

# Identification of Metabolic Pathway Dysregulation in Diabetic Retinopathy: A Propensity Score-Matched Metabolomic Study

Chengnan Guo (✉ [gcn757473437@163.com](mailto:gcn757473437@163.com))

Wenzhou Medical University <https://orcid.org/0000-0003-2275-3575>

**Depeng Jiang**

University of Manitoba

**Yixi Xu**

Wenzhou Medical University

**Fang Peng**

Wenzhou Medical University

**Shuzhen Zhao**

Wenzhou Medical University

**Huihui Li**

Wenzhou Medical University

**Dongzhen Jin**

Wenzhou Medical University

**Xin Xu**

Wenzhou Medical University

**Zhezheng Xia**

Wenzhou Medical University

**Mingzhu Che**

Wenzhou Medical University

**Mengyuan Lai**

Wenzhou Medical University

**Ruogu Huang**

Wenzhou Medical University

**Chao Zheng**

Zhejiang University

**Guangyun Mao**

Wenzhou Medical University

---

Original investigation

**Keywords:** Diabetic retinopathy, Metabolomics study, Metabolic alteration, Machine learning, Propensity score matching

**Posted Date:** July 9th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-684065/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

## Background

Diabetic retinopathy (DR) is a major diabetes-related disease linked to metabolism. However, scientifically assessment of serum metabolic alterations in DR is scarce. We aimed to investigate the changes in metabolic coregulation from type 2 diabetic patients (T2DM) to DR and identify corresponding metabolite predictors via a widely targeted metabolomics approach.

## Methods

In this case-control study, we tested 613 serum metabolites in 69 pairs of T2DM with DR (case) and propensity score-matched T2DM without DR (control) utilizing the ultra-performance liquid chromatography-electrospray ionization-tandem mass spectrometry system. The discrimination capability of differentially expressed metabolites (DEMs) in DR identification was also evaluated using a least absolute shrinkage and selection operator (LASSO) regression-based linear support vector machine (SVM) classifier. Metabolic pathway dysregulation in DR were comprehensively investigated by metabolic pathway analysis, chemical similarity enrichment analysis and MetaMapp approaches.

## Results

A total of 89 DEMs were identified after paired univariate analysis and partial least squares discriminant analysis. The linear-SVM model based on LASSO regression selected DEMs had an excellent discrimination with an area under the ROC curve (AUC) as 0.99 (95% confidence interval: 0.95, 1.00). The biosynthesis of polyunsaturated fatty acids (PUFAs), thiamine metabolism, amino acids (mainly glycine, serine and threonine metabolism), hydroxyeicosatetraenoic acid (HETE), disaccharides, indoles and nucleotides were significantly enriched in DR.

## Conclusions

This study systematically demonstrates that distinct metabolic alterations are linked to DR initiation. n-3 PUFAs, trehalose and vitamin B1 play an important role in inhibiting DR progression.

## Background

Despite being possibly preventable and treatable, diabetic retinopathy (DR) is still the major microvascular complication of diabetes mellitus (DM) and the chief reason for vision impairment and blindness around the working-age population (1). World Health Organization (WHO) has considered DR to be prevented or treated as one of the principal eye conditions (2). From 1980 to 2018, the annual incidence and progression of DR have increased from 2.2% and 3.4–12.7% and 12.3%, respectively (3). A recent meta-analysis reveals that DR is the only cause for the global growth of blindness in age-standardized prevalence between 1990 and 2020, especially in many parts of Asia and sub-Saharan Africa (4). According to the Handan Eye Study, about 43.1% of people over 30 years old suffering from

DM has been diagnosed as DR (5). So, it is crucial to achieve the early diagnosis of DR and formulate a comprehensively DR pathogenesis.

It is believed that neural and retinal vascular detriments in DR patients can appear before evident clinical manifestation. In the past several decades, glycosylated hemoglobin (HbA1c) has been demonstrated as a biomarker for the recognition of DR progression. Nevertheless, it is not enough to serve as an ideal biomarker since it can explain only a 6.6% variation in the risk of DR (6). Currently, there are several ways of detecting DR in its early stage, mainly retinal imaging techniques (e.g. standard retinal imaging, optical coherence tomography, etc.), to reduce the rate of DR-related severe vision impairments (7). However, the majority of the work-aged population will not universally take these tests, which will inevitably miss a good chance of early identification of DR. Therefore, there is an urgent need to achieve a simple but effective approach to detect the initiation and progression of DR as early as possible.

Metabolomics, which reflects cellular functions and phenotypes, mainly focuses on investigating the full complements of metabolic spectrum in the organism of interest (8). The significant metabolic disorders linked to DR have been reported in previous research. Available evidence reveals that DR-altered metabolites are concentrated in pentose phosphate pathway, arginine to proline pathway, polyol pathway, and ascorbic acidic pathways (9). However, DR-related metabolomics study is still in its infancy (9; 10). Although previous studies have tried to develop effective biomarkers of DR recognition using vitreous samples(9; 11), their sample sizes are greatly restricted because the vitreous sampling is invasive, which inevitably affects the robustness of their results. Fortunately, Li *et al* (12) and Chen *et al* (13) have suggested that serum metabolic profiles will alter the metabolic pathways associated with DR. Furthermore, Xuan *et al* (10) report that 12-hydroxyeicosatetraenoic acid (12-HETE) and 2-piperidone may serve as diagnostic markers of DR and further emphasize the value of serum metabolomics studies for ascertaining its pathogenesis. Though above-mentioned metabolomics studies are devoted to finding DR-related biomarkers, research on the mechanism of DR is still limited, which might be due to the incomplete existing metabolic pathways map as well as improper application of threshold-based pathway analysis. Hence, the metabolic pathway modules perturbed by DR need to be further investigated.

With the rapid development of detection technology, some metabolomic platforms, including nuclear magnetic resonance spectroscopy (NMR), liquid chromatography-mass spectrometry (LC-MS) and gas chromatography-mass spectrometry (GC-MS), have been developed when studying DR in the past several decades (10; 14). In practice, although NMR is simple and highly reproducible, it has limitations such as the unsatisfied sensitivity and incomplete metabolome detection (10). Meanwhile, LC-MS system is more generally applied for metabolome detection than other platforms because of its comprehensive detection and high sensitivity (14). So, it may be a good option to elucidate the abnormal metabolism associated with DR using a widely-targeted metabolomics strategy, an approach developed recently and better than traditional metabolomics methods, by the LC-MS tool.

Herein, we comprehensively investigated DR-related serum metabolome changes in a propensity score matching (PSM)-designed case-control study using metabolic profiling data obtained from the ultra-performance liquid chromatography - tandem mass spectrometry (UPLC-MS/MS) platform. This study aimed to (1) assess the metabolome patterns of DR after adjusting hyperglycemia state (HbA1c), (2) identify valuable metabolic biomarkers that could enhance discrimination ability of DR beyond traditional risk factors, and (3) systematically evaluate metabolic pathway dysregulation during the occurrence and development of DR.

## Research Design And Methods

### Study population

This was a two-center, PSM-based case-control study. The rationale and study design of the study have been previously reported (14; 15). Briefly, from August 2017 to June 2018, we enrolled a total of 195 T2DM patients (112 without DR and 83 with DR) aged over 35 years from the endocrinology departments of affiliated hospitals of two medical universities in Wenzhou and Anhui provinces, China. All enrolled participants had no histories of the following diseases such as any other eye diseases, type 1 diabetes, cardiovascular disease, heart failure, cancer, infectious disease or other chronic systemic diseases. The diagnosis of T2DM was applied strictly according to the standard criteria recommended by WHO since 1999. Two trained ophthalmologists independently performed meticulous eye examinations on participating patients and made DR diagnoses based on their retinal pictures. Inconsistent results of them on the DR diagnosis would be separately determined by another ophthalmologist.

To eliminate the impact due to major known confounding bias caused by the demographic and clinical characteristics of participants and improve the stability of our findings, a PSM approach was applied in the design process (16; 17). Baseline variables including age, gender, body mass index (BMI) and HbA1c were included in a multivariable logistic model for the propensity score, and the nearest neighbor algorithm was used in the matching process at a ratio of 1:1. Finally, 69 pairs of recruited patients with T2DM alone (control) and T2DM combined with DR (case) were successfully matched. Details of the study design have shown in Supplementary Fig. 1.

### Ethics statement

The protocol of the current study had been approved in advance by the Ethics Committee of the Eye Hospital of Wenzhou Medical University (Number: KYK [2017] 46). Written informed consent was obtained from each participant before data collection. The examination was performed per the tenets of the Declaration of Helsinki.

### Demographic and clinical data collection

Standardized structure questionnaires, containing information on age, gender, height, weight, duration of diabetes, occupation, life habits as well as histories of hypertension, tobacco and alcohol consumption,

treatment and family, were used to collect all participants' demographic characteristics by a face-to-face interview. BMI was calculated as weight (kg) / (height (m) <sup>2</sup>).

Features for clinical manifestation and biochemistry, including fasting blood glucose (FPG), HbA1c, total cholesterol (TC), triglyceride (TG), high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C) as well as systolic blood pressure (SBP) and diastolic blood pressure (DBP), were determined by two systematically trained investigators. The associated standardized operation procedures (SOP) of this study were strictly followed in the process.

## **Metabolomics analyses**

The venous blood samples of all participants were gathered after fasting for more than 8 hours at their enrollments. The samples were then prepared in conical polypropylene centrifuge tubes and centrifuged (2000 rpm, \*4°C, \*10 min) for separating the serum. After that, 1.5 mL serum sample was stored using sterile tubes at -86°C deep low temperature refrigerator for further metabolomic assessment. All steps of the following sample preprocessing and metabolic analyses were carefully carried out by a professional technician in the central laboratory of Metware Inc., a professional metabolomics institution in China. Details could be found in the supplemental materials.

## **Data processing**

After the determination of widely targeted UPLC-MS/MS-based metabolic profiles, metabolic profile data were obtained by Analyst Software V.1.6.3 (AB Sciex) and processed in an untargeted manner using MultiQuant Software (AB Sciex) to convert, peak detects, retention time correct and peak align.

Metabolites with a coefficient of variation (CV) larger than 30% in the quality control (QC) samples were discarded. The metabolites with a missing ratio of over 20% in the cases or controls were also removed. Besides, for those with a missing ratio of less than 20%, they would be separately imputed by half of the lowest detected peak areas. Afterwards, log transformation and Pareto scaling were individually utilized to improve the normality of associated data and make them more comparable. The distributions of related metabolites for the pre- and post-transformation were presented in Supplementary Fig. 2.

## **Statistical analysis**

The sample size estimation has been reported in our previous study, with 90% power to detect the difference of metabolites in DR and DM groups (14). Normally or approximately normally distributed continuous variables were described as mean ± SD and compared by the paired t-test. Otherwise, variables with obviously skewed distribution would be presented as median (1st quartile, 3rd quartile) and compared by Wilcoxon signed-rank test. Categorical variables were reported as frequency (percentage) and McNemar-Bowker test or Wilcoxon signed-rank test would be applied for the comparisons. A directed acyclic graph (DAG) was applied to evaluate the causal relationship between DR and potential predictor variables according to the criteria of p-value less than 0.1, and to decide which indicators would be included in the final classifier (Supplementary Fig. 3) (18).

Paired t-test with Benjamini-Hochberg false positive rate (FDR) correction was performed to identify the differentially expressed metabolites (DEMs) when comparing DR with DM. Furthermore, to markedly increase the reliability of detected DEMs, a partial least squares discriminant analysis (PLS-DA) model was constructed by 10-fold cross-validation to obtain the variable importance for the projection (VIP) of metabolites. Thereafter, 1000-times permutation test would be further applied to avoid potential overfitting of the PLS-DA model. Finally, the criteria of DEMs screening were determined as FDR-adjusted q-value < 0.05, fold changes (FC) > 1.2 or FC < 0.8, and VIP > 1.

Relationships of the top 25 DEMs, which had the lowest q-value, were additionally visualized via a hierarchical clustering heatmap depending on the Euclidean distance metric and Ward's clustering method. To explore the potential collinearity among the top 25 DEMs, a spearman correlation coefficient matrix was plotted. Least absolute shrinkage and selection operator (LASSO) regression was further utilized to minimize the potential collinearity and over-fitting of DEMs. 10-fold cross-validation via minimum ( $\lambda$ ) criteria was applied to select metabolic predictors. Afterwards, the normalized peak areas of the selected DEMs were visualized with violin charts.

To further validate the likelihood of DEMs as potential metabolic markers for distinguishing DR from T2DM, we proposed a support vector machine (SVM) classifier depending on the linear kernel algorithm using Python version 3.7.3 (Copyright © 2001–2021 Python Software Foundation). Then, 10-fold cross-validation approach was applied to avoid overfitting and the area under the receiver operating characteristic curve (AUC under ROC) with 95% confidence interval was utilized to assess the classification effect. All above-mentioned data management and analysis were jointly implemented by RStudio version 1.2.5042 (© 2009–2020 RStudio, Inc) and MetaboAnalyst version 5.0 (<https://www.metaboanalyst.ca/>) (19).

## **Biological Interpretation**

### **Metabolic Pathway Analysis (MetPA)**

This study chose over-representation analysis (ORA) via hypergeometric test and pathway topology (PT) via out-degree centrality (ODC) to explore the metabolic pathway dysregulation in DR (20). Above MetPA methods were both based on Kyoto Encyclopedia of Genes and Genomes (KEGG) database (21), which was one of the most useful metabolomics database worldwide. Depending on hypergeometric test, metabolic pathways with p-value less than 0.05 were thought to have been disturbed. The enriched metabolic pathways were visualized by bubble chart and interpreted using KEGG Mapper (<https://www.genome.jp/kegg/mapper.html>). To further verify the dysregulated metabolic pathways, a DR-altered KEGG global metabolic network were also presented using iPath version 3.0 (<https://pathways.embl.de/>) (22).

### **Chemical Similarity Enrichment Analysis (ChemRICH) for metabolites**

Since the existing metabolic knowledge annotation is sparse, ChemRICH (23), a novel pathway mapping tool based on chemical similarity, was applied as a supplement of MetPA. The significantly disturbed metabolite clusters ( $p$ -values  $< 0.05$ ) were visualized by a bubble plot. And then, MetaMapp associated graph (<http://metamapp.fiehnlab.ucdavis.edu/>) were applied for visualizing clearer metabolic network in DR-disturbed metabolomes and module identification (24), and shown in Cytoscape with a circular layout (8; 25) since Cytoscape was broadly applied in omics network visualization and subnetwork identification. Details of ChemRICH and MetaMapp were in the supplemental materials.

## Results

### Characteristics of the study participants

A total of 69 pairs of DM and DR, including 60 with non-proliferative DR (NPDR: 9 mild, 31 moderate and 20 severe) and 9 with proliferative DR (PDR), were included in the current study. The demographic and clinical variables of participants were described in Table 1. As compared to DM, DR patients tended to have a longer duration of T2DM ( $p = 0.002$ ) and higher SBP ( $p = 0.003$ ). Compared to DM, DR patients were more likely to receive treatments ( $p < 0.001$ ), but slightly less likely to receive insulin medication since being diagnosed with DM ( $p = 0.093$ ). DR patients were more likely to have vision damage ( $p = 0.005$ ) than DM patients.

Table 1  
Demographics and clinical indicators of participants included in the study.

Variable	DM	DR	P
<b>Continuous variable</b>			
Age, years	53.0(48.0,61.0)	56.0(51.0,65.0)	0.022
BMI, Kg/m <sup>2</sup>	24.4 ± 3.2	24.6 ± 3.5	0.773
FPG, mmol/L	8.4(6.9,12.0)	8.5(6.3,10.2)	0.225
HbA1c, %	10.1 ± 2.3	9.9 ± 1.9	0.500
LDL, mmol/L	2.7 ± 1.0	2.6 ± 1.1	0.617
HDL, mmol/L	1.0(0.8,1.3)	1.1(0.9,1.3)	0.703
TG, mmol/L	1.6(1.0,2.2)	1.4(1.0,1.9)	0.184
TC, mmol/L	4.7 ± 1.1	4.5 ± 1.4	0.337
SBP, mmHg	124.0(118.0,139.0)	135.0(122.0,148.0)	0.003
DBP, mmHg	79.0(74.0,86.0)	76.0(70.0,85.0)	0.449
Duration of diabetes, years	8.0(4.0,13.0)	12.0(8.0,17.0)	0.002
<b>Category variable, n/N</b>			
Gender			0.625
Male	38/69	36/69	
Female	31/69	33/69	
Occupation			0.825
Manual workers	31/65	34/64	
Mental worker	15/65	11/64	
Both	19/65	19/64	
Center			0.074
Wenzhou	36/69	48/69	
Hefei	33/69	21/69	
Hypertension			0.078

*Abbreviations: BMI: body mass index; FPG: fasting plasma glucose; HbA1c: glycated hemoglobin; LDL: low density lipoprotein; HDL: high density lipoprotein; TG: triglyceride; TC: total cholesterol; SBP: systolic blood pressure; DBP: diastolic blood pressure; DM: type 2 diabetes mellitus (T2DM) without diabetic retinopathy; DR: T2DM with diabetic retinopathy.*

Variable	DM	DR	P
No	47/66	37/65	
Yes	19/66	28/65	
Smoking habits			0.530
Non-smokers	41/66	36/65	
Ex-smokers	6/66	8/65	
Current smokers	19/66	21/65	
Alcohol consumption			0.921
Non-drinkers	33/66	29/65	
Ex-drinkers	3/66	9/65	
Current drinkers	30/66	27/65	
Three-generation family history			0.556
No	33/63	29/66	
Yes	30/63	37/66	
Therapy or not			< 0.001
No	14/58	1/63	
Yes	44/58	62/63	
Insulin medication			0.093
No	46/65	55/65	
Yes	19/65	10/65	
Heel pain			0.134
No	55/66	46/65	
Yes	11/66	19/65	
Vision loss			0.005
No	37/66	19/65	
Yes	29/66	46/65	

*Abbreviations: BMI: body mass index; FPG: fasting plasma glucose; HbA1c: glycated hemoglobin; LDL: low density lipoprotein; HDL: high density lipoprotein; TG: triglyceride; TC: total cholesterol; SBP: systolic blood pressure; DBP: diastolic blood pressure; DM: type 2 diabetes mellitus (T2DM) without diabetic retinopathy; DR: T2DM with diabetic retinopathy.*

## Primary metabolomics analyses via UPLC-MS/MS

A total of 613 identified metabolites, including 318 in positive ion modes and 295 in negative ion modes, were detected via the UPLC-MS/MS system. After the necessary data preprocessing, 461 metabolites participated in the subsequent analysis. The PCA score plot demonstrated that QC samples had good consistency (black nodes), which meant the detection system was robust (Supplementary Fig. 4A). To further distinguish DR from T2DM and obtain VIP of metabolites, PLS-DA model was applied and the results showed that the model could well classify the cases and controls (Supplementary Fig. 4B). Established using three principal components, the cumulative  $R^2$ ,  $Q^2$  and accuracy respectively achieved 90.2%, 77.9% and 97.2% combined with p-values of 1000 times permutation test less than 0.001 (Supplementary Fig. 4C).

### **Identification of DEMs in DR participants**

According to the criteria ( $q$ -value  $< 0.05$ ,  $FC < 0.8$  or  $FC > 1.2$  and  $VIP > 1$ ) of DEMs detection, a total of 89 metabolites were determined as DEMs. Among them, 34 were apparently under-regulated and the other 55 were obviously increased in DR when comparing with DM. Based on the top 25 DEMs determined by 25 smallest  $q$ -values, the heatmap clearly exhibited distinct patterns between DR and DM (Fig. 1A). Metabolites in DR were obviously altered when comparing with DM (Fig. 1A, 1B). Figure 1C showed that the top 25 DEMs were highly correlated to each other, especially among fatty acids. After LASSO regression selection (Supplementary Fig. 5), 10 DEMs were retained. The violin plots of LASSO-selected DEMs revealed significant up- or downregulation in the DR group (Fig. 1D). In addition, the heatmap and volcano plot of all 89 DEMs revealed the same patterns of metabolic alteration in DR (Supplementary Fig. 6).

### **Assessment of the value of DEMs in DR detection by linear SVM model**

As could be seen in Fig. 2, the LASSO-selected DEMs had excellent discrimination performance in DR detection. The model 1 included traditional clinical features (the duration of T2DM, SBP and insulin medication) screened by previous studies and DAG (Supplementary Fig. 3). The discrimination performance of DEMs-added model 2 [AUC (95% CI): 0.99 (0.95, 1.00)] was statistically better than the model 1 [AUC (95% CI): 0.71 (0.51, 0.91)]. It strongly suggested that DEMs might be optimal biomarkers of DR identification and would play a critical role when distinguishing DR from T2DM.

### **MetPA plots of DR-altered metabolic pathways**

To better understand the biological meaning of DR-related metabolic pathways, a bubble plot was utilized, in which Y-axis and X-axis respectively meant the  $-\lg(p)$  and ODC. The size and color of each bubble represented its ODC and metabolic pathway categories, respectively. Among all enriched pathways, the biosynthesis of unsaturated fatty acids, thiamine metabolism, and glycine, serine and

threonine metabolism were more important than others ( $p$ -values  $< 0.05$ , Fig. 3B). And the hub metabolite (linoleate) of linoleic acid metabolism was manifestly decreased as ODC equals to 0.75. The notably altered metabolic pathways were further interpreted in Fig. 3C. As could be seen in this figure, the biosynthesis of unsaturated fatty acids in DR was evidently decreased in metabolites of n-6 and n-3 polyunsaturated fatty acids (PUFAs) families. The elevation of L-cysteine and reduction in thiamine triphosphate constitutes an overt alteration of thiamine metabolism. Meanwhile, the glycine, serine and threonine metabolism was also remarkably altered by increased L-cysteine and decreased  $\beta$ -hydroxypyruvic acid, creatine and sarcosine. In addition, the altered metabolic pathways enriched by KEGG global metabolic network were basically consistent with previous results (Supplementary Fig. 7 and Supplementary Table 1). MetPA suggested that DR would mainly disturb the metabolisms of lipid, amino acid and some cofactors as well as vitamins.

### ChemRICH plots of DR-altered metabolite clusters

Identified metabolites by the PubChem compound database were summarized by ChemRICH (Fig. 4A). Based on  $q$ -value  $< 0.01$ , the most interesting aspect of this graph was that 11 clusters (unsaturated FA, disaccharides, ethanolamines, HETE, dipeptides, indoleacetic acids, pyridines, hydroxybutyrates, saturated FA, glutamates, and amino acids, sulfur) were successfully enriched. Several metabolite clusters (key compound) of disaccharides (D-trehalose), ethanolamines (choline) and HETE (12-HETE) were notably altered ( $q$ -value  $< 0.001$ ) in DR (Fig. 4B). Among them, the most notably decreased cluster was unsaturated FA ( $q$ -value =  $8.7 \times 10^{-6}$ ), and linoleic acid (C18:2N6C) was detected as the key compound.

In addition, several key points in DR-altered metabolome were discovered and presented in Fig. 4C. The orange and green nodes respectively indicated apparently increased and decreased metabolites ( $p < 0.05$ ) in DR patients. Meanwhile, the size of nodes represented the magnitude of FC. Purple edges symbolized KEGG reactant pair links, while the grey ones symbolized Tanimoto chemical similarity over 700. According to the same figure, we observed that DR-altered metabolites were mainly gathered into 5 modules by chemical similarity: PUFAs (a), amino acids (b), nucleotides (c), indoles (d) and benzene homologues (e). According to ChemRICH, PUFAs module was the vital one owing to linoleic acid (C18:2N6C) and 12-HETE. And amino acids module was the central hub in DR-altered network mentioned above and connected with other modules through KEGG reactant pairs. The inside and outside of the modules were mainly connected by chemical similarity and prior knowledge, respectively.

## Discussion

In this PSM-based case-control study, we detect disrupted metabolic pathways and central clusters in DR serum specimens via UPLC-MS/MS platform combined with prior-knowledge-based metabolic pathway

analysis and chemical-based ChemRICH. Our findings suggest that DR-altered metabolites are mainly associated with biosynthesis of unsaturated fatty acids, thiamine metabolism, and glycine, serine and threonine metabolism. In addition, unsaturated FA, disaccharides, ethanolamines and HETE are the greatly disturbed metabolite clusters according to ChemRICH. More importantly, the metabolic network built in this study displays that some serum metabolites are apparently altered in DR and predominantly gathered into five modules, including PUFAs, amino acids, indoles, nucleotides and benzene homologues. We also reveal a fact that current knowledge on the metabolome of DR progression is far from enough.

In the present study, we observe that biosynthesis of PUFAs may play a vital role in DR occurrence and linoleic acid (LA) is the essential compound of unsaturated FA cluster. PUFAs, especially n-3 and n-6 series, are apparently decreased in DR relative to DM, which is consistent with previous reports (9; 15). Available evidence suggests that LA is the main dietary n-6 PUFA and has a forceful inhibitory impact on the progress of DR (26). Meanwhile, supplementations of docosahexaenoic acid (DHA) and eicosapentanoic acid (EPA) have also been suggested as alternative diet therapeutic strategies to prevent DR (27). Growing evidence reveals that PUFAs, known as bioactive lipids, have critical abilities in angiogenesis, regulation of inflammation and homeostasis maintenance (2; 10; 14). They generally work in a two-step reaction sequence. First, PUFAs, such as LA, AA, DHA and EPA, will be over-released by the action of phospholipase A2 (PLA2) enzyme from the cell membrane lipid layers in a hyperglycemic or hypoxic environment. Second, the released PUFAs are metabolized by either Cytochrome P450 (CYP), lipoxygenases, or Cyclooxygenases (Cox) enzymes (2). Through above-mentioned three enzymatic reactions, metabolites of the n-6 PUFA (e.g., AA) will induce proangiogenic and proinflammatory effects on the promotion of DR occurrence. On the other hand, these effects are majorly offset by the metabolites derived from n-3 PUFAs, such as DHA and EPA (Supplementary Fig. 8).

As the key compound of hydroxyeicosatetraenoic acid (HETE) cluster, 12-HETE is observed to be positively associated with the presence of DR. Existing evidence manifests that 12- and 15-HETE are the primary metabolite of 12- and 15-Lipoxygenase (LOX) in human (2), and has the same relationship with PUFAs (Fig. 4C(a)). It is also considered to evoke balanced autocrine and paracrine proangiogenic pathways through activation of retinal endothelial as well as Muller cells, respectively, when there is ischemic retinopathy (2). In retinal endothelial cells, 12- and 15-HETE may activate vascular endothelial growth factor receptor 2 (VEGFR2) by evoking endoplasmic reticulum (ER) stress and NADPH oxidase, which eventually leads to DR (2; 10; 28). In this condition,  $Ca^{2+}$  homeostasis disorder may be a pivotal step to promote the autocrine loop (29). While in Muller cells, endothelial-derived HETEs may induce retinal vascular dysfunction via upregulating inflammatory cytokines and VEGF, and downregulate pigment epithelium-derived factor (PEDF) (2) (Supplementary Fig. 8C).

In the present study, amino acid metabolism serves as the connecting bridge of DR-altered metabolic network and many amino acids are altered significantly in DR patients. It is reported that the glycine, serine and threonine metabolism is the most enriched pathway, which reproduces the results of a recent metabolomics research (10). According to Rajala et al (30), serine and glycine residues can stabilize the Warburg-like effect by enriching the photoreceptor protein rhodopsin, which is closely associated with the

phosphorylation of pyruvate kinase M2. The decrease in creatine is also reported to be linked to vascular proliferation, and supplementing creatine may be helpful to suppress retinal neovascularization in DR (11). Furthermore, the levels of L-cysteine and glutamate (glutamic acid) are also evidently increased in DR patients. It is considered as the result of glutamate-induced oxidative stress (OS), which is happened via suppression of cysteine/glutamate antiporter, leading to consumption of glutathione (GSH) levels and accumulation of reactive oxygen species (ROS) (31), and finally results in DR (32).

Thiamine metabolism is also significantly enriched in this study, mainly due to the decreasing of Thiamine Triphosphate (ThTP), which exists as the non-coenzyme form of thiamine (Vitamin B1) in all living organisms (33). Although the mechanism is currently unclear, ThTP is considered as the allosteric activator of glutamate dehydrogenase (GDH), which promotes the metabolism of glutamate to alpha ketoglutarate (essential substances for the tricarboxylic acid cycle) (34). This further explains the growth of glutamate. In the other case, ThTP has a specific neurophysiological role and can phosphorylate rapsyn (a protein organizing acetylcholine receptors in the synapses), which may be linked to the facilitation of acetylcholinergic neurotransmission (33). The enriched ethanolamines cluster (choline as key compound) in this study may corroborate this view.

As the key compound of disaccharides, D-Trehalose is notably decreased in DR when comparing with T2DM. To the best of our knowledge, this is the first report to detect a connection of trehalose with DR. In fact, trehalose cannot be synthesized in human body and mainly used in Asian countries as a food stabilizer (35). Existing evidence reveals that trehalose can effectively govern hyperglycemia of diabetic patients via relieving impaired glucose tolerance, mitigating insulin resistance, and reducing postmeal insulin bursts (35). Taya *et al* (36) disclose that trehalose is capable of reducing the production of inflammatory cytokines by protecting I $\kappa$ B- $\alpha$  reduction in vivo. Since hyperglycemia and inflammation are essential pathogenic factors contributing to DR and the safety of trehalose has been confirmed (35), we presume that trehalose supplementation may be beneficial to DR prevention and control. Further studies should be conducted to verify this presumption.

In the present study, indoles are also obviously enriched in the DR-related metabolites network, in which indoleacetic acids (IAA) is detected as the mainly disturbed metabolite cluster. As a major metabolite of tryptophan, tryptamin can be transformed to IAA or serotonin, leading to IAA and serotonin synthesis as a competitive relationship (37), which is consistent with our results. Meanwhile, as a metabolite of serotonin, melatonin can inhibit the progress of DR through regulating the circadian rhythm, anti-angiogenic and anti-inflammatory (38).

The abnormal elevation of most nucleotides and their derivatives observed in this study may be related to OS. A recent study believes that deoxyguanosine is an oxidative DNA breakdown product, which has been proved to be a valuable biomarker of DR identification (39). Its considerable elevation reflects the high level of OS in DR patients. Though benzene homologues are not notably enriched in metabolic pathways, they are observed to be abnormally increased in DR-related metabolic network. Considering the existence potentially due to external sources, the module is not fully elaborated in this study (40).

Our study has several strengths. First, the PSM approach is applied to adjust for the impacts of the potential confounders including HbA1c. This process allows us to look at metabolic pathway dysregulation in addition to energy metabolism, which has been thoroughly studied. Second, a widely targeted metabolomics platform with UPLC-MS/MS system, rather than traditional detection strategies, is used to precisely and efficiently acquire accurate data and determine the structure of metabolites, which improves the credibility of our results. Third, the application of ChemRICH well makes up for the limitations of incomplete metabolomics database and the neglects of those marginally significant metabolites due to traditional MetPA. Fourth, depending on prior knowledge or chemical similarity, the MetaMapp-integrated network clearly discloses the relationships of connected metabolites, which is highly beneficial to spy the global metabolic network perturbed by DR.

This study also has some limitations. First, the sample size is not large enough to detect as more differential metabolites as possible. Nevertheless, small sample size does not necessarily mean that it is insufficient. The current sample size is sufficient yet since it has been thoroughly considered during study design. Second, the enzymatic reactions among metabolites cannot be directly calculated to map to the MetaMapp-integrated network since adding too many metabolites will cause blurring. Third, though we think that our findings are robust because of careful design, accurate measurements, and comprehensive data analysis, they need to be verified by more prospective cohorts or experimental studies.

## Conclusions

In conclusion, the aforementioned results strongly suggest that metabolomics profiling is promising for clarifying the pathogenesis of DR from a three-level perspective of metabolite-pathway-network and screening biomarkers for DR early identification. Our results reveal that DR initiation and development are evidently associated with the disturbances of the biosynthesis of PUFAs and HETE, thiamine metabolism, amino acids (mainly glycine, serine and threonine metabolism), D-Trehalose, indoles, and nucleotides.

## Abbreviations

**MetPA:** Metabolic Pathway Analysis

**ORA:** Over-representation analysis

**PT:** Pathway topology

**ODC:** Out-degree centrality

**KEGG:** Kyoto Encyclopedia of Genes and Genomes

**ChemRICH:** Chemical Similarity Enrichment Analysis

**NPDR:** Non-proliferative diabetic retinopathy

**PDR:** Proliferative diabetic retinopathy

**LA:** Linoleic acid

**DHA:** Docosahexaenoic acid

**EPA:** Eicosapentanoic acid

**CYP:** Cytochrome P450

**Cox:** Cyclooxygenases

**LOX:** Lipoxygenase

**VEGFR2:** Vascular endothelial growth factor receptor 2

**ER:** Endoplasmic reticulum

**PEDF:** Pigment epithelium-derived factor

**OS:** Oxidative stress

**GSH:** Glutathione

**ROS:** Reactive oxygen species

**ThTP:** Thiamine Triphosphate

**GDH:** Glutamate dehydrogenase

**IAA:** Indoleacetic acid

## **Declarations**

## **Ethics approval and consent to participate**

The protocol of the current study had been approved in advance by the Ethics Committee of the Eye Hospital of Wenzhou Medical University (Number: KYK [2017] 46). Written informed consent was obtained from each participant before data collection. The examination was performed per the tenets of the Declaration of Helsinki.

## **Consent for publication**

Not applicable.

# Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

## Competing interests

The authors declare that they have no conflict of interest.

## Funding

This work was supported by National Nature Science Foundation of China (82070833), Zhejiang Basic Public Welfare Research Project (LGF19H260011) and Major Project of the Eye Hospital of Wenzhou Medical University (YNZD201602). Part of this work was also funded by Zhejiang Provincial Key Research and Development Program (2021C03070) and Natural Science Foundation of Zhejiang Province (LZ19H020001).

## Authors' contributions

Chengnan Guo: Conceptualization, Data analysis, Methodology, Writing - original draft, Writing - review & editing. Yixi Xu: Data analysis, Writing - review & editing, Supervision. Shuzhen Zhao, Huihui Li, Dongzhen Jin, Xin Xu, Fang Peng, Zhezheng Xia, Mingzhu Che, Mengyuan Lai and Ruogu Huang: Contributed to the epidemiological investigation, sample handling, data management and analysis. Depeng Jiang and Chao Zheng: Writing - review & editing. Guangyun Mao: Conceptualization, Funding acquisition, Supervision, Writing - review & editing.

## Acknowledgements

The authors must express utmost thanks to participants. Without their supports we would not have achieved any of the goals we have set out to attain. We would like to take this opportunity to acknowledge and thank the colleagues from Wenzhou Medical University for their hard works and collaborations.

## References

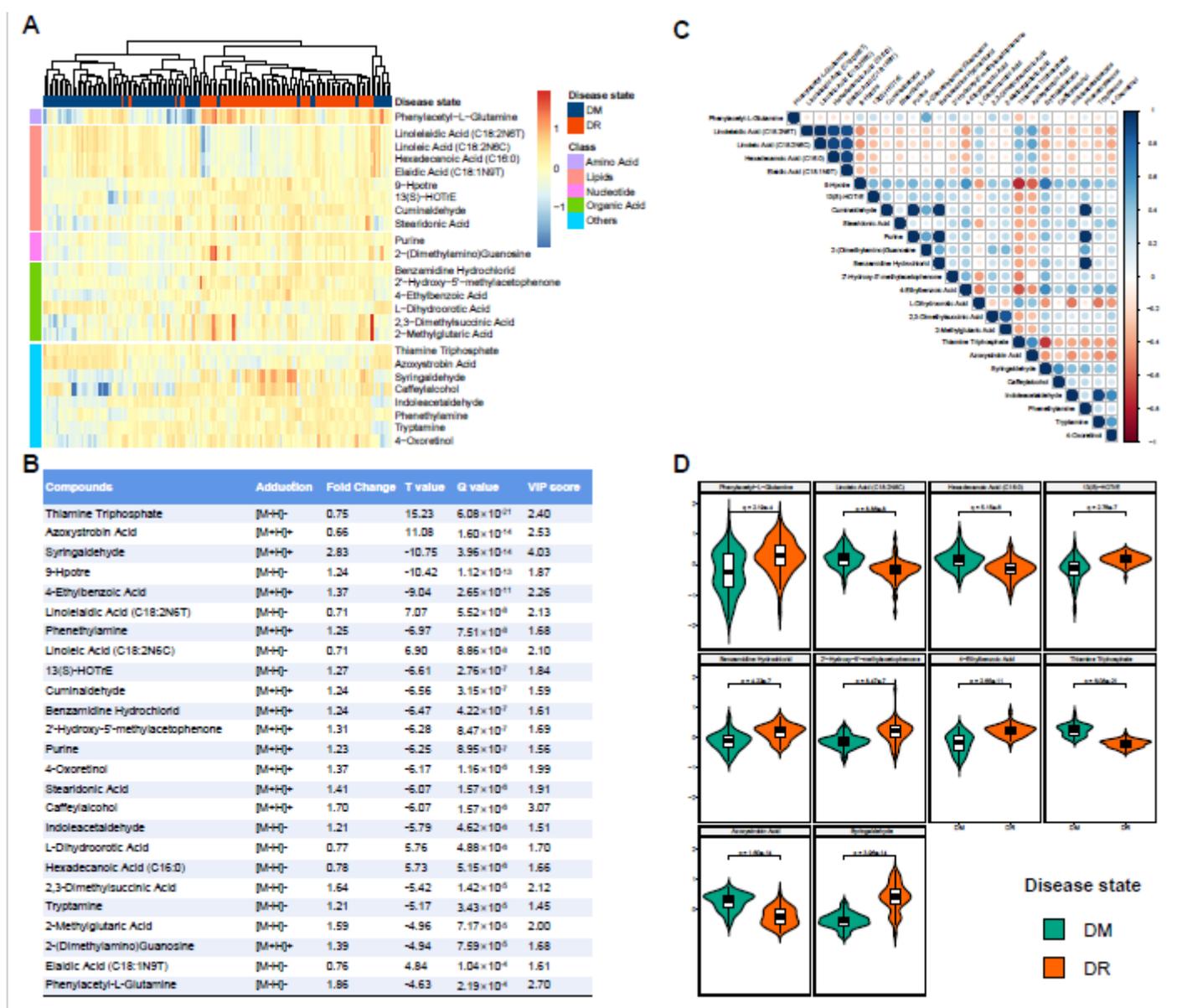
1. IDF Diabetes Atlas, 9th edn [article online]. 2019. Available from .
2. Elmasry K, Ibrahim AS, Abdulmoneim S, et al. Bioactive lipids and pathological retinal angiogenesis. *Br J Pharmacol*. 2019;176:93–109.
3. Sabanayagam C, Banu R, Chee ML, et al. Incidence and progression of diabetic retinopathy: a systematic review. *Lancet Diabetes Endocrinol*. 2019;7:140–9.
4. Steinmetz JD, Bourne RRA, Briant PS, et al. Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to

- Sight: an analysis for the Global Burden of Disease Study. *The Lancet Global Health*. 2021;9:e144–60.
5. Wang FH, Liang YB, Zhang F, et al. Prevalence of diabetic retinopathy in rural China: the Handan Eye Study. *Ophthalmology*. 2009;116:461–7.
  6. Hirsch I, Brownlee M. Beyond hemoglobin A1c—need for additional markers of risk for diabetic microvascular complications. *JAMA*. 2010;303:2291–2.
  7. Liew G, Michaelides M, Bunce C. A comparison of the causes of blindness certifications in England and Wales in working age adults (16–64 years), 1999–2000 with 2009–2010. *BMJ open*. 2014;4:e004015.
  8. Fan S, Shahid M, Jin P, et al. Identification of Metabolic Alterations in Breast Cancer Using Mass Spectrometry-Based Metabolomic Analysis. *Metabolites* 2020;10.
  9. Liew G, Lei Z, Tan G, et al. Metabolomics of Diabetic Retinopathy. *Curr Diab Rep*. 2017;17:102.
  10. Xuan Q, Ouyang Y, Wang Y, et al. Multiplatform Metabolomics Reveals Novel Serum Metabolite Biomarkers in Diabetic Retinopathy Subjects. *Adv Sci (Weinh)*. 2020;7:2001714.
  11. Tomita Y, Cagnone G, Fu Z, et al. Vitreous metabolomics profiling of proliferative diabetic retinopathy. *Diabetologia*. 2021;64:70–82.
  12. Li X, Luo X, Lu X, et al. Metabolomics study of diabetic retinopathy using gas chromatography-mass spectrometry: a comparison of stages and subtypes diagnosed by Western and Chinese medicine. *Mol Biosyst*. 2011;7:2228–37.
  13. Chen L, Cheng C, Choi H, et al. Plasma Metabonomic Profiling of Diabetic Retinopathy Diabetes. 2016;65:1099–108.
  14. Zuo J, Lan Y, Hu H, et al. Metabolomics-based multidimensional network biomarkers for diabetic retinopathy identification in patients with type 2 diabetes mellitus. *BMJ Open Diabetes Res Care* 2021;9.
  15. Li J-s, Wang T, Zuo J-j, et al. Association of n-6 PUFAs with the risk of diabetic retinopathy in diabetic patients. *Endocrine Connections* 2020.
  16. ROSENBAUM PR, RUBIN DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70:41–55.
  17. Li C, Sun YD, Yu GY, et al. Integrated Omics of Metastatic Colorectal Cancer. *Cancer Cell*. 2020;38:734–47. e739.
  18. Daltveit DS, Klungsoyr K, Engeland A, et al. Cancer risk in individuals with major birth defects: large Nordic population based case-control study among children, adolescents, and adults. *BMJ*. 2020;371:m4060.
  19. Xia J, Psychogios N, Young N, et al. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Res*. 2009;37:W652–60.
  20. Xia J, Wishart DS. MetPA: a web-based metabolomics tool for pathway analysis and visualization. *Bioinformatics*. 2010;26:2342–4.

21. Kanehisa M, Furumichi M, Sato Y, et al. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* 2021;49:D545–51.
22. Darzi Y, Letunic I, Bork P, et al. iPath3.0: interactive pathways explorer v3. *Nucleic Acids Res.* 2018;46:W510–3.
23. Barupal DK, Fiehn O. Chemical Similarity Enrichment Analysis (ChemRICH) as alternative to biochemical pathway mapping for metabolomic datasets. *Sci Rep.* 2017;7:14567.
24. Barupal DK, Haldiya PK, Wohlgemuth G, et al. MetaMapp: mapping and visualizing metabolomic data by integrating information from biochemical pathways and chemical and mass spectral similarity. *BMC Bioinformatics.* 2012;13:99.
25. Saito R, Smoot ME, Ono K, et al. A travel guide to Cytoscape plugins. *Nat Methods.* 2012;9:1069–76.
26. Houtsmuller A, van Hal-Ferwerda J, Zahn K, et al. Favorable influences of linoleic acid on the progression of diabetic micro- and macroangiopathy in adult onset diabetes mellitus. *Progress in lipid research.* 1981;20:377–86.
27. Tikhonenko M, Lydic TA, Wang Y, et al. Remodeling of retinal Fatty acids in an animal model of diabetes: a decrease in long-chain polyunsaturated fatty acids is associated with a decrease in fatty acid elongases Elov12 and Elov14. *Diabetes.* 2010;59:219–27.
28. Ibrahim AS, Elshafey S, Sellak H, et al. A lipidomic screen of hyperglycemia-treated HRECs links 12/15-Lipoxygenase to microvascular dysfunction during diabetic retinopathy via NADPH oxidase. *J Lipid Res.* 2015;56:599–611.
29. Elmasry K, Ibrahim AS, Saleh H, et al. Role of endoplasmic reticulum stress in 12/15-lipoxygenase-induced retinal microvascular dysfunction in a mouse model of diabetic retinopathy. *Diabetologia.* 2018;61:1220–32.
30. Rajala RV, Rajala A, Kooker C, et al. The Warburg Effect Mediator Pyruvate Kinase M2 Expression and Regulation in the Retina. *Sci Rep.* 2016;6:37727.
31. Xia Y, Xing JZ, Krukoff TL. Neuroprotective effects of R,R-tetrahydrochrysenes against glutamate-induced cell death through anti-excitotoxic and antioxidant actions involving estrogen receptor-dependent and -independent pathways. *Neuroscience.* 2009;162:292–306.
32. Brownlee M. Biochemistry and molecular cell biology of diabetic complications. *Nature.* 2001;414:813–20.
33. Bettendorff L, Wins P. Thiamin diphosphate in biological chemistry: new aspects of thiamin metabolism, especially triphosphate derivatives acting other than as cofactors. *FEBS J.* 2009;276:2917–25.
34. Bunik V, Artiukhov A, Aleshin V, et al. Multiple Forms of Glutamate Dehydrogenase in Animals: Structural Determinants and Physiological Implications. *Biology (Basel)* 2016;5.
35. Sokolowska E, Sadowska A, Sawicka D, et al. A head-to-head comparison review of biological and toxicological studies of isomaltulose, d-tagatose, and trehalose on glycemic control. *Crit Rev Food Sci Nutr* 2021:1–26.

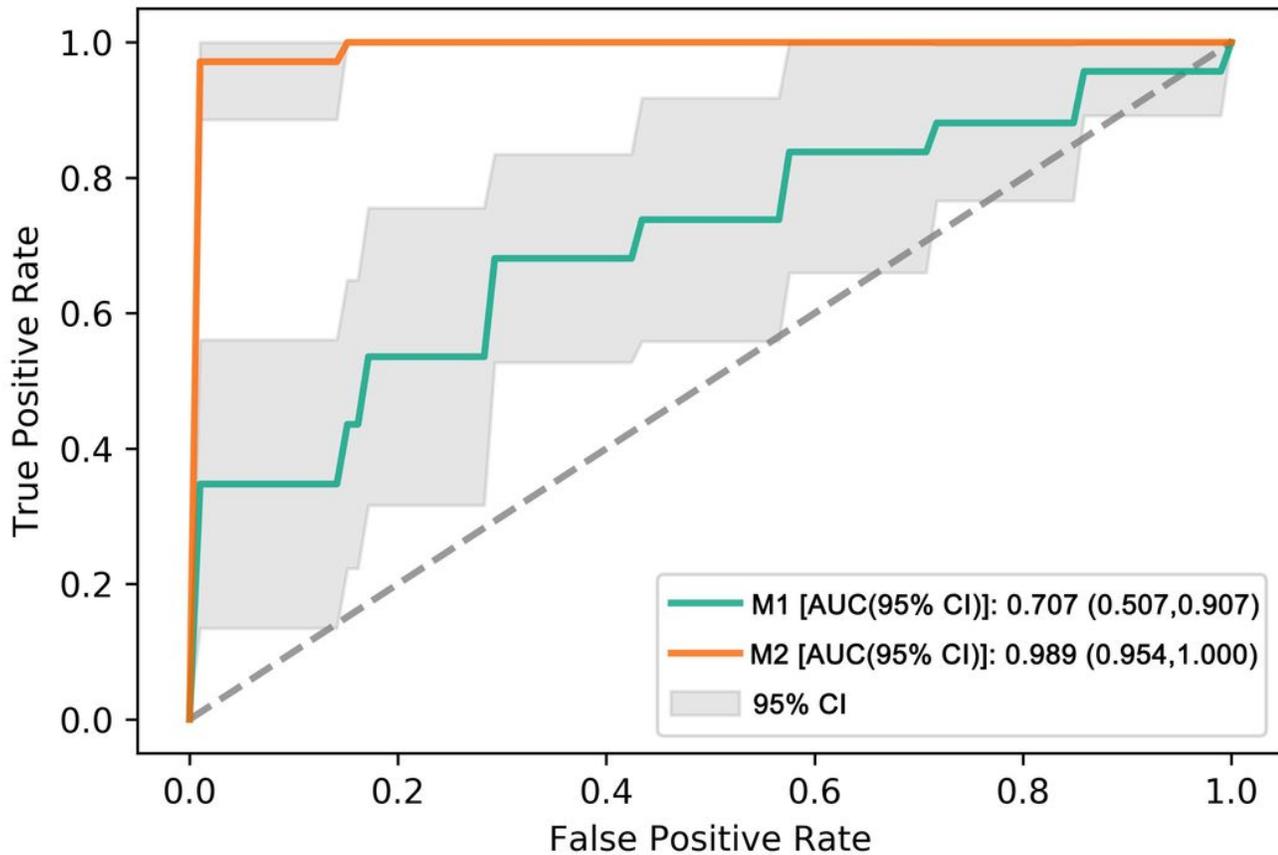
36. Taya K, Hirose K, Hamada S. Trehalose inhibits inflammatory cytokine production by protecting I kappaB-alpha reduction in mouse peritoneal macrophages. Arch Oral Biol. 2009;54:749–56.
37. Posmyk MM, Janas KM. Melatonin in plants. Acta Physiol Plant. 2008;31:1–11.
38. Crooke A, Huete-Toral F, Colligris B, et al. The role and therapeutic potential of melatonin in age-related ocular diseases. J Pineal Res 2017;63.
39. Hainsworth DP, Gangula A, Ghoshdastidar S, et al. Diabetic Retinopathy Screening Using a Gold Nanoparticle-Based Paper Strip Assay for the At-Home Detection of the Urinary Biomarker 8-Hydroxy-2'-Deoxyguanosine. Am J Ophthalmol. 2020;213:306–19.
40. Sukul P, Schubert JK, Zanaty K, et al. Exhaled breath compositions under varying respiratory rhythms reflects ventilatory variations: translating breathomics towards respiratory medicine. Sci Rep. 2020;10:14109.

## Figures



**Figure 1**

(A) Heatmap for intensities of top 25 DEMs between DM and DR patients with the smallest paired t-test FDR-adjusted q value. Euclidean distance metric and Ward's clustering method were applied for the hierarchical clustering of patients. Red represents increased intensities and blue decreased intensities. Metabolites of the same category are expressed together. (B) Statistics table for the top 25 DEMs. Fold changes are presented as the average of DR divided by DM. Q value represents the p-value after Benjamini–Hochberg FDR correction. VIP scores are calculated by the PLS-DA model. (C) Correlation matrix of top 25 DEMs between DM and DR patients with Spearman correlation coefficient. Blue represents positive correlations and red represents negative correlations. The magnitude of the correlation coefficient is proportional to color intensity and diameter. (D) Violin plots of DEMs selected by LASSO regression showing significant differences between DM and DR patients. Abbreviations: DEMs: differentially expressed metabolites; DM: type 2 diabetes mellitus (T2DM) without diabetic retinopathy; DR: T2DM with diabetic retinopathy; FDR: false discovery rate; VIP, variable importance in the project; PLS-DA: partial least squares discriminant analysis; LASSO: least absolute shrinkage and selection operator.



**Figure 2**

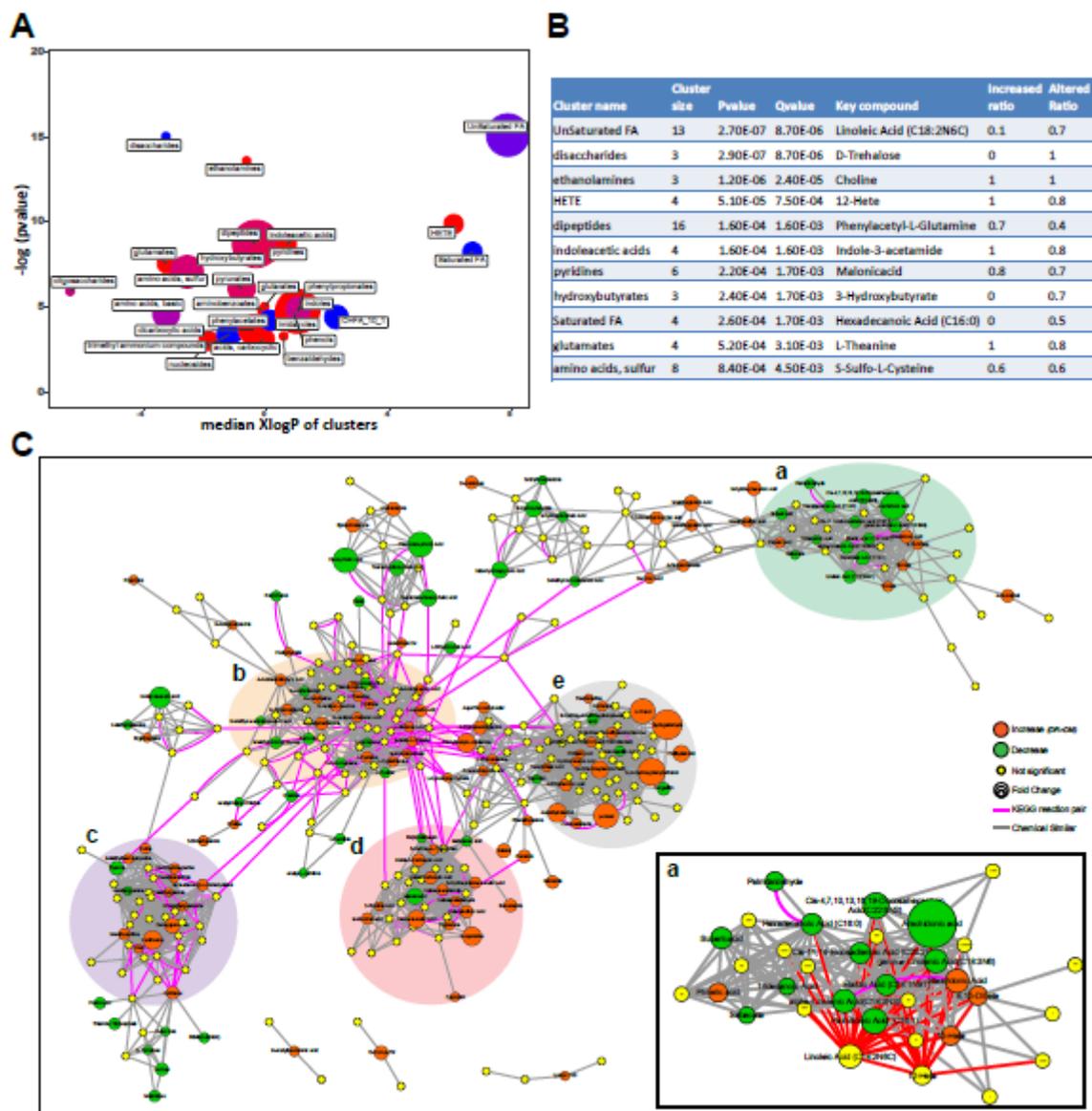
ROC curves distinguishing DM versus DR patients using linear SVM model with 10-fold cross-validation. The diagnosis performance of models is described by AUC (95% credible interval). The gray shading represents the 95% credible interval. Model 1: adjusted for duration of diabetes, SBP and insulin first; Model 2: adjusted for M1 plus DEMs selected by LASSO regression. Abbreviations: ROC curve: receiver operating characteristic curve; DM: type 2 diabetes mellitus (T2DM) without diabetic retinopathy; DR: T2DM with diabetic retinopathy; linear SVM: linear support vector machine; AUC: the area under the curve; SBP: systolic blood pressure; DEMs: differentially expressed metabolites; LASSO: least absolute shrinkage and selection operator.



**Figure 3**

(A) Bubble plot for MetPA on DEMs between DM and DR patients. Y-axis shows  $-\lg(p)$  calculated by hypergeometric test using over-representation analysis. X-axis and the bubble size show ODC using pathway topology. The bubble color represents different categories of metabolic pathways. (B) Three significant metabolic pathways with p-value less than 0.05 and one metabolic pathway with ODC larger than 0.5. (C) Metabolite changes in the three significant metabolic pathways. Red nodes indicate increased metabolites in DR patients as compared to DM patients, while the blue indicates a decrease.

Abbreviations: MetPA: metabolomics pathway analysis; DEMs: differentially expressed metabolites; DM: type 2 diabetes mellitus (T2DM) without diabetic retinopathy; DR: T2DM with diabetic retinopathy; ODC: out-degree centrality.



**Figure 4**

(A) ChemRICH analysis showed the most significantly altered metabolite clusters based on chemical similarity. Cluster size indicates the number of metabolites in each cluster. The proportion of increased or decreased metabolites compared to DM patients are shown by color (red = increased, purple = partly decreased, blue = decreased). Chemical enrichment statistics were calculated by the Kolmogorov–Smirnov test and only enrichment clusters with  $p < 0.05$  are shown in the bubble plot. (B) Statistics table for significant altered metabolite clusters (adjusted  $q$  value  $< 0.01$ ). (C) Metabolite network visualizing by MetaMapp. Orange nodes indicate increased metabolites in DR patients compared to DM patients, while the green nodes indicate a decrease. Node size indicates the magnitude of fold change. Purple edges denote KEGG reactant pair links, and grey edges symbolize Tanimoto chemical similarity over 700. The

graphic is mainly gathered into five modules: a mainly includes PUFAs; b mainly includes amino acids; c mainly includes nucleotides; d mainly includes indoles; e mainly includes benzene homologues. Abbreviations: ChemRICH: Chemical Similarity Enrichment Analysis for Metabolites; DM: type 2 diabetes mellitus (T2DM) without diabetic retinopathy; DR: T2DM with diabetic retinopathy; KEGG: Kyoto Encyclopedia of Genes and Genomes database.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplement.docx](#)