692) had normal glucose regulation (NGR) with fasting glucose 5.23 (SD=0.39) mmol/l; 48.91% (n=1418) had IGR with fasting glucose 5.90 (SD=0.51) mmol/l; and 27.2% (n=890) had type 2 diabetes with fasting glucose 7.15 (SD=1.39) mmol/l [14]. For the integration of non-omics data such as health status, lifestyle and medication with metabolomics, advanced statistical techniques were applied to analyse the data (see Methods). Beyond multivariate and association analyses we performed causal mediation analysis to evaluate potential causal roles of mediators on outcome [15, 16]. A study on drug–omics associations in type 2 diabetes [17] used an unsupervised deep learning framework of multi-omics variational autoencoders (MOVE) to extract significant drug response patterns from 789 individuals newly diagnosed with type 2 diabetes in the IMI-DIRECT cohort. We integrated the polypharmacy effect on metabolomics knowledge from MOVE and compared with our molecular findings in this study.

Our aims in this study were as follows: (1) to characterise 911 small molecular (132 targeted, 779 untargeted metabolomics analysis approach) features associated with prediabetes/IGR and type 2 diabetes; (2) to identify baseline metabolites associated with progression rate estimated from cross-sectional data; (3) to investigate potential mediation effects of metabolites from baseline glycaemic status to follow-up using mediation analysis; and (4) to identify causal relationships between metabolites and type 2 diabetes using genetics drivers using two-sample Mendelian randomisation (2SMR) tests.

Methods

DIRECT cohort

The Diabetes Research on Patient Stratification (DIRECT) cohort encompasses 24,682 European participants at varying risk of glycaemic deterioration, identified and enrolled into a prospective cohort (study 1) of prediabetes (n=2235) and type 2 diabetes (n=830). Using ADA 2011 glycaemic categories in study 1, 33% (n=692) of cohort 1 (prediabetes risk) had NGR, 67% (n=1418) had IGR and 108 were excluded. In study 2, 789 samples were included and 41 samples were excluded. From study 1, 101 excluded samples entered study 2 (n=890). The ratio of self-reported sex varied in each study. Detailed characteristics on inclusion and exclusion criteria, along with the protocol timeline for visits and tests for both studies, have been described elsewhere [14, 18]. In summary, venous blood fasting samples were obtained, followed by performance of DNA extractions and additional biochemical analyses. Metabolomics measurements for distinct samples at the baseline is considered in this study.

Targeted metabolomics (Absolute*IDQ* p150 Kit)

Blood samples in the study were analysed with the Absolute*IDQ* p150 Kit (BIOCRATES Life Sciences, Innsbruck, Austria) (see electronic supplementary material [ESM] Methods for details) [19]. After data export, lower and upper outliers were defined as samples with >33% of metabolite concentrations below 25% quantile ($\pm 1.5 \times$ IQR). Metabolite traits with too many zero-concentration samples and unidentified metabolites (NAs, >50%) were excluded (none). The CV was calculated in reference samples for each metabolite over all plates. Metabolite traits with CV>0.25 were excluded. After quality control, 132 metabolites were included in this study (ESM Table 1). Metabolite concentrations were $\log_e$-transformed and scaled (mean=0, SD=1) to ensure comparability between the metabolites.

Untargeted metabolomics (Metabolon platform)

Untargeted LC/MS-based techniques covers a broad spectrum of metabolites, in contrast to the targeted techniques wherein metabolites are limited to a predefined set of molecules. For details on sample preparation, measurement and identification of metabolites, see ESM Methods. Incomplete databases and the presence of unknown or novel metabolites have been reported with an asterisk (*) against the metabolite name. The measured volume of the datasets contained 12% missing values. We screened for outlier remover (see ESM Fig. 1 for an example), which added 4% more missing values onto existing missing values (ESM Table 2). Peaks were quantified using AUC. For studies spanning multiple days, a data normalisation step was performed to correct variation resulting from instrument inter-day tuning differences. Essentially, each compound was corrected in run-day blocks by registering the medians to equal one and normalising each data point proportionately (termed the 'block correction'; ESM Fig. 2). Principal component analysis was performed on the metabolite dataset and checked for technical effects such as centre and sex (see ESM Fig. 3). The data missing pattern was tested using logistic regression considering missing as 0 and non-missing as 1; there was no significant association between missing and regressors indicating the missing-at-random pattern. The K-nearest neighbour (KNN)-based

imputation method was applied using K=10 as suggested and optimised from German Cohort KORA F4 [20].

Statistics

Multivariable logistic regression and linear regression Identifying metabolites specifically associated with the presence of IGR and type 2 diabetes, we ran the logistic regression with adjustment for age, sex, BMI and centre as the basic model, and adjusted additionally for alcohol consumption, smoking, BP, fasting HDL-cholesterol and fasting triacylglycerol as the full model. The concentration of each metabolite was $\log_e$-transformed and scaled to have a mean of zero and an SD of 1. Each metabolite was taken as exposure and a binary NGR-IGR, NGR-type 2 diabetes (NGR-T2D) or IGR-type 2 diabetes (IGR-T2D) variable as an outcome. The OR of outcomes was calculated using the β coefficient from logistic regression, where OR>1 indicates higher odds of outcome and OR<0 shows lower odds of outcome. To account for multiple testing, the *p* values from regression analyses were adjusted for multiple testing using the Bonferroni correction (p_{fdr} values). To stratify sex-dependent metabolites, men and women were separated to test the associations by performing the logistic regression full models.

For incidents of IGR and type 2 diabetes analysis, a binary NGR-IGR, NGR-T2D or IGR-T2D variable at follow-up times of 18 months and 48 months was taken as the outcome; transformed metabolites and the same risk factors in the full model were taken as exposure and covariates, respectively. The same *p* correction method was adopted.

The linear regression model was used to explore the association between HbA_{1c} progression rate and metabolites at the baseline. HbA_{1c} progression rate was computed with a conditional linear mixed effect model and adjusted for changes in BMI and diabetes medications [21]. Each transformed metabolite was taken as the independent variable and HbA_{1c} concentration as the dependent variable, with adjustment for age and sex. Bonferroni correction was performed for *p* correction.

Mediation analysis Mediation analysis followed the basic steps suggested by Baron and Kenny [22], and the significance of the mediation effect was tested with a non-parametric causal mediation analysis [22, 23]. Each identified metabolite was taken as a mediator, glycaemic category status at the baseline as the independent variable and glycaemic category at the follow-up (18 months and 48 months) as the dependent variable. R package 'mediation (4.5.0)' was used to calculate the *p* value and proportion of the mediation effect by bootstrapping with 1000 resamples.

Mendelian randomisation We used 2SMR approaches from the MRInstruments (0.3.2) and TwoSampleMR library (v0.5.6) to check causal inference [24]. The 2SMR technique enables the establishment of a causal relationship between two observational studies (ESM Fig. 4), solely relying on summary statistics obtained from GWAS [24, 25]. To evaluate the influence of type 2 diabetes on metabolite levels, we conducted a 2SMR examination. Type 2 diabetes instruments were obtained from the genome-wide genotyping study [26] and the corresponding SNP estimates on metabolites were extracted from the metabolite-GWAS [10, 27]. Prior to performing Mendelian randomisation (MR) analysis, exposure and outcome data were harmonised by aligning the SNPs on the same effect allele. We employed the inverse-variance weighting [10, 26, 27] to estimate the causal effect.

Results

Study populations

After stringent quality control (see ESM Methods), we identified 132 (ESM Table 1) and 779 (ESM Table 2) metabolites from targeted and untargeted metabolomics measurements, respectively, that were profiled for 3000 samples (ESM Table 3) [28]. Baseline characteristics (Table 1) revealed that there were significant differences in BMI, fasting variables and health status observed between NGR, IGR and type 2 diabetes groups. No significant differences in age and smoking status were observed between these three groups. In addition, the study was conducted across seven countries; type 2 diabetes participants were recruited in all centres while participants with NGR or IGR were only recruited in the Amsterdam, Copenhagen, Kuopio and Lund centres.

Metabolites associated with prediabetes and diabetes from targeted metabolomics measurements

A multivariable logistic regression model was used with known diabetes-related variables as covariates to identify significant metabolites. Study centre, sex, age and BMI were covariates in the basic model while the additional variables systolic BP, fasting HDL-cholesterol, fasting triacylglycerol, smoking status, alcohol status and health status were added in the full model. Based on the full model, four metabolites differed significantly between the NGR and IGR groups (Fig. 1a). Of these, hexoses (H1) showed the strongest association (OR 1.81 [95% CI 1.59, 2.06], $p_{fdr}=3.97\times10^{-17}$) and served as a positive control throughout our analysis. Thirty-four and 50 metabolites differed significantly between NGR and IGR vs type 2 diabetes, respectively (Fig. 1b,c). As a general pattern, phosphatidylcholines (PCs) and lysophosphatidylcholine (lysoPC) were negatively associated with progression to type 2 diabetes, while branched-chain and aromatic amino acids as

Table 1 Baseline characteristics of the DIRECT participants based on their glycaemic category

Characteristic	NGR	IGR	T2D	*p* value
Sample size	692	1418	890	
Male sex	519 (75.0)	1074 (75.7)	525 (59.0)	<0.001
Centre				<0.001
Amsterdam	167 (24.1)	300 (21.2)	183 (20.6)	
Copenhagen	54 (7.8)	223 (15.7)	97 (10.9)	
Dundee	0	0	164 (18.4)	
Exeter	0	0	142 (16.0)	
Kuopio	407 (58.8)	820 (57.8)	34 (3.8)	
Lund	64 (9.2)	75 (5.3)	104 (11.7)	
Newcastle	0	0	166 (18.7)	
Age, years	62.15±6.43	62.08±6.19	61.99±7.96	0.894
BMI, kg/m^2	27.15±3.65	28.33±4.06	30.59±4.92	<0.001
Systolic BP, mmHg	128.48±15.21	131.62±15.20	132.02±15.78	<0.001
Diastolic BP, mmHg	79.18±8.73	81.20±8.97	76.48±9.88	<0.001
Fasting glucose, mmol/l	5.23±0.39	5.90±0.51	7.13±1.39	<0.001
Fasting HDL-cholesterol, mmol/l	1.37±0.35	1.30±0.36	1.18±0.38	<0.001
Fasting LDL-cholesterol, mmol/l	3.21±0.90	3.19±0.95	2.43±1.00	<0.001
Fasting TG, mmol/l	1.22±0.53	1.44±0.66	1.56±0.88	<0.001
Fasting cholesterol, mmol/l	5.14±0.97	5.15±1.01	4.33±1.17	<0.001
Fasting HbA_{1c}, mmol/mol	35.34±2.22	37.86±2.88	45.86±5.94	<0.001
Fasting HbA_{1c}, %	5.38±0.20	5.61±0.26	6.35±0.54	<0.001
Fasting insulin, pmol/l	50.84±30.90	72.42±50.22	96.56±72.69	<0.001
Smoking status				0.717
Current smoker	93 (13.4)	215 (15.2)	117 (13.2)	
Ex-smoker	326 (47.1)	681 (48.0)	445 (50.1)	
Never	272 (39.3)	520 (36.7)	326 (36.7)	
Not Known	1 (0.1)	2 (0.1)	1 (0.1)	
Alcohol consumption status				0.004
Never	96 (13.9)	166 (11.7)	140 (15.7)	
Occasionally	134 (19.4)	282 (19.9)	214 (24.1)	
Regularly	462 (66.8)	968 (68.3)	534 (60.1)	
Not known	0	2 (0.1)	1 (0.1)	
Health status				<0.001
Poor	1 (0.1)	10 (0.7)	28 (3.1)	
Fair	49 (7.1)	74 (5.2)	34 (3.8)	
Good	331 (47.8)	744 (52.5)	428 (48.1)	
Very good	213 (30.8)	396 (27.9)	239 (26.9)	
Excellent	49 (7.1)	74 (5.2)	34 (3.8)	
Not known	4 (0.6)	11 (0.8)	19 (2.1)	

Quantitative variables are expressed as mean ± SD; categorical variables are expressed as *n* (%)

The significant difference of population characteristics between the individuals with IGR/type 2 diabetes and the normal participants (NGR) was calculated. Test statistics for categorical variables were calculated via the χ^2 test and Student's *t* test for continuous variables

T2D, type 2 diabetes; TG, triacylglycerol

well as valeryl/glutaryl-related acylcarnitines were positively associated with type 2 diabetes.

H1 (OR 9.67 [95% CI 6.54, 14.32], p_{fdr}=1.13×10^{-27}) also had the strongest associations in NGR-T2D while C5-M-DC (OR=5.31 [95% CI 4.16, 6.77], p_{fdr}=1.07×10^{-38}) had the strongest association in IGR-T2D. Three metabolites (H1, lysoPC a C17:0, lysoPC a C18:0) were significantly different in all comparisons (NGR-IGR, NGR-T2D and IGR-T2D), suggesting their important roles in diabetes indication and severity. Detailed statistics for the basic model and full model are shown in ESM Tables 3–8. As there were many more male participants than female participants

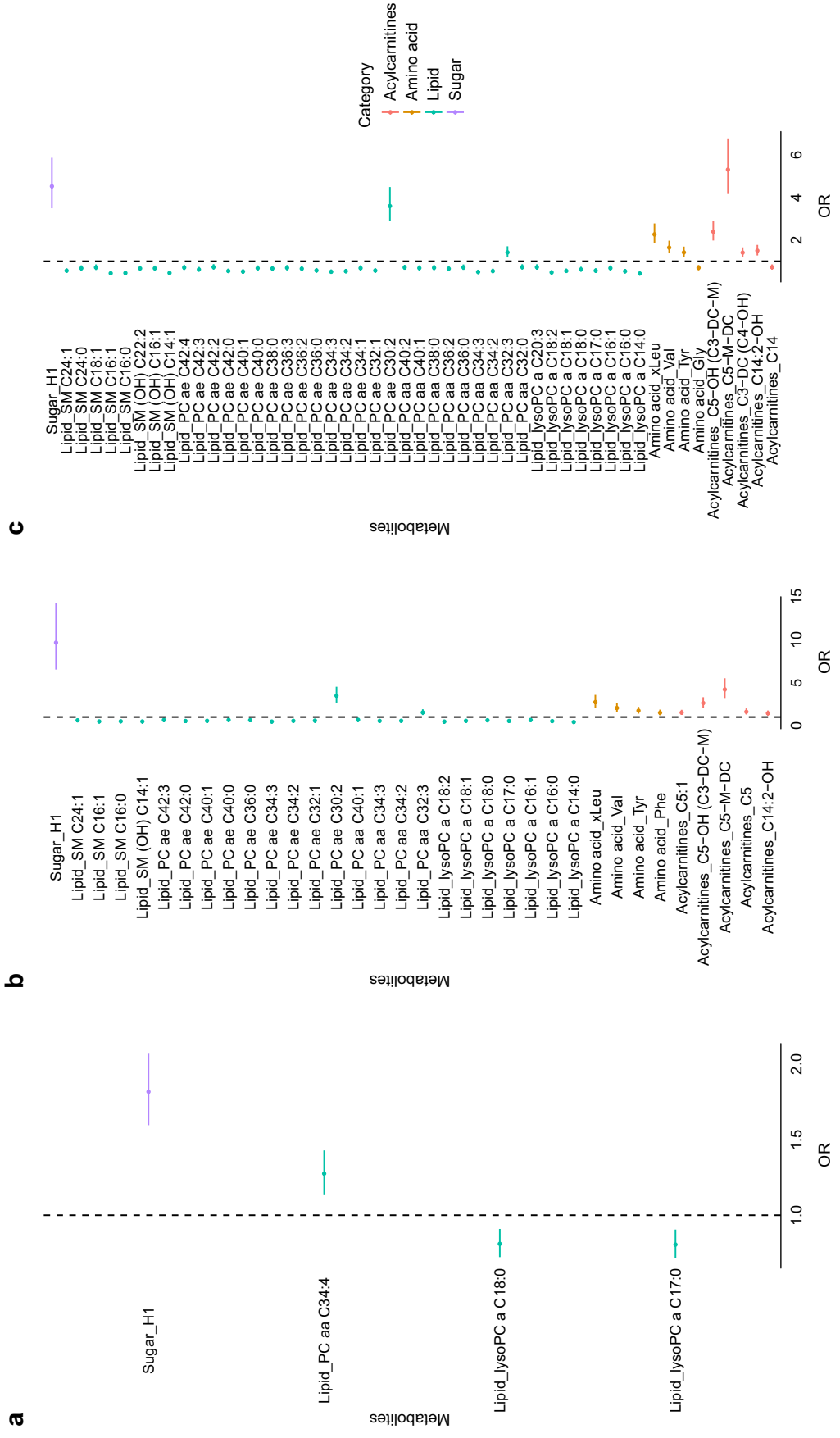


Fig. 1 Flag plots representing the results of the multivariable logistic regression models for NGR vs IGR (**a**), NGR vs type 2 diabetes (**b**) and IGR vs type 2 diabetes (**c**) as dependent variables and the metabolites as independent variables, adjusted for study centre, sex, age, BMI, BP, fasting HDL-cholesterol, fasting triacylglycerol, smoking status, alcohol status and health status. The *x*-axis shows OR (95% CI) and the *y*-axis shows each significant metabolite; metabolite classes are represented by different colours. SM, sphingomyelin

enrolled in the study, a sensitivity analysis stratified by sex was conducted, and is reported in ESM Results, ESM Tables 9–14 and ESM Fig. 5.

Metabolites associated with prediabetes and diabetes from untargeted metabolomics measurements

Fifteen metabolites were significantly changed between NGR and IGR based on the logistic regression analyses in the full model (Fig. 2a). Fructosyl lysine had the highest statistically significant association with progression to IGR (OR 1.53 [95% CI 1.37, 1.71], p_{fdr}=8.64×10^{-12}). Similarly, 99 and 108 metabolites differed significantly between NGR or IGR and type 2 diabetes, respectively (Fig. 2b,c). As a general pattern, lipids were negatively associated and amino acids were positively associated with progression to type 2 diabetes. 1-(1-Enyl-palmitoyl)-2-oleoyl-GPC (P-16:0_18:1)* (OR 0.23 [95% CI 0.17, 0.31], p_{fdr}=3.48×10^{-18}) had the strongest association for the NGR-T2D comparison, while cysteine-*S*-sulphate (OR 3.25 [95% CI 2.55, 4.15], p_{fdr}=3.11×10^{-18}) was significantly associated in the IGR-T2D comparison. Seven metabolites (fructosyl lysine, glutamate, 1-stearoyl-GPC (18:0), *N*-lactoylphenylalanine, *N*-lactoylvaline, picolinoyl glycine, mannonate) appeared significant in all comparison groups, suggesting their important roles as diabetes risk indicators. Detailed statistics are presented in ESM Tables 15–20. A sex-based sensitivity analysis of metabolomics data from the untargeted measurements is reported in ESM Results, ESM Table 21–26, ESM Fig. 6.

Metabolites associated with HbA_{1c} progression rate

HbA_{1c} progression rate was computed with a conditional linear mixed effect model and adjusted for changes in BMI and diabetes medications [21]. In multivariable linear regression analysis, lysoPC a C17:0 (β −0.0535 [95% CI −0.08, −0.0269], p_{fdr}=0.0109), glycine (Gly) (β −0.0509 [95% CI −0.0782, −0.0236], p_{fdr}=0.0347) and H1 (β 0.0481 [95% CI 0.0218, 0.0745], p_{fdr}=0.0452) were significantly correlated with HbA_{1c} progression rate and all were related to glycaemic-deterioration traits as well. In untargeted metabolomic profiling, 20 metabolites were significantly related to HbA_{1c} progression rate, with pyruvate (β 0.0877 [95% CI 0.0609, 0.114], p_{fdr}=1.28×10^{-7}) showing the strongest association. Besides pyruvate, *N*-lactoylleucine, lactate, *N*-lactoylphenylalanine, X-15245, *N*-lactoylisoleucine, *N*-lactoylvaline, 1-(1-enyl-palmitoyl)-2-oleoyl-GPC (P-16:0/18:1)*, cortolone glucuronide, X-24295, formiminoglutamate and *N*-lactoyltyrosine were also significantly associated with glycaemic categories. Tables 2 and 3 show the metabolites with significant associations, while the complete results are reported in ESM Tables 27–28.

Metabolite association with incident diabetes (IGR/type 2 diabetes)

Several metabolites were identified to be significantly associated with HbA_{1c} progression rate as well as glycaemic category: three targeted metabolites (lysoPC a C17:0; glycine, H1); and 12 untargeted metabolites (pyruvate, *N*-lactoylleucine, lactate, *N*-lactoylphenylalanine, X-15245, *N*-lactoylisoleucine, *N*-lactoylvaline, 1-[1-enyl-palmitoyl[-2-oleoyl-GPC* [PC(P-16:0/18:1)], cortolone glucuronide, X-24295, formiminoglutamate, *N*-lactoyltyrosine). Next, we investigated their predictive value for IGR and type 2 diabetes by including baseline metabolite concentrations and incident IGT or type 2 diabetes in follow-up timelines in multivariable logistic regression. As shown in Table 4, lysoPC a C17:0 concentration at baseline was observed to significantly differ in 244 incident IGR individuals compared with 398 NGR control individuals after 18 months. The sum of H1 at baseline concentrations showed significant differences between incident IGR (at 48 month follow-up) and NGR or incident type 2 diabetes and IGR at both the 18 month and the 48 month follow-up.

In untargeted metabolomic profiling, lactate and X-24295 baseline concentrations were significantly correlated with IGR or type 2 diabetes incidence at the 18 month and 48 month follow-up (Table 5). Formiminoglutamate, *N*-lactoylleucine and *N*-lactoylvaline significantly differed in 244 incident IGT individuals compared with 398 NGT control individuals after 18 months. We did not find any significant metabolites from untargeted measurements to predict the incidence of IGR from NGR at 48 months.

Mediation analysis

Causal mediation analysis was employed to explore the potential mediation effects of the identified metabolites from baseline glycaemic status to follow-up. Consistent with incidence results, lysoPC a C17:0 showed strong significance (proportion of mediation by 13%, mediation effect p=0.034, Fig. 3a), indicating that this metabolite partially mediated the glycaemic deterioration from NGR to IGR at 18 months. The positive control H1 exhibited significant mediation effects in all groups (between 6% and 9%) as it is mainly represented by blood glucose.

N-Lactoylvaline (proportion of mediation 24%, mediation effect p<2×10^{-16}), lactate (proportion of mediation 22%, mediation effect p=0.002), *N*-lactoylleucine (proportion of mediation 20%, mediation effect p=0.006), formiminoglutamate (proportion of mediation 11%, mediation effect

Fig. 2 Flag plots representing the results of the multivariable logistic regression models for NGR vs IGR (**a**), NGR vs type 2 diabetes (**b**) and IGR vs type 2 diabetes (**c**) as dependent variables and the metabolites as independent variables, adjusted for study centre, sex, age, BMI, BP, fasting HDL-cholesterol, fasting triacylglycerol, smoking status, alcohol status and health status. The *x*-axis shows OR (95% CI) and the *y*-axis shows each significant metabolite; metabolite classes are represented by different colours. Asterisks (*) indicate the presence of unknown or novel metabolites

Table 2 Metabolites from targeted measurements significantly associated with HbA_{1c} progression rate from a linear regression model

Metabolite	β (95% CI)	*p* value	p_{fdr} value
LysoPC a C17:0	−0.053 (−0.080, −0.027)	8.25×10^{-5}	0.011
Gly	−0.051 (−0.078, −0.024)	2.63×10^{-4}	0.0345
H1	0.048 (0.022, 0.075)	3.42×10^{-4}	0.045

The dependent variable is the HbA_{1c} progression rate while the independent variable is the $\log_e$-transformed and standardised baseline concentration of a given metabolite, adjusted by age and sex

The p_{fdr} values represent the adjusted *p* value for multiple testing by Bonferroni correction

p=0.034) and X-24295 (proportion of mediation 11%, mediation effect *p*=0.042) were all observed to show significant mediation effects from baseline NGR to IGR at 18 months' follow-up (Fig. 3b). Furthermore, formiminoglutamate (proportion of mediation 23%, mediation effect *p*=0.006) showed a significant mediation effect from NGR to IGR at 48 months. These results suggest that these metabolites own a significant mediation effect on glycaemic deterioration.

MR

The availability of genetic data on type 2 diabetes makes the use of MR particularly compelling. To assess bidirectional causal relationships between type 2 diabetes and metabolites (Fig. 4), we employed 2SMR tests. After multiple testing correction only the concentration of the sum of H1 was determined by type 2 diabetes ($p<0.05/117=0.00042$). For untargeted metabolites we found instruments for only 19% of the metabolites (i.e. 151 out of 779). For example, instruments are from genes *TCF7L2*, *IGF2BP2*, *NOTCH2*, *CDKAL1*, *PABPC4*, *FTO* and *JAZF1*, known to be associated with diabetes and that have been further significantly associated with the metabolites. Following multiple testing correction, it suggests that the change in an amino acid (glutamate) and a lipid (caproate, FA C6:0) was caused by change in type 2 diabetes status ($p<0.05/151=0.000331$). However, metabolites that are causal for type 2 diabetes (meaning that the change in metabolite caused change in the disease status) included several phosphatidylcholines, namely PC aa C36:2, PC aa C36:5, PC ae C36:3 and PC ae C34:3, from the targeted metabolomics dataset. From the untargeted metabolomics

Table 3 Metabolites from untargeted metabolomics measurements significantly associated with HbA_{1c} progression rate from a linear regression model

Metabolite	β (95% CI)	*p* value	p_{fdr} value
Pyruvate	0.087 (0.060, 0.114)	1.65×10^{-10}	1.28×10^{-7}
N-Lactoylleucine	0.082 (0.056, 0.109)	8.43×10^{-10}	6.57×10^{-7}
Lactate	0.075 (0.049, 0.102)	3.30×10^{-8}	2.57×10^{-5}
N-Lactoylphenylalanine	0.074 (0.048, 0.100)	3.66×10^{-8}	2.85×10^{-5}
X-15245	0.074 (0.047, 0.100)	6.24×10^{-8}	4.86×10^{-5}
N-Lactoylisoleucine	0.068 (0.042, 0.095)	3.11×10^{-7}	2.42×10^{-4}
N-Lactoylvaline	0.067 (0.041, 0.094)	5.69×10^{-7}	4.43×10^{-4}
X-11444	0.068 (0.041, 0.094)	6.22×10^{-7}	4.84×10^{-4}
Orotidine	0.065 (0.038, 0.091)	1.74×10^{-6}	1.35×10^{-3}
Metabolonic lactone sulphate	0.063 (0.036, 0.089)	2.9×10^{-6}	2.28×10^{-3}
3,4-Dihydroxybutyrate	0.060 (0.033, 0.087)	1.11×10^{-5}	8.64×10^{-3}
*N*4-Acetylcytidine	0.059 (0.033, 0.085)	1.16×10^{-5}	9.06×10^{-3}
X-24337	0.058 (0.032, 0.085)	1.47×10^{-5}	0.011
1-(1-Enyl-palmitoyl)-2-oleoyl-GPC(P-16:0/18:1)*	−0.058 (−0.084, −0.032)	1.49×10^{-5}	0.016
X-25828	−0.058 (−0.085, −0.032)	1.50×10^{-5}	0.017
Cortolone glucuronide	0.058 (0.032, 0.085)	1.73×10^{-5}	0.013
X-24295	0.057 (0.031, 0.084)	1.77×10^{-5}	0.014
Formiminoglutamate	0.059 (0.032, 0.088)	2.75×10^{-5}	0.021
1-Palmitoyl-2-oleoyl-GPE (16:0/18:1)	0.056 (0.029, 0.082)	3.59×10^{-5}	0.028
N-Lactoyltyrosine	0.055 (0.029, 0.082)	3.98×10^{-5}	0.031

The dependent variable is the HbA_{1c} progression rate while the independent variable is the $\log_e$-transformed and standardised baseline concentration of a given metabolite, adjusted by age and sex. The p_{fdr} are adjusted *p* for multiple testing by Bonferroni correction

Table 4 Metabolites from targeted measurements that were significantly associated with incidence of IGR and type 2 diabetes in different pairwise comparisons

Comparison	OR (95% CI)	*p* value
18 months		
398 NGR vs 244 IGR		
lysoPC a C17:0	−0.246 (−0.452, −0.043)	0.018
897 IGR vs 71 T2D		
H1	0.545 (0.164, 0.945)	0.006
48 months		
244 NGR vs 295 IGR		
H1	0.433 (0.189, 0.690)	$7x10^{-3}$
821 IGR vs 128 T2D		
H1	0.347 (0.064, 0.642)	0.018

Baseline metabolites were taken as the independent variables with glycaemic category in different timelines (18 months and 48 months) as the dependent variables, adjusted by study centre, sex, age, BMI, BP, fasting HDL-cholesterol, fasting triacylglycerol, smoking status, alcohol status and health status

ORs and *p* values were calculated from the logistic regression model

T2D, type 2 diabetes

Table 5 Metabolites from untargeted measurements that were significantly associated with incidence of IGR and type 2 diabetes in different pairwise comparisons

Comparison	OR (95% CI)	*p* value
18 months		
398 NGR vs 244 IGR		
Formiminoglutamate	0.369 (0.157, 0.588)	7.7×10^{-4}
Lactate	0.373 (0.143, 0.557)	0.002
N-Lactoylleucine	0.294 (0.079, 0.514)	0.008
N-Lactoylvaline	0.248 (0.039, 0.460)	0.021
X-24295	0.225 (0.022, 0.432)	0.031
897 IGR vs 71 T2D		
X-24295	0.474 (0.162, 0.801)	$3.6x10^{-3}$
Lactate	0.409 (0.077, 0.747)	$1.6x10^{-2}$
48 months		
821 IGR vs 128 T2D		
X-24295	0.474 (0.162, 0.801)	$3.6x10^{-3}$
Lactate	0.409 (0.077, 0.747)	$1.6x10^{-2}$

Baseline metabolites were taken as the independent variables with glycaemic category in different timelines (18 months and 48 months) as the dependent variables, adjusted by study centre, sex, age, BMI, BP, fasting HDL-cholesterol, fasting triacylglycerol, smoking status, alcohol status and health status

ORs and *p* values were calculated from the logistic regression model

T2D, type 2 diabetes

dataset, two *n*-3 fatty acids, namely stearidonate (18:4n3) and docosapentaenoate (n3 DPA; 22:5n3), were identified to be causal for type 2 diabetes. Detailed statistics of our MR analysis are presented in ESM Tables 29–32.

Discussion

In this study, we used untargeted metabolomics to provide semi-quantitative global screening of metabolites in the development of a disease whereas targeted metabolomics was used to quantify a pre-selected subset of metabolites with absolute concentrations. However, the overlap between the two metabolomic techniques was limited to a few amino acids and lipids. In the current study we report 19 metabolites (three from targeted and 14 from global profiling, plus one common lysoPC a C18:0 / 1-stearoyl-GPC [18:0]) that were significantly associated with prediabetes in the DIRECT cohort. The advantages of global profiling become evident as it allows for the identification of a broader spectrum of metabolites. Few notable examples are given here. First, picolinoylglycine (HMDB0059766), which is potentially a phase II product of picolinic acid, a degradation product of tryptophan [29] and glycine [30], and shows potential as a novel marker for glycaemic deterioration. Prediabetes is often associated with dyslipidaemia, marked by an imbalanced lipid profile compared with individuals with NGR [24]. Second, *N*-lactoyl amino acids are not infrequently observed in metabolomic datasets. In fact it has come to light that *N*-lactoyl amino acids were misidentified in some metabolomic studies and were erroneously reported as 1-carboxyethyl amino acids. In particular, *N*-lactoyl-phenylalanine (Lac-Phe) is known to act as an appetite suppressant when given to obese mice [31]. However, in humans Lac-Phe concentrations were observed to rise following vigorous exercise [32]. In fact, the most recent study shows that Lac-Phe facilitates the impact of metformin on both food intake and body weight [33, 34]. It seems that the exact role of Lac-Phe in the human body and pathways downstream, such as energy metabolism, insulin signalling, exercise-induced pathways, are unclear and needs further research.

We are aware of several limitations to our study. Although metabolomics screening showcases numerous valuable attributes in health science, challenges inherent to this approach continue to exist, especially in the accurate identification of metabolites which is crucial for the biological interpretation and validation of metabolomics

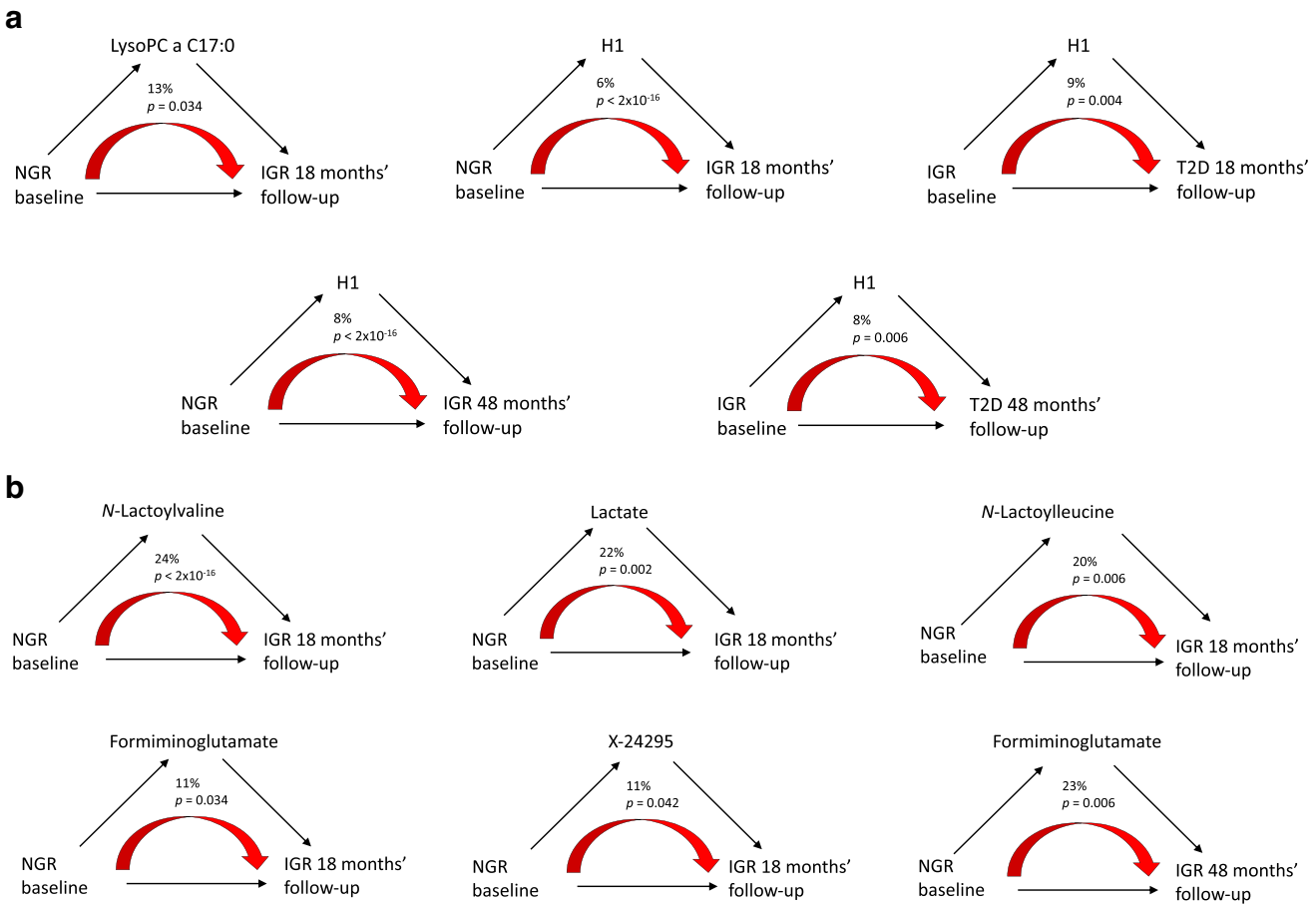


Fig. 3 Schematic overview of mediation analysis with lysoPC a C17:0 and hexoses (**a**) or *N*-lactoylvaline, lactate, *N*-lactoylleucine, formiminoglutamate and X-24295 (**b**) as mediators. Numbers above the red arrows indicate the percentage and significance of mediation effects. T2D, type 2 diabetes

data [35]. Variability in sample collection, preparation and analytical techniques can impact the reproducibility and comparability of results across different studies. Standardisation efforts are ongoing but may not fully address all sources of variation. The identification of metabolites, especially in untargeted metabolomics, can be challenging. Incomplete databases and the presence of unknown or novel metabolites have been reported with a metabolite name with an asterisk (*) sign. However, ongoing advancements in technology, methodology and standardisation efforts aim to enhance the robustness and applicability of metabolomics studies [35]. The current study is predominantly based on White male participants from the Kuopio region of Europe, and for this reason an additional sex-based sensitivity analysis has been performed and reported separately (ESM Results 1 and 2). Challenges in MR studies include limited statistical power, potential reverse causation, confounding and pleiotropy [36]. Caution is advised in interpreting causality inference, considering the various limitations mentioned in the methods, and precautionary measures were taken by using valid MR instruments and reporting Bonferroni significance.

A drug–metabolomics associations study [17] was examined to determine whether or not metabolites linked to type 2 diabetes from the DIRECT study were also associated with a particular drug. Looking at our results and those of Allesøe et al [17], we found that 44% (15 out of 34) of targeted metabolites and 3% (three out of 99) of non-targeted metabolites that were significantly associated with type 2 diabetes also showed a significant association with at least one of the 20 drugs. This suggests that metabolites linked to type 2 diabetes may be confounded by polypharmacy.

However, metabolite association with incident prediabetes or diabetes (IGR-T2D) showed that lysoPC a C17:0 could predict the risk of developing IGR at 18 months and 48 months. It has already been shown that lysoPCs differ significantly between individuals with incident IGT or type 2 diabetes and individuals with NGR in the KORA study

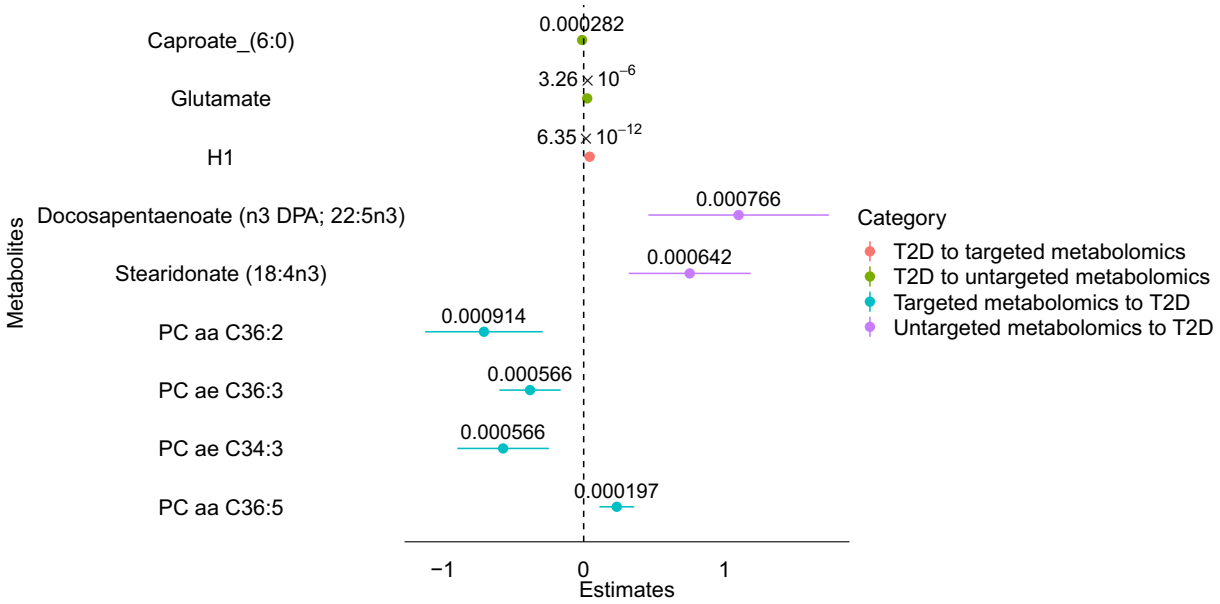


Fig. 4 Forest plot representing causal estimates of type 2 diabetes on targeted and untargeted metabolites in the two-sample MR test. T2D, type 2 diabetes

[37]. LysoPC a C17:0 was negatively associated with diabetes, a finding that was confirmed in several studies [38, 39]. The aforementioned drug–metabolomics association study [17] showed that lysoPC a 17:0 was not associated with the drugs. However, the origin of odd-chain fatty acids (mainly C15:0 and C17:0) remains elusive. Jenkins et al [40] investigated the origin of circulating odd-chain fatty acids (C17:0, C15:0) through a combination of animal and human studies to determine possible contributions of fatty acids from the gut-microbiota, diet and novel endogenous biosynthesis [41]. The findings suggested that C15:0 was linked to dietary intake, while C17:0 was predominantly biosynthesised, indicating independent origins and non-homologous roles in disease causation.

Causal mediation analysis indicated that plasma lactate strongly mediates the effects of identified metabolites in the transition from baseline glycaemic status to follow-up [42]. In a longitudinal study of Swedish men, elevated serum lactate was independently linked to a higher incidence of type 2 diabetes, irrespective of obesity measures [43]. Formiminoglutamate was confirmed to be associated with a higher risk of incident type 2 diabetes in older Puerto Ricans [44]. *N*-lactoylleucine and *N*-lactoylvaline, derivatives of leucine and valine, respectively, are ubiquitous pseudodipeptides of lactic acid and amino acids that are formed by reverse proteolysis [32] and are correlated with underivatised amino acids in human plasma. The Microbiome and Insulin Longitudinal Evaluation Study (MILES) [45] investigated the association between *ABO* haplotypes and insulin-related characteristics, and explored possible pathways that could mediate these associations. The study showed that the A1 haplotype potentially enhances favourable insulin sensitivity in non-Hispanic White individuals, with lactate likely influencing this mechanism, while gut bacteria are not believed to be a contributing factor.

In MR, causality signifies that modifying exposure leads to a predictable change in the outcome. Our 2SMR analysis suggests that the metabolites causal for type 2 diabetes are PC aa C36:2, PC aa C36:5, PC ae C34:3 and PC ae C36:3 and all these metabolites are significantly associated with drug–metabolomics. However, from untargeted metabolomics two *n*-3 fatty acids, namely stearidonate (18:4n3) and docosapentaenoate DPA 22:5n3), are not further associated with drugs. In 2012, Banz et al [46] explored the therapeutic implications of stearidonate acid in preventing or managing type 2 diabetes. The Fatty Acids and Outcomes Research Consortium (FORCE) [47] found that higher circulating biomarkers of seafood-derived *n*-3 fatty acids were associated with lower type 2 diabetes risk. On the contrary, branched-chain amino acids [48] and sphingomyelin [15] have been shown to have a causal role in type 2 diabetes development, a correlation not observed in the DIRECT study.

Conclusions

Our study demonstrates that alteration in blood plasma metabolites is associated with glycaemic deterioration. The progression from prediabetes to diabetes is mediated by novel

metabolites such as picolinoylglycine and *N*-lactoyl-amino acids, as demonstrated by evidence from the DIRECT study. *N*-lactoyl-amino acids are known to be exercise-induced metabolites that suppress food intake and influence glucose homeostasis. Additional functional research and quantification are needed to advance the identification of early metabolic biomarkers such as *N*-lactoyl-amino acids, which have the potential to forecast the onset of type 2 diabetes. Collectively, these findings direct attention towards novel metabolic signatures associated with glycaemic deterioration.

Supplementary Information The online version contains peer-reviewed but unedited supplementary material available at https://doi.org/10.1007/s00125-024-06282-6.

Acknowledgements We extend our gratitude to the IMI-DIRECT study participants who willingly participated in phenotyping as well as to the clinical and technical staff across European study centres for their contributions to participant recruitment and clinical assessment. This publication's development has been supported by the Innovative Medicines Initiative Joint Undertaking under grant agreement 115317 (DIRECT), with resources derived from the European Union's Seventh Framework Programme (FP7/2007–2013) and EFPIA companies' in-kind contributions. Special thanks go to the study centre team, L. M. 't Hart, F. Rutters, J. Vangipurapu and T. H. Hansen, for providing internal review on this manuscript.

Data availability Access to the molecular and clinical raw data, as well as the processed data, is restricted. This is in accordance with the informed consent provided by study participants, the various national ethical approvals obtained for the study, and compliance with the European General Data Protection Regulation (GDPR). Individual-level clinical and molecular data cannot be transferred from the centralised IMI-DIRECT repository. Requests for access will receive guidance on accessing data through the DIRECT secure analysis platform after submitting an appropriate application. The IMI-DIRECT data access policy and additional information about the IMI-DIRECT research consortium's initiatives and activities can be found at https://directdiabetes.org.

Code used for MR in the study is included as ESM.

Funding Open Access funding enabled and organized by Projekt DEAL. We would like to thank Helmholtz Munich, German Diabetes Center (DZD) for their support in current research and China Research Council (CRC) funding for a PhD student hosted by Helmholtz Munich.

Authors' relationships and activities The authors declare that there are no relationships or activities that might bias, or be perceived to bias, their work.

Contribution statement SS, QD, MH, JA and HG conceptualised the analysis plan. RWK, PWF, MW, IF, GG, IP, HR, MD, MM, OP, JS, KT, FDM, SB, AV, AM, TM, TK, JA, EP and HG were involved in conception and design of the DIRECT study. SS, QD, MH, JA, GK, AA, CP, RB, JFT, AJ and AT were involved in the data acquisition, pre-processing and interpretation of data. SS organised inclusion of outlined sections and, along with QD, wrote the original draft of the manuscript. All authors contributed to drafting the article or critically revising it for significant intellectual content and have provided approval to the final version to be published. SS and HG are the guarantors of this work and, as such, had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

References

1. Zimmet PZ (2017) Diabetes and its drivers: the largest epidemic in human history? Clin Diabetes Endocrinol 3:1. https://doi.org/10.1186/s40842-016-0039-3
2. Wesolowska-Andersen A, Brorsson CA, Bizzotto R et al (2022) Four groups of type 2 diabetes contribute to the etiological and clinical heterogeneity in newly diagnosed individuals: an IMI DIRECT study. Cell Rep Med 3(1):100477. https://doi.org/10.1016/j.xcrm.2021.100477
3. Emerging Risk Factors Collaboration, Sarwar N, Gao P et al (2010) Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. Lancet 375(9733):2215–2222. https://doi.org/10.1016/S0140-6736(10)60484-9
4. Emerging Risk Factors Collaboration, Di Angelantonio E, Kaptoge S et al (2015) Association of cardiometabolic multimorbidity with mortality. JAMA 314(1):52–60. https://doi.org/10.1001/jama.2015.7008
5. Rao KondapallySeshasai S, Kaptoge S, Thompson A et al (2011) Diabetes mellitus, fasting glucose, and risk of cause-specific death. N Engl J Med 364(9):829–841. https://doi.org/10.1056/NEJMoa1008862
6. Wurtz P, Kangas AJ, Soininen P, Lawlor DA, Davey Smith G, Ala-Korpela M (2017) Quantitative serum nuclear magnetic resonance metabolomics in large-scale epidemiology: a primer on -Omic technologies. Am J Epidemiol 186(9):1084–1096. https://doi.org/10.1093/aje/kwx016
7. Guasch-Ferre M, Hruby A, Toledo E et al (2016) Metabolomics in prediabetes and diabetes: a systematic review and meta-analysis. Diabetes Care 39(5):833–846. https://doi.org/10.2337/dc15-2251
8. Cai L, Wheeler E, Kerrison ND et al (2020) Genome-wide association analysis of type 2 diabetes in the EPIC-InterAct study. Sci Data 7(1):393. https://doi.org/10.1038/s41597-020-00716-7
9. Yengo L, Sidorenko J, Kemper KE et al (2018) Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. Hum Mol Genet 27(20):3641–3649. https://doi.org/10.1093/hmg/ddy271
10. Shin SY, Fauman EB, Petersen AK et al (2014) An atlas of genetic influences on human blood metabolites. Nat Genet 46(6):543–550. https://doi.org/10.1038/ng.2982
11. Gieger C, Geistlinger L, Altmaier E et al (2008) Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. PLoS Genet 4(11):e1000282. https://doi.org/10.1371/journal.pgen.1000282
12. Lanznaster D, Veyrat-Durebex C, Vourc'h P, Andres CR, Blasco H, Corcia P (2020) Metabolomics: a tool to understand the impact

of genetic mutations in amyotrophic lateral sclerosis. Genes (Basel) 11(5):537. https://doi.org/10.3390/genes11050537
13. Suhre K, Raffler J, Kastenmuller G (2016) Biochemical insights from population studies with genetics and metabolomics. Arch Biochem Biophys 589:168–176. https://doi.org/10.1016/j.abb.2015.09.023
14. Koivula RW, Forgie IM, Kurbasic A et al (2019) Discovery of biomarkers for glycaemic deterioration before and after the onset of type 2 diabetes: descriptive characteristics of the epidemiological studies within the IMI DIRECT Consortium. Diabetologia 62(9):1601–1615. https://doi.org/10.1007/s00125-019-4906-1
15. Zhang Z, Zheng C, Kim C, Van Poucke S, Lin S, Lan P (2016) Causal mediation analysis in the context of clinical research. Ann Transl Med 4(21):425. https://doi.org/10.21037/atm.2016.11.11
16. Dong Q, Sidra S, Gieger C et al (2023) Metabolic signatures elucidate the effect of body mass index on type 2 diabetes. Metabolites 13(2):227. https://doi.org/10.3390/metabo13020227
17. Allesøe RL, Lundgaard AT, Hernandez Medina R et al (2023) Discovery of drug-omics associations in type 2 diabetes with generative deep-learning models. Nat Biotechnol 41(3):399–408. https://doi.org/10.1038/s41587-022-01520-x
18. Koivula RW, Heggie A, Barnett A et al (2014) Discovery of biomarkers for glycaemic deterioration before and after the onset of type 2 diabetes: rationale and design of the epidemiological studies within the IMI DIRECT Consortium. Diabetologia 57(6):1132–1142. https://doi.org/10.1007/s00125-014-3216-x
19. Römisch-Margl W, Prehn C, Bogumil R, Röhring C, Suhre K, Adamski J (2012) Procedure for tissue sample preparation and metabolite extraction for high-throughput targeted metabolomics. Metabolomics. 8:133–142. https://doi.org/10.1007/s11306-011-0293-4
20. Do KT, Wahl S, Raffler J et al (2018) Characterization of missing values in untargeted MS-based metabolomics data and evaluation of missing data handling strategies. Metabolomics 14(10):128. https://doi.org/10.1007/s11306-018-1420-2
21. Bizzotto R, Jennison C, Jones AG et al (2021) Processes underlying glycemic deterioration in type 2 diabetes: an IMI DIRECT study. Diabetes Care 44(2):511–518. https://doi.org/10.2337/dc20-1567
22. Baron RM, Kenny DA (1986) The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol 51(6):1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173
23. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K (2014) mediation: R package for causal mediation analysis. J Stat Soft 59(5):1–38. https://doi.org/10.18637/jss.v059.i05
24. Hemani G, Zheng J, Elsworth B et al (2018) The MR-Base platform supports systematic causal inference across the human phenome. Elife 7. https://doi.org/10.7554/eLife.34408
25. Zaghlool SB, Sharma S, Molnar M et al (2021) Revealing the role of the human blood plasma proteome in obesity using genetic drivers. Nat Commun 12(1):1279. https://doi.org/10.1038/s41467-021-21542-4
26. Xue A, Wu Y, Zhu Z et al (2018) Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. Nat Commun 9(1):2941. https://doi.org/10.1038/s41467-018-04951-w
27. Draisma HHM, Pool R, Kobl M et al (2015) Genome-wide association study identifies novel genetic variants contributing to variation in blood metabolite levels. Nat Commun 6:7208. https://doi.org/10.1038/ncomms8208
28. Brown AA, Fernandez-Tajes JJ, Hong MG et al (2023) Genetic analysis of blood molecular phenotypes reveals common properties in the regulatory networks affecting complex traits. Nat Commun 14(1):5062. https://doi.org/10.1038/s41467-023-40569-3
29. Hu Q, Jin L, Zeng J et al (2020) Tryptophan metabolite-regulated Treg responses contribute to attenuation of airway inflammation during specific immunotherapy in a mouse asthma model. Hum Vaccin Immunother 16(8):1891–1899. https://doi.org/10.1080/21645515.2019.1698900
30. Choi JY, Kim SH, Kim JE et al (2019) Four amino acids as serum biomarkers for anti-asthma effects in the ovalbumin-induced asthma mouse model treated with extract of Asparagus cochinchinensis. Lab Anim Res 35:32. https://doi.org/10.1186/s42826-019-0033-x
31. Li VL, He Y, Contrepois K et al (2022) An exercise-inducible metabolite that suppresses feeding and obesity. Nature 606(7915):785–790. https://doi.org/10.1038/s41586-022-04828-5
32. Jansen RS, Addie R, Merkx R et al (2015) N-lactoyl-amino acids are ubiquitous metabolites that originate from CNDP2-mediated reverse proteolysis of lactate and amino acids. Proc Natl Acad Sci U S A 112(21):6601–6606. https://doi.org/10.1073/pnas.1424638112
33. Xiao S, Li VL, Lyu X et al (2024) Lac-Phe mediates the effects of metformin on food intake and body weight. Nat Metab 6(4):659–669. https://doi.org/10.1038/s42255-024-00999-9
34. Scott B, Day EA, O'Brien KL et al (2024) Metformin and feeding increase levels of the appetite-suppressing metabolite Lac-Phe in humans. Nat Metab 6(4):651–658. https://doi.org/10.1038/s42255-024-01018-7
35. Kirwan JA, Gika H, Beger RD et al (2022) Quality assurance and quality control reporting in untargeted metabolic phenotyping: mQACC recommendations for analytical quality management. Metabolomics 18(9):70. https://doi.org/10.1007/s11306-022-01926-3
36. Davey Smith G, Hemani G (2014) Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet 23(R1):R89-98. https://doi.org/10.1093/hmg/ddu328
37. Wang-Sattler R, Yu Z, Herder C et al (2012) Novel biomarkers for pre-diabetes identified by metabolomics. Mol Syst Biol 8:615. https://doi.org/10.1038/msb.2012.43
38. Floegel A, Stefan N, Yu Z et al (2013) Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. Diabetes 62(2):639–648. https://doi.org/10.2337/db12-0495
39. Lee H-S, Xu T, Lee Y et al (2016) Identification of putative biomarkers for type 2 diabetes using metabolomics in the Korea Association REsource (KARE) cohort. Metabolomics 12(12):178. https://doi.org/10.1007/s11306-016-1103-9
40. Jenkins BJ, Seyssel K, Chiu S et al (2017) Odd chain fatty acids; new insights of the relationship between the gut microbiota, dietary intake, biosynthesis and glucose intolerance. Sci Rep 7:44845. https://doi.org/10.1038/srep44845
41. Weitkunat K, Schumann S, Nickel D et al (2017) Odd-chain fatty acids as a biomarker for dietary fiber intake: a novel pathway for endogenous production from propionate. Am J Clin Nutr 105(6):1544–1551. https://doi.org/10.3945/ajcn.117.152702
42. Crawford SO, Hoogeveen RC, Brancati FL et al (2010) Association of blood lactate with type 2 diabetes: the Atherosclerosis Risk in Communities Carotid MRI Study. Int J Epidemiol 39(6):1647–1655. https://doi.org/10.1093/ije/dyq126
43. Ohlson LO, Larsson B, Bjorntorp P et al (1988) Risk factors for type 2 (non-insulin-dependent) diabetes mellitus. Thirteen and one-half years of follow-up of the participants in a study of Swedish men born in 1913. Diabetologia 31(11):798–805. https://doi.org/10.1007/BF00277480
44. Rivas-Tumanyan S, Pacheco LS, Haslam DE et al (2022) Novel plasma metabolomic markers associated with diabetes progression in older puerto ricans. Metabolites 12(6):513. https://doi.org/10.3390/metabo12060513
45. Li-Gao R, Grubbs K, Bertoni AG et al (2022) The roles of gut microbiome and plasma metabolites in the associations between ABO blood groups and insulin homeostasis: the Microbiome and Insulin Longitudinal Evaluation Study (MILES). Metabolites 12(9):787. https://doi.org/10.3390/metabo12090787

46. Banz WJ, Davis JE, Clough RW, Cheatwood JL (2012) Stearidonic acid: is there a role in the prevention and management of type 2 diabetes mellitus? J Nutr 142(3):635S-640S. https://doi.org/10.3945/jn.111.146829
47. Qian F, Ardisson Korat AV, Imamura F et al (2021) n-3 fatty acid biomarkers and incident type 2 diabetes: an individual participant-level pooling project of 20 prospective cohort studies. Diabetes Care 44(5):1133–1142. https://doi.org/10.2337/dc20-2426
48. Lotta LA, Scott RA, Sharp SJ et al (2016) Genetic predisposition to an impaired metabolism of the branched-chain amino acids and risk of type 2 diabetes: a mendelian randomisation analysis. PLoS Med 13(11):e1002179. https://doi.org/10.1371/journal.pmed.1002179

Authors and Affiliations

Sapna Sharma[1] · Qiuling Dong[1,2] · Mark Haid[3] · Jonathan Adam[1,4] · Roberto Bizzotto[5] · Juan J. Fernandez-Tajes[6] · Angus G. Jones[7] · Andrea Tura[5] · Anna Artati[3] · Cornelia Prehn[3] · Gabi Kastenmüller[8] · Robert W. Koivula[9] · Paul W. Franks[10] · Mark Walker[11] · Ian M. Forgie[12] · Giuseppe Giordano[13] · Imre Pavo[14] · Hartmut Ruetten[15] · Manolis Dermitzakis[16,17,18] · Mark I. McCarthy[6] · Oluf Pedersen[19,20] · Jochen M. Schwenk[21] · Konstantinos D. Tsirigos[22] · Federico De Masi[22] · Soren Brunak[20,23] · Ana Viñuela[24] · Andrea Mari[5] · Timothy J. McDonald[25] · Tarja Kokkola[26] · Jerzy Adamski[27,28,29] · Ewan R. Pearson[12] · Harald Grallert[1,4]

✉ Sapna Sharma
sapna.sharma@helmholtz-munich.de

✉ Harald Grallert
harald.grallert@helmholtz-munich.de

1 Research Unit of Molecular Epidemiology, Institute of Epidemiology, German Research Center for Environmental Health, Helmholtz Zentrum München, Neuherberg, Germany

2 Faculty of Medicine, Ludwig-Maximilians-University München, Munich, Germany

3 Metabolomics and Proteomics Core, German Research Center for Environmental Health, Helmholtz Zentrum München, Neuherberg, Germany

4 German Center for Diabetes Research (DZD), München Neuherberg, Germany

5 Institute of Neuroscience, National Research Council, Padova, Italy

6 Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK

7 Department of Clinical and Biomedical Sciences, University of Exeter College of Medicine & Health, Exeter, UK

8 Institute of Computational Biology, Helmholtz Zentrum München, Munich, Germany

9 Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK

10 Department of Clinical Science, Genetic and Molecular Epidemiology, Lund University Diabetes Centre, Malmö, Sweden

11 Translational and Clinical Research Institute, Faculty of Medical Sciences, University of Newcastle, Newcastle upon Tyne, UK

12 Population Health and Genomics, Ninewells Hospital and Medical School, University of Dundee, Dundee, UK

13 Department of Clinical Science, Genetic and Molecular Epidemiology, Lund University Diabetes Centre, Malmö, Sweden

14 Eli Lilly Regional Operations GmbH, Vienna, Austria

15 Sanofi Partnering, Sanofi-Aventis Deutschland GmbH, Frankfurt am Main, Germany

16 Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland

17 Institute for Genetics and Genomics in Geneva (iGE3), University of Geneva, Geneva, Switzerland

18 Swiss Institute of Bioinformatics, Geneva, Switzerland

19 Center for Clinical Metabolic Research, Herlev and Gentofte University Hospital, Copenhagen, Denmark

20 Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

21 Science for Life Laboratory, School of Biotechnology, KTH - Royal Institute of Technology, Solna, Sweden

22 Department of Health Technology, Technical University of Denmark, Kongens Lyngby, Denmark

23 Department of Health Technology, Technical University of Denmark, Kongens Lyngby, Denmark

24 Biosciences Institute, Faculty of Medical Sciences, University of Newcastle, Newcastle upon Tyne, UK

25 Blood Sciences, Royal Devon and Exeter NHS Foundation Trust, Exeter, UK

26 Internal Medicine, Institute of Clinical Medicine, University of Eastern Finland, Kuopio, Finland

27 Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

28 Institute of Experimental Genetics, German Research Center for Environmental Health, Helmholtz Zentrum München, Neuherberg, Germany

29 Institute of Biochemistry, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia

Paper III

Title: Subphenotypes of adult-onset diabetes: Data-driven clustering in the population-based KORA cohort

Authors: Qiuling Dong, Yue Xi, Stefan Brandmaier, Markéta Fuchs, Marie-Theres Huemer, Melanie Waldenberger, Jiefei Niu, Christian Herder, Wolfgang Rathmann, Michael Roden, Wolfgang Koenig, Gidon J. Bönhof, Christian Gieger, Barbara Thorand, Annette Peters, Susanne Rospleszcz, Harald Grallert

Journal: Diabetes, Obesity and Metabolism

Status: Published

Year: 2024

doi: https://doi.org/10.1111/dom.16022

Supplements:https://dom-pubs.onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1111%2Fdom.16022&file=dom16022-sup-0001-Supinfo.docx

Received: 10 August 2024 | Revised: 27 September 2024 | Accepted: 6 October 2024

DOI: 10.1111/dom.16022

ORIGINAL ARTICLE

WILEY

Subphenotypes of adult-onset diabetes: Data-driven clustering in the population-based KORA cohort

Qiuling Dong MSc[1,2,3] | Yue Xi MSc[2,3] | Stefan Brandmaier PhD[1,2] | Markéta Fuchs MSc[1,2] | Marie-Theres Huemer PhD[2] | Melanie Waldenberger PhD[1,2] | Jiefei Niu MSc[1,2] | Christian Herder PhD[4,5,6] | Wolfgang Rathmann MD[4,7] | Michael Roden MD[4,5,6] | Wolfgang Koenig MD[8,9,10] | Gidon J. Bönhof PhD[4,5] | Christian Gieger PhD[1,2,11] | Barbara Thorand PhD[2,4,12] | Annette Peters PhD[2,4,8,12] | Susanne Rospleszcz PhD[2,8,12,13] | Harald Grallert PhD[1,2,11]

[1]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany

[2]Institute of Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany

[3]Pettenkofer School of Public Health, Faculty of Medicine, LMU Munich, Munich, Germany

[4]German Center for Diabetes Research (DZD), Düsseldorf, Germany

[5]Institute for Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf, Düsseldorf, Germany

[6]Department of Endocrinology and Diabetology, Medical Faculty and University Hospital Dusseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

[7]Institute for Biometrics and Epidemiology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf, Düsseldorf, Germany

[8]German Research Center for Cardiovascular Disease (DZHK), Partner site Munich Heart Alliance, Munich, Germany

[9]Deutsches Herzzentrum München, Technische Universität München, Munich, Germany

[10]Institute of Epidemiology and Medical Biometry, University of Ulm, Ulm, Germany

[11]German Center for Diabetes Research (DZD), Neuherberg, Germany

[12]Chair of Epidemiology, Institute for Medical Information Processing, Biometry, and Epidemiology (IBE), Faculty of Medicine, Ludwig-Maximilians-University München, Munich, Germany

[13]Department of Diagnostic and Interventional Radiology, Medical Center – University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany

Correspondence
Qiuling Dong and Harald Grallert, Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, 85764 Neuherberg, Germany.
Email: qiuling.dong@helmholtz-munich.de and harald.grallert@helmholtz-munich.de

Funding information
The German Center for Diabetes Research (DZD); German Federal Ministry of Education and Research; Munich Center of Health Sciences; Ludwig-Maximilians-Universität München; Helmholtz Munich—German Research Center for Environmental Health;

Abstract

Aims: A data-driven cluster analysis in a cohort of European individuals with type 2 diabetes (T2D) has previously identified four subgroups based on clinical characteristics. In the current study, we performed a comprehensive statistical assessment to (1) replicate the above-mentioned original clusters; (2) derive de novo T2D subphenotypes in the Kooperative Gesundheitsforschung in der Region Augsburg (KORA) cohort and (3) describe underlying genetic risk and diabetes complications.

Methods: We used data from $n = 301$ individuals with T2D from KORA FF4 study (Southern Germany). Original cluster replication was assessed forcing $k = 4$ clusters

Susanne Rospleszcz and Harald Grallert contributed equally to this study.

China Scholarship Council, Grant/Award Number: 202008310176; State of Bavaria; German Federal Ministry of Health; Ministry of Culture and Science of the State North Rhine-Westphalia

using three different hyperparameter combinations. De novo clusters were derived by open *k*-means analysis. Stability of de novo clusters was assessed by assignment congruence over different variable sets and Jaccard indices. Distribution of polygenic risk scores and diabetes complications in the respective clusters were described as an indication of underlying heterogeneity.

Results: Original clusters did not replicate well, indicated by substantially different assignment frequencies and cluster characteristics between the original and current sample. De novo clustering using $k = 3$ clusters and including high sensitivity C-reactive protein in the variable set showed high stability (all Jaccard indices >0.75). The three de novo clusters ($n = 96$, $n = 172$, $n = 33$, respectively) adequately captured heterogeneity within the sample and showed different distributions of polygenic risk scores and diabetes complications, that is, cluster 1 was characterized by insulin resistance with high neuropathy prevalence, cluster 2 was defined as age-related diabetes and cluster 3 showed highest risk of genetic and obesity-related diabetes.

Conclusion: T2D subphenotyping based on its sample's own clinical characteristics leads to stable categorization and adequately reflects T2D heterogeneity.

KEYWORDS
clustering, cohort study, database research, diabetes complications, type 2 diabetes

1 | INTRODUCTION

Diabetes is a rapidly growing global health concern.[1,2] The underlying causes of pancreatic beta-cell dysfunction are heterogeneous, and individual trajectories of hyperglycaemia and subsequent diabetes complications vary widely.[3,4] Therefore, classifications of type 2 diabetes (T2D) that predict the risk of complications and provide options for a tailored treatment have been actively studied.[5–8]

Traditionally, diabetes is mainly classified into type 1 (T1D) and T2D, primarily determined by the presence (T1D) or absence (T2D) of autoantibodies. A novel approach to identify subphenotypes of diabetes was the hallmark study by Ahlqvist et al.[9] They used six diabetes-related variables including age at diagnosis, body mass index (BMI), haemoglobin A1c (HbA1c), homeostasis model assessment (HOMA) estimates of beta-cell function (HOMA2-B) and insulin resistance (HOMA2-IR) and glutamic acid decarboxylase antibodies (GADA) to categorize individuals with diabetes into five clusters. Thereby, four clusters mainly represent T2D subphenotypes and one cluster with severe autoimmune diabetes (SAID) mainly corresponds to the T1D subphenotype. The four T2D subphenotypes were labelled based on their distinctive features as severe insulin-deficient diabetes (SIDD), severe insulin resistant diabetes (SIRD), mild obesity-related diabetes (MOD) and mild age-related diabetes (MARD) and exhibited different risks of disease progression and diabetes complications. These clusters have been replicated in diverse ethnic groups such as British,[10] German,[11,12] American and Chinese,[13,14] Mexican,[15] Icelandic,[16] Japanese[17] and Asian Indian cohorts.[18] Recently, subphenotypes were characterized in more detail from a molecular perspective, including potential underlying genetic determinants[19,20] and cluster-specific signatures of metabolomics and proteomics.[21,22] There appear to be differences in biomarkers of inflammation between diabetes subphenotypes, which is in line with the involvement of inflammatory mechanisms, most often assessed by C-reactive protein (CRP), in the progression of diabetes.[12,23] Taken together, the current state of evidence suggests that diabetes subphenotyping, including deep molecular phenotyping, holds the potential to offer key insights into the underlying pathophysiology of glucose dysregulation and the onset of comorbidities among individuals with T2D, while it further enables the advancement of personalized treatment of diabetes.

In the current study, we aimed to perform a comprehensive statistical assessment of T2D subphenotyping in the Cooperative Health Research in the Region of Augsburg (KORA) FF4 cohort (Southern Germany). Our aims were threefold: (1) to investigate to which extent the original clusters from Ahlqvist et al.[9] could be replicated in the KORA sample; (2) to derive novel T2D subphenotypes based on data-driven clustering, also accounting for inflammation and (3) to investigate heterogeneity between the de novo derived subphenotypes by describing the distribution of genetically predicted risk as captured by a polygenic risk score (PRS), diabetes-related complications and parental history of diabetes. An overview of the study design is shown in Figure 1.

2 | METHODS

2.1 | Study population and clinical data

KORA comprises several deeply phenotyped population-based epidemiological surveys.[24] The current analysis is based on data from the KORA-FF4 study, conducted between 2013 and 2014. Details about

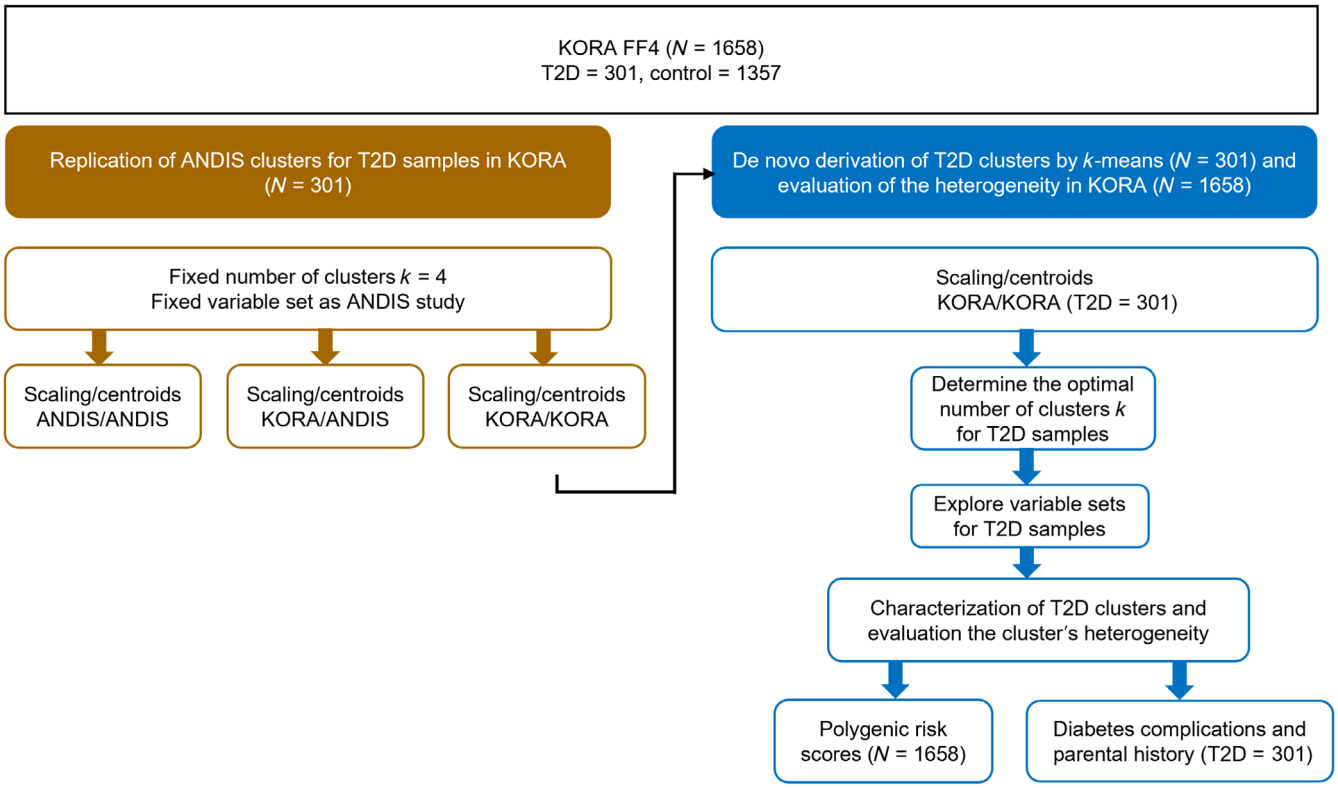


FIGURE 1 Study design. The left part in orange corresponds to aim (1) whereas the right part in blue corresponds to aims (2) and (3). The fixed variable set contained the basic variables: Age, body mass index, haemoglobin A1c, homeostasis model assessment (HOMA) estimates of beta-cell function (HOMA2-B) and insulin resistance (HOMA2-IR), corresponding to the original ANDIS study. Variable sets in KORA contained the basic variables plus one additional variable, respectively: High sensitivity C-reactive protein, triglycerides, HDL-cholesterol or systolic blood pressure. ANDIS, Swedish All New Diabetics in Scania cohort; KORA, 'Cooperative Health Research in the Region of Augsburg' cohort; T2D, type 2 diabetes; PRS, polygenic risk score.

TABLE 1 Characteristics of the KORA FF4 participants for men and women.

	Men ($N = 178$)	Women ($N = 123$)	p
Age at examination (years) mean (SD)	69.6 (10.0)	69.4 (10.2)	0.83
BMI (kg/m^2) mean (SD)	30.3 (4.9)	32.2 (5.7)	0.003
HbA1c (mmol/mol) mean (SD)	46.9 (11.7)	47.9 (10.9)	0.48
HOMA2-B % mean (SD)	72.3 (38.4)	71.5 (31.8)	0.85
HOMA2-IR (SD)	2.1 (1.2)	2.0 (1.0)	0.52
hsCRP (mg/L) mean (SD)	2.7 (3.3)	4.5 (6.0)	0.001
TG (mmol/L) mean (SD)	1.9 (1.1)	1.6 (0.8)	0.025
HDL-C (mmol/L) mean (SD)	1.4 (0.4)	1.6 (0.4)	<0.001
SBP (mmHg) mean (SD)	130.4 (17.8)	122.3 (19.9)	<0.001
Fasting glucose (mmol/L) mean (SD)	7.6 (2.0)	7.5 (1.9)	0.543
Use of metformin	84 (47.2)	57 (46.3)	0.94
Any oral antidiabetic medication or insulin treatment	96 (53.9)	64 (52.0)	0.80
Known diabetes (%)	123 (69.1%)	84 (68.3%)	0.982

Note: Mean and standard deviation (SD) are provided for quantitative variables and differences were evaluated by student's t test. Count and percentage are provided for categorical variables and differences were evaluated by chi square test.
Abbreviations: BMI, body mass index; HbA1C, haemoglobin A1c; HDL-C, high-density lipoprotein cholesterol; hsCRP, high sensitivity C-reactive protein; HOMA2-B, homeostasis model assessment estimates of beta-cell function; HOMA2-IR, homeostasis model assessment estimates of insulin resistance; Known diabetes, the diabetes diagnosis was known prior to the study; TG, triglycerides; SBP, systolic blood pressure.

the study sample and the assessment of clinical data are presented in Supplementary Material.

For the current analysis, only participants with T2D were included for cluster analysis. Participants with T1D ($n = 6$) were excluded from all analyses. Moreover, participants with missing values for clustering variables (described below) were excluded ($n = 20$). Finally, the cluster analysis comprised $N = 301$ individuals with T2D (Table 1). For the assessment of genetically T2D risk, a PRS was calculated for all individuals with T2D ($N = 301$) and without T2D ($N = 1357$) as the control group (Table S1).

2.2 | Genotyping and polygenetic risk score

Genetically predicted T2D risk was calculated by an established PRS, as described in Supplementary Material.

2.3 | Statistical analysis

Statistical analysis was conducted using R version 4.1.1. A two-sided p value <0.05 was considered statistically significant. A detailed description of (1) the replication of original clusters, including different combinations of scaling and centroid hyperparameters, (2) de novo cluster derivation in the KORA study and (3) assessment of differences between clusters with respect to PRS, parental history of diabetes and diabetes complications is presented in Supplementary Material.

3 | RESULTS

3.1 | Study sample

The final sample included 301 individuals with T2D, thereof 94 (31.2%) with newly detected diabetes by oral glucose tolerance test (oGTT). Comparison between women and men showed higher BMI, hsCRP and HDL-C values in women and higher TG and SBP levels in men, whereas medication intake (metformin and any other oral antidiabetic medication or insulin treatment) was similar (Table 1). Fasting glucose and HbA1c values over time are presented in Figure S1.

3.2 | Replication of the four ANDIS T2D clusters

3.2.1 | Assignment by using ANDIS scaling and ANDIS centroids

First, clinical variables of the KORA participants were scaled based on ANDIS's scaling parameters, and each participant was assigned to a single cluster based on the Euclidean distance to the ANDIS centroids.[25] The characteristics of four clusters are shown in Table S2 and Figure 2A. The SIDD cluster in KORA was characterized by a relatively younger age, lower insulin secretion (HOMA2-B) and highest HbA1c; the SIRD cluster had the highest level of insulin resistance (HOMA2-IR) and insulin secretion (HOMA2-B); the MOD cluster had a high BMI but younger age and the MARD cluster showed low insulin resistance, low BMI and older age. The relative cluster sizes in KORA were not comparable to those found in the ANDIS study. SIDD made up only 2% of the T2D cases in KORA compared to 17.5% in ANDIS. More than 80% of participants in KORA were assigned to the MARD cluster, compared to only around 40% in ANDIS.

3.2.2 | Assignment by using KORA scaling and ANDIS centroids

Second, the clinical variables of KORA participants were scaled based on own scaling parameters derived from the KORA sample and then assigned to a single cluster based on the Euclidean distance to the ANDIS centroids.[21] The characteristics of four clusters are shown in Table S2 and Figure S2A. The SIDD cluster in KORA was characterized by a relatively younger age, lower insulin secretion (HOMA2-B) and poorer glycaemic control (higher HbA1c); the SIRD cluster had the highest level of insulin resistance (HOMA2-IR) and insulin secretion (HOMA2-B); the MOD cluster had a high BMI and individuals were younger and the MARD cluster had low insulin resistance and low BMI, but an older age. All these variables followed the same trend in KORA and ANDIS. The relative cluster sizes in KORA were comparable to those found in the ANDIS study, for example, most participants were allocated to MARD for both KORA (46.8%) and ANDIS (39.1%), and 15.3% of individuals in KORA were assigned to SIDD which was similar to the ANDIS study (17.5%).

We then investigated the transfer of individuals when using ANDIS centroids, with either ANDIS scaling or KORA scaling. Sixty-five percent of participants were assigned to the same clusters (Figure S2B). Compared to ANDIS scaling, clusters were more evenly distributed when using KORA scaling. Most strikingly, a substantial part of the MARD cluster when using ANDIS scaling was allocated to the SIDD, SIRD and MOD clusters using KORA scaling.

3.2.3 | Assignment by using KORA scaling and KORA centroids

Third, clusters were derived based on hyperparameters from KORA data alone, using k-means clustering on the same variable set (age, BMI, HbA1c, HOMA2-B and HOMA2-IR) forcing the same number of clusters ($k = 4$) as in the ANDIS cohort. As shown in Figure S3, cluster 1 was characterized by low insulin secretion (low HOMA2-B), high BMI and poor metabolic control (high HbA1c); thus, we labelled cluster 1 as SIDD. Cluster 2 had insulin resistance as evidenced by a high HOMA2-IR which could be compared to SIRD. Cluster 3 featured

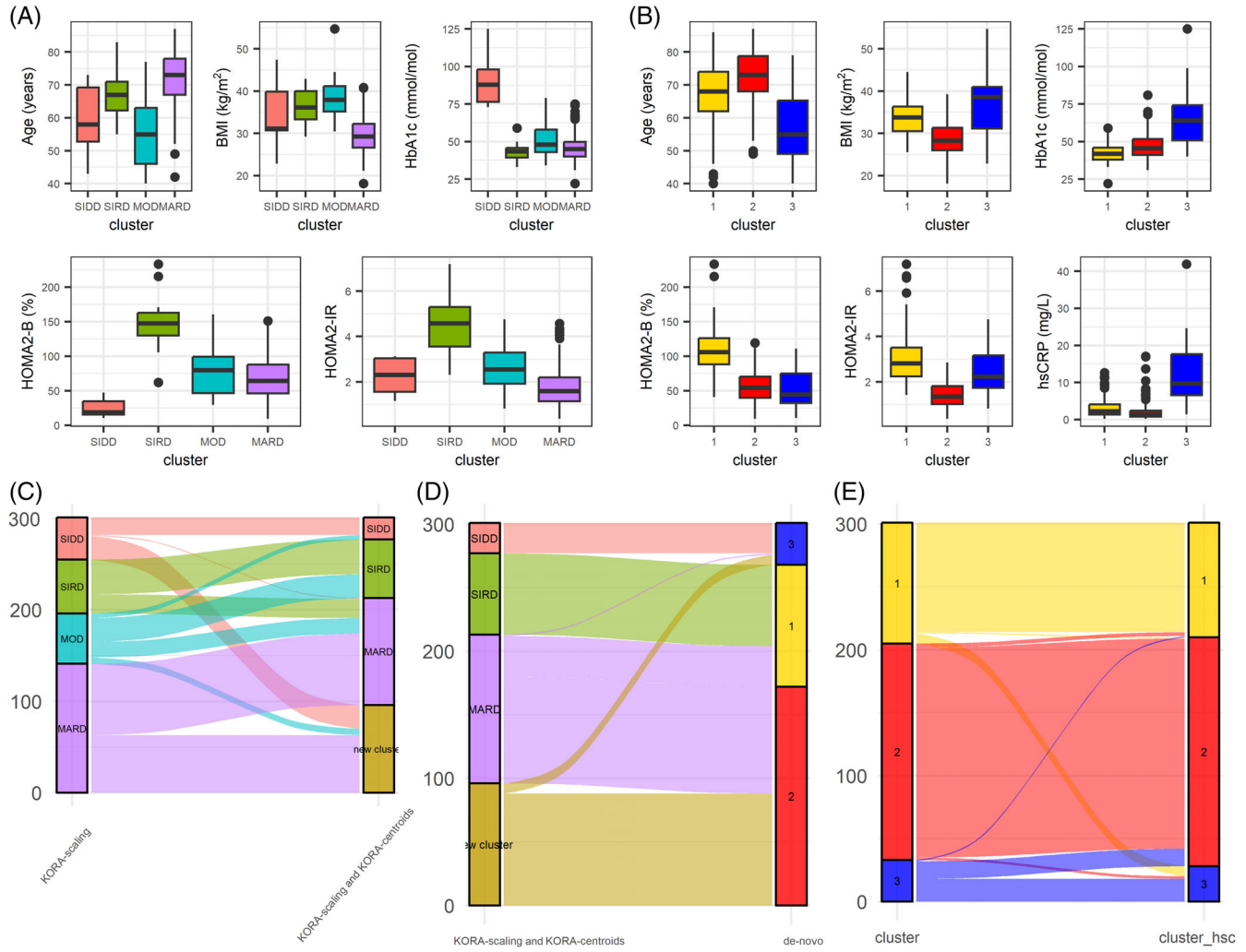


FIGURE 2 Distributions of age at examination, body mass index (BMI), haemoglobin A1c (HbA1c), homeostasis model assessment (HOMA) estimates of beta-cell function (HOMA2-B) and insulin resistance (HOMA2-IR) in the KORA FF4 cohort for each cluster (A) using ANDIS scaling and ANDIS centroids or (B) with additional high sensitivity C-reactive protein (hsCRP) derived from de novo *k*-means with $k = 3$. The upper and lower bounds of boxes represent the first and third quartiles, box centres represent the median values and circles represent outliers. (C) Sankey diagram displaying the transfer of individuals between the clusters identified using KORA scaling and ANDIS centroids (left side) and the clusters identified using KORA scaling and KORA centroids (right side), (D) transition of individuals between the clusters originally replicated using KORA scaling and KORA centroids (left side, corresponding to the right side of Figure 2C) and de novo derivation (right side) and (E) transition of individuals between the de novo clusters identified using basic variables (left side) and with the additional variable hsCRP (right side). MARD, mild age-related diabetes; MOD, mild obesity-related diabetes; SIDD, severe insulin-deficient diabetes; SIRD, severe insulin resistant diabetes.

elderly individuals with relatively mild metabolic irregularities which is similar to MARD in the ANDIS study. Cluster 4 represented a novel distinct subphenotype with the overall most modest metabolic impairments and low BMI and was thereby distinct from the ANDIS cluster MOD.

We also generated the Sankey diagram to visualize and compare the cluster assignment based on the second approach (using KORA scaling and ANDIS centroids) and the third approach (using KORA scaling and KORA centroids). We observed consistent cluster assignments for only 45% of the individuals between the second and third approach (Figure 2C). Taken together, these results suggest that the original ANDIS clusters do not fully reflect the characteristics of the KORA sample.

3.3 | De novo cluster derivation in KORA

3.3.1 | Determination of *k* and cluster derivation

Both silhouette width and the elbow plot methods agreed that $k = 3$ rather than $k = 4$ was the optimal number of clusters for the KORA data (Figure S4). Subsequently, *k*-means was used on the basic variable set to categorize KORA participants into three clusters representing three T2D subphenotypes. Clinical characteristics according to each subphenotype are shown in (Figure S5A). Cluster 1 ($n = 96$, 31.9%) was characterized by hyperinsulinemia and insulin resistance (most similar to the SIRD cluster in the ANDIS cohort); participants in cluster 2 ($n = 172$, 57.1%) had older age, low BMI and low insulin

resistance which could be compared to MARD in the ANDIS cohort; and cluster 3 ($n = 33$, 11.0%) showed insulin deficiency (low HOMA2-B), high BMI and poor glycaemic control (high HbA1c), which is a distinct cluster from those present in the ANDIS cohort. We then compared participant transitions from the original cluster replication using KORA centroids and KORA scaling (third approach as described above) with the de novo derived clusters (Figure 2D). Individuals previously allocated to the MARD subphenotype were reallocated to the new cluster 1 and cluster 2. Individuals previously allocated to the novel distinct subphenotype were mainly reallocated not only to the new cluster 2, but also to the new cluster 3 (distinct). Individuals previously allocated to the SIRD subphenotype were reallocated to the new cluster 3.

3.3.2 | Different variable sets and final clusters

We assessed the stability of cluster assignments when using different sets of variables for clustering: basic variables (age at examination, BMI, HbA1c, HOMA2-B and HOMA2-IR) plus hsCRP, TG, HDL-C or SBP, respectively. In general, the addition of these variables did not substantially influence the distribution of the basic variables between clusters and did not lead to substantial transition of participants between clusters (Figure 2B,D, Figures S5, S6 and S7). In detail, 90%, 93%, 90% and 98% of participants were allocated to the same cluster when using basic variables compared to when adding hsCRP, TG, HDL-C or SBP, respectively. To account for the role of systemic inflammation in diabetes differentiation, we defined the clusters derived from the variable set of age, BMI, HbA1c, HOMA2-B and HOMA2-IR plus hsCRP as the final subphenotypes, presented in Figure 2B, Tables S3 and S4. Cluster 1 included 91 participants (30.2%) and was characterized by insulin resistance (high HOMA2-IR) and hyperinsulinemia, with a high proportion of newly diagnosed diabetes cases (most similar to the SIRD cluster in ANDIS). Cluster 2 included 182 individuals (60.5%) and was characterized by high age, low BMI and low insulin resistance (most similar to the MARD cluster in ANDIS). Cluster 3 included 28 participants (9.3%) and was characterized by a high BMI, poor glycaemic control, high level of subclinical inflammation (high hsCRP) and relative insulin deficiency, broadly resembling a typical patient seen in clinical practice (most similar to SIDD/MOD cluster in ANDIS).

The assessment of cluster stability showed that Jaccard indices of all final clusters were above 0.75, indicating reasonably high cluster stability for the final variable set (Table S5). Of note, with additional variables TG, HDL-C or SBP, stability slightly decreased for all clusters and cluster 3 even showed Jaccard indices below 0.75 (Table S5). Besides, the majority of individuals (95%) were assigned to the same cluster as in the initial data analysis, and both men and women showed the same trend on the clinical variable distribution (Figure S8), suggesting a lack of substantial sex-specific effects.

3.4 | Cluster differences in genetic risk, diabetes-related complications and parental history

3.4.1 | Polygenic risk score

The overall distribution of the PRS in the KORA FF4 sample is given in Figure 3A. Participants with T2D had significantly higher PRS values ($p < 0.001$) compared to those without T2D and were

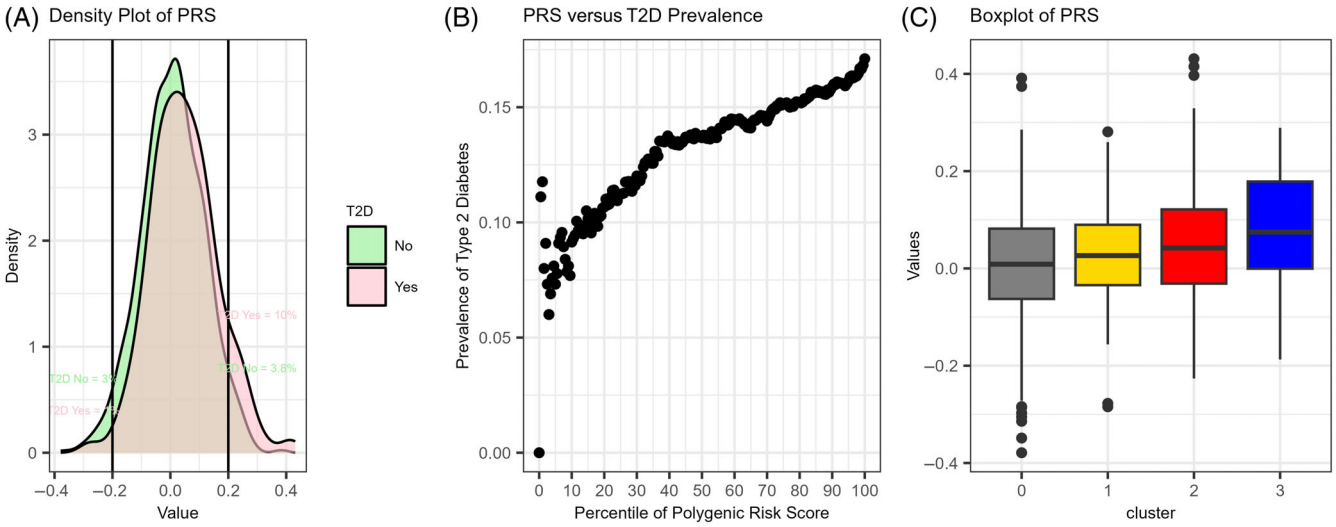


FIGURE 3 (A) Density plot shows the polygenic risk score (PRS) distribution in the KORA FF4 sample without (light green) and with (light red) type 2 diabetes (T2D). Data beyond the two vertical lines indicate extreme values of the PRS distribution, and the corresponding numbers reflect the proportion of individuals without (light green) and with (light red) T2D who showed extreme PRS values. (B) Percentile of increasing PRS (*x*-axis) versus the prevalence of T2D (*y*-axis). (C) Distributions of PRS in control group (marked as 0) and three clusters representing T2D subphenotypes.

overrepresented in the highest quantiles of the distribution (Figure 3A,B). When comparing the distribution of PRS in the respective clusters to individuals without diabetes (Figure 3C), the PRS in cluster 2 and cluster 3 was significantly different to the control group (both $p < 0.001$, respectively) but the PRS in cluster 1 was not different to the control group. An additional t test confirmed that cluster 3 had a significantly higher PRS than cluster 1 ($p = 0.034$), whereas there was no significant difference between cluster 1 and cluster 2.

3.4.2 | Diabetes-related complications and parental history of diabetes

We evaluated the prevalence of diabetes-related complications and parental history of T2D in the three clusters. As shown in Figure S9 and Table S7, in general, individuals in clusters 1 and 2 had a substantially higher prevalence of myocardial infarction, stroke and chronic kidney disease (CKD). Individuals in cluster 3 had a more frequently positive parental history of diabetes.

Moreover, compared to cluster 1, cluster 2 had a lower frequency of neuropathy ($p = 0.043$) but a higher prevalence of stroke (not significant) and CKD ($p = 0.030$).

4 | DISCUSSION

The T2D subphenotype classification scheme proposed by Ahlqvist et al.[9] has been replicated in different populations and has proven to be a useful tool to further characterize potential pathophysiological pathways and diabetes progression. Our study aimed at a comprehensive assessment of original cluster replication, including a systematic illustration of participant transitions between replicated clusters, de novo cluster derivation, including the assessment of cluster stability, and underlying genetic risk and complication distribution. We found that the original clusters only partially reflected the characteristics of individuals with T2D in the KORA sample, whereas de novo derived clusters showed excellent stability and captured the underlying heterogeneity between the T2D subphenotypes. Our results therefore underscore the importance of subphenotyping by illustrating the importance of individual study characteristics, and we contribute another potential T2D subphenotype to the existing panel.

Our results align with recent findings, which indicated that 11 of 18 studies either delineated distinct subphenotypes or failed to identify all ANDIS subphenotypes.[26] Part of the lack of replicability of the original clusters may be attributed to differences in the study setup and participants' characteristics. For example, we used age at examination for clustering, since age of diabetes onset for most T2D participants was not available. Therefore, the average age used in the KORA sample was significantly higher compared to the ANDIS cohort (Table S1), especially in the de novo cluster 2 (comparable to MARD). Moreover, individuals in KORA had better glycaemic control and less insulin resistance compared to the ANDIS sample (Table S1), indicating that KORA potentially included a larger proportion of T2D cases with less severe disease. Furthermore, our HOMA models were based on insulin instead of C-peptide, which might have led to differences in estimates. Some studies[27,28] suggested that C-peptide better reflected insulin secretion, while another study[29] suggested that both of them performed similar in evaluating beta cell function.

Employing different scaling parameters generated a big difference in cluster allocation, and different studies applied different approaches.[21,25] The incongruence of cluster assignment, together with the identification of a novel, distinct subphenotype not present in ANDIS when using KORA centroids, shows that the original clusters do not capture the characteristics of the KORA sample as well. We consider this finding important for personalized prevention. While the ANDIS cohort captured crucial subphentoypes, these clusters might take different shapes or not fully reflect the underlying sample in other cohorts with different characteristics. Contributions from multiple studies are therefore needed to expand and refine the current panel of T2D subphenotypes.

Determination of the optimal number of clusters k based on silhouette and elbow plot showed that in the KORA sample, $k = 3$ was the best number of clusters, which is consistent with the Danish DD2 study.[25] A head-to-head comparison between the clusters from KORA and DD2 revealed major similarities (Table S8). Consistent with the research from Safai et al.,[30] which did not identify an evident MOD-like cluster in their de novo cluster analysis (when using $k = 5$, including SAID), the clusters with the highest BMI also exhibited significantly higher insulin resistance. Besides the clinical characteristics used for clustering, multiple other factors are associated with T2D. We thus assessed cluster stability across different variable sets, additionally including hsCRP, HDL-C, TG or SBP, respectively. We found that these additional variables did not contribute much to the reallocation of individuals, as more than 90% individuals were still assigned to the same cluster, indicating high cluster stability and robustness towards different variable sets. One could thus hypothesize that the original variables already capture a major part of T2D heterogeneity and are adequate to identify clinically meaningful T2D subphenotypes. Other studies[18,30,31] also applied analytical approaches for a wider range of clusters or included different variables than the ANDIS study but did not systematically evaluate how participants were reallocated when using different clustering variables.

CRP is regulated by proinflammatory cytokines derived from adipose tissue.[32,33] In individuals with T2D, CRP levels are chronically elevated.[34] In the current analysis, we included hsCRP for clustering to account for the role of subclinical inflammation and assess potential differences according to subphenotypes. The de novo derived cluster 3 could not be mapped to one of the original ANDIS clusters and was characterized by high BMI, high hsCRP and relatively low HOMA2-B. Increased CRP levels have been linked to excess body weight since adipose tissue produces tumour necrosis factor α (TNF-α) and interleukin-6 (IL-6), which are pivotal factors for CRP stimulation.[32,33] We could thus hypothesize that cluster 3 represents a T2D subphenotype with chronic, obesity-induced subclinical inflammation. The PRS and the prevalence of self-reported parental history of diabetes were both the highest in cluster 3. So, cluster 3 could represent a T2D

subphenotype with higher genetically induced risk for both diabetes and obesity, resulting in chronic subclinical inflammation (Table S6 and Figure S10). We note that the use of a PRS to define subgroups of diabetes is still questionable and would render the algorithm less readily applicable in clinical practice and other studies, which is why we only use it descriptively. Since non-genetic risk factors might have even stronger unfavourable impacts in individuals with genetic predisposition, the group in cluster 3 would particularly benefit from rigorous weight control, either through lifestyle modifications or drug treatment. Moreover, these individuals should be monitored for potential other causes of inflammation, such as infections or wounds.

The analysis of diabetes complications showed that in cluster 2, there was a higher proportion of CKD cases and a relatively higher percentage of stroke (not significant) compared to cluster 3. This could be due to the higher average age in cluster 2, since it is well-established that age is a major risk factor for metabolic complications in T2D.[35,36] Because risk in cluster 2 is mainly conferred by aging processes, and age is a non-modifiable factor, for this cluster in particular, close monitoring of comorbidities and strict, potentially medication-based, control of, for example, blood pressure and renal function is advisable. Cluster 1 was characterized by hyperinsulinemia and a comparatively higher prevalence of neuropathy compared to cluster 2. Insulin dysregulation can contribute to neuropathic changes in sensory neurons, and the peripheral nervous system is one of several organ systems that are profoundly affected in diabetes.[37] Interestingly, HbA1c levels in cluster 1 were comparatively low, so it would be crucial to investigate the use of glucose-lowering therapy in this cluster to evaluate their role in the prevention of neuropathy in this subphenotype. Medication therapy in this cluster was comparatively low, likely due to the high proportion of newly diagnosed diabetes cases, so this would be an obvious target to tackle insulin resistance in these individuals. Moreover, lifestyle interventions would be beneficial, including dietary changes by reducing calorie intake and limiting high glycaemic index carbohydrates and regular physical activity which enhances calorie burning and increases insulin sensitivity in muscle tissue.[38,39] Evidence indicates that an increased level of hsCRP is linked with diabetes-related complications,[40,41] but cluster 3 with the highest hsCRP levels was not characterized by a high load of complications. This may be due to the younger age of individuals in cluster 3 (Figure 2B), since given the potential pathway discussed above about a genetic predisposition to obesity-induced inflammation, it would be possible that diabetes complications in cluster 3 have not yet developed.

We acknowledge the limitations of our current study. The sample size was relatively small compared to other population-based studies, and although unsupervised clustering does not have strict sample size requirements, the small number of individuals with diabetes-related complications and family history information impedes the interpretation of shared disease characteristics. While the clusters represent a true underlying structure in the data from a statistical perspective, this structure could also have emerged due to other shared characteristics of the respective individuals, for example, environmental factors, and do not necessarily represent shared pathophysiology. Moreover, our results regarding diabetes complications need to be interpreted with caution, since complications were self-reported, and the sample size was small. We were unable to model medication effects, since medication could not be included as a variable in the clustering procedure, and participants' individual medication regimes could not be disentangled. Moreover, our participants were exclusively of white European ethnicity, which limits the generalizability to other populations.

In conclusion, to exploit the full advantages of T2D subphenotyping, a potential mismatch between reported T2D clusters and the individual study characteristics has to be taken into account. Since adapting the clustering algorithm might not always be possible, further efforts should be undertaken to identify further subtypes from different well-characterized studies, in order to expand and refine the current panel of T2D subphenotypes.

AUTHOR CONTRIBUTIONS

Conceptualization: Qiuling Dong, Yue Xi, Harald Grallert and Susanne Rospleszcz. *Methodology*: Qiuling Dong and Yue Xi. *Software*: Qiuling Dong and Yue Xi. *Validation*: Qiuling Dong and Yue Xi. *Formal analysis*: Qiuling Dong. *Investigation*: Qiuling Dong and Yue Xi. *Resources*: Susanne Rospleszcz, Stefan Brandmaier, Barbara Thorand, Marie-Theres Huemer, Melanie Waldenberger, Christian Herder, Wolfgang Rathmann, Wolfgang Koenig, Gidon J. Bönhof, Christian Gieger, Annette Peters and Harald Grallert. *Data curation*: Christian Herder, Wolfgang Rathmann, Wolfgang Koenig, Gidon J. Bönhof and Christian Gieger. *Writing—original draft preparation*: Qiuling Dong. *Writing—review and editing*: Qiuling Dong, Yue Xi, Susanne Rospleszcz, Stefan Brandmaier, Barbara Thorand, Marie-Theres Huemer, Melanie Waldenberger, Jiefei Niu, Christian Herder, Wolfgang Rathmann, Wolfgang Koenig, Gidon J. Bönhof, Christian Gieger, Annette Peters and Harald Grallert. *Visualization*: Qiuling Dong. *Supervision*: Annette Peters, Susanne Rospleszcz and Harald Grallert. *Project administration*: Harald Grallert and Annette Peters. *Funding acquisition*: Harald Grallert and Annette Peters. All authors have read and agreed to the published version of the manuscript.

ACKNOWLEDGEMENTS

We express our appreciation to all KORA study participants for donating their blood and time. We thank the field staff in Augsburg conducting the KORA studies. We would like to thank the China Scholarship Council (CSC) for the financial support (No. 202008310176).

FUNDING INFORMATION

The KORA study was initiated and financed by Helmholtz Munich—German Research Center for Environmental Health, which is financed by the German Federal Ministry of Education and Research and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC-Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ. The German Diabetes Center is funded by the German Federal Ministry of Health (Berlin, Germany) and the Ministry of Culture and Science of the State North Rhine-Westphalia (Düsseldorf, Germany) and receives additional funding from the German Federal Ministry of Education and

Research (BMBF) through the German Center for Diabetes Research (DZD e.V.).

CONFLICT OF INTEREST STATEMENT

M.R. reports receipt of consulting fees by AstraZeneca, Boehringer Ingelheim, Echosens, Eli Lilly, Madrigal, NovoNordisk and institutional research grants from Boehringer Ingelheim, Novartis Pharma, NovoNordisk and Nutriticia/Danone outside of the topic of this publication. W.R. reports the receipt of consulting fees for attending educational sessions or advisory boards run by AstraZeneca, Boehringer Ingelheim and NovoNordisk and institutional research grants from NovoNordisk outside of the topic of this publication. The other authors declare no conflict of interest.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/dom.16022.

DATA AVAILABILITY STATEMENT

The KORA FF4 datasets are not publicly available but can be accessed upon application through the KORA-PASST (Project application self-service tool, https://www.helmholtz-munich.de/epi/research/cohorts/kora-cohort/data-use-and-access-via-korapasst/index.html).

INFORMED CONSENT STATEMENT

Written informed consent has been obtained from the study participants.

ORCID

Qiuling Dong https://orcid.org/0000-0002-3369-4120

REFERENCES

1. Collaboration NCDRF. Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. *Lancet*. 2016;387(10027):1513-1530. doi:10.1016/S0140-6736(16)00618-8
2. World Health Organization. April 5 2023. Diabetes https://www.who.int/news-room/fact-sheets/detail/diabetes. accessed at 2024.01.10
3. Davidson MB. Diagnosing diabetes with glucose criteria: worshiping a false god. *Diabetes Care*. 2011;34(2):524-526. doi:10.2337/dc10-1689
4. American Diabetes Association Professional Practice C. 2. Classification and diagnosis of diabetes: standards of medical care in diabetes-2022. *Diabetes Care*. 2022;45(suppl 1):S17-S38. doi:10.2337/dc22-S002
5. Schwartz SS, Epstein S, Corkey BE, Grant SF, Gavin JR 3rd, Aguilar RB. The time is right for a new classification system for diabetes: rationale and implications of the beta-cell-centric classification schema. *Diabetes Care*. 2016;39(2):179-186. doi:10.2337/dc15-1585
6. Bancks MP, Casanova R, Gregg EW, Bertoni AG. Epidemiology of diabetes phenotypes and prevalent cardiovascular risk factors and diabetes complications in the National Health and Nutrition Examination Survey 2003–2014. *Diabetes Res Clin Pract*. 2019;158:107915. doi:10.1016/j.diabres.2019.107915
7. Thorens B, Rodriguez A, Cruciani-Guglielmacci C, Wigger L, Ibberson M, Magnan C. Use of preclinical models to identify markers of type 2 diabetes susceptibility and novel regulators of insulin secretion – a step towards precision medicine. *Mol Metab*. 2019;27S (Suppl):S147-S154. doi:10.1016/j.molmet.2019.06.008
8. Herder C, Roden M. A novel diabetes typology: towards precision diabetology from pathogenesis to treatment. *Diabetologia*. 2022;65(11):1770-1781. doi:10.1007/s00125-021-05625-x
9. Ahlqvist E, Storm P, Karajamaki A, et al. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol*. 2018;6(5):361-369. doi:10.1016/S2213-8587(18)30051-2
10. Dennis JM, Shields BM, Henley WE, Jones AG, Hattersley AT. Disease progression and treatment response in data-driven subgroups of type 2 diabetes compared with models based on simple clinical features: an analysis using clinical trial data. *Lancet Diabetes Endocrinol*. 2019;7(6):442-451. doi:10.1016/S2213-8587(19)30087-7
11. Zaharia OP, Strassburger K, Strom A, et al. Risk of diabetes-associated diseases in subgroups of patients with recent-onset diabetes: a 5-year follow-up study. *Lancet Diabetes Endocrinol*. 2019;7(9):684-694. doi:10.1016/S2213-8587(19)30187-1
12. Herder C, Maalmi H, Strassburger K, et al. Differences in biomarkers of inflammation between novel subgroups of recent-onset diabetes. *Diabetes*. 2021;70(5):1198-1208. doi:10.2337/db20-1054
13. Zou X, Zhou X, Zhu Z, Ji L. Novel subgroups of patients with adult-onset diabetes in Chinese and US populations. *Lancet Diabetes Endocrinol*. 2019;7(1):9-11. doi:10.1016/S2213-8587(18)30316-4
14. Li X, Yang S, Cao C, et al. Validation of the Swedish diabetes re-grouping scheme in adult-onset diabetes in China. *J Clin Endocrinol Metab*. 2020;105(10):e3519-e3528. doi:10.1210/clinem/dgaa524
15. Bello-Chavolla OY, Bahena-Lopez JP, Vargas-Vazquez A, et al. Clinical characterization of data-driven diabetes subgroups in Mexicans using a reproducible machine learning approach. *BMJ Open Diabetes Res Care*. 2020;8(1):e001550. doi:10.1136/bmjdrc-2020-001550
16. Gudmundsdottir V, Zaghlool SB, Emilsson V, et al. Circulating protein signatures and causal candidates for type 2 diabetes. *Diabetes*. 2020;69(8):1843-1853. doi:10.2337/db19-1070
17. Tanabe H, Saito H, Kudo A, et al. Factors associated with risk of diabetic complications in novel cluster-based diabetes subgroups: a Japanese retrospective cohort study. *J Clin Med*. 2020;9(7):2083. doi:10.3390/jcm9072083
18. Anjana RM, Baskar V, Nair ATN, et al. Novel subgroups of type 2 diabetes and their association with microvascular outcomes in an Asian Indian population: a data-driven cluster analysis: the INSPIRED study. *BMJ Open Diabetes Res Care*. 2020;8(1):e001506.
19. Mansour Aly D, Dwivedi OP, Prasad RB, et al. Genome-wide association analyses highlight etiological differences underlying newly defined subtypes of diabetes. *Nat Genet*. 2021;53(11):1534-1542. doi:10.1038/s41588-021-00948-2
20. Zaharia OP, Strassburger K, Knebel B, et al. Role of patatin-like phospholipase domain-containing 3 gene for hepatic lipid content and insulin resistance in diabetes. *Diabetes Care*. 2020;43(9):2161-2168. doi:10.2337/dc20-0329
21. Zaghlool SB, Halama A, Stephan N, et al. Metabolic and proteomic signatures of type 2 diabetes subtypes in an Arab population. *Nat Commun*. 2022;13(1):7121. doi:10.1038/s41467-022-34754-z
22. Wesolowska-Andersen A, Brorsson CA, Bizzotto R, et al. Four groups of type 2 diabetes contribute to the etiological and clinical heterogeneity in newly diagnosed individuals: an IMI DIRECT study. *Cell Rep Med*. 2022;3(1):100477. doi:10.1016/j.xcrm.2021.100477
23. Stanimirovic J, Radovanovic J, Banjac K, et al. Role of C-reactive protein in diabetic inflammation. *Mediators Inflamm*. 2022;2022:3706508. doi:10.1155/2022/3706508

24. Holle R, Happich M, Lowel H, Wichmann HE; Group MKS. KORA – a research platform for population based health research. *Gesundheitswesen*. 2005;67(suppl 1):S19-S25. doi:10.1055/s-2005-858235
25. Christensen DH, Nicolaisen SK, Ahlqvist E, et al. Type 2 diabetes classification: a data-driven cluster study of the Danish Centre for Strategic Research in Type 2 Diabetes (DD2) cohort. *BMJ Open Diabetes Res Care*. 2022;10(2):e002731. doi:10.1136/bmjdrc-2021-002731
26. Varghese JS, Carrillo-Larco RM, Narayan KV. Achieving replicable subphenotypes of adult-onset diabetes. *Lancet Diabetes Endocrinol*. 2023;11(9):635-636. doi:10.1016/S2213-8587(23)00195-X
27. Song YS, Hwang YC, Ahn HY, Park CY. Comparison of the usefulness of the updated homeostasis model assessment (HOMA2) with the original HOMA1 in the prediction of type 2 diabetes mellitus in Koreans. *Diabetes Metab J*. 2016;40(4):318-325. doi:10.4093/dmj.2016.40.4.318
28. Caumo A, Perseghin G, Brunani A, Luzi L. New insights on the simultaneous assessment of insulin sensitivity and beta-cell function with the HOMA2 method. *Diabetes Care*. 2006;29(12):2733-2734. doi:10.2337/dc06-0070
29. Li X, Zhou ZG, Qi HY, Chen XY, Huang G. Replacement of insulin by fasting C-peptide in modified homeostasis model assessment to evaluate insulin resistance and islet beta cell function. *Zhong Nan Da Xue Xue Bao Yi Xue Ban*. 2004;29(4):419-423.
30. Safai N, Ali A, Rossing P, Ridderstrale M. Stratification of type 2 diabetes based on routine clinical markers. *Diabetes Res Clin Pract*. 2018;141:275-283. doi:10.1016/j.diabres.2018.05.014
31. Slieker RC, Donnelly LA, Fitipaldi H, et al. Replication and cross-validation of type 2 diabetes subtypes based on clinical variables: an IMI-RHAPSODY study. *Diabetologia*. 2021;64(9):1982-1989. doi:10.1007/s00125-021-05490-8
32. Freeman DJ, Norrie J, Caslake MJ, et al. C-reactive protein is an independent predictor of risk for the development of diabetes in the west of Scotland coronary prevention study. *Diabetes*. 2002;51(5):1596-1600. doi:10.2337/diabetes.51.5.1596
33. Sproston NR, Ashworth JJ. Role of C-reactive protein at sites of inflammation and infection. *Front Immunol*. 2018;9:754. doi:10.3389/fimmu.2018.00754
34. Kanmani S, Kwon M, Shin MK, Kim MK. Association of C-reactive protein with risk of developing type 2 diabetes mellitus, and role of obesity and hypertension: a large population-based Korean cohort study. *Sci Rep*. 2019;9(1):4573. doi:10.1038/s41598-019-40987-8
35. Bhatti GK, Bhadada SK, Vijayvergiya R, Mastana SS, Bhatti JS. Metabolic syndrome and risk of major coronary events among the urban diabetic patients: North Indian Diabetes and Cardiovascular Disease Study—NIDCVD-2. *J Diabetes Complications*. 2016;30(1):72-78.
36. Shamshirgaran SM, Mamaghanian A, Aliasgarzadeh A, Aiminisani N, Iranparvar-Alamdari M, Ataie J. Age differences in diabetes-related complications and glycemic control. *BMC Endocr Disord*. 2017;17(1):25. doi:10.1186/s12902-017-0175-5
37. Grote CW, Wright DE. A role for insulin in diabetic neuropathy. *Front Neurosci*. 2016;10:581. doi:10.3389/fnins.2016.00581
38. Racz O, Linkova M, Jakubowski K, Link R, Kuzmova D. Az inzulinkezeles elkezdesenek gyakorlati akadalyai 2-es tipusu cukorbetegekben – a "pszichologiai inzulinrezisztencia" lekuzdese (Barriers of the initiation of insulin treatment in type 2 diabetic patients – conquering the "psychological insulin resistance"). *Orv Hetil*. 2019;160(3):93-97. doi:10.1556/650.2019.31269
39. Yaribeygi H, Atkin SL, Simental-Mendia LE, Sahebkar A. Molecular mechanisms by which aerobic exercise induces insulin sensitivity. *J Cell Physiol*. 2019;234(8):12385-12392. doi:10.1002/jcp.28066
40. Mugabo Y, Li L, Renier G. The connection between C-reactive protein (CRP) and diabetic vasculopathy. Focus on preclinical findings. *Curr Diabetes Rev*. 2010;6(1):27-34. doi:10.2174/157339910790442628
41. Esser N, Paquot N, Scheen AJ. Inflammatory markers and cardiometabolic diseases. *Acta Clin Belg*. 2015;70(3):193-199. doi:10.1179/2295333715Y.0000000004

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Dong Q, Xi Y, Brandmaier S, et al. Subphenotypes of adult-onset diabetes: Data-driven clustering in the population-based KORA cohort. *Diabetes Obes Metab*. 2025;27(1):338-347. doi:10.1111/dom.16022

Acknowledgements

Firstly, I would like to express my gratitude to my direct supervisor, Prof. Dr. Annette Peters, Chair of Epidemiology of Faculty of Medicine from LMU Munich and Director of the Institute of Epidemiology at Helmholtz Zentrum München. Throughout the course of my studies, Prof. Dr. Peters offered invaluable guidance and thoughtful advice.

Next, I am deeply grateful to my second supervisor Dr. Harald Grallert, research group leader of Diabetes and related traits. Epidemiology was a totally new field to me, Dr. Grallert encouraged me and offered a valuable platform to learn. His productive discussions, support, and guidance were instrumental throughout my research.

My colleagues including Dr. Susanne Rospleszcz and Yue Xi, and my researcher team member including Dr. Stefan Brandmaier, Jonathan Adam and Jiefei Niu gave me invaluable assistance for statistical analysis and insightful advice during my research. I also would like to express my sincere gratitude to Dr. Esienanwan Esien Efiong for her valuable support in rephrasing and refining my thesis.

I am sincerely grateful to the Chinese Scholarship Council for their financial support throughout the duration of my research, and Helmholtz Graduate School Environmental Health (HELENA) for the scientific training and travel grant.

My parents and my husband have provided unwavering emotional support throughout my educational journey. Additionally, the birth of my son, Felix, in June 2024 has further inspired me to continue progressing.

Learning is a lifelong journey, and my PhD studies in Germany have been the most unforgettable part of it. I would like to extend my heartfelt thanks to everyone who has contributed to my PhD projects.

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Dekanat Medizinische Fakultät
Promotionsbüro

Confirmation of congruency between printed and electronic version of the doctoral thesis

Dong, Qiuling

name, first name

I hereby declare that the electronic version of the submitted thesis, entitled:

Prediction and progression of Type 2 Diabetes using molecular characterized signatures

is congruent with the printed version both in content and format.

Munich, 18.11.2024
Place, Date

Qiuling Dong
Signature doctoral candidate