Integrative Metabolomics of Targeted and Non-targeted Analyses in T2D Progression

Running title: Metabolic Shifts in Diabetes Progression

Jianhong Ge[1,2,3], Siyu Han[2], Mengya Shi[1,2,3], Makoto Harada[2], Shixiang Yu[1,2,3], Jiaqi Zheng[2,3], Cornelia Prehn[4], Jerzy Adamski[5,6,7], Gabi Kastenmüller[8], Sabrina Schlesinger[9], Wolfgang Koenig[10], Birgit Linkohr[11,12], Barbara Thorand[3,11,13] , Karsten Suhre[14,15], Christian Gieger[11,16], Annette Peters[3,11,12,13], Rui Wang-Sattler[2,3,13,*]

[1] TUM School of Medicine and Health, Technical University of Munich, Munich, Germany.
[2] Institute of Translational Genomics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany.
[3] German Center for Diabetes Research (DZD), Partner Neuherberg, Neuherberg, Germany.
[4] Metabolomics and Proteomics Core, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany.
[5] Institute of Experimental Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany
[6] Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
[7] Institute of Biochemistry, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia
[8] Institute of Computational Biology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany
[9] Institute for Biometrics and Epidemiology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University, Düsseldorf, Germany.
[10] Deutsches Herzzentrum München, Technische Universität München, München, Germany
[11] Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany.
[12] Munich Heart Alliance, German Center for Cardiovascular Health (DZHK E.V., Partner-Site Munich), Munich, Germany
[13] Institute for Medical Information Processing, Biometry, and Epidemiology (IBE), Faculty of Medicine, Ludwig-Maximilians-Universität München, Pettenkofer School of Public Health, Munich, Germany
[14] Bioinformatics Core, Weill Cornell Medicine-Qatar, Education City, Doha, Qatar
[15] Englander Institute for Precision Medicine, Weill Cornell Medicine, New York, NY USA
[16] Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany

* Corresponding author: Rui Wang-Sattler, Email: rui.wang-sattler@helmholtz-munich.de, Phone number: +49 89 3187 3978

A Twitter Summary*:*

Stage-specific metabolic reprogramming over 14 years in the KORA cohort reveals early energy imbalance and later amino acid perturbations, offering insights into diabetes progression and prevention.

Abstract

Objective

This study aimed to identify metabolites characterizing the progression from normal glucose metabolism (NORM) to prediabetes (PreT2D) and type 2 diabetes (T2D), focusing on stage-specific metabolic shifts (early: NORM to PreT2D; late: PreT2D to T2D) and mechanistic relevance.

Research Design and Methods

We analyzed 8,240 observations from the KORA cohort, profiling 104 targeted and 312 non-targeted metabolites across three time points: baseline (S4) and follow-ups (F4 and FF4) spanning 14 years. Trajectory analyses of 1,050 individuals identified 211 incident PreT2D and 112 incident T2D cases. Linear mixed-effects models (basic: adjusted for age, sex, BMI, lifestyle; sensitivity: additionally adjusted for glycemic factors like fasting glucose, and cardiovascular factors such as systolic blood pressure (BP) were used to evaluate metabolic differences across glycemic states. Mediation and Mendelian randomization (MR) analyses examined mechanistic and causal relationships.

Results

We identified 140 Bonferroni-significant metabolites (45 targeted, 109 non-targeted, 14 overlapping), including 68 early-stage metabolites (significant in PreT2D/T2D vs. NORM), primarily energy metabolism markers such as fatty acid oxidation metabolites (e.g., 37 lipids) and TCA cycle metabolites (e.g., citrate). Twenty late-stage metabolites (significant in T2D vs. PreT2D/NORM) included amino acids like BCAAs and γ-glutamyl derivatives. Fewer significant associations were observed in incident cases. Sensitivity models validated 50% of early-stage but not late-stage metabolites. Fasting glucose mediated 35.1% of the γ-glutamyl-valine-T2D association, while MR analysis found no causal roles for C2, BCAAs, or γ-glutamyl-valine.

Conclusions

Energy metabolism shifts occur early, while amino acid alterations emerge later stages. These stage-specific signatures may guide diabetes prevention strategies.

Article highlights

• *Why did we undertake this study?*

To identify stage-specific metabolic shifts in diabetes progression using targeted and non-targeted metabolomics analyses of a cohort followed over 14 years.

• *What is the specific question(s) we wanted to answer?*

Can we characterize diabetes progression and its underlying metabolic mechanisms?

• *What did we find?*

Early-stage: Fatty acid oxidation (lipids) and TCA cycle metabolites (e.g., citrate).

Late-stage: Amino acids (e.g., BCAAs, γ-glutamyl derivatives).

Fewer significant associations were found in incident prediabetes and diabetes cases.

MR analysis showed no causal roles for C2, BCAAs, or γ-glutamyl-valine.

• *What are the implications of our findings?*

Energy metabolism shifts occur early, while amino acid alterations arise later. These stage-specific signatures may inform diabetes prevention strategies.

Introduction

Type 2 diabetes (T2D) is characterized by elevated blood glucose levels resulting from insulin resistance, pancreatic β-cell dysfunction, and widespread metabolic disruptions (1). Prediabetes (PreT2D), an intermediate state of impaired glucose regulation, offers a key opportunity for prevention (2). However, without timely intervention, over 70% of individuals with PreT2D progress to T2D (3).

Metabolomics enables a comprehensive approach to study metabolites involved in metabolic dysregulation. Prior studies using targeted metabolomics have linked branched-chain amino acids (BCAAs), aromatic amino acids (AAAs), and lipid intermediates to insulin resistance and mitochondrial dysfunction, key features of T2D (4, 5). Elevated BCAAs and acylcarnitines have been associated with β-oxidation defects, oxidative stress, and altered glucose metabolism, suggesting early metabolic signatures of disease (6). In addition to targeted approaches, non-targeted metabolomics offers the detection of a broad spectrum of metabolites (7). For instance, bile acids and phospholipids have been linked to incident T2D (8). However, few studies have systematically explored how these metabolites evolve during the progression from normal glucose metabolism (NORM) to PreT2D and T2D.

Especially for PreT2D, it can be further classified into isolated impaired fasting glucose (i-IFG) and impaired glucose tolerance (IGT). They differ fundamentally in their underlying physiological mechanisms, i-IFG is primarily characterized by elevated fasting glucose levels and hepatic insulin resistance, while IGT is associated with postprandial hyperglycemia due to peripheral insulin resistance and impaired muscle glucose uptake (9). Previous studies have reported distinct cardiometabolic risk profiles and progression rates to T2D between these subtypes (2, 10), underscoring the importance of understanding their metabolic differences during the prediabetic stage. In addition, recent cluster-based analyses in older adults without T2D have revealed further heterogeneity in cardiometabolic risk and subclinical inflammation within the prediabetic stage (11).

Emerging evidence suggests that oxidative stress, disruptions in amino acid and glutathione metabolism (12, 13), and impairments in the TCA cycle and redox pathways contribute to metabolic shifts during the progression from prediabetes to T2D, i.e. in the late-stage (14). However, these mechanisms remain insufficiently characterized across different glycemic stages.

To address these gaps, we leveraged data from the population-based Cooperative Health Research in the Region of Augsburg (KORA) cohort, encompassing three examination phases: baseline survey number 4 (S4), 1st follow-up (F4), and 2nd follow-up (FF4) (15). This longitudinal dataset allowed us to track metabolic shifts over time. By integrating targeted and non-targeted metabolomics data, we aimed to identify stage-specific metabolic alterations associated with T2D progression and subgroups within PreT2D. Linear mixed-effects models (LMEMs), adjusted for physiological, lifestyle, glycemic, and cardiovascular risk factors, were applied to evaluate these metabolic differences. Additionally, mediation analysis and two-sample Mendelian randomization (MR) were conducted to investigate the causal relevance and underlying mechanisms of the observed metabolites.

Research design and methods

Study design and participants

This study utilized data from the KORA cohort, which was approved by the Ethics Committee of the Bavarian Chamber of Physicians, Germany, with informed consent obtained from all participants. Data were collected from three studies: baseline (S4, 1999–2001; N = 4,261), its 1st follow-up (F4, 2006–2008; N = 3,080), and 2nd follow-up (FF4, 2013–2014; N = 2,279) (16).

Glycemic states were classified according to the Standards of Care in Diabetes – 2025 (17), with the WHO criteria applied for individuals with IFG. Specifically, glycemic status was defined at each time point using physician-validated self-reports, fasting glucose levels, and 2-hour post-OGTT glucose values (2-h PG): NORM was defined as fasting glucose < 110 mg/dL and 2-h PG < 140 mg/dL; PreT2D included individuals with IFG (fasting glucose 110 – 125 mg/dL) and/or IGT (2-h PG 140 – 199 mg/dL); and T2D was defined as either fasting glucose ≥ 126 mg/dL or 2-h PG ≥ 200 mg/dL, or a physician-confirmed diagnosis of T2D (2).

Participants with unclear glycemic states or those who were non-fasting at any time point were excluded from the analysis (Fig. S1). Furthermore, individuals with T2D receiving oral antidiabetic medication or insulin therapy were excluded, as previous studies have shown that treatments such as metformin and sodium-glucose-cotransporter-2 inhibitors (SGLT2i), significantly influences both targeted (18) and non-targeted metabolite profiles (19).

Blood samples used for metabolomics analyses

Blood samples were collected in the morning between 8:00 AM and 10:30 AM after at least 8 hours of overnight fasting. Serum samples were processed using standard procedures as previously described in detail (2, 20) and used for both targeted and non-targeted analyses.

Targeted metabolite quantification and normalization

Targeted metabolite profiling was performed using the Absolute*IDQ*™ p180 or p150 Kits (BIOCRATES Life Sciences AG, Innsbruck, Austria). In KORA S4 (analyzed Mar. - Apr. 2011, N = 1,615) and FF4 (Feb. - Oct. 2019, N = 2,218), 188 metabolites were measured with the p180 kit, which included five identical quality controls (QCs, defined as reference samples) per plate, three manufacture-provided QCs and three zero samples (21). In KORA F4 (Aug. 2008 - Mar. 2009, N = 3,056), 163 metabolites were measured using the p150 kit. To ensure consistency, 288 F4 samples were re-analyzed with the p180 kit (Sep. – Oct. 2019) (16).

Identical QC procedures were applied across three time-points. Metabolites were retained if they met the following criteria: 1) The average coefficient of variation (CV) was <25% in the reference or QC samples. 2) At least 50% of measured concentrations were above the limit of detection (LOD), defined as three times the median of zero samples. 3) The proportion of missing values was <5%. Non-detectable values were randomly imputed within 75–125% of half the lowest measured value per plate. Additionally, a non-parametric machine learning method, TIGER, was used to normalize the data across the three time-points to enable joint analysis (16). After QC, 104 metabolites overlapping across S4, F4, and FF4 were included (Table S1).

Non-Targeted metabolite quantification and normalization

Non-targeted metabolite profiles were analyzed using the Metabolon analytical platform (Metabolon Inc., Durham, North Carolina, USA). Serum samples from the KORA S4 and F4 studies were analyzed using the same HD2 platform in 2011 and 2009, respectively (7, 8, 19).

QC procedures followed the same criteria described previousl. Metabolites with >20% missing values in each study were excluded, along with samples with containing >10% missing metabolites. Normalized relative ion counts were log-transformed, and missing data were imputed. After QC, 312 overlapping metabolites between KORA S4 and F4 were retained. However, the identities of metabolite ID M32654 and the molecule labeled "3-dehydrocarnitine*" could not be confirmed and were therefore denoted as X-21365 (Table S1). Z-score standardization was applied to all metabolites within each S4 and F4 studies prior to statistical analysis.

Statistical analysis

Linear mixed effect model (LMEM)

To appropriately account for the repeated measures of the same study participants, we employed LMEMs with a random intercept for participant ID, which allowed for within-individual dependence and prevented observations from being analyzed as independent. Basic model was adjusted for study, age, sex, BMI, physical activity, smoking status, and alcohol intake. For example, of comparison between NORM and PreT2D:

$$Y_{\text{metabolite}} \sim \text{glycemic state}_{(1 = \text{PreT2D vs. } 0 = \text{NORM})} + \text{study}_{(\text{categorical, S4} = 0, \text{F4} = 1, \text{FF4} = 2)} + \text{age}_{(\text{years at corresponding time point})} + \text{sex}_{(\text{female} = 0, \text{male} = 1)} + \text{BMI}_{(\text{kg/m\^2})} + \text{smoking status}_{(\text{non-smoker} = 0, \text{ex-smoker} = 1, \text{current smoker} = 2)} + \text{physical activity}_{(\text{low} = 0, \text{medium} = 1; \text{high} = 2; \text{irregular exercise equal or less than 1 h per week is considered as low; regular exercise 1 h} < \text{per week} < 2 \text{ h is considered medium; regular exercise per week} > 2 \text{ h is considered high})} + \text{alcohol intake}_{(\text{g/day})} + (1 \mid \text{ID})$$

To assess robustness, further confounders were adjusted in the basic model, incorporating glycemic and cardiometabolic factors, including fasting glucose, HbA_{1C}, HDL cholesterol, and systolic BP (sensitivity model, SM). Separate sensitivity analyses were performed for fasting glucose (SM1), HbA_{1C} (SM2), HDL cholesterol (SM3), systolic BP (SM4), and waist circumference (SM5) (Fig. S1).

Further robustness was evaluated through two additional analyses: 1) Glycemic trajectories: Participants were classified: NORM (normal glucose at both time points), incident PreT2D (NORM at S4, progressing to PreT2D at FF4 / F4), PreT2D (PreT2D at both time points), incident T2D (progression to T2D at FF4 or F4), and T2D (T2D at both time points). 2) Subclassification of PreT2D: PreT2D was further subdivided into isolated impaired fasting glucose (i-IFG) and impaired glucose tolerance (IGT) to explore metabolic differences (Fig. S1).

To account for multiple testing in the LMEM models, Bonferroni correction was applied. Metabolites with a *P*-value below the threshold of $P = 0.05$ /'Number of metabolites after QC' were considered statistically significant in all analyses.

To further characterize the metabolic alterations across the spectrum of glycemic states, Bonferroni-significant metabolites were classified into three comparator groups based on their association patterns: 1. Persistent metabolites: Metabolites demonstrating progressive differences

across all glycemic states, from NORM to PreT2D and to T2D. 2. Early metabolites: Metabolites show significant differences between both PreT2D/T2D and NORM, but no significant difference between T2D and PreT2D. 3. Late metabolites: Metabolites exhibiting significant differences exclusively between T2D and PreT2D/NORM (Fig. S1).

Mediation analysis

The mediation analysis was conducted following the standard steps outlined by Baron and Kenny (22). Fasting glucose, HbA_{1C}, HDL cholesterol, and systolic BP were separately considered as mediators. Glycemic states were set as the independent variable, and one identified metabolite was used as the dependent variable. The analysis was performed using the R package 'mediation' (version 4.5.0), with *P*-values and the proportion of mediation effects calculated via nonparametric bootstrapping with 1,000 resamples.

Mendelian randomization methods

Two-sample MR analyses were performed to evaluate potential causal associations between selected metabolites and T2D using summary-level data from the IEU OpenGWAS database via the TwoSampleMR R package (version 0.5.6) (23).

Independent SNPs were selected at genome-wide significance ($P < 5 \times 10^{-8}$) after clumping ($r^2 < 0.001$, 10,000 kb window). Mean F-statistics > 10 were used to confirm adequate instrument strength.

The primary analysis employed the inverse-variance weighted (IVW) method. Sensitivity analyses included MR-Egger, weighted median, weighted mode, Cochran's Q test, MR-Egger intercept, and MR-PRESSO global test to assess heterogeneity and pleiotropy (24). Leave-one-out analyses were performed to evaluate the impact of individual SNPs.

All statistical analyses were conducted using R software (v4.3.2) (25).

Data and resource availability

KORA datasets are not publicly available because of data protection agreements. However, datasets can be provided upon request through the KORA-PASST (Project application self-service tool, https://helmholtz-muenchen.managed-otrs.com/external).

Results

Characterization of study participants

This longitudinal analysis included 8,240 samples from 3,183 participants across three time points (S4, F4, FF4) using targeted and non-targeted metabolomics (Table 1, Table S2). Progression from NORM to PreT2D and T2D was associated with significant shifts in cardiometabolic risk profiles, characterized by older age; higher BMI, waist circumference, fasting glucose, 2-h PG, HbA_{1C}, systolic BP, and gamma-glutamyl transferase (GGT); but lower HDL cholesterol and reduced physical activity.

Trajectory analysis included 1,050 individuals: 576 NORM, 118 PreT2D, 33 T2D, 211 incident PreT2D, and 112 incident T2D cases (Tables S3 for targeted and Table S4 for non-targeted datasets). Compared to individuals who remained NORM throughout the study, participants with

incident PreT2D and incident T2D were more often male (57.9% for 114 individuals with incident PreT2D and 58.8% for 68 incident T2D, compared to 48% for 225 NORM in the targeted dataset, Table S3). These participants also had higher BMI, waist circumference, fasting glucose, HbA_{1C}, and cardiovascular risk factors, including lower HDL cholesterol and higher systolic BP, at both baseline and follow-up. These trends were consistent across the longitudinal datasets.

Metabolites differentiating glycemic states in targeted analysis

Of 104 targeted metabolites, 45 revealed Bonferroni-significant differences between 3,867 NORM and 381 T2D individuals in LMEMs adjusted for the basic model (Fig. 1A, Table S5). Three metabolites, hexose (H1), valine, and SM (OH) C22:2, demonstrated progressive changes across all glycemic states, classifying them as persistent metabolites (Fig. 1B). 29 metabolites, including glycine and 28 lipids (e.g., 3 carnitines: C2, C16 and C18:1; 4 PC aas; 10 PC aes; 5 LPCs; 6 SMs), were identified as early-stage specific, differing between PreT2D/T2D and NORM but not between T2D and PreT2D. Conversely, 13 metabolites, including tyrosine, showed Bonferroni-significant differences only between T2D and NORM, without stage-specific changes (Fig. S2A, Table S5).

Trajectory analysis (S4, FF4) identified H1 as the only Bonferroni-significant metabolite across seven pairwise LMEMs adjusted for the basic model. H1 displayed significant differences across all glycemic states, except between established T2D and incident T2D (Fig. 1C).

Metabolites differentiating glycemic states in non-targeted analysis

Among 312 non-targeted metabolites, 109 showed Bonferroni-significant differences between T2D and NORM, including 11 persistent, 40 early-stage, and 20 late-stage metabolites (Figs. 2A, S3A, Table S6). Persistent metabolites included glucose, α-hydroxybutyrate (AHB), glutamate, and glycine. Early-stage metabolites were C2, 7 long-chain fatty acids (LCFAs), pyruvate, and lactate (Figs. S4A & S4B). Late-stage metabolites, such as BCAAs (valine, isoleucine, leucine), 1,5-anhydroglucitol (1,5-AG), four γ-glutamyl derivatives, erythrose and erythronate, differed between T2D and both NORM and PreT2D, but not between PreT2D and NORM (Figs. 2A & 2B, Fig. S3A, Table S6).

Among 140 identified metabolites (45 targeted, 109 non-targeted, 14 overlapping), three metabolites revealed consistent patterns: glucose (persistent), C2 (early-stage), and tyrosine (non-stage specific). However, valine, glycine and four carnitines exhibited contrasting classifications between targeted and non-targeted datasets despite consistent directional changes. For example, valine was borderline significant in PreT2D vs. NORM (P = 3.18E-04) in the targeted analysis and classified as persistent, while in the non-targeted analysis, it was grouped with isoleucine and leucine as late-stage metabolites (Figs. 1B & 2B, Tables S5 & S6).

Trajectory analysis (S4 to F4; Fig. 2C) identified fewer Bonferroni-significant metabolites from seven pairwise LMEMs. AHB and 3-methyl-2-oxobutyrate were consistently differential in both incident and established T2D compared to NORM. Two unknown metabolites (X-11470, X-12039) were also identified: X-11470 increased in incident PreT2D, while X-12039 was depleted in T2D, both compared to NORM.

Sensitivity and subgroup analyses reveal stage-specific metabolic patterns and glycemic dependencies

Sensitivity analyses of six additional LMEM models confirmed stage-specific metabolic patterns (Figs. S2B-S2G & S3B-S3G, Tables S5 & S6).

Among 13 persistent metabolites (3 targeted, 11 non-targeted, 1 overlapping), the overlapping hexose/glucose lost significance in SM and SM1 (adjusted for fasting glucose) but remained significant in SM2, SM3, SM4, and SM5 across all pairwise comparisons (PreT2D vs. NORM, T2D vs. NORM, T2D vs. PreT2D) The only exception was glucose, which showed borderline significance in PreT2D vs. NORM (P = 2.03E-05 in SM and P = 4.84E-06 in SM1), compared to highly significant results in the basic model of the non-targeted analyses (P = 2.01E-37). These observations confirm the glycemic dependent of hexose/glucose and their role as positive controls for glycemic progression.

SM (OH) C22:2 remained a persistent metabolite in SM4 and SM5 but shifted to early-stage in SM, SM1, SM2, and SM3 (targeted analysis, Fig. S2). In both targeted and non-targeted analyses, valine was not significant in distinguishing NORM from PreT2D across all six sensitivity analyses. However, valine consistently shifted to later-stage in SM3, SM4, and SM5 (Figs. S2, S3).

Of the 68 early-stage metabolites identified (29 targeted, 40 non-targeted, 1 overlapping), 34 remained significant across all six sensitivity models. These included C2 (overlapping), LPC 18:2 (targeted), and pyruvate, lactate, and 3-methyl-2-oxobutyrate (non-targeted). In contract, citrate was the only exception, classified as a non–stage-specific metabolite in the basic model, just below the borderline significance threshold in PreT2D vs. NORM (P = 1.74E-04, cutoff: P = 1.60E-04). However, citrate shifted to an early-stage metabolite with consistently higher levels in PreT2D and T2D compared to NORM across all five sensitivity models (SM, SM1, SM2, SM3, and SM4). P values hovered near borderline significance (ranging from 8.42E-06 to 5.06E-04 in both PreT2D/T2D vs. NORM comparisons, while showing no significance between T2D and PreT2D (non-targeted, Fig. S3). Based on these observations, citrate is most likely an early-stage metabolite and appears to be independent of established cardiometabolic risk factors.

None of the 20 late-stage metabolites identified (non-targeted) remained significant in SM, with associations for BCAAs and γ-glutamyl derivatives attenuated in SM1 and SM2, and erythrose in SM2, SM3, SM4, and SM5 (Fig. S3, Table S6).

Subgroup analysis of PreT2D into i-IFG and IGT revealed clearer metabolic shifts (Figs. S5A & S5B). H1 and glucose consistently showed higher levels across glycemic stages (e.g., i-IFG/IGT vs. NORM, T2D vs. i-IFG/IGT). Targeted analysis identified 7 IGT-specific metabolites, including C2 and LPC 18:2 (Fig. S5A). Non-targeted analysis showed lactate remained significant for i-IFG/IGT vs. NORM, while AHB and glycine were significant for IGT vs. NORM and T2D vs. i-IFG. Late-stage γ-glutamyl-leucine was significant for T2D vs. IGT (Fig. S5B).

Mediation but no direct causal effect of glycemic states on metabolic alterations

Mediation analysis of γ-glutamyl-valine revealed that fasting glucose and HbA_{1C} significantly mediated its association with late-stage glycemic status (e.g., T2D vs. PreT2D), accounting for 35.1% and 31.0% of the effect, respectively (Fig. 3A). HDL cholesterol had a small but significant mediation effect (6.4%, P = 0.020), while systolic BP showed no significant effect (Fig. 3A).

MR analyses for C2, BCAAs, and γ-glutamyl-valine showed no statistically significant causal effects on T2D using the IVW method (Fig. 3B, Fig. S6). Sensitivity analyses, including MR-Egger intercept, Cochran's Q test, MR-PRESSO, and leave-one-out, identified no evidence of pleiotropy or heterogeneity (Fig. S7, Tables S7 & S8).

Discussion

This study identified distinct stage-specific metabolic shifts during the progression from normoglycemia to incident and established PreT2D and T2D using longitudinal data from the KORA S4/F4/FF4 cohort. Early-stage alterations primarily involved energy metabolism (e.g., C2, lactate, acylcarnitines, LCFAs), whereas later-stage changes were dominated by amino acid-related pathways, including BCAAs and γ-glutamyl derivatives. These patterns were consistent across glycemic stages (NORM, PreT2D, T2D), incident PreT2D and T2D cases, and PreT2D subtypes (i-IFG and IGT). Robust integration of targeted and non-targeted metabolomics enabled the identification and internal validation of stage-specific metabolites, independent of physiological, lifestyle, glycemic, and cardiovascular covariates. Subsequent mediation and MR analyses indicated that glycemic markers, particularly fasting glucose and HbA_{1C}, may partially mediate the associations between BCAA-related metabolites and T2D risk, although no evidence of a direct causal relationship was observed.

Our findings align with prior research, confirming glucose and hexose (H1) as consistent markers across glycemic stages and platforms, serving as positive controls for diabetes research (2). Early-stage metabolites, such as C2 and LPC 18:2, exhibited significant elevation from NORM to PreT2D, particularly in IGT, reflecting enhanced fatty acid oxidation and mitochondrial stress due to reduced insulin regulation of lipolysis (2, 26). Late-stage metabolites, including AHB, mannose, and 1,5-AG showed reproducible associations with established T2D (27, 28).

We also observed significant metabolic disruptions in PreT2D, marked by higher citrate levels indicating impaired TCA cycle activity and lower glycine levels reflecting increased oxidative stress. These alterations align with a metabolic shift toward anaerobic glycolysis and reduced efficiency in energy metabolism (Fig. 3C). Supporting this, an imbalance between pyruvate and lactate further highlights impaired oxidative metabolism, with increased diversion of pyruvate toward lactate production (Fig. S4C). Collectively, these findings suggest a redox state in which mitochondrial oxidation is constrained, favoring cytosolic glycolysis.

Consistently, glycine depletion, accompanied by elevated glutamate and γ-glutamyl dipeptides, underscores oxidative stress and accelerated glutathione (GSH) turnover (29). At later stages, elevations in erythrose and erythronate levels suggest activation of the non-oxidative pentose phosphate pathway (PPP), likely to supply NADPH for GSH regeneration and lipid synthesis (30, 31). Overall, these findings indicate a tight coupling between energy metabolic inefficiency and oxidative stress throughout the progression of T2D (Fig. 3C).

Rising BCAA levels, particularly between PreT2D and T2D, further highlight their dysregulation as a hallmark of advanced disease stages, consistent with prior studies (32). Similarly, γ-glutamyl derivatives, including γ-glutamyl-valine, exhibited minor changes in prediabetes but larger increases in T2D, suggesting their potential as stage-specific markers of amino acid overload and redox stress (33).

Among the 14 overlapping metabolites measured by both platforms, valine and glycine exhibited some discrepancies, which may partly reflect differences in participant composition or platform sensitivity. Prior research comparing Biocrates and Metabolon technologies reported correlation coefficients of 0.661 for valine, and 0.747 for glycine, and 0.891 for C2 (34). This may explain the consistent results for C2 across platforms. Despite platform-specific variations, glycine was consistently identified as an early-stage metabolite across all 6 LMEM models (targeted), and the SM (non-targeted).

Interestingly, early-stage metabolites (e.g., citrate, lactate) remained robustly associated with glycemic status, even after adjusting for glucose, HbA_{1C}, HDL cholesterol, systolic BP, and waist circumference, implying that they may reflect broader metabolic disturbances rather than hyperglycemia alone. Conversely, late-stage metabolites (e.g., BCAAs, γ-glutamyl derivatives) were more sensitive to glycemic adjustments, indicating their dependency on glucose regulation.

Trajectory-based subgroup analyses identified fewer significant metabolites, likely due to the smaller number of incident cases and the stringent Bonferroni threshold. However, incident cases shared similarities with preceding glycemic states rather than established T2D. For example, elevated levels of AHB and 3-methyl-2-oxobutyrate were observed in both incident T2D and established T2D, consistent with previous findings. AHB is a known biomarker for insulin resistance and IGT (35, 36), while 3-methyl-2-oxobutyrate has been linked to incident T2D (8). In our longitudinal analyses, 3-methyl-2-oxobutyrate remained significant across all six sensitivity models as early-stage specific, suggesting independence from physiological, lifestyle, and cardiovascular risk factors. In the PreT2D subgroup analyses, there was also a significant difference between IGT and NORM. As a branched-chain keto-acid (BCKA) derivative of valine, 3-methyl-2-oxobutyrate may reflect reduced mitochondrial capacity in PreT2D and T2D to break down BCAAs, as suggested by Fall et al (8).

Mediation analyses supported the sensitivity findings, showing that glycemic markers, particularly fasting glucose and HbA_{1C}, mediated over 30% of the association with γ-glutamyl-valine, suggesting these derivatives are largely glycemia-driven. This aligns with the known role of GGT in oxidative stress and cardiometabolic risk (12), linking the γ-glutamyl pathway to amino acid dysregulation and hyperglycemia-related stress.

To explore causality, MR analyses found no evidence that BCAAs, C2 or γ-glutamyl-valine directly drive T2D. Given longstanding hypotheses that BCAAs may promote insulin resistance, we further repeated MR focusing on insulin resistance–predominated T2D. This additional analysis again showed no causal effect of BCAAs (Fig. 3B). Taken together with the observed rise in BCAAs during T2D progression in our studies, these results suggest that elevations of BCAAs, C2, and γ-glutamyl-valine are more likely downstream manifestations of metabolic dysregulation rather than causal drivers of T2D. However, limited genetic instruments and the absence of GWAS data for prediabetes restrict the causal assessment of the PreT2D-to-T2D transition.

This study has several limitations. Residual confounding from unmeasured lifestyle factors, such as diet, cannot be ruled out, although fasting serum samples were used in the current study. Diabetes duration was not considered, potentially influencing metabolomic profiles within T2D participants. Furthermore, the predominantly European population restricts the generalizability of findings to other ethnic groups.

We propose a hypothetical sequence of metabolic disturbances (Fig. 3C). Insulin resistance initiates lipid oxidation changes, reflected by elevated acylcarnitines like C2 and LCFAs. These

impairments disrupt the TCA cycle, elevating citrate and lactate causing a pyruvate–lactate imbalance indicative of impaired oxidative metabolism, while altering α-ketoglutarate metabolism, leading to higher glutamate (5, 37), which may reflect enhanced glutaminolysis, providing α-KG to compensate for reduced TCA cycle efficiency. Glycine depletion likely reflects increased GSH synthesis in response to oxidative stress, driven by chronic insulin resistance and low-grade inflammation (38). As disease progresses, impaired amino acid clearance elevates circulating BCAAs, fueling γ-glutamyl derivative formation (39). The observed increases in γ-glutamyl-valine and related derivatives may represent an adaptive response facilitated by GGT to buffer amino acid and redox imbalance (40).

In conclusion, we identified distinct stage-specific metabolomic signatures along the glycemic continuum. Early-stage metabolites (e.g., C2, lactate) reflect energy metabolism, while late-stage metabolites (e.g., BCAAs, γ-glutamyl derivatives) characterize advanced metabolic disruption. Our systematic analysis of glycemic states exhibited unique metabolic shifts, distinguishing them from established T2D and highlighting potential windows for early detection and intervention. The reproducibility of findings across platforms and time points strengthens their credibility, offering a foundation for developing stage-specific biomarkers and precision strategies for diabetes prevention.

Acknowledgements

We sincerely thank all KORA study participants for their enduring commitment to the study and the dedicated staff for their efforts in data collection and management. We also extend our appreciation to the KORA Study Group (https://www.helmholtz-munich.de/en/epi/cohort/kora) for their essential role in the design and execution of the study.

Funding

J.G. was supported by a scholarship from the China Scholarship Council (CSC), under the State Scholarship Fund (File No. 202206010083). Part of this study was supported by funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No. 821508 (CARDIATEAM). The JU receives support from the European Union's Horizon 2020 research and innovation programme and the European Federation of Pharmaceutical Industries and Associations (EFPIA). The German Diabetes Center is supported by the German Federal Ministry of Health (Berlin, Germany) and the Ministry of Science and Culture in North-Rhine Westphalia (Düsseldorf, Germany). The KORA study was initiated and financed by the Helmholtz Zentrum Munchen, German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Data collection in the KORA study is done in cooperation with the University Hospital of Augsburg.

Declaration of competing interest

No potential conflicts of interest relevant to this article were reported.

Author contributions

J.G. and R.W-S. designed and implemented the study. J.G. performed the statistical analyses. C.P., J.A., G.B., K.S., C.G., A.P. and R.W-S. generated the metabolomics data. S.S., W.K., B.L., B.T., and A.P. contributed to generating KORA phenotypes and clinical variables. S.H., M.S., M.H., S.Y., and J.Z. carried out formal analyses. J.G. and R.W-S. drafted the manuscript, and all authors critically revised it for important intellectual content. R.W-S. is guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Declaration of generative AI and AI-assisted technologies

We would like to acknowledge the valuable support of OpenAI's ChatGPT language model in refining the English grammar and clarity of this manuscript.

Reference

1. Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet* 2023;402:203-234

2. Wang-Sattler R, Yu Z, Herder C, Messias AC, Floegel A, He Y, Heim K, Campillos M, Holzapfel C, Thorand B, Grallert H, Xu T, Bader E, Huth C, Mittelstrass K, Döring A, Meisinger C, Gieger C, Prehn C, Roemisch-Margl W, Carstensen M, Xie L, Yamanaka-Okumura H, Xing G, Ceglarek U, Thiery J, Giani G, Lickert H, Lin X, Li Y, Boeing H, Joost HG, de Angelis MH, Rathmann W, Suhre K, Prokisch H, Peters A, Meitinger T, Roden M, Wichmann HE, Pischon T, Adamski J, Illig T. Novel biomarkers for pre-diabetes identified by metabolomics. *Mol Syst Biol* 2012;8:615

3. Tabák AG, Herder C, Rathmann W, Brunner EJ, Kivimäki M. Prediabetes: a high-risk state for diabetes development. *Lancet* 2012;379:2279-2290

4. Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E, Lewis GD, Fox CS, Jacques PF, Fernandez C, O'Donnell CJ, Carr SA, Mootha VK, Florez JC, Souza A, Melander O, Clish CB, Gerszten RE. Metabolite profiles and the risk of developing diabetes. *Nat Med* 2011;17:448-453

5. Newgard CB, An J, Bain JR, Muehlbauer MJ, Stevens RD, Lien LF, Haqq AM, Shah SH, Arlotto M, Slentz CA, Rochon J, Gallup D, Ilkayeva O, Wenner BR, Yancy WS, Jr., Eisenson H, Musante G, Surwit RS, Millington DS, Butler MD, Svetkey LP. A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell Metab* 2009;9:311-326

6. Stumvoll M, Goldstein BJ, van Haeften TW. Type 2 diabetes: principles of pathogenesis and therapy. *Lancet* 2005;365:1333-1346

7. Suhre K, Shin SY, Petersen AK, Mohney RP, Meredith D, Wägele B, Altmaier E, Deloukas P, Erdmann J, Grundberg E, Hammond CJ, de Angelis MH, Kastenmüller G, Köttgen A, Kronenberg F, Mangino M, Meisinger C, Meitinger T, Mewes HW, Milburn MV, Prehn C, Raffler J, Ried JS, Römisch-Margl W, Samani NJ, Small KS, Wichmann HE, Zhai G, Illig T, Spector TD, Adamski J, Soranzo N, Gieger C. Human metabolic individuality in biomedical and pharmaceutical research. *Nature* 2011;477:54-60

8. Fall T, Salihovic S, Brandmaier S, Nowak C, Ganna A, Gustafsson S, Broeckling CD, Prenni JE, Kastenmüller G, Peters A, Magnusson PK, Wang-Sattler R, Giedraitis V, Berne C, Gieger C, Pedersen NL, Ingelsson E, Lind L. Non-targeted metabolomics combined with genetic analyses identifies bile acid synthesis and phospholipid metabolism as being associated with incident type 2 diabetes. *Diabetologia* 2016;59:2114-2124

9. Nathan DM, Davidson MB, DeFronzo RA, Heine RJ, Henry RR, Pratley R, Zinman B. Impaired Fasting Glucose and Impaired Glucose Tolerance: Implications for care. *Diabetes Care* 2007;30:753-759

10. Luo H, Bauer A, Nano J, Petrera A, Rathmann W, Herder C, Hauck SM, Sun BB, Hoyer A, Peters A, Thorand B. Associations of plasma proteomics with type 2 diabetes and related traits: results from the longitudinal KORA S4/F4/FF4 Study. *Diabetologia* 2023;66:1655-1668

11. Huemer MT, Spagnuolo MC, Maalmi H, Wagner R, Bönhof GJ, Heier M, Koenig W, Rathmann W, Prystupa K, Nano J, Ziegler D, Peters A, Roden M, Thorand B, Herder C. Phenotype-based clusters, inflammation and cardiometabolic complications in older people before the diagnosis of type 2 diabetes: KORA F4/FF4 cohort study. *Cardiovasc Diabetol* 2025;24:83

12. Wu Q, Li J, Zhu J, Sun X, He D, Li J, Cheng Z, Zhang X, Xu Y, Chen Q, Zhu Y, Lai M. Gamma-glutamyl-leucine levels are causally associated with elevated cardio-metabolic risks. *Front Nutr* 2022;9:936220

13. Mitrić A, Castellano I. Targeting gamma-glutamyl transpeptidase: A pleiotropic enzyme involved in glutathione metabolism and in the control of redox homeostasis. *Free Radic Biol Med* 2023;208:672-683

14. Haythorne E, Rohm M, van de Bunt M, Brereton MF, Tarasov AI, Blacker TS, Sachse G, Silva dos Santos M, Terron Exposito R, Davis S, Baba O, Fischer R, Duchen MR, Rorsman P, MacRae JI, Ashcroft FM. Diabetes causes marked inhibition of mitochondrial metabolism in pancreatic β-cells. *Nat Commun* 2019;10:2474

15. Han S, Yu S, Shi M, Harada M, Ge J, Lin J, Prehn C, Petrera A, Li Y, Sam F, Matullo G, Adamski J, Suhre K, Gieger C, Hauck SM, Herder C, Roden M, Casale FP, Cai N, Peters A, Wang-Sattler R. LEOPARD: missing view completion for multi-timepoint omics data via representation disentanglement and temporal knowledge transfer. *Nat Commun* 2025;16:3278

16. Han S, Huang J, Foppiano F, Prehn C, Adamski J, Suhre K, Li Y, Matullo G, Schliess F, Gieger C, Peters A, Wang-Sattler R. TIGER: technical variation elimination for metabolomics data using ensemble learning architecture. *Brief Bioinform* 2022;23

17. American Diabetes Association Professional Practice C. 2. Diagnosis and Classification of Diabetes: Standards of Care in Diabetes—2025. *Diabetes Care* 2024;48:S27-S49

18. Xu T, Brandmaier S, Messias AC, Herder C, Draisma HH, Demirkan A, Yu Z, Ried JS, Haller T, Heier M, Campillos M, Fobo G, Stark R, Holzapfel C, Adam J, Chi S, Rotter M, Panni T, Quante AS, He Y, Prehn C, Roemisch-Margl W, Kastenmüller G, Willemsen G, Pool R, Kasa K, van Dijk KW, Hankemeier T, Meisinger C, Thorand B, Ruepp A, Hrabé de Angelis M, Li Y, Wichmann HE, Stratmann B, Strauch K, Metspalu A, Gieger C, Suhre K, Adamski J, Illig T, Rathmann W, Roden M, Peters A, van Duijn CM, Boomsma DI, Meitinger T, Wang-Sattler R. Effects of metformin on metabolite profiles and LDL cholesterol in patients with type 2 diabetes. *Diabetes Care* 2015;38:1858-1867

19. Harada M, Adam J, Covic M, Ge J, Brandmaier S, Muschet C, Huang J, Han S, Rommel M, Rotter M, Heier M, Mohney RP, Krumsiek J, Kastenmüller G, Rathmann W, Zou Z, Zukunft S, Scheerer MF, Neschen S, Adamski J, Gieger C, Peters A, Ankerst DP, Meitinger T, Alderete TL, de Angelis MH, Suhre K, Wang-Sattler R. Bidirectional modulation of TCA cycle metabolites and anaplerosis by metformin and its combination with SGLT2i. *Cardiovasc Diabetol* 2024;23:199

20. Huang J, Huth C, Covic M, Troll M, Adam J, Zukunft S, Prehn C, Wang L, Nano J, Scheerer MF, Neschen S, Kastenmüller G, Suhre K, Laxy M, Schliess F, Gieger C, Adamski J, Hrabe de Angelis M, Peters A, Wang-Sattler R. Machine Learning Approaches Reveal Metabolic Signatures of Incident Chronic Kidney Disease in Individuals With Prediabetes and Type 2 Diabetes. *Diabetes* 2020;69:2756-2765

21. Haid M, Muschet C, Wahl S, Römisch-Margl W, Prehn C, Möller G, Adamski J. Long-Term Stability of Human Plasma Metabolites during Storage at -80 °C. *J Proteome Res* 2018;17:203-211

22. Baron RM, Kenny DA. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J Pers Soc Psychol* 1986;51:1173-1182

23. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess S, Bowden J, Langdon R, Tan VY, Yarmolinsky J, Shihab HA, Timpson NJ, Evans DM, Relton C, Martin RM, Davey Smith G, Gaunt TR, Haycock PC. The MR-Base platform supports systematic causal inference across the human phenome. *Elife* 2018;7

24. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet* 2018;50:693-698

25. RCoreTeam. R: A Language and Environment for Statistical Computing. Vienna, Austria, R Foundation for Statistical Computing, 2023

26. Hosseinkhani S, Arjmand B, Dilmaghani-Marand A, Mohammadi Fateh S, Dehghanbanadaki H, Najjar N, Alavi-Moghadam S, Ghodssi-Ghassemabadi R, Nasli-Esfahani E, Farzadfar F, Larijani B, Razi F. Targeted metabolomics analysis of amino acids and acylcarnitines as risk markers for diabetes by LC-MS/MS technique. *Sci Rep* 2022;12:8418

27. Kim MJ, Jung HS, Hwang-Bo Y, Cho SW, Jang HC, Kim SY, Park KS. Evaluation of 1,5-anhydroglucitol as a marker for glycemic variability in patients with type 2 diabetes mellitus. *Acta Diabetol* 2013;50:505-510

28. Steiner B, Leitner C, Stadler D, Prugger EM, Magnes C, Herzog PL. Enzymatic detection of α-hydroxybutyrate, an important marker of insulin resistance, and comparison with LC-MS/MS detection. *Pract Lab Med* 2024;40:e00398

29. Sekhar RV, Patel SG, Guthikonda AP, Reid M, Balasubramanyam A, Taffet GE, Jahoor F. Deficient synthesis of glutathione underlies oxidative stress in aging and can be corrected by dietary cysteine and glycine supplementation. *Am J Clin Nutr* 2011;94:847-853

30. Lund J, Ouwens DM, Wettergreen M, Bakke SS, Thoresen GH, Aas V. Increased Glycolysis and Higher Lactate Production in Hyperglycemic Myotubes. *Cells* 2019;8

31. Moon DO. NADPH Dynamics: Linking Insulin Resistance and β-Cells Ferroptosis in Diabetes Mellitus. *Int J Mol Sci* 2023;25

32. Choi BH, Hyun S, Koo SH. The role of BCAA metabolism in metabolic health and disease. *Exp Mol Med* 2024;56:1552-1559

33. Zaghlool SB, Halama A, Stephan N, Gudmundsdottir V, Gudnason V, Jennings LL, Thangam M, Ahlqvist E, Malik RA, Albagha OME, Abou-Samra AB, Suhre K. Metabolic and proteomic signatures of type 2 diabetes subtypes in an Arab population. *Nat Commun* 2022;13:7121

34. Suhre K, Meisinger C, Döring A, Altmaier E, Belcredi P, Gieger C, Chang D, Milburn MV, Gall WE, Weinberger KM, Mewes HW, Hrabé de Angelis M, Wichmann HE, Kronenberg F, Adamski J, Illig T. Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting. *PLoS One* 2010;5:e13953

35. Cobb J, Eckhart A, Motsinger-Reif A, Carr B, Groop L, Ferrannini E. α-Hydroxybutyric Acid Is a Selective Metabolite Biomarker of Impaired Glucose Tolerance. *Diabetes Care* 2016;39:988-995

36. Gall WE, Beebe K, Lawton KA, Adam KP, Mitchell MW, Nakhle PJ, Ryals JA, Milburn MV, Nannipieri M, Camastra S, Natali A, Ferrannini E. alpha-hydroxybutyrate is an early biomarker of insulin resistance and glucose intolerance in a nondiabetic population. *PLoS One* 2010;5:e10883

37. Adams SH. Emerging perspectives on essential amino acid metabolism in obesity and the insulin-resistant state. *Adv Nutr* 2011;2:445-456

38. Dawi J, Misakyan Y, Affa S, Kades S, Narasimhan A, Hajjar F, Besser M, Tumanyan K, Venketaraman V. Oxidative Stress, Glutathione Insufficiency, and Inflammatory Pathways in Type 2 Diabetes Mellitus: Implications for Therapeutic Interventions. *Biomedicines* 2024;13

39. Lee DS, Evans JC, Robins SJ, Wilson PW, Albano I, Fox CS, Wang TJ, Benjamin EJ, D'Agostino RB, Vasan RS. Gamma glutamyl transferase and metabolic syndrome, cardiovascular disease, and mortality risk: the Framingham Heart Study. *Arterioscler Thromb Vasc Biol* 2007;27:127-133

40. Fraser A, Harris R, Sattar N, Ebrahim S, Davey Smith G, Lawlor DA. Alanine aminotransferase, gamma-glutamyltransferase, and incident diabetes: the British Women's Heart and Health Study and meta-analysis. *Diabetes Care* 2009;32:741-750

Table 1. Characteristics of study participants across glycemic states in the targeted metabolomics study

	S4 (N= 1,361)				F4 (N = 1,965)				FF4 (N = 1,959)			
	NORM	**PreT2D**	**T2D**	***P***	**NORM**	**PreT2D**	**T2D**	***P***	**NORM**	**PreT2D**	**T2D**	***P***
N	867	348	146		1571	298	96		1429	391	139	
Age, years	63.47 ± 5.55	64.95 ± 5.17	65.05 ± 5.33	<0.001	51.64 ± 11.73	60.90 ± 10.54	63.80 ± 9.06	<0.001	56.83 ± 11.49	65.28 ± 11.12	69.74 ± 10.36	<0.001
Male sex, %	47.8	59.2	58.9	<0.001	46.3	55.4	60.4	0.001	44.5	56.3	56.8	<0.001
BMI, kg/m^2	27.67 ± 4.09	29.52 ± 4.05	30.33 ± 4.40	<0.001	26.51 ± 4.21	29.54 ± 4.99	30.40 ± 4.15	<0.001	26.62 ± 4.42	29.73 ± 4.78	30.91 ± 5.36	<0.001
Waist circumference, cm	93.30 ± 11.15	99.30 ± 10.35	101.92 ± 10.97	<0.001	90.20 ± 12.81	99.66 ± 13.74	103.22 ± 10.46	<0.001	93.00 ± 12.95	103.13 ± 12.60	105.99 ± 13.05	<0.001
Fasting glucose, mg/dL	95.63 ± 7.08	107.68 ± 9.54	134.02 ± 32.98	<0.001	91.62 ± 7.55	103.31 ± 11.02	124.62 ± 20.28	<0.001	94.20 ± 7.34	106.64 ± 9.59	126.02 ± 29.26	<0.001
2-h PG, mg/dL	101.98 ± 21.02	147.21 ± 30.41	232.63 ± 67.44	<0.001	97.18 ± 20.75	149.80 ± 27.07	218.92 ± 42.46	<0.001	96.54 ± 20.10	147.50 ± 27.28	224.78 ± 64.71	<0.001
HbA_{1C}, %	5.56 ± 0.33	5.66 ± 0.38	6.39 ± 1.20	<0.001	5.35 ± 0.30	5.63 ± 0.35	6.24 ± 0.61	<0.001	5.32 ± 0.32	5.60 ± 0.35	6.14 ± 0.81	<0.001
HbA_{1C}, mmol/mol	37.26 ± 3.65	38.32 ± 4.17	46.37 ± 13.13	<0.001	34.93 ± 3.30	38.09 ± 3.85	44.66 ± 6.68	<0.001	34.65 ± 3.55	37.73 ± 3.81	43.63 ± 8.81	<0.001
HDL cholesterol, mg/dL	60.34 ± 16.38	55.50 ± 15.33	50.86 ± 15.40	<0.001	57.68 ± 14.56	52.57 ± 13.41	48.09 ± 11.85	<0.001	68.35 ± 18.67	60.95 ± 17.95	57.24 ± 16.23	<0.001
Systolic BP, mmHg	131.60 ± 18.92	140.27 ± 19.14	146.47 ± 21.28	<0.001	118.70 ± 16.93	127.54 ± 18.02	135.67 ± 18.32	<0.001	116.01 ± 16.22	123.54 ± 17.98	125.68 ± 21.10	<0.001
Physical activity, %				0.019				0.001				<0.001
Low	53.1	61.1	65.1		39.1	50.7	46.9		36.5	48.6	60.4	
Medium	27.3	21.9	21.2		33.0	30.5	27.1		34.3	30.2	20.1	
High	19.6	17.0	13.7		27.8	18.8	26.0		29.3	21.2	19.4	
Smoking, %				0.226				<0.001				0.003
Non-smoker	48.3	47.1	41.8		41.7	47.0	35.4		41.3	42.7	54.0	
Ex-smoker	37.1	41.7	43.8		39.6	43.6	54.2		41.4	43.5	39.6	
Current smoker	14.7	11.2	14.4		18.8	9.4	10.4		17.3	13.8	6.5	
Alcohol intake, g/day	14.86 ± 18.50	19.60 ± 25.30	15.74 ± 21.60	0.002	14.42 ± 19.51	16.69 ± 21.35	16.89 ± 23.48	0.122	13.92 ± 18.09	17.13 ± 22.95	15.83 ± 24.61	0.013
GGT, U/l					35.55 ± 34.87	51.42 ± 65.09	56.05 ± 54.17	<0.001	30.73 ± 32.09	41.94 ± 34.44	54.12 ± 71.29	<0.001

Data are means ± SD for continuous variables and % for categorical variables. *P*-values were calculated using one-way ANOVA for continuous variables and chi-square test for categorical variables, comparing across the three groups (NORM, PreT2D, T2D). Abbreviations: BP, blood pressure; BMI, body mass index; GGT, gamma-glutamyl transferase; HDL, high density lipoprotein; NORM, normal glucose metabolism; PG, post-OGTT values; PreT2D, prediabetes; T2D, type 2 diabetes; GGT, gamma-glutamyl-transferase.

Figure Legends

Figure 1. Targeted analysis of glycemic states. **A**: Volcano plots illustrating differences in metabolite concentrations for the pairwise comparisons: PreT2D vs. NORM, T2D vs. NORM, and T2D vs. PreT2D. The dashed line indicates the Bonferroni correction threshold, with metabolites above this threshold considered Bonferroni-significant. **B**: Box plots showing group differences for notable metabolites, including hexose (H1), SM (OH) C22:2, valine (Val) and glycine (Gly). Persistent metabolites (e.g., H1 and valine) demonstrated consistent changes across glycemic states, while early metabolites (e.g., C2 and Glycine) were altered in PreT2D and T2D compared to NORM but not between PreT2D and T2D. **C:** Summary of trajectory analysis across NORM, incident PreT2D, PreT2D, incident T2D, and T2D. All analyses were conducted using linear mixed-effects models (LMEM) adjusted for age, sex, BMI, and lifestyle factors (physical activity, smoking status, and alcohol intake). Metabolites with Bonferroni-significant are shown in plot C.

Fig. 1

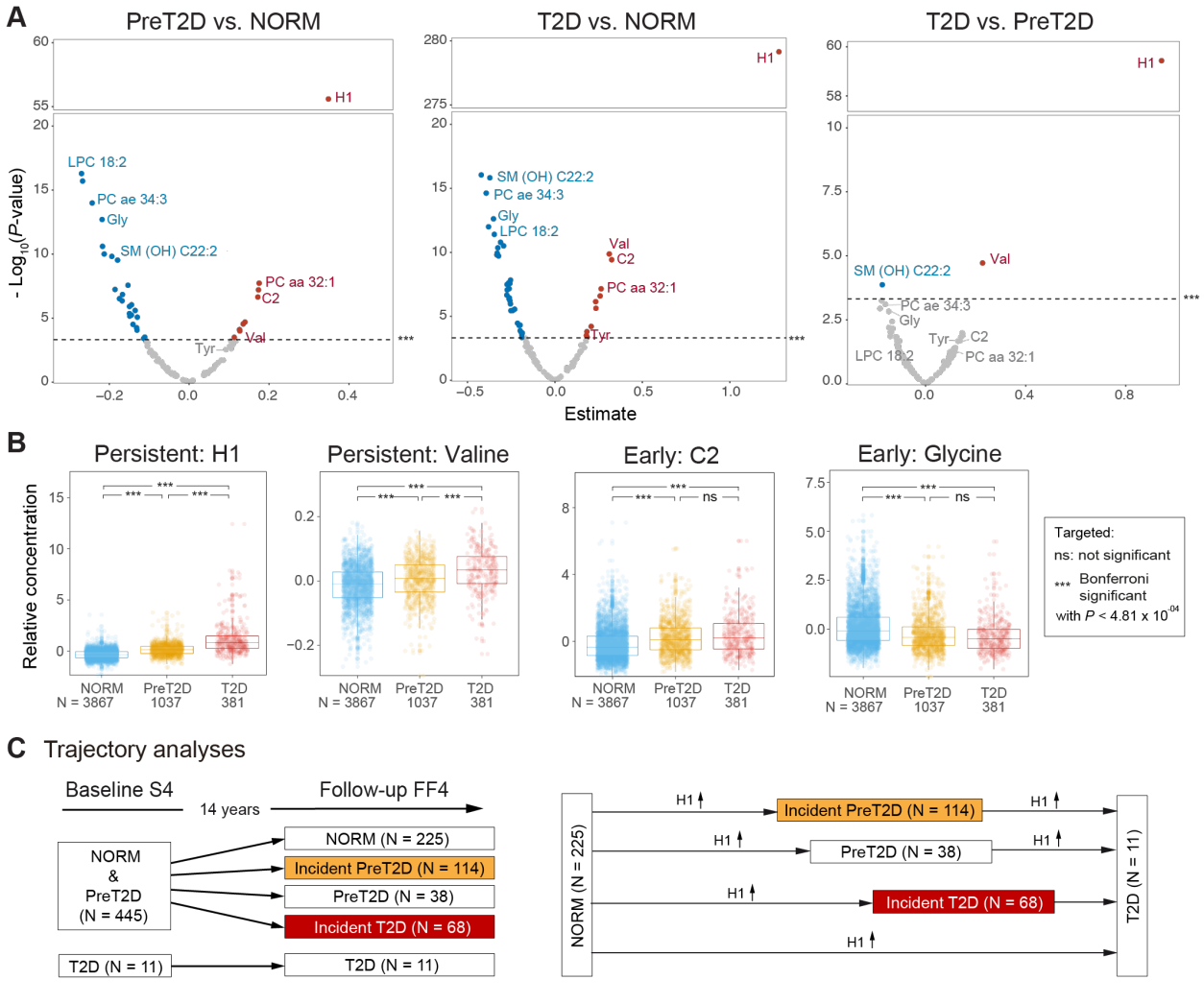

Figure 2. Non-targeted metabolomic analysis of glycemic states. A: Volcano plots illustrating differences in metabolite ion counts for the three pairwise comparisons: PreT2D vs. NORM, T2D vs. NORM, and T2D vs. PreT2D. The dashed line indicates the Bonferroni correction threshold ($P < 1.60E-04$), with metabolites above this threshold considered Bonferroni-significant (***). Note that the y-axis presents different values as indicated. **B**: Box plots showing relative ion counts of five metabolites across the glycemic states. **C:** Summary of trajectory analysis across NORM, incident PreT2D, PreT2D, incident T2D, and T2D. Unknown metabolites are also shown. The analysis was conducted using linear mixed-effects models adjusted for age, sex, BMI, and lifestyle factors (physical activity, smoking status, and alcohol intake), and Bonferroni-significant metabolites are shown in plot C. **Abbreviations:** AHB, α-hydroxybutyrate; Val, valine; Gly, glycine; γ-Glu-Val, γ-glutamyl-valine.

Fig. 2

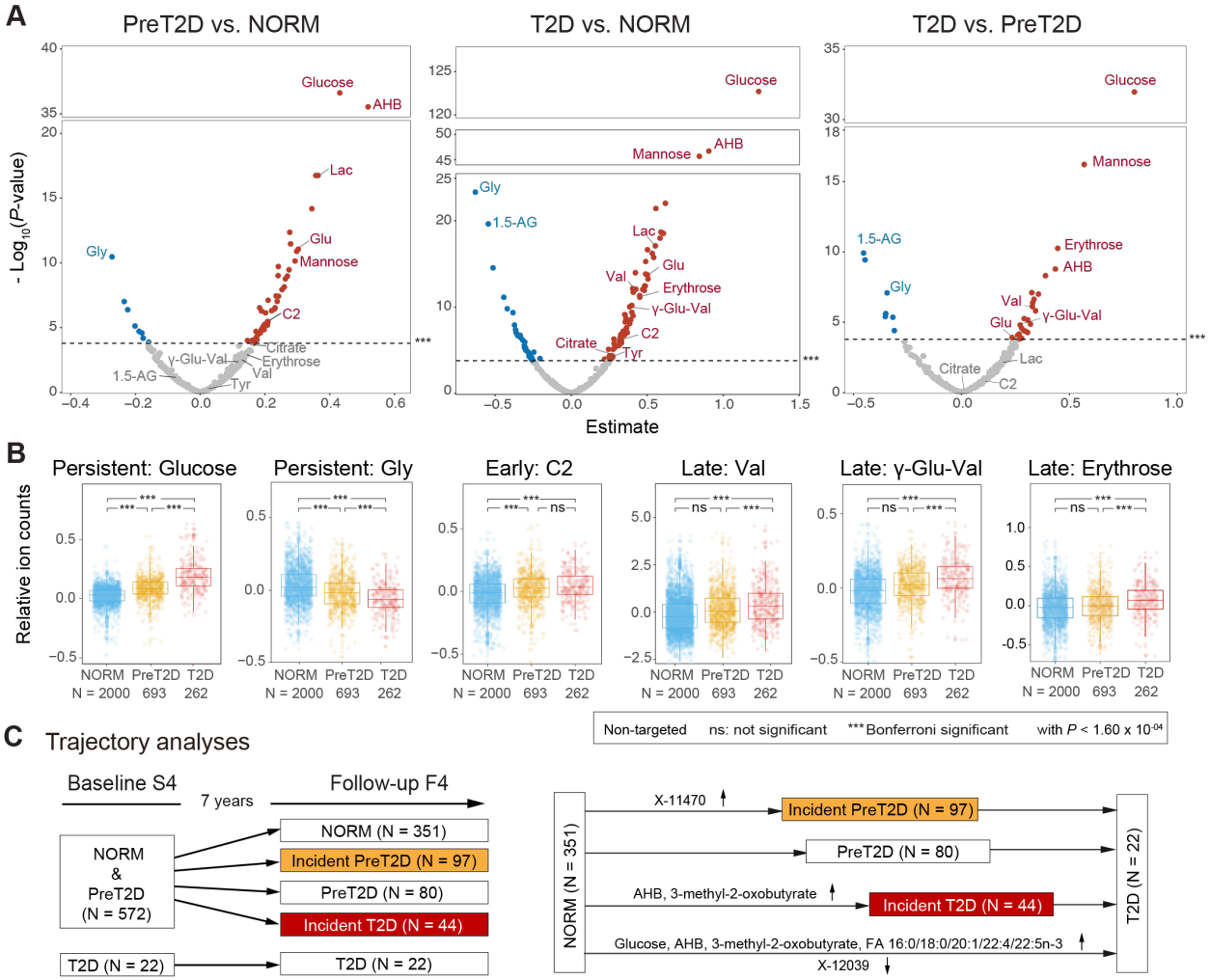

Figure 3. Metabolic pathways linking glycemic progression to T2D: insights from sensitivity, mediation, and causal analyses. Analyses of metabolic alterations across glycemic states and potential mechanistic pathways. **A:** Mediation analysis of γ-glutamyl-valine with fasting glucose, HbA1C, HDL cholesterol, and systolic BP as mediators in comparison between T2D and PreT2D. **B:** Mendelian randomization (MR) analysis of BCAA, C2, and γ-glutamyl-valine, with odds ratios and 95% CIs across MR methods. **C:** Proposed trajectory of metabolic changes across stages of T2D progression. **Abbreviations:** AAs, amino acids; ACME, average causal mediation effect; ADE, average direct effect; α-KG, alpha-ketoglutarate; BCAA, branched chain amino acid; LCFAs, long-chain fatty acids; BP, blood pressure; GGT, γ-glutamyl transferase; γ-Glu-Val, γ-Glutamyl-Valine; GSH, glutathione; HDL, high density lipoprotein; IVW, inverse variance weighted method; LDH, lactate dehydrogenase; LPC, lysophosphatidylcholine; NADH/NAD$^+$, nicotinamide adenine dinucleotide (reduced/oxidized form); PC, phosphatidylcholine; PPP, pentose phosphate pathway; SM, sphingomyelin; TCA, tricarboxylic acid cycle; T2D, type 2 diabetes.

Fig. 3