Received: 1 December 2024 | Revised: 27 February 2025 | Accepted: 8 March 2025

DOI: 10.1111/dom.16351

ORIGINAL ARTICLE

WILEY

Metabolomic network reveals novel biomarkers for type 2 diabetes mellitus in the UK Biobank study

Jiahao Liu BBS[1,2] | Xianwen Shang MPH[2,3] | Xueli Zhang[3] | Yutong Chen BSc[4] | Beiou Zhang[5] | Wentao Tang[6] | Li Li MD[1,2] | Ruiye Chen MD[1,2] | Catherine Jan BOptom[1,2] | Wenyi Hu MD[1,2] | Mayinuer Yusufu MTI[1,2] | Yujie Wang MD[1,2] | Zhuoting Zhu MD[2] | Mingguang He MD[2,7,8] | Lei Zhang PhD[9,10,11,12]

[1]Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Melbourne, Victoria, Australia

[2]Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, Melbourne, Victoria, Australia

[3]Department of Ophthalmology, Guangdong Eye Institute, Guangdong Provincial People's Hospital, Guang-dong Academy of Medical Sciences, Guangzhou, China

[4]Faculty of Medicine, Nursing and Health Science, Monash University, Clayton, Victoria, Australia

[5]Institute for Health and Sport, Victoria University, Melbourne, Victoria, Australia

[6]Faculty of Medicine and Health, University of Sydney, Sydney, New South Wales, Australia

[7]Ophthalmology, Department of Surgery, University of Melbourne, Melbourne, Victoria, Australia

[8]State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guang-Zhou, China

[9]Phase I Clinical Trial Research Ward, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi Province, China

[10]China-Australia Joint Research Center for Infectious Diseases, School of Public Health, Xi'an Jiaotong University Health Science Center, Xi'an, Shaanxi, China

[11]Artificial Intelligence and Modelling in Epidemiology Program, Melbourne Sexual Health Centre, Alfred Health, Melbourne, Victoria, Australia

[12]School of Translational Medicine, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Victoria, Australia

Correspondence
Lei Zhang, Level 10, 200 Victoria Parade, East Melbourne, VIC 3002, Australia.
Email: lei.zhang1@monash.edu

Funding information
Australian Government: the National Critical Research Infrastructure Initiative, Medical Research Future Fund, Grant/Award Number: MRFAI00035; NHMRC Investigator Gran, Grant/Award Number: APP1175405,2010072; Global STEM Professorship Scheme, Grant/Award Number: P0046113; Fundamental Research Funds of the State Key Laboratory of Ophthalmology, Project of Investigation on Health Status of Employees in Financial Industry in Guangzhou, China, Grant/Award Number: Z012014075; Victorian State Government; Melbourne Research Scholarship established by the University of Melbourne

Abstract

Aims: To identify hub metabolic biomarkers that constructively shape the type 2 diabetes mellitus (T2DM) risk network.

Materials and Methods: We analysed data from 98 831 UK Biobank participants, confirming T2DM diagnoses via medical records and International Classification of Diseases codes. Totally 168 circulating metabolites were quantified by nuclear magnetic resonance at baseline. Metabolome-wide association studies with Cox proportional hazards models were performed to identify statistically significant metabolites. Network analysis was applied to compute topological attributes (degree, betweenness, closeness and eigencentrality) and to detect small-world features (high clustering, short path lengths). Identified metabolites were used with XGBoost models to assess risk prediction performance.

Results: Over a median 12-year follow-up, 114 metabolites were significantly associated with T2DM risk and clustered into three distinct small-world modules. Total triglycerides and large high-density lipoprotein (HDL) cholesterol emerged as the

pivotal biomarkers in the 'risk' and 'protective' modules, respectively, as evidenced by their high eigencentrality. Moreover, total branched-chain amino acids (BCAAs) exhibited small-world network characteristics exclusively in pre-T2DM individuals, suggesting them as a potent early indicators. GlycA demonstrated high closeness centrality in females, implying a female-specific risk biomarker.

Conclusions: By constructing a metabolic network that captures the complex interrelationships among circulating metabolites, our study identified total triglycerides and large HDL cholesterol as central hubs in the T2DM risk metabolome network. BCAA and GlycA emerged as alarm indicators for pre-T2DM individuals and females, respectively. Network analysis not only elucidates the topological functional roles of biomarkers but also addresses the limitations of false positives and collinearity in single-metabolite studies, offering insights for metabolic pathway research and precision interventions.

KEYWORDS
metabolite, metabolome-wide association study, network analysis, type 2 diabetes mellitus

1 | INTRODUCTION

Type 2 diabetes mellitus (T2DM) is a global health crisis characterized by metabolic dysfunctions, including insulin resistance, dyslipidaemia, diminished pancreatic β-cell functionality and elevated oxidative stress, all contributing to chronic low-level inflammation.[1–3] According to the International Diabetes Federation (IDF), approximately 537 million adults aged 20–79 will be living with diabetes in 2021, with estimates forecasting a rise to 643 million by 2030 and 783 million by 2045.[4] Despite being a preventable condition, the complexity of the pathogenesis of T2DM continues to hinder the development of effective screening strategies.[5]

Circulating metabolites provide critical insights into metabolic alterations that precede T2DM onset.[6] These metabolites are intricately linked to T2DM through both direct and indirect metabolic pathways and processes.[1,7,8] Branched-chain and aromatic amino acids have been identified as early indicators of insulin resistance,[9] yet focusing on individual metabolites risks oversimplifying the intricate metabolic interactions underlying T2DM pathogenesis. A comprehensive understanding of the metabolic network remains elusive.

Network medicine has emerged as a powerful tool, integrating systems biology and network science to unravel the complex interactions underlying disease pathogenesis.[10] By mapping networks where nodes represent biomarkers (e.g., genes, metabolites) and edges denote relationships, this approach captures both direct and indirect influences on disease states.[10] Furthermore, when mapping disease networks, both disease-essential and non-essential networks may be included. Identifying central 'hubs' can reveal valuable diagnostic or therapeutic biomarkers.[11] Previous research in cardiovascular disease has leveraged untargeted metabolomics data on network analysis to reveal disease-associated molecular pathways that uncover hidden components of signalling and regulatory networks.[12] Similarly, genetic expression network analysis in glioblastoma research unveiled the critical role of gene ASPM in tumour cell division.[13] Given the complex aetiology of T2DM, metabolic network analysis offers a promising approach to identifying disease biomarkers, therapeutic targets and understanding the intricate pathways involved in disease development.

Leveraging the large-scale populational cohort UK Biobank data, which includes health information from over 500 000 participants followed up for more than 12 years.[14] This study aims to employ advanced network analysis techniques to identify key circulating biomarkers for T2DM risk. By deciphering the interrelationships between key metabolites and providing a holistic view of the metabolic network involved in T2DM, ultimately contributing to the prevention strategy and therapeutic targets of T2DM.

2 | MATERIALS AND METHODS

This study utilized data from the UK Biobank cohort study, which recruited 502 534 individuals aged 40–69 years in the United Kingdom between 2006 and 2010.[14] Demographic information and physical measurements were collected by touch-screen self-report questionnaires, online computer-assisted assessments or by attending one of the 22 assessment centres in England, Scotland and Wales.[14] The selection criteria are illustrated in Figure S1: (1) participants with completed nuclear magnetic resonance (NMR) metabolomic data; (2) participants with completed baseline and follow-up T2DM status records; (3) participants diagnosed as T2DM cases at baseline; (4) participants with completed baseline glycated haemoglobin (HbA1c) records; and (5) participants with HbA1c levels ≥48 mmol/mol were excluded to minimize the inclusion of undiagnosed T2DM cases. A diagnostic threshold was established by the 2021 guidelines of the American Diabetes Association.[15] The remaining participants comprised the overall population used for subsequent analyses.

2.1 | Diagnosis of T2DM

Both baseline and incident diagnostic criteria in our study were ascertained by examining participants' hospital inpatient records and death registers and employing the International Classification of Diseases (ICD) coding system. For the ICD-9 classification, codes 25 001, 25 003, 25 011, 25 013, 25 021, 25 023, 25 031, 25 033, 25 041, 25 043, 25 051, 25 053, 25 061, 25 063, 25 071, 25 073, 25 081, 25 083, 25 091 and 25 093 were incorporated. Correspondingly, for the ICD-10 system, we utilized codes E10, E11, E13 and E14.[16]

2.2 | Circulating metabolites

The sample collection and metabolomic quantification methods employed in the study are elaborately documented.[17] Briefly, plasma samples treated with ethylenediaminetetraacetic acid (EDTA) were collected in a non-fasting state from a randomized cohort during the initial recruitment phase of the UK Biobank. These samples were transported under cryogenic conditions to the Nightingale Health laboratories in Finland for biochemical analysis.[14] A total of 168 directly measured metabolites were included in our analysis (Supplementary methods—Circulating metabolites and Table S1).[14]

2.3 | Metabolome-wide association study

We used five multiple imputations with a random forest algorithm, conducting 50 iterations separately for each subgroup to address missing data.[18] All imputed variables were grouped into quintiles. To maintain an unbiased evaluation, the dataset was randomly divided into training and testing subsets with a 1:1 ratio.

Pseudo-numerical multivariate Cox regression was then used to assess associations between metabolite profiles and incident T2DM. Model 1 accounted for the primary demographic factors of baseline age and gender, aiming to capture essential variations without over-adjustment. Model 2 extended these adjustments to include ethnicity, socioeconomic variables (education level, assessment centre location, Townsend index), lifestyle factors (smoking status, physical activity) and albumin level. Model 3 incorporated additional clinical covariates such as body mass index (BMI), HbA1c, blood pressure medication, exogenous hormone use and a history of hypertension (Supplementary methods—Covariates), providing a comprehensive assessment that controls for established T2DM risk determinants.

To mitigate the risk of multiple testing, a Bonferroni correction (significance threshold <0.05) was applied in Model 1. Only metabolites that remained significantly associated ($p < 0.05$) across all three models, in both the training and testing datasets, as well as the combined dataset, were deemed robust findings. These results are presented in a forest plot (Figure S3). For stratification analysis, the overall population was classified based on pre-T2DM status (<39; ≥39 and <48 mmol/mL), gender and family history of T2DM (Figures S2 and S3).

2.4 | Network analysis

In network analysis, nodes symbolize significant metabolites, and edges denote the correlations among nodes based on Spearman's correlation analysis.[19] Key topological attributes—degree, betweenness, closeness, eigencentrality, transitivity and modularity—were quantified to elucidate the structural and functional intricacies of the metabolic network (Supplementary methods—Topological attributes).[20]

Non-directional networks were generated for various metabolite characteristics, including hazard ratio (HR), metabolite categories and subcategories. Network visualizations were created using the ggraph package in R, employing the force-directed Fruchterman–Reingold algorithm.[21] This approach treated the included nodes as repelling particles, with edges acting as springs that pull connected nodes closer. The intensity of the node, node colours and edge colours represented the metabolites characteristics and correlation strength, respectively.

Modules within these networks were defined by clustering analysis based on their modularity, facilitating the understanding of network organization and functional grouping. Distinct colours differentiated modules in the network graphs, while scatter plots explored the relationships between HR and topological features across modules.

2.5 | Small-world network

The small-world network property, defined by high clustering and low path lengths, was assessed by calculating the clustering coefficient, which quantifies the tendency of nodes to form clusters.[10] The average path length, representing the mean number of steps along the shortest paths between all node pairs, was also evaluated to gauge network navigability.

Uniform analytic methodologies, including visualization tools, layout algorithms and topological property computations, were employed across all identified small-world networks. Heatmaps and Venn diagrams were used to contrast the topological characteristics across subpopulations.

2.6 | Machine learning prediction of T2DM incidence

To evaluate prediction performance and the feature importance, we applied machine learning models (Extreme Gradient Booting [XGBoost]) with hyperparameters to assess the area under the curve (AUC).[22] Two models were developed: Model 1 utilized metabolites identified within small-world networks as predictive features. Model 2 additionally included age at baseline, gender, family history, BMI and hypertension status. Prediction performance was visualized using receiver operating characteristic (ROC) curves.[22] The contribution of each metabolite and covariate to T2DM risk was further analysed and visualized through histograms.

All statistical analyses were operated on by RStudio Desktop 4.3.0.

2.7 | Ethics

The ethical approval for the UK Biobank Study was granted by the National Information Governance Board for Health and Social Care and the NHS North West Multicentre Research Ethics Committee. Signing the consent form is mandatory for participants at recruitment. The establishment of this study was authorized under application number 62443 of the UK Biobank resource.

3 | RESULTS

3.1 | Overall population and T2DM incidence

Of 502 534 participants, 98 831 subjects were screened for the overall population analysis. A median follow-up of 12 years in 95 363 non-T2DM individuals at baseline and 3468 subsequent T2DM onset revealed a T2DM incidence rate of 2.97 per 1000 person-years. Significant differences ($p < 0.001$) were observed between those who developed T2DM and those who did not, including gender distribution, age at baseline and socioeconomic status measured by the Townsend Index. Males and older individuals were more likely to develop T2DM. Additionally, lower physical activity levels and higher BMI were associated with increased T2DM risk. Clinical measures, including HbA1c and cholesterol levels, further distinguished individuals with T2DM onset from those without (Table S2, all $p < 0.001$).

3.2 | Metabolome-wide association study

Our metabolome-wide association study (MWAS) analysis identified 114 of 168 evaluated metabolites that were significantly associated with T2DM risk in the overall population (Figure 1, Table S4). Of these, 58 were identified as risk factors, such as S-HDL-Triglycerides (Node 168, HR = 1.20, 1.17–1.24, $p = 2.0 \times 10^{-39}$), XXL-VLDL-Phospholipids (Node 73, HR = 1.19, 1.16–1.22, $p = 3.8 \times 10^{-35}$) and the concentration of XXL-VLDL particles (Node 71, HR = 1.18, 1.15–1.22, $p = 5.1 \times 10^{-35}$). The remaining 56 metabolites functioned as protective factors, such as the degree of unsaturation (Node 42, HR = 0.83, 0.80–0.85, $p = 7.2 \times 10^{-43}$), L-HDL-cholesteryl esters (Node 152, HR = 0.83, 0.81–0.86, $p = 3.0 \times 10^{-31}$) and L-HDL-Cholesterol (Node 151, HR = 0.84, 0.81–0.86, $p = 1.1 \times 10^{-28}$) (Figure S3).

3.3 | Network analysis of significant metabolites

Visualizing the metabolome network with the Fruchterman–Reingold force-directed layout revealed four distinct modules (Figure 1 and Figure S5). The network analysis of these significant metabolites revealed a transitivity of 0.85 and a modularity of 0.46 (Table S5), indicating a tightly interconnected yet modularly diverse metabolic network. Clustering analysis identified three functional modules (Figure 2 and Figure S8). Module 1 (M1) encompassed all identified risk factors, representing metabolites associated with elevated T2DM risk. Module 2 (M2) comprised metabolites with moderate protective effects, while Module 3 (M3) included all protective factors. Module 4 (M4) consists exclusively of amino acids, yet it did not meet the criteria for classification as a small-world network.

Metabolites in M1 exhibited a higher degree centrality compared with those in M2 and M3, including VLDL-lipoprotein (Node 29, degree = 74), total lipids in M-VLDL (Node 93, degree = 73) and total fatty acids (Node 41, degree = 71), all integral to M1 (Figures S6 and S7). This indicates that these metabolites are extensively connected to other nodes and may therefore exert a broad influence on metabolic interactions. Sphingomyelins (Node 38, betweenness = 1304) exhibited the highest betweenness centrality in the overall population network, forming extensive connections within Module 2 (M2) and serving as a bridge between M2 and M3. As a 'connector' node, it facilitates critical pathways across different metabolic clusters. This was followed by saturated fatty acids (Node 47, betweenness = 805) in M1 and the total concentration of lipoprotein particles (Node 28, betweenness = 445) in M3. Metabolites with high closeness centrality included amino acids such as valine (Node 57), tyrosine (Node 59) and total branched-chain amino acids (BCAAs) (Node 54), indicating their tight interconnections and swift information relay capacity within the network.

3.4 | Network analysis of modules with small-world characteristic

In M1, which consists entirely of risk-factor metabolites (Figure 1 and Figures S6 and S8), various triglycerides and VLDL components of different sizes were identified (Figure 2 and Figure S8). Notably, total triglycerides (Node 8) occupy a central position within M1, as demonstrated by their prioritized eigencentrality yet low closeness. Nodes with high eigencentrality function as central hubs, exerting widespread influence despite not being the quickest in connecting with others. Conversely, GlycA (Node 70), XL-HDL-Triglycerides (Node 147), M-HDL-TG (161) and saturated fatty acids (47) are positioned peripherally in M1, with a high closeness but low eigencentrality, indicating that although they can rapidly interact with other metabolites, they are less integral to the network's overall structure.

In M2, most LDL and IDL components were included. L-LDL-phospholipid (Node 122) consistently exhibited the prioritized eigencentrality but the lowest closeness centrality, while sphingomyelins (Node 38) showed the opposite pattern, with high closeness but the lowest eigencentrality within M2.

In M3, which is composed solely of protective-factor metabolites, HDL components predominated. L-HDL Cholesterols (Node 153) and the concentration of L-HDL (Node 148) demonstrated the highest eigencentrality across all populations.

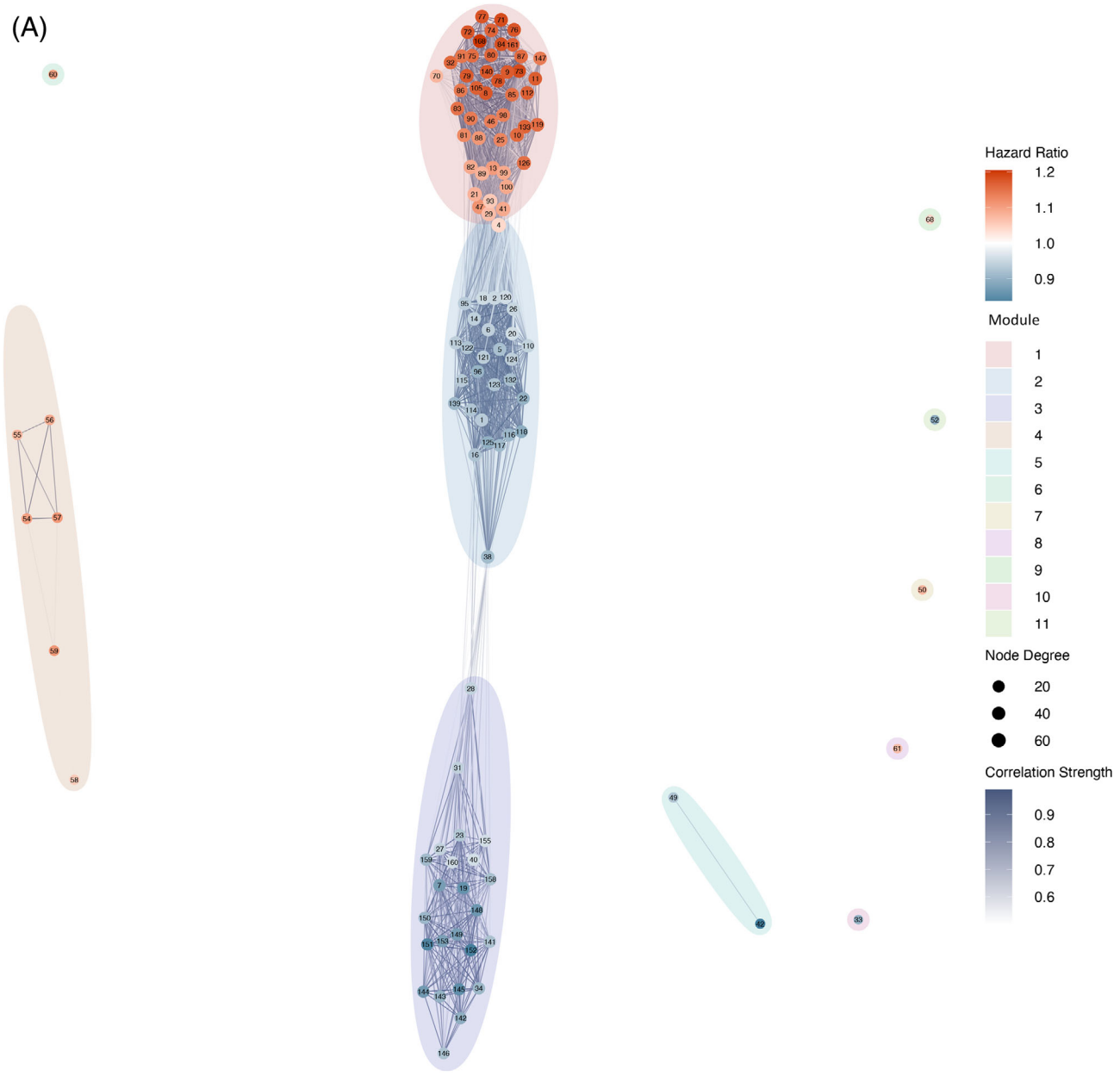

FIGURE 1 Metabolome network of the overall population. The non-directional metabolome networks were generated based on metabolites that are significantly associated with type 2 diabetes mellitus risk. Nodes represent significant metabolites, node size corresponds to degree (number of correlated connections) and edges signify the correlation strength between them (darker edges indicate stronger correlations). Non-directional networks were generated for three main metabolite characteristics: (A) hazard ratio (HR) (darker nodes represent higher absolute HR), (B) metabolite categories and (C) subcategories. The layout of the networks was constructed based on Fruchterman-Reingold force-directed layout. Distinct colours identify different modules, determined by clustering analysis to maximize modularity.

3.5 | Machine learning prediction of T2DM incidence

To evaluate prognostic precision, the metabolite-only model achieved an AUC of 0.73 (Figure 3), which improved to 0.79 with the addition of demographic and clinical parameters. Notably, M-VLDL-Cholesteryl esters (Node 96, M2), S-HDL-triglycerides (Node 168, Module 1), glycoprotein acetyls (Node 70, M1), the L-HDL-Cholesteryl ester (Node 152, M3) and XL-HDL-free cholesterols (Node 146, M3) were consistently assigned high importance, highlighting their potential as biomarkers in metabolic profiling.

3.6 | Stratification analysis

Since only females reported using exogenous hormones, a gender-specific analysis became essential (Table S3). In females, Module 3 (M3) appeared disconnected from M1 and M2, yet M1 and M2

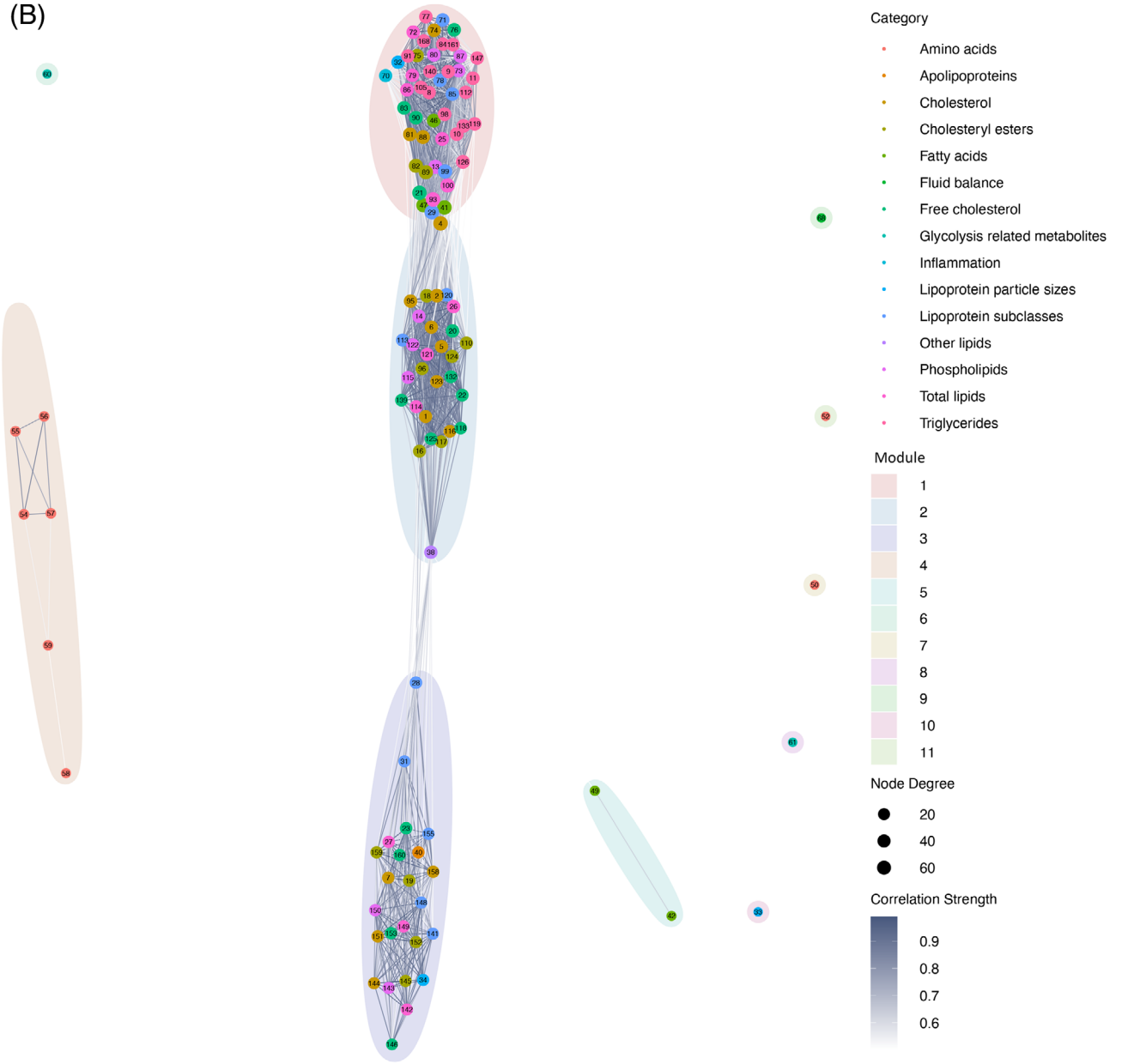

FIGURE 1 (Continued)

exhibited a closer interplay (Figures S5g–i). This observation suggests underlying, gender-specific metabolic interactions that may influence T2DM risk. Of particular note, GlycA (Node 70) was a significant risk metabolite only in females (Figure S3) and demonstrated the highest closeness but low eigencentrality in M1, implying a capacity for rapid interaction with other risk-factor metabolites but less central to the network structure. The female-specific predictive model achieved an AUC of 0.76 using only metabolites, rising to 0.80 with additional covariates (Figure S10f), highlighting its robust performance in gender-tailored T2DM risk stratification. Furthermore, GlycA was the top predictor in Model 1 and remained highly ranked in Model 2 (Figures S11g–h), reinforcing its importance as a female-specific biomarker for T2DM.

The population with pre-T2DM exhibited a ninefold higher incidence rate of T2DM compared with the baseline healthy population (13.18 vs. 1.46 per 1000 person-years) (Table S3). Notably, metabolites in Module 4 (M4) exhibited small-world characteristics exclusively in the pre-T2DM population (Figures S5 and S8), including total BCAAs, leucine, valine and tyrosine (Nodes 56, 57, 54 and 59, respectively). Three of these four metabolites are BCAAs, which show high closeness in the complete metabolome network (Figure S6c) and elevated eigencentrality in the small-world network M4. Their high closeness implies an ability to rapidly interact with a broad range of other metabolites, while their high eigencentrality underscores their structural significance within the module, suggesting that BCAAs may be

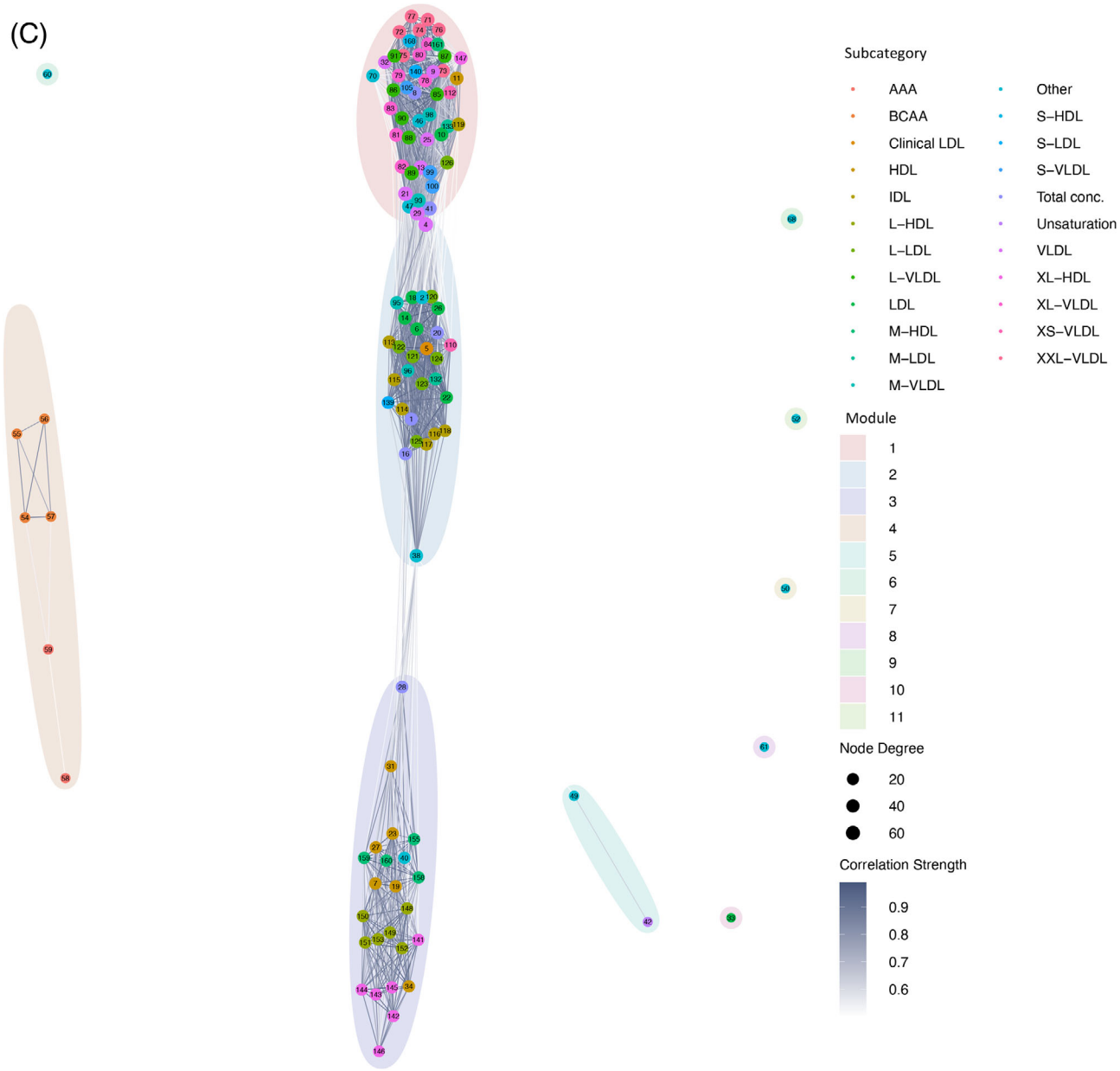

FIGURE 1 (Continued)

early metabolic signals of disease progression in pre-T2DM populations. Moreover, protective metabolites in Module 3 (M3) were fewer in healthy individuals compared with the pre-T2DM population (14 vs. 20; Table S4), suggesting that those with pre-T2DM may mount a more active defensive response to slow T2DM development.

The incidence rate was approximately doubled in populations with a family history of T2DM compared with those without a family history (5.02 vs. 2.46 per 1000 person-years) (Table S3). Interestingly, individuals without a family history had a greater number of T2DM-associated metabolites (99 vs. 74) and more connections between metabolites in M1 and M2 (Table S4). Conversely, the population with a T2DM family history exhibited significantly lower network modularity (Table S5).

4 | DISCUSSION

The integration of large populational cohort metabolomic profiling and network analysis provided a holistic view of the metabolome network of T2DM risk. Analysis revealed 114 metabolites significantly associated with T2DM risk. Clustering analysis identified three main modules within this network, with M1 encompassing all identified risk factors and M3 including metabolites with a strong protective effect on T2DM risk. Total triglycerides in M1 and L-HDL-Cholesterol in M3 emerged as the core prevention disease biomarkers of T2DM risk. Total BCAA was identified as a biomarker for elevating HbA1c, and GlycA emerged as a female-specific T2DM risk signalling biomarker. The stronger connections between risk and protective factors in the

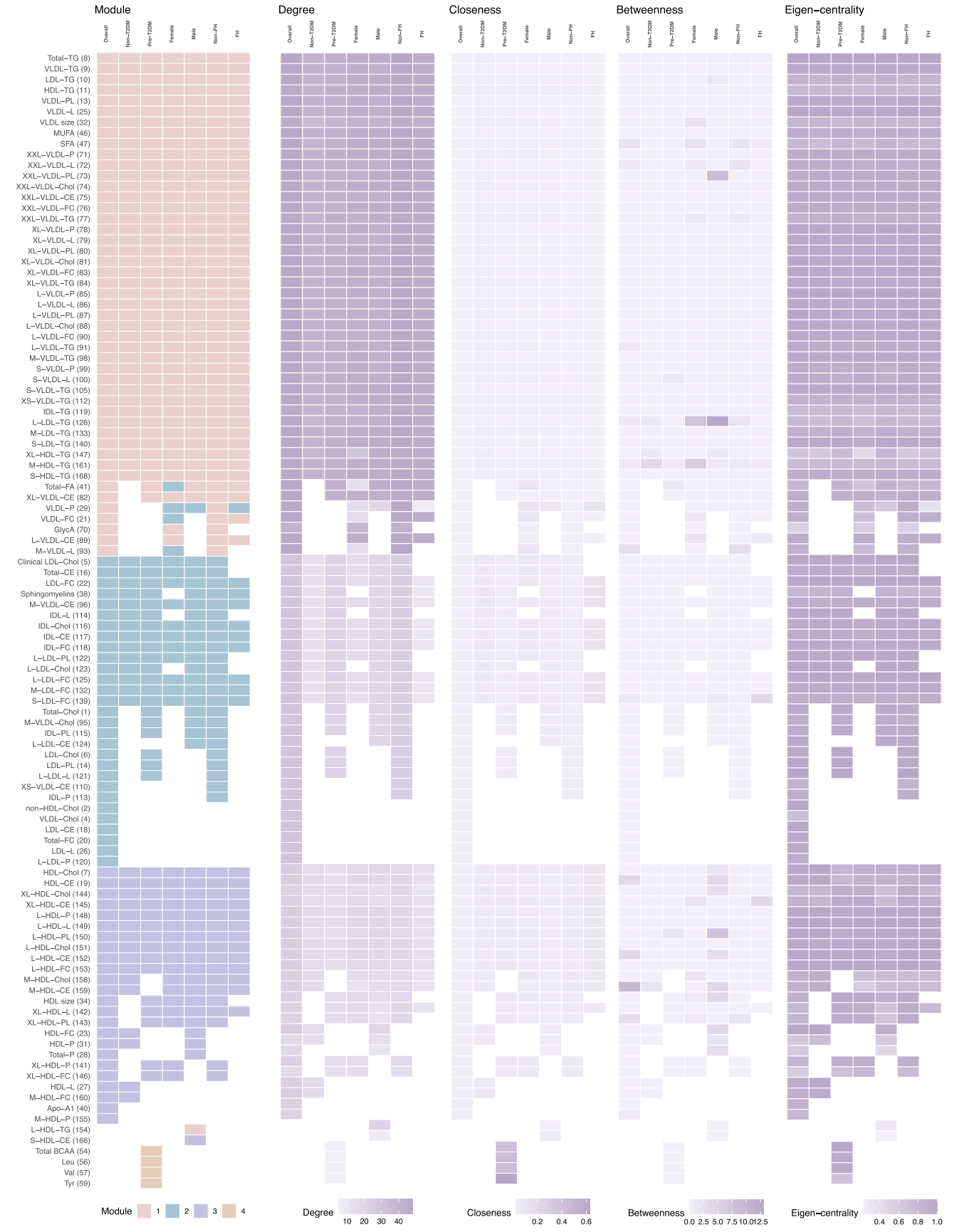

FIGURE 2 Legend on next page.

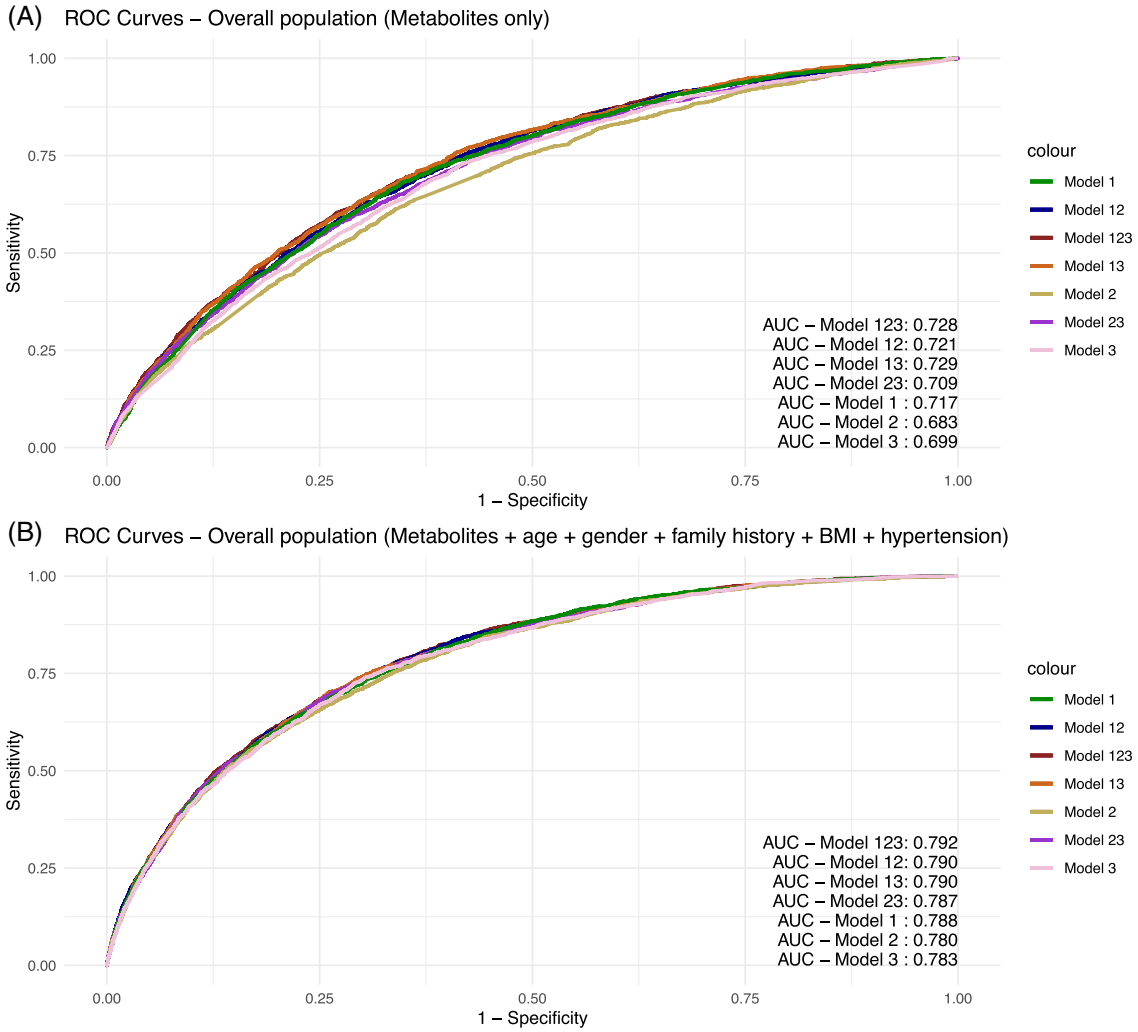

FIGURE 3 This figure shows receiver operating characteristic (ROC) curves and corresponding area under the curve (AUC) values for type 2 diabetes mellitus (T2DM) prediction models in the overall population. Both panels plot the true positive rate (sensitivity) on the *y*-axis and the false positive rate (1—specificity) on the *x*-axis, ranging from 0.0 to 1.0, to illustrate the performance of various T2DM prediction models. Panel A presents the prediction performance of models derived solely from metabolites clustered in distinct small-world modules (Models 1, 2, 3 or their combinations). Model numbers indicate the specific modules included (for example, Model 123 integrates metabolites from Modules 1, 2 and 3). Panel B displays the predictive values after integrates additional key covariates, including baseline age, gender, T2DM family history, body mass index and hypertension status.

female population underscored gender-specific differences in metabolic signalling. The pivotal metric of eigencentrality further emphasized the central role of these metabolites in shaping the network's topology and their potential as critical therapeutic targets for T2DM.

Total triglyceride levels increased T2DM risk by 18% in the overall population and were identified as the central hub within the 'risk' module, underscoring their potential as a primary biomarker for T2DM risk. This finding aligns with previous studies that associate elevated triglyceride levels with increased T2DM risk.[23] A longitudinal study established that every 10 mg/dL increase in triglyceride level increased the T2DM risk by 4% and impaired fasting glucose by 2%.[23] Evidence also suggests that traditional biomarkers such as glucose and HbA1c levels are influenced by triglycerides, highlighting triglyceride modulation as a potential intervention point. In M1, we

FIGURE 2 Heatmap of four network metrics (degree, closeness, betweenness and eigencentrality) across modules with small-world characteristics in various population subsets. Each row corresponds to a metabolite, and numbers in parentheses alongside the metabolite names indicate their respective network codes. The belonging module number is colour-coded on the left. For degree, closeness, betweenness and eigencentrality values, darker shading in the heatmap denotes higher metric values. This visualization highlights how different metabolites shape module structure and underscores their relative importance within the network. FH, family history; T2DM, type 2 diabetes mellitus.

identified various triglycerides and VLDL components of different sizes. This observation is consistent with the mechanistic link between elevated triglyceride levels and T2DM risk, which is rooted in obesity-associated metabolic disturbances that promote insulin resistance via inflammation and oxidative stress pathways.[24] Triglyceride accumulation impairs insulin signalling, reducing glucose uptake by cells and consequently elevating blood glucose levels.[24] Additionally, lipoprotein lipase, critical for triglyceride metabolism, often exhibits reduced activity in insulin-deficient states, further contributing to elevated triglycerides.[25] The importance of BMI in our prediction models underscores its relationship with triglycerides. Recent studies have introduced the triglyceride-glucose index and the triglyceride-HDL cholesterol ratio as surrogate biomarkers for assessing insulin resistance and predicting T2DM risk,[26,27] reinforcing the essential role of triglycerides in early prevention strategies for T2DM.

Within the metabolic landscape of M3, L-HDL-Cholesterol emerges with pronounced eigencentrality, identifying it as a key component in T2DM risk prevention. Its strong correlation with L-HDL-Cholesteryl ester (correlation coefficient = 0.99) further accentuates its critical role in lipid metabolism and T2DM risk prediction. Additionally, M3 comprises various types of HDL components. HDL plays a central role in lipid transport; cholesterol esterification within HDL generates cholesteryl ester transfer protein, which facilitates the exchange of esterified cholesterol and triglycerides between HDLs and LDLs.[28] HDL is ultimately reclaimed by the liver, where triglyceride-enriched HDL is degraded by hepatic lipase to clear excess triglycerides. Diabetic dyslipidaemia disrupts cholesterol efflux, a hallmark of the metabolic dysregulation seen in T2DM.[29] This reverse cholesterol transportation system is endowed with vascular protective effects, including anti-inflammatory, anti-oxidative and anti-thrombotic actions, embodying the multifaceted benefits beyond lipid transport.[30] Evidence supports the protective role of L-HDL-Cholesterol against T2DM.[31] A longitudinal Finnish population study of individuals aged 24–45 revealed an inverse relationship between L-HDL-Cholesterol and T2DM risk.[31] Additionally, a lower Triglyceride/HDL-Cholesterol ratio has been linked to a reduced risk of insulin resistance, highlighting the integral role of L-HDL-Cholesterol levels in T2DM risk prevention.[32] These findings collectively underscore the significance of L-HDL-Cholesterol in the metabolic network's architecture. The evidence advocates for a deeper exploration of HDL's role in metabolic regulation, offering insights into potential therapeutic interventions.

Our analysis identified metabolites with higher closeness centrality as crucial T2DM risk biomarkers in prediction models. Specifically, valine, tyrosine, leucine, and total BCAA may serve as alarm indicators of unhealthy HbA1c. These metabolites, which exhibit a small-world network structure unique to pre-T2DM populations and high closeness centrality, were prioritized in importance analyses. This observation aligns with studies linking elevated BCAAs and tyrosine levels to both increased T2DM risk and HbA1c levels, underscoring their utility as early markers for deteriorating glycaemic control.[33] The role of BCAAs in insulin resistance, mediated through mTOR signalling pathway activation, further exacerbates glucose homeostasis dysregulation.[9] A prior longitudinal study demonstrated that individuals with elevated BCAAs and tyrosine levels co-currently were more likely to progress from normoglycaemia to impaired glucose tolerance or T2DM,[34] signifying the pivotal role of these M4 metabolic indicators in the early detection and management of pre-T2DM.

Intriguingly, in female populations, M3 does not connect to M1 and M2 modules, indicating a disconnect between protective and risk factors. This phenomenon might stem from the relatively weaker protective factors within M2 and M3, suggesting that risk factors in M1 do not directly impact those in M2 and M3 but may instead activate the inflammatory system. This is corroborated by the significant association between the inflammatory biomarker GlycA and T2DM in females, not in males. The node is within M1 with a high prioritized importance in the females' prediction model and the AUC is higher than that in males. Elevated triglyceride levels have been implicated in the induction of systemic inflammation, serving as a mechanistic bridge to increased GlycA levels.[35] Triglycerides can exacerbate inflammation through various pathways, including the activation of pro-inflammatory cytokines and the enhancement of adipose tissue inflammation.[36] Elevated GlycA levels reflect an increase in the glycosylation of acute-phase inflammatory proteins in serum, which are sensitive markers of systemic inflammation.[35,37] Prior studies have delineated a robust correlation between GlycA levels, the insulin sensitivity index and C-reactive protein, a marker of chronic subclinical inflammation intimately linked with insulin resistance.[37] Given these observations, it becomes evident that understanding the underlying distinct metabolic and inflammatory profiles in females necessitates further research to unravel the gender-specific mechanisms.

The present study provides a topological perspective on the interconnected metabolites involved in the development of T2DM, presenting a comprehensive approach that transcends traditional methods typically focused on individual biomarkers. The study's strength lies in its extensive sample size and prolonged follow-up period, ensuring robust results. By incorporating MWAS with the Bonferroni correction method, we have significantly reduced the probability of false positives, bolstering the reliability of our findings. Additionally, our application of detailed stratification analyses across various population segments has yielded customized insights. While conventional tests for insulin resistance, such as fasting blood glucose levels, are indispensable in diabetes management, they are prone to instability and fluctuations. The HbA1c test, commonly used for diagnosing pre-T2DM, often implies that some degree of physiological alteration has already occurred. Our study endeavours to bridge the gap in early T2DM prevention, aiming to identify both treatment and signalling biomarkers before more definitive signs of disease emerge.

However, several limitations warrant consideration. First, the use of non-fasting plasma samples may introduce misclassification bias. Postprandial increases in serum metabolite concentrations vary among metabolites; for instance, BCAAs typically peak 1–2 h after a meal,[38] whereas triglyceride levels may peak 3–5 h postprandially.[39] Consequently, measurements taken shortly after meals could bias the associations, potentially overestimating the risk. Second, due to

limited baseline data on medication use and comorbidities that affect glycaemia and lipid levels, we were unable to exclude these participants upfront. This gap may have biased our associations. We partially addressed this limitation by adjusting for relevant covariates in our Model 3 s. Nonetheless, future studies with more comprehensive data could better clarify how such factors modify T2DM risk. Third, the external validity of our findings may be constrained, as all the subjects are only representative of the UK population. The metabolic profiling conducted via the NMR metabolomics platform, while comprehensive, covers only a limited array of metabolic pathways. This limitation suggests that future investigations employing more sensitive metabolomics techniques, such as mass spectrometry, could enhance metabolome coverage and unearth insights into a broader spectrum of biological processes. Furthermore, the inherent collinearity among metabolites in relation to T2DM risk represents an unavoidable challenge, complicating the disentanglement of individual metabolites' effects on disease risk. These limitations underscore the need for cautious interpretation of the results and highlight the potential avenues for further research to refine our understanding of T2DM's metabolic underpinnings.

5 | CONCLUSION

This study demonstrates that total triglycerides and L-HDL-Cholesterol constitute pivotal biomarkers for the prevention of T2DM risk. Concurrently, total BCAA has been identified as markers indicative of elevated HbA1c levels, whereas GlycA emerges as a distinct signalling biomarker of T2DM risk, specifically within female cohorts.

AUTHOR CONTRIBUTIONS

Conceived the study: J.L., X.S., M.H. and L.Z. *Conceptualization*: X.S., M.H. and L.Z. *Methodology*: J.L., X.Z., X.S. and L.Z. *Software*: J.L., X.Z. and Y.C. *Formal analysis*: J.L., X.Z., M.Y. *Investigation*: J.L., B.Z., W.T. and Y.C. *Writing original draft*: J.L. *Writing review and editing*: X.S., X.Z., B.Z., W.T., C.J., W.H., R.C., L.L., Y.W., Z.Z., M.H. and L.Z. *Supervision*: X.S., M.H. and L.Z. *Project administration*: M.H., Z.Z. and L.Z. All authors have read and agreed to the published version of the manuscript.

ACKNOWLEDGEMENTS

This research was conducted using the UK Biobank resource. We thank the participants of the UK Biobank. Open access publishing facilitated by Monash University, as part of the Wiley - Monash University agreement via the Council of Australian University Librarians.

FUNDING INFORMATION

This project received grant funding from the Australian Government: the National Critical Research Infrastructure Initiative, Medical Research Future Fund (MRFAI00035) and the NHMRC Investigator Grant (APP1175405, 2010072). The contents of the published material are solely the responsibility of the Administering Institution, a participating institution or individual authors and do not reflect the views of the NHMRC. This work was supported by the Global STEM Professorship Scheme (P0046113), the Fundamental Research Funds of the State Key Laboratory of Ophthalmology and Project of Investigation on Health Status of Employees in Financial Industry in Guangzhou, China (Z012014075). The Centre for Eye Research Australia receives Operational Infrastructure Support from the Victorian State Government. J.L. is supported by the Melbourne Research Scholarship established by the University of Melbourne. The funding source had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Restrictions apply to the availability of these data. Data were obtained from UK Biobank and are available at https://www.ukbiobank.ac.uk (accessed on 17 November 2022) with the permission of UK Biobank.

ORCID

Jiahao Liu https://orcid.org/0000-0002-9979-5805
Zhuoting Zhu https://orcid.org/0000-0002-9897-1192
Lei Zhang https://orcid.org/0000-0003-2343-084X

REFERENCES

1. Galicia-Garcia U, Benito-Vicente A, Jebari S, et al. Pathophysiology of type 2 diabetes mellitus. *Int J Mol Sci*. 2020;21:1-9.
2. Abdul-Ghani M, Maffei P, DeFronzo RA. Managing insulin resistance: the forgotten pathophysiological component of type 2 diabetes. *Lancet Diabetes Endocrinol*. 2024;12:674-680.
3. Ahmad E, Lim S, Lamptey R, Webb DR, Davies MJ. Type 2 diabetes. *Lancet*. 2022;400:1803-1820.
4. Sun H, Saeedi P, Karuranga S, et al. IDF diabetes atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. *Diabetes Res Clin Pract*. 2022;183:109119.
5. Zhu L, Huang Q, Li X, et al. Serological phenotyping analysis uncovers a unique metabolomic pattern associated with early onset of type 2 diabetes mellitus. *Front Mol Biosci*. 2022;9:841209.
6. Morze J, Wittenbecher C, Schwingshackl L, et al. Metabolomics and type 2 diabetes risk: an updated systematic review and meta-analysis of prospective cohort studies. *Diabetes Care*. 2022;45:1013-1024.
7. Yang Q, Vijayakumar A, Kahn BB. Metabolites as regulators of insulin sensitivity and metabolism. *Nat Rev Mol Cell Biol*. 2018;19:654-672.
8. Mahjoub S, Masrour-Roudsari J. Role of oxidative stress in pathogenesis of metabolic syndrome. *Caspian J Intern Med*. 2012;3:386-396.
9. Vanweert F, Schrauwen P, Phielix E. Role of branched-chain amino acid metabolism in the pathogenesis of obesity and type 2 diabetes-related metabolic disturbances BCAA metabolism in type 2 diabetes. *Nutr Diabetes*. 2022;12:35.
10. Barabási A-L, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet*. 2011;12:56-68.
11. Loscalzo J, Barabasi AL. Systems biology and the future of medicine. *Wiley Interdiscip Rev Syst Biol Med*. 2011;3:619-627.
12. Pirhaji L, Milani P, Leidl M, et al. Revealing disease-associated pathways by network integration of untargeted metabolomics. *Nat Methods*. 2016;13:770-776.

13. Horvath S, Zhang B, Carlson M, et al. Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. *Proc Natl Acad Sci*. 2006;103:17402-17407.
14. UK Biobank. *UK Biobank: Protocol for a Large-Scale Prospective Epidemiological Resource*; 2021. UK Biobank.
15. Association AD. Standards of medical care in diabetes—2021 abridged for primary care providers. *Clin Diabetes*. 2021;39:14-43.
16. Eastwood SV, Mathur R, Atkinson M, et al. Algorithms for the capture and adjudication of prevalent and incident diabetes in UK Biobank. *PLoS One*. 2016;11:e0162388.
17. Soininen P, Kangas AJ, Würtz P, et al. High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism. *Analyst*. 2009;134:1781-1785.
18. Van Buuren S, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. *J Stat Softw*. 2011;45:1-67.
19. Goldberg RB, Bray GA, Marcovina SM, et al. Non-traditional biomarkers and incident diabetes in the diabetes prevention program: comparative effects of lifestyle and metformin interventions. *Diabetologia*. 2019;62:58-69.
20. Lin W, Wang M, Chen M, et al. Metabolomics and correlation network analyses of core biomarkers in type 2 diabetes. *Amino Acids*. 2020;52:1307-1317.
21. Fruchterman TM, Reingold EM. Graph drawing by force-directed placement. *Softw Pract Exp*. 1991;21:1129-1164.
22. Buergel T, Steinfeldt J, Ruyoga G, et al. Metabolomic profiles predict individual multidisease outcomes. *Nat Med*. 2022;28:2309-2320.
23. Beshara A, Cohen E, Goldberg E, Lilos P, Garty M, Krause I. Triglyceride levels and risk of type 2 diabetes mellitus: a longitudinal large study. *J Invest Med*. 2016;64:383-387.
24. Li M, Chi X, Wang Y, Setrerrahmane S, Xie W, Xu H. Trends in insulin resistance: insights into mechanisms and therapeutic strategy. *Signal Transduct Target Ther*. 2022;7:216.
25. Boren J, Taskinen M-R, Björnson E, Packard CJ. Metabolism of triglyceride-rich lipoproteins in health and dyslipidaemia. *Nat Rev Cardiol*. 2022;19:577-592.
26. Selvi NMK, Nandhini S, Sakthivadivel V, Lokesh S, Srinivasan AR, Sumathi S. Association of triglyceride-glucose index (TyG index) with HbA1c and insulin resistance in type 2 diabetes mellitus. *Maedica (Bucur)*. 2021;16:375-381.
27. Kim J, Shin S-J, Kim Y-S, Kang H-T. Positive association between the ratio of triglycerides to high-density lipoprotein cholesterol and diabetes incidence in Korean adults. *Cardiovasc Diabetol*. 2021;20:1-10.
28. Xepapadaki E, Nikdima I, Sagiadinou EC, Zvintzou E, Kypreos KE. HDL and type 2 diabetes: the chicken or the egg? *Diabetologia*. 2021;64:1917-1926.
29. Goldberg IJ. Diabetic dyslipidemia: causes and consequences. *J Clin Endocrinol Metab*. 2001;86:965-971.
30. Ng DS, Wong NC, Hegele RA. HDL—is it too big to fail? *Nat Rev Endocrinol*. 2013;9:308-312.
31. Ahola-Olli AV, Mustelin L, Kalimeri M, et al. Circulating metabolites and the risk of type 2 diabetes: a prospective study of 11,896 young adults from four Finnish cohorts. *Diabetologia*. 2019;62:2298-2309.
32. Yang T, Liu Y, Li L, et al. Correlation between the triglyceride-to-high-density lipoprotein cholesterol ratio and other unconventional lipid parameters with the risk of prediabetes and type 2 diabetes in patients with coronary heart disease: a RCSCD-TCM study in China. *Cardiovasc Diabetol*. 2022;21:93.
33. Florez JC, Jablonski KA, Bayley N, et al. TCF7L2 polymorphisms and progression to diabetes in the diabetes prevention program. *N Engl J Med*. 2006;355:241-250.
34. Wang TJ, Larson MG, Vasan RS, et al. Metabolite profiles and the risk of developing diabetes. *Nat Med*. 2011;17:448-453.
35. Puig-Jové C, Julve J, Castelblanco E, et al. The novel inflammatory biomarker GlycA and triglyceride-rich lipoproteins are associated with the presence of subclinical myocardial dysfunction in subjects with type 1 diabetes mellitus. *Cardiovasc Diabetol*. 2022;21:257.
36. Mehta NN, Dey AK, Maddineni R, Kraus WE, Huffman KM. GlycA measured by NMR spectroscopy is associated with disease activity and cardiovascular disease risk in chronic inflammatory diseases. *Am J Prev Cardiol*. 2020;4:100120.
37. Lorenzo C, Festa A, Hanley AJ, Rewers MJ, Escalante A, Haffner SM. Novel protein glycan–derived markers of systemic inflammation and C-reactive protein in relation to glycemia, insulin resistance, and insulin secretion. *Diabetes Care*. 2017;40:375-382.
38. Hagen IV, Helland A, Bratlie M, et al. Serum concentrations of amino acids and tryptophan metabolites are affected by consumption of a light breakfast: a clinical intervention study in adults with overweight or obesity. *BMC Nutr*. 2023;9:10.
39. Keirns BH, Sciarrillo CM, Koemel NA, Emerson SR. Fasting, nonfasting and postprandial triglycerides for screening cardiometabolic risk. *J Nutr Sci*. 2021;10:e75.

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: Liu J, Shang X, Zhang X, et al. Metabolomic network reveals novel biomarkers for type 2 diabetes mellitus in the UK Biobank study. *Diabetes Obes Metab*. 2025;27(6):3335-3346. doi:10.1111/dom.16351