

Metabolomics-based multidimensional network biomarkers for diabetic retinopathy identification in patients with type 2 diabetes mellitus

Jingjing Zuo ^{1,2,3}, Yuan Lan,⁴ Honglin Hu,⁵ Xiangqing Hou,^{3,6} Jushuang Li,^{3,7} Tao Wang,^{3,7} Hang Zhang,⁸ Nana Zhang,⁵ Chengnan Guo,^{3,7} Fang Peng,^{3,7} Shuzhen Zhao,^{3,7} Yaping Wei,⁹ Chaonan Jia,¹⁰ Chao Zheng,¹¹ Guangyun Mao^{1,2,3,7}

To cite: 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 Diab Res Care* 2021;**9**:e001443. doi:10.1136/bmjdr-2020-001443

► Supplemental material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bmjdr-2020-001443>).

JZ, YL and HH contributed equally.

Received 5 April 2020
Revised 3 January 2021
Accepted 26 January 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to
Dr Guangyun Mao;
mgy@wmu.edu.cn and
Professor Chao Zheng;
chao_zheng@zju.edu.cn

ABSTRACT

Introduction Despite advances in diabetic retinopathy (DR) medications, early identification is vitally important for DR administration and remains a major challenge. This study aims to develop a novel system of multidimensional network biomarkers (MDNBs) based on a widely targeted metabolomics approach to detect DR among patients with type 2 diabetes mellitus (T2DM) efficiently.

Research design and methods In this propensity score matching-based case-control study, we used ultra-performance liquid chromatography-electrospray ionization-tandem mass spectrometry system for serum metabolites assessment of 69 pairs of patients with T2DM with DR (cases) and without DR (controls). Comprehensive analysis, including principal component analysis, orthogonal partial least squares discriminant analysis, generalized linear regression models and a 1000-times permutation test on metabolomics characteristics were conducted to detect candidate MDNBs depending on the discovery set. Receiver operating characteristic analysis was applied for the validation of capability and feasibility of MDNBs based on a separate validation set.

Results We detected 613 features (318 in positive and 295 in negative ESI modes) in which 63 metabolites were highly relevant to the presence of DR. A panel of MDNBs containing linoleic acid, nicotinic acid, ornithine and phenylacetylglutamine was determined based on the discovery set. Depending on the separate validation set, the area under the curve (95% CI), sensitivity and specificity of this MDNBs system were 0.92 (0.84 to 1.0), 96% and 78%, respectively.

Conclusions This study demonstrates that metabolomics-based MDNBs are associated with the presence of DR and capable of distinguishing DR from T2DM efficiently. Our data also provide new insights into the mechanisms of DR and the potential value for new treatment targets development. Additional studies are needed to confirm our findings.

INTRODUCTION

It is reported that the number of people with diabetes mellitus (DM) has reached 415 million worldwide by 2015 and is expected to rise to 642 million in 2040.¹ With the

Significance of this study

What is already known about this subject?

- Diabetic retinopathy (DR) is a leading microvascular complication of diabetes mellitus, a major disease linked to metabolism.
- Optimal biomarkers with sufficient sensitivity and specificity are vitally important for the effective prevention and control of DR.

What are the new findings?

- We discovered that a highly specific and sensitive multidimensional network biomarkers (MDNBs) system containing linoleic acid, nicotinic acid, ornithine and phenylacetylglutamine is capable of detecting patients with DR with high sensitivity and specificity.
- Pathways, including linoleic acid metabolism, alanine, aspartate and glutamate metabolism and phenylalanine metabolism, may play a vital role in the initiation and development of DR.

How might these results change the focus of research or clinical practice?

- Our findings may provide new insights into better understanding of the pathogenesis of DR and hold translational value in DR administration and the development of novel treatments.

advancement of DM treatment and more patients with DM living longer, the number of people with diabetic retinopathy (DR) is increasing rapidly worldwide, especially in Asia, including China.²⁻⁴ The annual incidence and progression of DR ranged from 2.2% and 3.4% in 1980 to 12.7% and 12.3% in 2018, respectively.⁵ Yau *et al*² believed that the overall prevalence of all types of DR was 34.6% and about 93 million people were affected by DR worldwide in 2012. A meta-analysis revealed that the overall prevalence of DR in China was estimated to be 1.14%.⁶ According to Handan Eye Study, about 9.2

million Chinese rural people were suffering from DR.⁷ DR has become one of the most common microvascular complications of DM and remains a leading cause of vision impairment and blindness among working-age adults.⁵

As a devastating and progressive disease, DR can be generally divided into two stages according to its severity: non-proliferative DR (NPDR) and proliferative DR (PDR).⁴ NPDR is the earlier stage than PDR and characterized by microaneurysms, cotton-wool spots, intraretinal microvascular abnormalities, hard exudates as well as venous beading, whereas PDR is hallmarked by neovascularization of the optic disc or elsewhere, preretinal and vitreous hemorrhage.³ These clinical manifestations can only be detected by careful examination of the fundus, especially the peripheral fundus examination after mydriasis, which is often rejected by patients. In practice, NPDR usually remains undetected until it progresses to more advanced stages.⁵ Although current therapies including laser photocoagulation and intravitreal anti-vascular endothelial growth factor (anti-VEGF) injection can prevent visual loss when undertaken timely and appropriately, laser photocoagulation is inherently destructive with unavoidable side effects and anti-VEGF injection is expensive and needs repeated applications.⁴ Therefore, it is a pressing need to detect DR as early as possible in order to delay its progression, improve DR administration and significantly increase the effectiveness of therapy.

Metabolome analysis is a powerful approach to identify and quantify the entire collection of living systems' intracellular and extracellular low molecular-weight metabolites response to biological stimuli or genetic modification.^{8,9} The dynamic metabolites mentioned above are considered the downstream products from proteome and transcriptome, and their changes may reveal crucial information relevant to the individual's health status. Recently, several hospital-based metabolomics studies have been conducted to find efficient biomarkers and discover underlying mechanisms in diseases, including DM,¹⁰ cardiovascular disease,¹¹ Alzheimer's disease¹² and retinal diseases.¹³ Specific technique platforms containing nuclear magnetic resonance spectroscopy, gas chromatography-mass spectrometry (GC-MS) and liquid chromatography-mass spectrometry (LC-MS) have been developed for metabolomics studies in the past several decades.⁸ Among them, the LC-MS system is more commonly used for metabolomic profiling than others because of its high sensitivity and selectivity. Although previous metabolomics studies have been applied in DR,^{14,15} few focused on the metabolomic underpinnings of DR early identification, which is widely accepted to be vitally necessary for DR administration efficiently. Abundant evidence has disclosed that serum or plasma metabolites and lipids are associated with diabetes and the development of retina hyperpermeability.^{14,16-18} We assume that there might be a panel of metabolomic multidimensional network biomarkers

(MDNBs) for distinguishing DR from patients with type 2 diabetes mellitus (T2DM) efficiently.

To address this gap and thoroughly understand the unique, complicated metabolic state of DR, we, therefore, performed the current study using a widely targeted metabolomics approach by ion-pair reversed-phase ultra-performance liquid chromatography-electrospray ionization-tandem mass spectrometry (UPLC-ESI-MS/MS) in the serum of patients with T2DM with DR and those without DR, seeking to develop a new system of MDNBs with relatively high sensitivity and specificity for DR identification. We also aim to find underlying pathways for further understanding of the disease mechanisms.

MATERIALS AND METHODS

Study design and population

This was a propensity score matching (PSM)-based case-control study. To decrease the potential impact of the selection bias, we selected participants from the outpatient or inpatient of endocrinology departments in the Second Affiliated Hospital of Wenzhou Medical University (WMU) and the First Affiliated Hospital of Anhui Medical University in China. The diagnosis of DM was based on standard criteria recommended by WHO since 1999.

The inclusion criteria were as follows: (1) T2DM; (2) ≥ 35 years old. Participants with following situation would be excluded: (1) any other eye diseases or history of eye surgery; (2) cancer, infectious disease, mental disorder, heart failure, severe hypertension (systolic blood pressure ≥ 180 mm Hg or diastolic blood pressure ≥ 110 mm Hg) and any other severe chronic systemic disease; (3) poor quality of fundus photographs, which were not clear for DR diagnosis. Only those following each of the inclusion criteria and none of the exclusion criteria were potential participants.

All participants received detailed ophthalmic examinations and were separately assessed based on their retinal photographs by two experienced ophthalmologists (more details in the online supplemental methods). Among them, patients with T2DM with DR in at least one eye were determined as the cases, and others without any DR were defined as the controls. To reduce the impacts of potential confounders on our conclusions, we used a one-to-one optimal PSM method¹⁹ to minimize the difference within pairs based on a multiple logistic model containing age, sex, body mass index (BMI) and glycated hemoglobin. Ultimately, we included 69 pairs of participants in the current study. Besides, a total of 69 healthy volunteers without diabetes were also matched from the routine physical examination cohort in the Second Affiliated Hospital of WMU, using PSM by another multiple logistic regression model containing age, sex and BMI. The flowchart of the recruitment procedure is shown in online supplemental figure 1.

Data collection

Quality assurance and quality control were well considered to increase the quality of our data. All subjects were invited to complete a semi-structured questionnaire by a face-to-face interview. Age was calculated with the formula: (investigation date–birthday)/365. Information on the occupation, education level, disease history, therapy history for DM medication, smoking habits, alcohol consumption and other demographic features was also obtained. The collection of other data containing clinical manifestation and biochemical assessments was performed by two systematically trained investigators strictly following the specific standardized operation procedures of the current study.

Sample collection, storage and assessment

After at least 8-hour fasting, 6 mL of venous blood was collected under complete aseptic precautions from all the subjects and stored at 4°C. The serum was separated by centrifugation at 2000 rpm for 10 min within 30 min, then transferred into a 1.5 mL sterile tube and stored at –86°C ultra-low temperature freeze immediately. After the enrollment of all participants, serum samples were shipped in dry ice to the central laboratory of Metware, a professional and experienced metabolomics institution in Wuhan, China. Samples arrived in <24 hours and were stored immediately at –86°C until further processing. The qualitative and relative quantitative assessments of widely targeted metabolites in the serum samples were performed using UPLC-ESI-MS/MS system. The MS2 spectral tag library including the MS and MS/MS spectra was used.²⁰ Modified protocols for serum samples analysis were according to the method described previously²¹ and summarized in the online supplemental methods.

Data processing

After the serum metabolites assessment, the UPLC-ESI-MS/MS data acquired by Analyst Software V.1.6.3 (AB Sciex) were preprocessed (conversion, peak detection, retention time correction and peak alignment) by MultiQuant Software (AB Sciex), and then processed using MetaboAnalyst V.4.0 (<https://www.metaboanalyst.ca/>) and STATA MP V.15.0 (StataCorp, College Station, Texas, USA) including missing value processing,²² filtering²³ and normalization. Details are available in the online supplemental methods.

Statistical analysis

The sample size of this study was determined using STATA MP V.15.0, referring to a previous study.¹⁴ The current sample size (69 participants per group) was sufficient for the detection of differential metabolites between the two groups, with 90% of the power and a type I error of 0.05 (two-sided). Demographic and clinical data were presented as the mean±SD or median (the first quartile, the third quartile) depending on their distributions for continuous data and frequency (percentage) for categorical data. Differences between the two groups were

compared using paired t-test, Mann-Whitney U test or χ^2 test based on the requirements of each statistical approach.

Due to the multidimensional nature of metabolomics data and the number of metabolites being much larger than that of the samples, a series of univariate and multivariate analysis was applied as follows: the univariate methods containing paired t-test and fold change (FC, defined as the ratio of cases by controls) analysis to represent the differences between the cases and controls. The false discovery rate (FDR) adjustment was also performed to reduce the probability of false-positive results using the Benjamini-Hochberg method.

To achieve more robust findings, we first randomly shuffled the PSM dataset to avoid a correlation between consequent elements.²⁴ Then, the dataset was randomly split at a ratio of 2:1 into a discovery set (46 pairs) for training the model and a separate validation set (23 pairs) to test its performance. Participants in each set would never be included in another one, and both the discovery and validation sets were consistent with the original distribution. For multivariate analysis, principal component analysis (PCA) was applied to find a small set of variables that represent and explain the original data and simplify a dataset by reducing dimensionality. Meanwhile, orthogonal partial least squares discriminant analysis (OPLS-DA) was conducted to find a linear combination of features for the classification of cases and controls. To evaluate the robustness of the detected OPLS-DA model, we calculated R^2Y (cum) and Q^2 (cum) by 10-fold cross-validation, which independently represents the explanatory power and the predictive power of the model. An additional permutation test (1000 times) was also carried out to ensure the model without overfitting. Finally, the score value of variable importance for the projection (VIP) was acquired from the OPLS-DA model to individually select variables having a large influence on the contribution to the explanatory power of the model. Depending on the discovery set, the criteria of differential metabolites screening were defined as q value <0.05, FC value >1.2 or <0.8 as well as VIP >1. Meanwhile, a heatmap was also drawn to show the patterns of differential metabolites between the cases and controls.

Multivariable generalized linear regression models were performed to comprehensively investigate the associations between each differential metabolite and the presence of DR, in which ORs with 95% CI were separately calculated with the metabolite intensity as a continuous variable (IQR) and categorical variable (quartile). The capability and feasibility of each detected metabolite were evaluated using receiver operating characteristic (ROC) analysis based on the validation set. An optimal MDNBs system with higher sensitivity and specificity was screened depending on the area under the curve (AUC) from different combinations of detected differential metabolites.

All statistical tests were two-sided, and the p value ≤ 0.05 was set as the significant level.

Table 1 Clinical and demographic characteristics of the study population

Variables	Discovery set			Validation set		
	DM (n=46)	DR (n=46)	P value	DM (n=23)	DR (n=23)	P value
Age, years	55.3±11.1	58.1±10.5	0.205	53.0 (46.0, 55.0)	55.0 (50.0, 65.0)	0.073
Male, # (%)	24 (52)	22 (48)	0.677	14 (61)	14 (61)	1.000
BMI, kg/m ²	24.3±3.4	24.5±3.8	0.794	24.7±2.8	24.8±2.9	0.882
FPG, mmol/L	9.3±3.6	8.8±3.4	0.537	7.8 (7.0, 12.8)	8.5 (6.4, 9.6)	0.676
HbA1c, %	10.2±2.3	10.0±2.0	0.609	9.9±2.3	9.7±1.5	0.705
HDL, mmol/L	1.2±0.5	1.2±0.4	0.931	1.0±0.3	1.0±0.3	0.955
LDL, mmol/L	2.7±1.0	2.6±1.2	0.722	2.6±1.0	2.5±0.8	0.613
TG, mmol/L	1.6 (1.0, 2.2)	1.4 (1.0, 1.8)	0.234	1.6 (1.2, 2.0)	1.6 (1.2, 2.1)	0.684
TC, mmol/L	4.8±1.2	4.6±1.6	0.452	4.6±1.1	4.4±1.0	0.540
DM duration, years	7.5±6.5	11.7±6.4	0.002	7.8±6.1	11.3±7.3	0.083
Therapy history, # (%)			<0.001			0.574
No	17 (37)	2 (4)		3 (13)	1 (4)	
Yes	25 (54)	41 (89)		17 (74)	19 (83)	
Unknown	4 (9)	3 (7)		3 (13)	3 (13)	
Hypertension, # (%)			0.524			0.200
No	29 (63)	24 (52)		18 (78)	14 (61)	
Yes	17 (37)	22 (48)		5 (22)	9 (39)	
Smoking habits, # (%)			0.152			0.267
Non-smokers	30 (65)	25 (54)		11 (48)	11 (48)	
Current smokers	13 (28)	13 (28)		6 (26)	8 (35)	
Ex-smokers	1 (2)	7 (15)		5 (22)	1 (4)	
Unavailable	2 (4)	1 (2)		1 (4)	3 (13)	
Alcohol consumption, # (%)			0.346			0.665
Non-drinkers	21 (46)	19 (41)		12 (52)	10 (44)	
Current drinkers	21 (46)	19 (41)		9 (39)	8 (35)	
Ex-drinkers	2 (4)	7 (15)		1 (4)	2 (9)	
Unavailable	2 (4)	1 (2)		1 (4)	3 (13)	

Smoking habits were classified into three categories: (1) non-smokers, who never smoke; (2) ex-smokers, who had smoked before but quit smoking for at least 1 month; (3) current smokers, currently smoking over 3 months.

The alcohol consumption was also determined as three classes: (1) non-drinkers, who never drink; (2) ex-drinkers, who had drunk before but quit drinking for at least 1 month; (3) current drinkers, who drink for over 3 months currently.

Continuous data obeying normal or similar normal distribution were described as mean±SD, and the paired t-test was applied to compare the differences between the two groups. Otherwise, median (first quartile, third quartile) and Mann-Whitney U tests were used. Categorical data were presented as the number of cases (%), and the χ^2 test was used to compare the differences between the cases and controls.

All the percentage presentations were kept as integer accuracy.

Therapy history is for diabetes medications.

BMI, body mass index; DM, diabetes mellitus; DR, diabetic retinopathy; FPG, fasting plasma glucose; HbA1c, glycated hemoglobin; HDL, high-density lipoprotein; LDL, low-density lipoprotein; TC, total cholesterol; TG, triglyceride.

RESULTS

Characteristics of participants

Among 138 participants (69 pairs) in the cases and controls, 74 were men (46 in the discovery set and 28 in the validation set). The clinical and demographic characteristics of the participants stratified by the discovery and validation sets were presented in [table 1](#). In this PSM-based case-control study, DR cases were more likely to have longer duration suffering from T2DM and higher

proportion receiving medications than those of controls. In the DR group, there were 9 (13%) participants with mild NPDR, 31 (45%) with moderate NPDR, 20 (29%) with severe NPDR and 9 (13%) with PDR. Other characteristics of participants between the two groups in both the discovery and validation sets were comparable, indicating that the potential confounding impacts on our findings had been well adjusted for depending on the PSM approach.

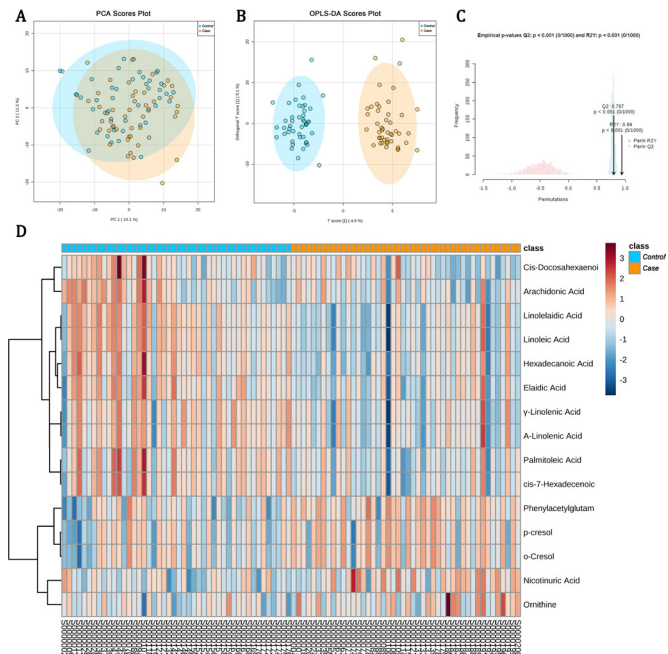


Figure 1 Metabolomic data analyses based on the discovery set. (A) Score plot of the principal component analysis (PCA) model; (B) score plot of the orthogonal partial least squares discriminant analysis (OPLS-DA) model, and the R^2 was 0.97 and the Q^2 was 0.65 calculated by 10-fold cross-validation; (C) results of 1000-time permutation test of the OPLS-DA model, and the empirical p values for R^2Y and Q^2 were all <0.001 . The blue dots represent the samples of type 2 diabetes mellitus (T2DM) without diabetic retinopathy (DR) (controls), and the yellow dots indicate the samples of DR (cases); (D) heatmap for intensities of metabolomics-based markers identified in the discovery set. The heatmap presents relative peak areas of the final 15 detected differential metabolites between DR and T2DM without DR based on the discovery set. With colors changing from red to blue, red and blue colors mean upregulated and downregulated metabolites in the serum sample of patients with DR, respectively.

PCA and OPLS-DA

A total of 613 features (318 in ESI^+ and 295 in ESI^- modes) were detected depending on the UPLC-ESI-MS/MS data (more details in the online supplemental results) in the discovery set after being processed. As can be seen clearly in the PCA score plot (figure 1A), we observed a shift between the cases (the yellow dots) and controls (the blue dots). Figure 1B also clearly revealed that the cases and controls could be successfully distinguished from each other depending on the OPLS-DA scores plot. The parameters of the OPLS-DA model (R^2Y (cum)=0.97, Q^2 (cum)=0.65) calculated by 10-fold cross-validation also suggested that the model had satisfied robustness. In the validation of this OPLS-DA model based on a 1000-time permutation test, the empirical p values for R^2Y and Q^2 were all <0.001 (figure 1C), which indicated again that the observed statistic is not part of the distribution formed by those from the permuted data, thereby suggesting that the OPLS-DA model can be used to detect

markers differentiating between DR and T2DM without DR successfully.

Differential metabolites detection

Under the criteria of detection which was defined as the combination of VIP >1 , FC value <0.8 or >1.2 and FDR-adjusted p value <0.05 , the intensity of 63 unique metabolites in the discovery set was significantly different between the cases and controls. Of the 63 unique metabolites, 42 metabolites in the cases were significantly increased, and the other 21 were undoubtedly decreased as compared with the controls. As we were mainly interested in the patterns of endogenous metabolites which could be driving systemic biological factors, a total of 48 metabolites that might be exogenous to humans (including medications, food additives and buffering agents) were excluded from the following data analysis. The remaining 15 metabolites are presented in table 2 and figure 1D. Also, pathway analysis was performed and strongly indicated that the occurrence and development of DR would be associated with the metabolisms of linoleic acid, alpha-linolenic acid, cysteine and methionine, arginine and proline, arachidonic acid and phenylalanine (online supplemental figure 2).

Association of metabolite intensity with the presence of DR

Depending on the discovery set, the associations of each detected differential metabolite intensity with the presence of DR could be found in table 3 and online supplemental table 1. After adjusting for the impacts due to some potential confounding factors containing age, therapy history and smoking habits, OR (95% CI) for four distinct metabolites including linoleic acid, nicotinic acid, ornithine and phenylacetylglutamine (PAG) were 0.22 (0.11 to 0.45), 2.50 (1.47 to 4.25), 1.80 (1.12 to 2.89), 4.64 (1.98 to 10.88) with per IQR elevation, respectively. These all revealed an individual independent association of DR occurrence with them. Meanwhile, the same associations were also observed between the odds of DR and quartile-classified metabolites intensities, and apparent linear trend between the outcome and four exposures could be observed individually (p for trend <0.05 , table 3).

In addition, online supplemental figure 3 showed the changes in the 15 metabolites mentioned above among the cases, T2DM controls and healthy controls. The intensities of PAG, p-cresol, o-cresol, nicotinic acid and ornithine revealed obvious accumulating trends from healthy volunteers to patients with DM and DR cases. In contrast, the levels of linoleic acid (C18:2N6C), linolelaidic acid (C18:2N6T), palmitoleic acid (C16:1), gamma-linolenic acid (C18:3N6), alpha-linolenic acid (C18:3N3), cis-7-hexadecenoic acid, hexadecanoic acid (C16:0), elaidic acid (C18:1N9T), cis-4,7,10,13,16,19-docosahexaenoic acid (C22:6N3) and arachidonic acid, all of which were polyunsaturated fatty acids (PUFAs), disclosed apparent decreasing trends from healthy volunteers to patients with DM and DR cases.

Table 2 Differential metabolites depending on the discovery set

Metabolites	Adduction	Q1 (Da)	Q3 (Da)	RT (min)	m/z	HMDB ID	P value	FDR adjusted		Trends	AUC (95% CI)	
								FC	VIP			
Phenylacetylglutamine	[M+H] ⁺	265.10	130.00	3.06	264.111	HMDB06344	5.60E-05	0.000898	2.01	2.85	↑	0.74 (0.64 to 0.84)
p-Cresol	[M-H] ⁻	107.06	92.00	4.07	108.058	HMDB01858	0.002617	0.020616	2.07	2.29	↑	0.68 (0.57 to 0.79)
o-Cresol	[M-H] ⁻	107.06	92.00	4.07	108.058	HMDB02055	0.002912	0.020112	2.00	2.22	↑	0.68 (0.57 to 0.79)
Linoleic acid (C18:2N6T)	[M-H] ⁻	279.20	279.20	11.10	280.240	HMDB06270	3.57E-07	1.23E-05	0.71	1.89	↓	0.80 (0.71 to 0.90)
Linoleic acid (C18:2N6C)	[M-H] ⁻	279.20	279.20	11.10	280.240	HMDB00673	5.91E-07	1.77E-05	0.72	1.85	↓	0.81 (0.72 to 0.90)
Palmitoleic acid (C16:1)	[M-H] ⁻	253.20	253.20	10.87	254.200	HMDB03229	0.000291	0.003535	0.64	1.74	↓	0.70 (0.60 to 0.81)
gamma-Linolenic acid (C18:3N6)	[M-H] ⁻	277.20	233.00	10.54	278.220	HMDB03073	3.89E-05	0.000647	0.71	1.70	↓	0.74 (0.64 to 0.85)
alpha-Linolenic acid (C18:3N3)	[M-H] ⁻	277.20	127.00	10.54	278.220	HMDB01388	7.45E-05	0.001114	0.72	1.64	↓	0.73 (0.63 to 0.84)
cis-7-Hexadecenoic acid	[M-H] ⁻	253.22	253.40	10.87	254.225	HMDB02186	0.000382	0.004289	0.68	1.62	↓	0.70 (0.60 to 0.81)
Hexadecanoic acid (C16:0)	[M-H] ⁻	255.20	255.20	11.65	256.240	HMDB00220	1.52E-05	0.000296	0.79	1.43	↓	0.76 (0.66 to 0.85)
Elaidic acid (C18:1N9T)	[M-H] ⁻	281.30	262.90	11.85	282.250	HMDB00573	0.000322	0.003705	0.77	1.38	↓	0.71 (0.60 to 0.81)
cis-4,7,10,13,16,19-Docosahexaenoic acid (C22:6N3)	[M-H] ⁻	327.20	283.30	10.81	328.240	HMDB02183	0.004646	0.028190	0.67	1.36	↓	0.67 (0.56 to 0.78)
Nicotinic acid	[M+H] ⁺	181.10	135.00	0.85	180.054	HMDB03269	0.000189	0.002571	1.23	1.24	↑	0.71 (0.61 to 0.82)
Arachidonic acid	[M-H] ⁻	303.20	259.30	10.97	304.200	HMDB01043	0.005288	0.030056	0.34	2.27	↓	0.66 (0.55 to 0.77)
Ornithine	[M+H] ⁺	133.09	70.10	0.64	132.090	HMDB00214	0.006361	0.033599	1.24	1.02	↑	0.63 (0.51 to 0.74)

AUC, area under the curve; FC, fold change; FDR, false discovery rate; HMDB, Human Metabolome Database; m/z, actual mass-to-charge ratio; RT, retention time; VIP, variable importance in the project.

Table 3 Association between the peak intensity of the four metabolites and the odds of diabetic retinopathy

Metabolites	N	Cases (%)	Crude		Adjusted*	
			OR (95% CI)	P value	OR (95% CI)	P value
Phenylacetylglutamine						
Per IQR			4.45 (2.10 to 9.43)	<0.001	4.64 (1.98 to 10.88)	<0.001
Quartiles						
Q ₁	23	3 (13)	1.00 (1.00 to 1.00)	Ref.	1.00 (1.00 to 1.00)	Ref.
Q ₂	23	13 (57)	8.67 (2.00 to 37.58)	0.004	12.05 (2.34 to 61.98)	0.003
Q ₃	23	13 (57)	8.67 (2.00 to 37.58)	0.004	11.78 (2.27 to 61.00)	0.003
Q ₄	23	17 (74)	18.89 (4.09 to 87.17)	<0.001	23.01 (3.98 to 133.20)	<0.001
P for trend				<0.001		0.001
Linoleic acid						
Per IQR			0.26 (0.14 to 0.49)	<0.001	0.22 (0.11 to 0.45)	<0.001
Quartiles						
Q ₁	23	19 (83)	1.00 (1.00 to 1.00)	Ref.	1.00 (1.00 to 1.00)	Ref.
Q ₂	23	15 (65)	0.40 (0.10 to 1.57)	0.186	0.30 (0.07 to 1.42)	0.129
Q ₃	23	9 (39)	0.14 (0.04 to 0.53)	0.004	0.11 (0.02 to 0.54)	0.006
Q ₄	23	3 (13)	0.03 (0.01 to 0.16)	<0.001	0.02 (0.00 to 0.14)	<0.001
P for trend				<0.001		<0.001
Nicotinuric acid						
Per IQR			2.32 (1.42 to 3.78)	0.001	2.50 (1.47 to 4.25)	0.001
Quartiles						
Q ₁	23	7 (30)	1.00 (1.00 to 1.00)	Ref.	1.00 (1.00 to 1.00)	Ref.
Q ₂	23	10 (44)	1.76 (0.52 to 5.91)	0.361	1.45 (0.39 to 5.41)	0.578
Q ₃	23	11 (48)	2.10 (0.63 to 7.01)	0.230	1.68 (0.48 to 5.84)	0.415
Q ₄	23	18 (78)	8.23 (2.18 to 31.13)	0.002	13.02 (2.79 to 60.71)	0.001
P for trend				0.002		0.002
Ornithine						
Per IQR			1.71 (1.09 to 2.69)	0.020	1.80 (1.12 to 2.89)	0.015
Quartiles						
Q ₁	23	10 (44)	1.00 (1.00 to 1.00)	Ref.	1.00 (1.00 to 1.00)	Ref.
Q ₂	23	9 (39)	0.84 (0.26 to 2.71)	0.765	0.74 (0.21 to 2.61)	0.643
Q ₃	23	10 (44)	1.00 (0.31 to 3.21)	1.000	1.04 (0.30 to 3.64)	0.948
Q ₄	23	17 (74)	3.68 (1.06 to 12.77)	0.040	4.46 (1.14 to 17.39)	0.031
P for trend				0.043		0.028

In the present study, the T2DM duration for each participant was obtained mainly by the self-report and highly correlated to age. Furthermore, it might be inaccurate since many people may be suffering from unnoticed T2DM for a long time. So, to avoid collinearity and overfitting, it was not included in the multiple generalized linear regression models, although it was not comparable between the cases and controls.

*Adjusted for age, smoking habits and therapy history.

Q₁, the first quartile; Q₂, the second quartile; Q₃, the third quartile; Q₄, the fourth quartile; Ref., reference; T2DM, type 2 diabetes mellitus.

MDNBs screening and evaluation of its capability on DR identification

Table 2 presented the capability of 15 metabolites on DR recognition depending on ROC analysis in the discovery set, respectively. In a series of different combinations of the 15 metabolites, which were developed to overcome the limitation of a single biomarker, we compared all other models with the total model, including all the 15

metabolites, as the reference to screen an optimal panel of MDNBs. According to the theory of Occam's razor,²⁵ entities should not be multiplied unnecessarily to avoid overfitting and potential collinearities. Therefore, we finally selected four uncorrelated metabolites based on the results of correlation analysis to fit two combined models with fewer variables but similar efficiency of distinguishing DR from their counterparts. As could be

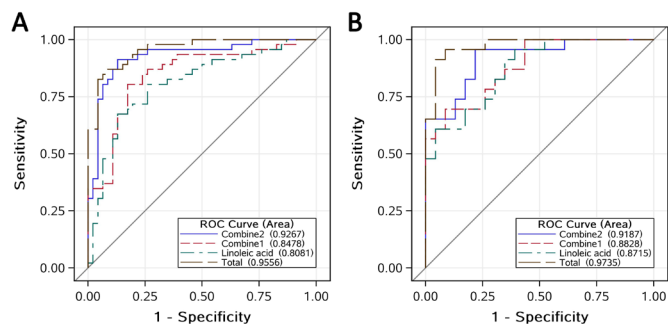


Figure 2 Combined receiver operating characteristic (ROC) curves of multidimensional network biomarkers (MDNBs) distinguishing subjects with diabetic retinopathy (DR) versus subjects with type 2 diabetes mellitus without DR in the discovery set (A) and validation set (B). The combined2 model (MDNBs) contains linoleic acid, nicotinuric acid, ornithine and phenylacetylglutamine; the combined1 model contains linoleic acid and nicotinuric acid; the total model contains all the 15 potential metabolite biomarkers.

seen in [figure 2A](#), the AUC (95% CI) for linoleic acid, combined1, combined2 and the total model were 0.81 (0.72 to 0.90), 0.85 (0.77 to 0.93), 0.93 (0.87 to 0.98) and 0.96 (0.92 to 0.99), respectively. As no significant difference was found between the combined2 model and the total model (online supplemental table 2), the combined2 model including linoleic acid, nicotinuric acid, ornithine and PAG was determined as the candidate MDNBs system with sufficient sensitivity (91%) and specificity (87%).

The evaluation of the candidate MDNBs system based on the validation set is presented in [figure 2B](#). Its AUC (95% CI), sensitivity and specificity were 0.92 (0.84 to 1.0), 96% and 78%, respectively. Moreover, no significant difference between the MDNBs and the total model was detected (online supplemental table 2). This strongly suggested that the panel of linoleic acid, nicotinuric acid, ornithine and PAG could be used as a sensitive and specific system of MDNBs to distinguish DR from their counterparts in the population with T2DM.

Sensitive analysis

After excluding nine PDR and their paired controls, the remaining 60 pairs of participants were also randomly split into a discovery set (40 pairs) and a validation set (20 pairs) at the ratio of 2:1, which has been widely applied in data mining practice including machine learning. Then we repeated the validation of this detected MDNBs system. The associated AUC (95% CI), sensitivity and specificity were 0.93 (0.88 to 0.99), 90% and 88% in the discovery set, and 0.90 (0.81 to 0.99), 75% and 90% in the validation set, respectively (online supplemental figure 4). The capability of the MDNBs on DR identification was consistent with the former findings.

DISCUSSION

In this PSM-based, multicenter case-control study, we extensively assessed the serum metabolomics profiles on

the presence of DR depending on the widely targeted metabolomics approach. A total of 15 endogenous metabolites were detected to be significant with the odds of DR. The results of pathway analysis revealed that DR occurrence was mainly involved in metabolisms of linoleic acid, alpha-linolenic acid, cysteine and methionine, arginine and proline, arachidonic acid as well as phenylalanine, which suggested that metabolic disorder would play an important role in the initiation and development of DR. Among them, the panel of linoleic acid, nicotinuric acid, ornithine and PAG were highly specific and sensitive to distinguishing DR from T2DM and could be an optimal MDNBs system for DR identification.

Available evidence suggests that PUFAs have a unique ability to modulate capillary integrity, neovascularization and inflammation in the retina.²⁶ PUFAs deficiency may lead to increased inflammation in patients with diabetes and primarily increase the likelihood of the presence, development and severity of DR.²⁷ In the present study, 10 PUFAs containing linoleic acid (C18:2n6C), arachidonic acid and others revealed an apparent decreasing trend from healthy participants to T2DM and patients with DR. Our findings are consistent with the studies mentioned above and suggest that elevated PUFAs may effectively decrease the likelihood of DR and postpone its progression. A similar result is also observed in a plasma GC-MS-based metabolomics study.²⁸ However, in another plasma metabolomics study, Chen *et al*¹⁴ report that no significant differences of linoleic acid and arachidonic acid intensities exist between patients with DR and controls with diabetes. They suppose that the low level of arachidonic acid might be modified by the use of antihypertensive medications. To clarify this potential effect of the antihypertensive medications on the PUFAs levels, we carried out a subanalysis in all participants with antihypertensive medications between DR cases and the controls. We observed that the intensities of these two metabolites were also significantly lower in DR cases than those of the controls (28 cases vs 18 control, $p < 0.001$ for linoleic acid and $p = 0.012$ for arachidonic acid). The inconsistency in the findings by Chen *et al* as compared with other studies may be owing to the different metabolomics-based approaches and distinct nationalities of the participants. Additional studies are needed to confirm this.

PAG is a metabolite from amino acid fermentation and results from glutamine conjugation of phenylacetic acid. The previous study suggests that elevated serum PAG is a strong and independent risk factor for overall mortality and cardiovascular disease, pointing to the relevance of microbial metabolism and tubular secretion, irrespective of protein binding.²⁹ However, no report has investigated the relationship between PAG and DR. And in the present study, we observed that elevated serum PAG would significantly increase the odds of DR (OR 4.64, 95% CI 1.98 to 10.88, with per IQR increase of PAG). As a non-proteinogenic amino acid and an intermediate in metabolic processes, ornithine is a central part of the urea cycle and essential on the protective effect of insulin

in DM through ornithine decarboxylase/polyamine systems.³⁰ However, no reports of the role of ornithine in DR can be found. In the current study, we detected that the intensity of serum ornithine was positively linked to elevated odds of DR (OR 1.80, 95% CI 1.12 to 2.89, with per IQR increase of ornithine).

Of the potential biomarkers found in the current study, nicotinic acid is the primary catabolic product of nicotinic acid, which is mainly related to the production of nicotinamide adenine dinucleotide (NAD⁺) and nicotinamide adenine dinucleotide phosphate (NADP⁺), the core enzymes for oxidation-reduction reactions.³¹ Huang *et al*³² have reported that the urine nicotinic acid intensity in diabetes (n=25) was higher than those without DM (n=37) in a urinary metabolomics study depending on the high-performance liquid chromatography-time-of-flight mass spectrometry. Furthermore, nicotinic acid has also been reported to alter lipid metabolism and increase the possibility of insulin resistance,³³ which is widely considered as the vital pathological mechanism of metabolic syndrome. In the present study, the highest level of serum nicotinic acid was discovered in DR cases followed by T2DM without DR and healthy controls (online supplemental figure 3). To the best of our knowledge, this is the first report to suggest that serum nicotinic acid level is accumulating during the development of DR, which may contribute to the understanding of the underlying pathological mechanisms of DR initiation and progression.

Several strengths can be found in the current study as compared with previous studies. First, the structure and intensities of 613 features are detected using a widely targeted metabolomics approach, which is believed to obtain more precise and accurate data on predefined metabolites as compared with traditional untargeted metabolomics study. Second, participants included in the current study have been matched by the PSM method, in which some potential confounding factors will be more comparable between the cases and controls. Third, a comprehensive analysis of the metabolomics characteristics of DR is conducted using several appropriate statistical methods. Finally, a sensitive and specific panel of MDNBs for DR screening, instead of a single biomarker, is trained depending on the discovery set and validated by both a 10-fold cross-validation test and a 1000-times permutation test, which suggests that our findings are robust and credible. Additional rigorous, designed studies would be needed to verify our findings.

This study also has limitations, such as the relatively small sample size and the participants only from two hospitals in China. Although a small sample size will affect the reliability of the results to some extent since it will lead to insufficient power to reject the null hypothesis, this does not mean that the current sample size is insufficient since it has been carefully estimated during the study design period. Furthermore, the participants of this study are enrolled from two top hospitals in both Zhejiang and Anhui provinces, which cover >120 million people. This

will decrease the potential impact of the selection bias on our findings to some extent. Another limitation is that our data of metabolites might be affected by the current body nutrition and metabolic status of each individual patient since this is a case-control study. Further longitudinal studies are needed to confirm our findings.

In conclusion, to the best of our knowledge, this is the first report to suggest that a system of metabolomics-based MDNBs containing linoleic acid, nicotinic acid, ornithine and PAG are associated with the presence of DR and capable of distinguishing DR from T2DM efficiently. Based on our results, we demonstrate that metabolomics-based MDNBs would like to be optimal biomarkers of DR and very suitable for distinguishing DR from T2DM efficiently, especially in those at the early stage of DR or without some typical signs and symptoms to be timely diagnosed via traditional routine techniques in clinical practice. Our findings will provide a new option to forecast some patients with T2DM at the high risk to develop DR before the occurrence and development, to detect many early stage patients with DR among T2DM, to perform early diagnosis of DR and provide efficient medications to patients with DR in time. All these applications will be very beneficial to DR administration and guide ophthalmologists to select better medications for their patients, which will provide a solid foundation for novel treatment targets screening in future studies. In a word, our findings emphasize the clinical and public health policy relevance of precise DR administration. Our findings also contribute to the current knowledge of DR pathophysiology by highlighting the role of lipid metabolism, especially the linoleic acid and arachidonic acid pathway, which indicates potential novel targets for DR treatment.

Author affiliations

¹Eye Hospital and School of Ophthalmology and Optometry, Wenzhou Medical University, Wenzhou, Zhejiang, China

²National Clinical Research Center for Ocular Diseases, Wenzhou, Zhejiang, China

³Center on Evidence-Based Medicine & Clinical Epidemiological Research, School of Public Health & Management, Wenzhou Medical University, Wenzhou, Zhejiang, China

⁴Department of Ophthalmology, Pingxiang People's Hospital, Pingxiang, Jiangxi, China

⁵Department of Endocrinology, First Affiliated Hospital of Anhui Medical University, Hefei, Anhui, China

⁶Faculty of Health Sciences, University of Macau, Macau, China

⁷Department of Preventive Medicine, School of Public Health & Management, Wenzhou Medical University, Wenzhou, Zhejiang, China

⁸Department of Endocrinology, Wenzhou Medical University Second Affiliated Hospital, Wenzhou, Zhejiang, China

⁹Beijing Advanced Innovation Center for Food Nutrition and Human Health, College of Food Science and Nutritional Engineering, China Agricultural University, Beijing, China

¹⁰Taizhou Municipal Center for Disease Control and Prevention, Taizhou, Zhejiang, China

¹¹The Second Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou, Zhejiang, China

Acknowledgements The authors would like to thank the colleagues from Wenzhou Medical University, Anhui Medical University and Zhejiang University School of Medicine for their hard works and collaborations. The authors would

like to thank Metware, Inc., for their careful assay of all the serum samples and comments on the data analysis. The authors would like to thank all study participants for donating their contribution and time. The authors would also like to thank Dr Clara Pak from the University of Rochester Medical Center for her valuable suggestions and edits to the manuscript.

Contributors JZ and YL conducted the research, analyzed the data and wrote the manuscript. XH and YW analyzed the data. JL, TW, HZ, NZ, CG, FP and SZ contributed to the epidemiological investigation, sample handling, data management and analysis. CJ repeated the data analysis independently. GM, CZ and HH designed the study, thoroughly reviewed and edited the manuscript.

Funding This study was funded by Zhejiang Basic Public Welfare Research Project (LGF19H260011), Wenzhou Basic Public Welfare Research Project (Y20180201), the Initial Scientific Research Fund (KYQD170301), the Major Project of the Eye Hospital of Wenzhou Medical University (YNZD201602) and the academician's science and technology innovation program in Zhejiang province (2018R413182). Part of this work was also funded by the National Nature Science Foundation of China (81670777) and Natural Science Foundation of Zhejiang Province (LZ19H020001).

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval The protocol had been carefully reviewed and approved by the Ethics Committee of the Eye hospital of WMU before the study (Number: KYK (2017) 46). The procedures strictly followed the tenets of the Declaration of Helsinki. All participants in the current study were voluntary and provided written informed consents.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. The raw/processed data cannot be shared at this time as the data also form part of an ongoing study but are available from corresponding author on reasonable request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Jingjing Zuo <http://orcid.org/0000-0001-6093-2684>

REFERENCES

- Ogurtsova K, da Rocha Fernandes JD, Huang Y, *et al.* IDF diabetes atlas: global estimates for the prevalence of diabetes for 2015 and 2040. *Diabetes Res Clin Pract* 2017;128:40–50.
- Yau JWY, Rogers SL, Kawasaki R, *et al.* Global prevalence and major risk factors of diabetic retinopathy. *Diabetes Care* 2012;35:556–64.
- Cheung N, Mitchell P, Wong TY. Diabetic retinopathy. *Lancet* 2010;376:124–36.
- Wong TY, Cheung CMG, Larsen M, *et al.* Diabetic retinopathy. *Nat Rev Dis Primers* 2016;2:16012.
- Sabanayagam C, Banu R, Chee ML, *et al.* Incidence and progression of diabetic retinopathy: a systematic review. *Lancet Diabetes Endocrinol* 2019;7:140–9.
- Song P, Yu J, Chan KY, *et al.* Prevalence, risk factors and burden of diabetic retinopathy in China: a systematic review and meta-analysis. *J Glob Health* 2018;8:010803.
- Wang FH, Liang YB, Zhang F, *et al.* Prevalence of diabetic retinopathy in rural China: the Handan eye study. *Ophthalmology* 2009;116:461–7.
- Nicholson JK, Lindon JC. Systems biology: metabolomics. *Nature* 2008;455:1054–6.
- Fiehn O. Metabolomics--the link between genotypes and phenotypes. *Plant Mol Biol* 2002;48:155–71.
- Wang TJ, Larson MG, Vasan RS, *et al.* Metabolite profiles and the risk of developing diabetes. *Nat Med* 2011;17:448–53.
- Alshehry ZH, Mundra PA, Barlow CK, *et al.* Plasma lipidomic profiles improve on traditional risk factors for the prediction of cardiovascular events in type 2 diabetes mellitus. *Circulation* 2016;134:1637–50.
- Mapstone M, Cheema AK, Fiandaca MS, *et al.* Plasma phospholipids identify antecedent memory impairment in older adults. *Nat Med* 2014;20:415–8.
- Lains I, Gantner M, Murinello S, *et al.* Metabolomics in the study of retinal health and disease. *Prog Retin Eye Res* 2019;69:57–79.
- Chen L, Cheng C-Y, Choi H, *et al.* Plasma metabolomic profiling of diabetic retinopathy. *Diabetes* 2016;65:1099–108.
- Barba I, Garcia-Ramirez M, Hernandez C, *et al.* Metabolic fingerprints of proliferative diabetic retinopathy: an 1H-NMR-based metabolomic approach using vitreous humor. *Invest Ophthalmol Vis Sci* 2010;51:4416–21.
- 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.
- Lains I, Gantner M, Murinello S, *et al.* Metabolomics in the study of retinal health and disease. *Prog Retin Eye Res* 2019;69:57–79.
- Peng L, Sun B, Liu M, *et al.* Plasma metabolic profile reveals PGF2 α protecting against non-proliferative diabetic retinopathy in patients with type 2 diabetes. *Biochem Biophys Res Commun* 2018;496:1276–83.
- Austin PC. A comparison of 12 algorithms for matching on the propensity score. *Stat Med* 2014;33:1057–69.
- Zhu G, Wang S, Huang Z, *et al.* Rewiring of the fruit metabolome in tomato breeding. *Cell* 2018;172:249–61. e12.
- Chen W, Gong L, Guo Z, *et al.* A novel integrated method for large-scale detection, identification, and quantification of widely targeted metabolites: application in the study of rice metabolomics. *Mol Plant* 2013;6:1769–80.
- Xia J, Wishart DS. Metabolomic data processing, analysis, and interpretation using MetaboAnalyst. *Curr Protoc Bioinformatics* 2011;Chapter 14:Unit 14.10.
- Hackstadt AJ, Hess AM. Filtering for increased power for microarray data analysis. *BMC Bioinformatics* 2009;10:11.
- Bonaccorso G. *Machine learning algorithms*. second edition. Birmingham, UK: Packt Publishing Ltd, 2018.
- Van Den Berg HA. Occam's razor: from Ockham's *via moderna* to modern data science. *Sci Prog* 2018;101:261–72.
- Tikhonenko M, Lydic TA, Opreanu M, *et al.* N-3 polyunsaturated fatty acids prevent diabetic retinopathy by inhibition of retinal vascular damage and enhanced endothelial progenitor cell reparative function. *PLoS One* 2013;8:e55177.
- Sasaki M, Kawasaki R, Rogers S, *et al.* The associations of dietary intake of polyunsaturated fatty acids with diabetic retinopathy in well-controlled diabetes. *Invest Ophthalmol Vis Sci* 2015;56:7473–9.
- 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.
- Poesen R, Claes K, Evenepoel P, *et al.* Microbiota-Derived phenylacetylglutamine associates with overall mortality and cardiovascular disease in patients with CKD. *J Am Soc Nephrol* 2016;27:3479–87.
- Tong F, Liu S, Yan B, *et al.* Endogenous ornithine decarboxylase/polyamine system mediated the antagonist role of insulin/PEG-CMCS preconditioning against heart ischemia/reperfusion injury in diabetes mellitus. *Int J Nanomedicine* 2018;13:2507–20.
- Rawlings BJ. Biosynthesis of fatty acids and related metabolites. *Nat Prod Rep* 1998;15:275–308.
- Huang C-F, Cheng M-L, Fan C-M, *et al.* Nicotinic acid: a potential marker of metabolic syndrome through a metabolomics-based approach. *Diabetes Care* 2013;36:1729–31.
- Chang AM, Smith MJ, Galecki AT, *et al.* Impaired beta-cell function in human aging: response to nicotinic acid-induced insulin resistance. *J Clin Endocrinol Metab* 2006;91:3303–9.