

RESEARCH ARTICLE

Open Access



Plasma metabolomic profiles of dementia: a prospective study of 110,655 participants in the UK Biobank

Xinyu Zhang^{1,2†}, Wenyi Hu^{1†}, Yueye Wang^{3†}, Wei Wang^{3†}, Huan Liao⁴, Xiayin Zhang¹, Katerina V. Kiburg⁵, Xianwen Shang¹, Gabriella Bulloch⁵, Yu Huang¹, Xueli Zhang¹, Shulin Tang¹, Yijun Hu^{6,7}, Honghua Yu¹, Xiaohong Yang¹, Mingguang He^{1,3,5} and Zhuoting Zhu^{1,5*}

Abstract

Background: Plasma metabolomic profile is disturbed in dementia patients, but previous studies have discordant conclusions.

Methods: Circulating metabolomic data of 110,655 people in the UK Biobank study were measured with nuclear magnetic resonance technique, and incident dementia records were obtained from national health registers. The associations between plasma metabolites and dementia were estimated using Cox proportional hazard models. The 10-fold cross-validation elastic net regression models selected metabolites that predicted incident dementia, and a 10-year prediction model for dementia was constructed by multivariable logistic regression. The predictive values of the conventional risk model, the metabolites model, and the combined model were discriminated by comparison of area under the receiver operating characteristic curves (AUCs). Net reclassification improvement (NRI) was used to estimate the change of reclassification ability when adding metabolites into the conventional prediction model.

Results: Amongst 110,655 participants, the mean (standard deviation) age was 56.5 (8.1) years, and 51 186 (46.3%) were male. A total of 1439 (13.0%) developed dementia during a median follow-up of 12.2 years (interquartile range: 11.5–12.9 years). A total of 38 metabolites, including lipids and lipoproteins, ketone bodies, glycolysis-related metabolites, and amino acids, were found to be significantly associated with incident dementia. Adding selected metabolites ($n=24$) to the conventional dementia risk prediction model significantly improved the prediction for incident dementia (AUC: 0.824 versus 0.817, $p=0.042$) and reclassification ability (NRI = 4.97%, $P=0.009$) for identifying high risk groups.

Conclusions: Our analysis identified various metabolomic biomarkers which were significantly associated with incident dementia. Metabolomic profiles also provided opportunities for dementia risk reclassification. These findings may help explain the biological mechanisms underlying dementia and improve dementia prediction.

Keywords: Metabolites, Dementia, UK Biobank

[†]Xinyu Zhang, Wenyi Hu, Yueye Wang and Wei Wang contributed equally to this work.

*Correspondence: zhuoting_zhu@hotmail.com

⁵ Centre for Eye Research, University of Melbourne, East Melbourne, Victoria, Australia

Full list of author information is available at the end of the article

Background

Dementia is a leading cause of disability in people over 65 years worldwide [1]. It is expected to affect over 131.5 million people and cost over a trillion dollars by 2050 [2]. As no effective treatments for dementia are currently available, early identification of patients at high risk is



a public health priority in efforts to delay disease progression and alleviate the burden of disease on patients, policy makers, and healthcare providers [3]. Despite the plethora of research, current screening tools and prediction models are insufficiently accurate and are associated with high costs, invasive tests, and complex technicalities. Unfortunately, the early detection of dementia is a continually evolving field of research that frequently underdelivers to effectively target the needs and necessities of population-based screening.

Metabolomics comprehensively analyses small molecular metabolites in targeted tissues or biofluids, to indicate genetic, environmental, and pathological changes during disease development. Previous studies have revealed deranged metabolomic profiles in dementia patients, but further studies have provided conflicting conclusions [1, 4–6]. Such inconsistencies may be due to the small sample size, cross-sectional design, short follow-up time, limited adjustment for confounding factors, and different analytical chemistry techniques used to measure metabolomics in these studies. Furthermore, only few studies have investigated the additive value of plasma metabolites in dementia prediction, with inconsistent conclusions [7–13].

The UK Biobank Study is a prospective study that measured 249 metabolomic biomarkers in approximately 118,000 EDTA plasma samples using high-throughput nuclear magnetic resonance (NMR). Its large sample size, long-term follow-up, standardized platform of metabolomics profiling, and systematic ascertainment of incident dementia provide a unique opportunity to investigate the metabolomic profile of incident dementia and assess the additive values of metabolites for incident dementia.

Methods

Study participants

We studied participants in the prospective UK Biobank Study cohort that met the inclusion criteria of this study. The UK Biobank Study includes over 500,000 people of middle and old age that were recruited from 2006 to 2010 across the UK. Their baseline demographic, phenotypic, and genotypic characteristics were collected in 22 assessment centres at enrolment, and further collections were accumulated after follow-up intervals of 6 months to 3 years. Detailed protocols of the UK Biobank Study are described elsewhere [14].

For the identification of metabolomic biomarkers associated with incident dementia, the present analyses included participants without prior dementia at baseline with available metabolomic data. A total of 110,730 participants had metabolites data, of which 75 participants with a history of dementia were excluded. Finally,

110,655 participants were included in the current analysis. To develop a prediction model for 10-year incident dementia risk, participants were randomly divided into a discovery group ($n = 55,328$) and a replication group ($n = 55,327$). The workflow of the analyses is presented in Fig. 1. Baseline characteristics of study participants in each dataset are described in Additional file 1: Table S1.

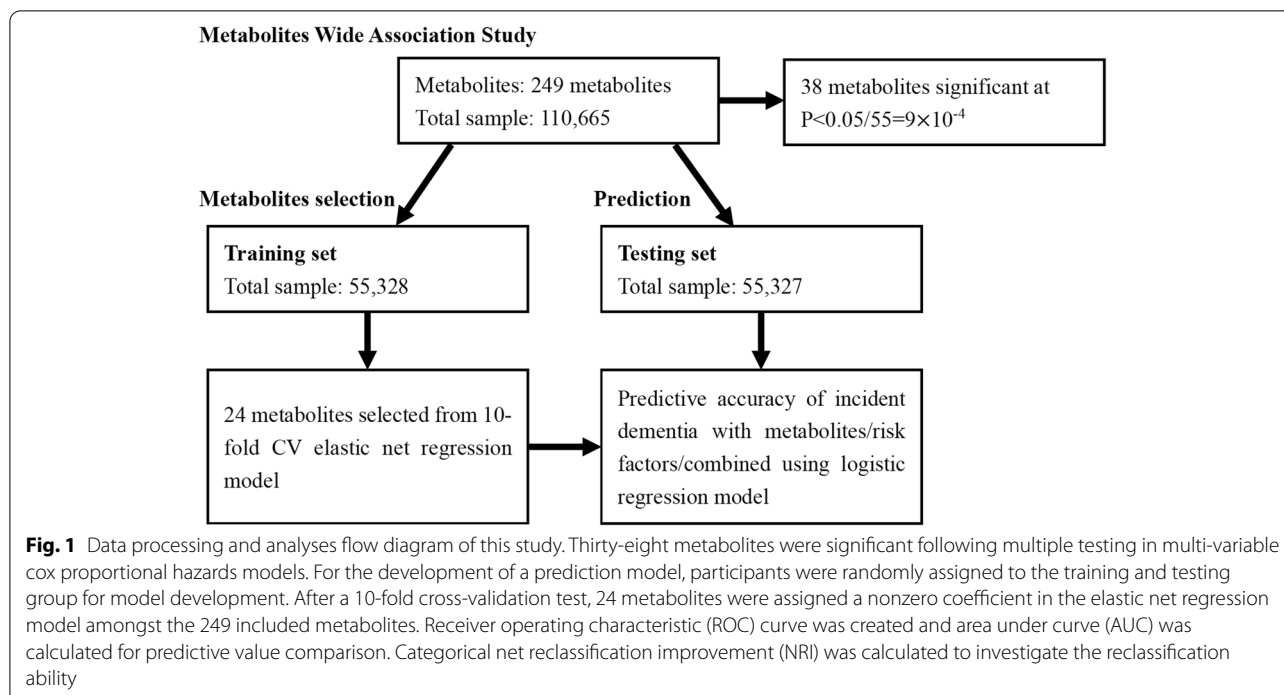
Ethics for the UK Biobank was approved by North West Multi-Centre Research Ethics Committee across the United Kingdom, and the Human Tissue Authority license was also approved for the UK Biobank. We gained access to the UK Biobank data through application. All participants submitted signed informed consent in written form. The Declarations of Helsinki were complied with throughout this study.

Metabolite quantification

Detailed protocols on sample collection and metabolomic quantification are presented elsewhere [15–17]. In brief, EDTA plasma samples were collected at baseline recruitment (118,000 samples) and repeat assessment (5000 samples). Samples were prepared directly in 96-well plates by UK Biobank, with each plate containing a serum mimic as a quantification consistency monitor and a mixture of 2 low-molecular-weight metabolite as a technical reference. These samples were shipped to Nightingale Health's laboratories in Finland on dry ice and measured between June 2019 and April 2020. In the lab, samples were prepared with an automated liquid handler, automatically analysed with spectrometers and a robotic sample changer, and quantified with Nightingale Health's proprietary software (Nightingale Health Biomarker quantification library 2020). Accredited quality control was done during the whole process to eliminate systemic and technical variance, and only samples and biomarkers that underwent the quality control process were stored in the UK Biobank dataset and used in our present study. Each sample included 168 metabolites in absolute level (mmol/L) spanning fatty acids, glycolysis metabolites, ketone bodies, amino acids, lipids, and lipoproteins, and 81 in ratio measurement.

Ascertainment of dementia

Dementia incidence data were collected through hospital in-patient admission records and death registries. The identification of dementia was based on the International Classification of Diseases (ICD) code, including 290.0–290.4, 294.1, 331.0–331.2, and 331.5 in ICD-9 and A81.0, F00, F01, F02, F03, F05.1, F10.6, G30, G31.0, G31.1, and G31.8 in ICD-10, covering Alzheimer's Disease dementia, vascular dementia, and dementias of other causes. The follow-up period was defined from baseline to the earliest one amongst the incident dementia date,



lost-of-follow-up date, death date, or the last date of data update, which was April 28, 2021.

Traditional risk factors

Age [18], gender [18], education level [18], systolic pressure [19], anti-hypertension treatment [20], diabetes mellitus [21], smoking status [22], history of stroke [23], history of coronary heart disease [24], and APOE $\epsilon 4$ allele were established risk factors of dementia [25], and therefore were used as covariates in Cox regression analysis and constituted conventional prediction model for dementia. Age (UKB Field 21022) and systolic pressure (UKB Filed 4080) were continuous variables and gender (UKB Field 31, female or male), education level (UKB Field 6138, college/university degree or others), anti-hypertension treatment (UKB Field 6153, no or yes), diabetes mellitus (no or yes, including UKB Field 2443, doctor-diagnosed diabetes, UKB Field 6153, insulin treatment, UKB Field 20003, diabetes-related medication and UKB Field 30750, plasma HbA1c level of or over 48 mmol/mol), smoking status (UKB Field 20116-0.0, never or former/current), history of stroke (UKB Field 4056, no or yes), history of coronary heart disease (UKB Field 3627,3894, no or yes), and APOE $\epsilon 4$ carrier (no or yes) were redefined as categorical variables.

Statistical analyses

Continuous variables were described with mean (standard deviation, SD) or median (interquartile range, IQR),

and categorical variables were described with number and percentage. The values of all metabolites were first transformed using natural logarithmic transformation ($\ln[x+1]$) and then Z-transformed. Our present study included two separate analyses (Fig. 1). In the first analysis, the associations between metabolites and dementia were estimated using cox proportional-hazard models, with confounders including age, sex, education level, systolic pressure, anti-hypertension treatment, diabetes mellitus, smoking status, history of stroke, history of coronary heart disease, APOE $\epsilon 4$ allele adjusted. A P value less than 0.05 was set as nominal significance. The corrected P-value was estimated through a principal component analysis developed by Gao et al. [26]. Strong correlations were considered, and 55 parameters explained 99.5% of metabolites variations. Therefore, 55 independent tests were conducted for correlation ascertainment, and P value significance was set at 9×10^{-4} ($0.05/55$) or less. β coefficients of adjusted hazard ratio (HR), 95% coefficient interval (CI) and P values of all 249 metabolites in the cox model are presented in Additional file 1: Table S2.

In the second analysis, participants were randomly assigned to a training set ($n=55,328$) and a testing set ($n=55,327$) to develop and validate the 10-year incident dementia risk prediction model. We used elastic net regularized logistic regression models to select metabolomic predictors. The elastic net regularized logistic regression model is a straightforward supervised

machine learning algorithm combining least absolute shrinkage and selection operator (LASSO) regression and Ridge regression with a good probabilistic interpretation of variables suitable for disease prediction. The LASSO penalty selected variables by reducing the absolute value of weight, while the Ridge penalty further reducing the extremities of weights. Details of the model have been described elsewhere [27]. Two tuning parameters for elastic net regression model were used, including α (representing the weight of the penalty) and λ (representing the complexity of the penalty). Of note, α controls the balance between LASSO and Ridge, with $\alpha(1)$ corresponding to the lasso (the default estimator) and $\alpha(0)$ corresponding to ridge regression [28]. To achieve the sparsity of the model and select core metabolomic predictors, we tested α of 0.5, 0.75, and 1. We then used 10-fold cross-validation to select optimal λ and β coefficients for elastic net regression models in seek of minimum to minimize cross-validation prediction error, and order to achieve optimal robustness of the model. Each combination of α and λ on a two-dimensional grid underwent 10-fold cross-validation for elastic net logistic regression to compute cross-validation prediction error and assign β coefficient for each metabolite to achieve the minimum cross-validation function. In the final model, $\alpha=1$ and $\lambda=0.0003648$ were selected by cross-validation. Minimum cross-validation mean deviance was 0.0887588. 24 metabolites were given nonzero coefficient (Additional file 1: Table S3). In the testing set, we applied three logistic regression models to estimate the predictive values of different parameters. Model 1 included conventional risk factors, including age, gender, education level, systolic pressure, anti-hypertension treatment, diabetes mellitus, smoking status, history of stroke, history of coronary heart disease, and APOE $\epsilon 4$ allele; Model 2 included selected metabolomic biomarkers; and Model 3 included combined conventional risk factors and selected metabolomic biomarkers. Coefficients (95% CI) and P value for each exposure parameter are presented in Additional file 2: Table S4. The predictive value of the selected 24 metabolites was assessed through two methods. Firstly, the receiver operating characteristic (ROC) curve was constructed and the areas under the curves (AUCs) were compared amongst different models. Secondly, using $\geq 5\%$ as the threshold for the group at high risk to develop dementia in 10 years [29], categorical net reclassification improvement (NRI) was estimated for the added value of selected metabolomic biomarkers for risk stratification over conventional risk factors.

All analyses were performed using Stata version 13 (Stata Corp) and R (version 3.3.1, R Project for Statistical Computing, Vienna, Austria).

Results

Baseline characteristics of the study participants

Our analysis included 110,655 participants without dementia at baseline, with an average (SD) age of 56.5 (8.1), of which 59,469 (53.7%) were female. After a median follow-up of 12.2 years (IQR: 11.5–12.9 years), 1439 (1.30%) participants developed dementia. Baseline characteristics of all participants stratified by incident dementia are summarized in Table 1. Participants with incident dementia were often older, male, APOE $\epsilon 4$ carriers, with lower education, higher systolic pressures, a history of diabetes mellitus, anti-hypertensive medication use, former or current smokers, and had a history of stroke or coronary heart disease.

Associations of baseline circulating metabolites with incident dementia

Amongst the 249 metabolomic biomarkers tested by UK Biobank Study with NMR technique, after controlling for multiple testing, 38 metabolites remained significantly associated with dementia incidence (Fig. 2). These metabolites included amino acids, fatty acids, glycolysis-related metabolites, ketone bodies, and lipids and lipoproteins categories. Only eight metabolomic biomarkers, including citrate (HR=1.09 [95% CI: 1.04–1.14], $P=1.52\times 10^{-4}$), three ketone bodies (e.g., HR_{Acetoacetate}=1.11 [95%CI: 1.07–1.15], $P=8.78\times 10^{-8}$), phospholipids to total lipids ratio in intermediate-density lipoproteins (IDL), small low-density lipoproteins (LDL), and very small very-low-density lipoproteins (VLDL) (e.g., HR_{s-LDL-PL-%}=1.12 [95% CI: 1.06–1.18], $P=1.54\times 10^{-5}$), and free cholesterol to total lipids ratio in very large VLDL (HR_{VL-VLDL-FC-%}=1.08 [95% CI: 1.03–1.12], $P=5.62\times 10^{-4}$), were positively associated with dementia, while 30 other metabolomic biomarkers from amino acids, fatty acids, lipids and lipoprotein subclasses, were inversely associated with the incident dementia.

Metabolite selection, prediction, and reclassification of the incident possibility of dementia

Elastic net regularized logistic regression analysis identified metabolites for the construction of a dementia prediction model, and 24 metabolites (Additional file 1: Table S3) were selected for inclusion in the training set. The prediction ability of these metabolites (Xb2) was worse than the conventional prediction model (Xb1) (AUC: 0.677 versus 0.817), although the addition of these metabolites to the conventional risk factors-based model (Xb3) improved the prediction precision (AUC: 0.824 versus 0.817, $p=0.042$) (Fig. 3). Categorical net reclassification improvement analysis showed significant benefit in reclassification ability through the addition of metabolomic biomarkers into the conventional prediction model (NRI=4.97%, SE=0.009, $p<0.001$) (Table 2).

Table 1 Baseline characteristics of study participants in the prospective study of dementia

Baseline Characteristics	Overall (N= 110,655)	Individuals with incident dementia (n=1439)	Individuals without incident dementia (n=109,216)	P value
Age, mean (SD), years	56.5 (8.10)	64.2 (4.88)	56.4 (8.08)	<0.001
Gender, No. (%)				<0.001
Female	59,469 (53.7)	652 (45.3)	58,817 (53.9)	
Male	51,186 (46.3)	787 (54.7)	50,399 (46.1)	
Education level, No. (%)				<0.001
College or university degree	35,744 (32.3)	301 (20.9)	35,443 (32.5)	
Others	74,911 (67.7)	1138 (79.1)	73,773 (67.5)	
Systolic pressure, mean (SD), mmHg	137 (18.5)	143 (19.3)	138 (18.5)	<0.001
Anti-hypertension treatment, No. (%)				<0.001
No	100,148 (90.5)	1211 (84.2)	98,937 (90.6)	
Yes	10,507 (9.50)	228 (15.8)	10,279 (9.41)	
Diabetes mellitus, No. (%)				<0.001
No	103,950 (93.9)	1207 (83.9)	102,743 (94.1)	
Yes	6705 (6.06)	232 (16.1)	6473 (5.93)	
Smoking status, No. (%)				<0.001
Never	60,195 (54.7)	666 (46.6)	59,529 (54.8)	
Former/current	49,896 (45.3)	762 (53.4)	49,134 (45.2)	
History of stroke, No. (%)				<0.001
No	109,108 (98.6)	1361 (94.6)	107,747 (98.7)	
Yes	1547 (1.40)	78 (5.42)	1469 (1.35)	
History of coronary heart disease, No. (%)				<0.001
No	106,067 (95.9)	1267 (88.1)	104,800 (96.0)	
Yes	4588 (4.15)	172 (12.0)	4416 (4.04)	
APOE ε4 carrier, No. (%)				<0.001
No	83,454 (75.8)	795 (55.6)	82,649 (76.0)	
Yes	26,675 (24.2)	635 (44.4)	26,040 (24.0)	

SD standard deviation, No. number

Discussion

Using data from over 100,000 plasma samples from UK Biobank, we discovered that 38 serum metabolites including amino acids, ketone bodies, glycolysis metabolites, lipids, and lipoproteins were significantly associated with an increased risk of incident dementia. Our findings suggest the addition of metabolites into dementia risk stratification models could improve its prediction and reclassification for at-risk individuals. These metabolites also further insights to biological mechanisms which may underly dementia, and may assist in the prediction of dementia.

Our results showed branched-chain amino acids (BCAAs) are negatively associated with an elevated risk of incident dementia, which is consistent with several

previous studies [1, 30]. It is suggested circulatory BCAAs may pass the blood-brain barrier (BBB) [1] and participate in the synthesis of key neurotransmitters, proteins, and energy [31], which may exert neuroprotective benefits for the ageing brain. A recent randomized clinical trial also suggested dietary BCAA supplements improved cognitive function in middle-aged and older adults [32]. As such, this study supports the role of BCAAs for dementia prevention, and this should be further investigated as a possible modifiable factor for delaying dementia onset.

Our analysis also implicated ketone bodies and citrate as metabolites positively associated with incident dementia. Elevation of plasma ketone bodies infer a switch from glucose-dependent substrates as energy, which occurs

(See figure on next page.)

Fig. 2 Adjusted HR (95% CI) of incident dementia for metabolites after multiple testing. Hazard ratios (HR) are per 1 standard deviation (SD) higher of Z-transformed metabolic marker and are adjusted for age, gender, education level, systolic pressure, anti-hypertension treatment, diabetes mellitus, smoking status, history of stroke, history of coronary heart disease, and APOE ε4 allele. CI, confidence interval; LDL, low-density lipoprotein; HDL, high-density lipoprotein; VLDL, very-low-density lipoprotein; IDL, intermediate-density lipoprotein

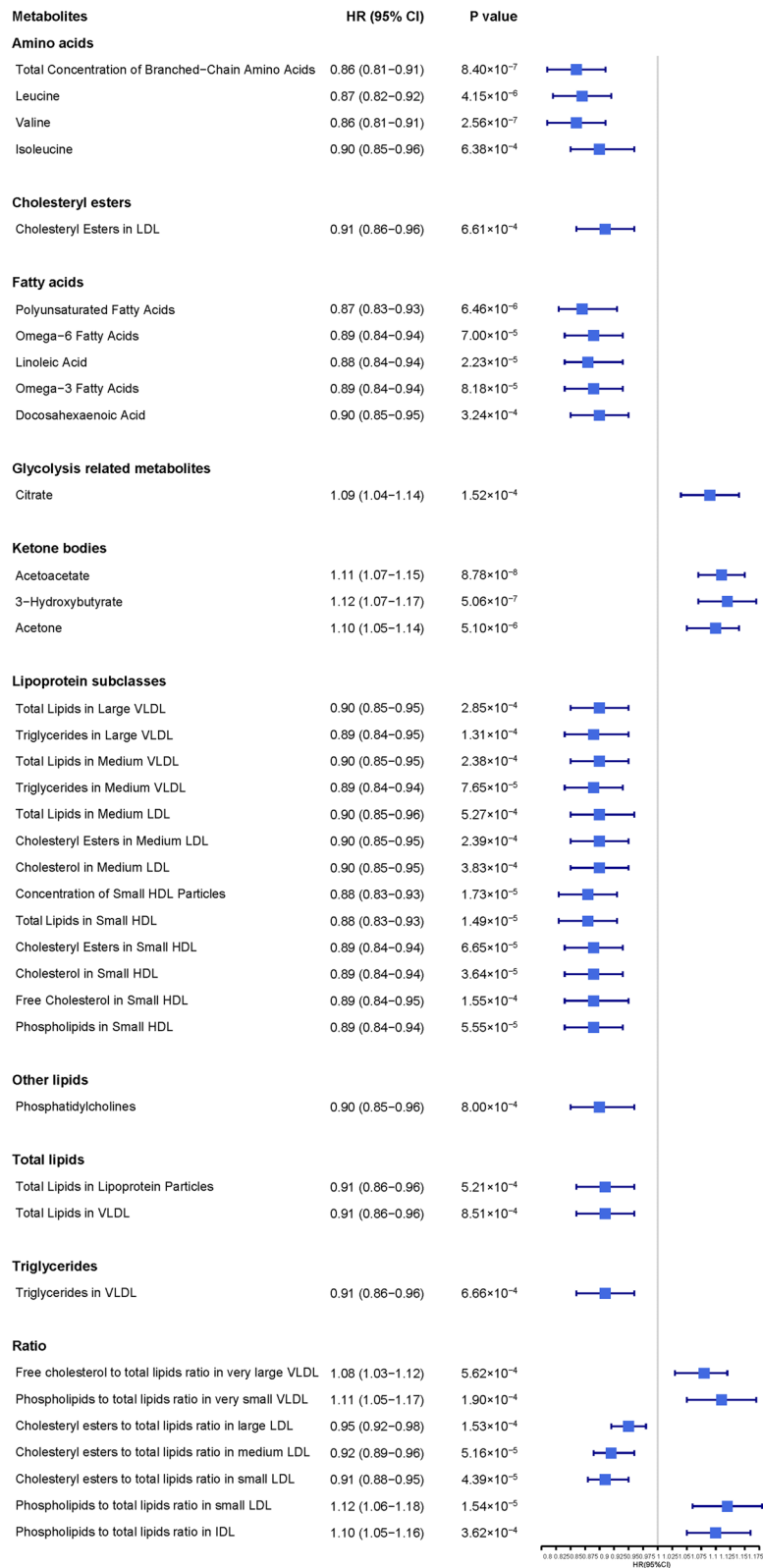


Fig. 2 (See legend on previous page.)

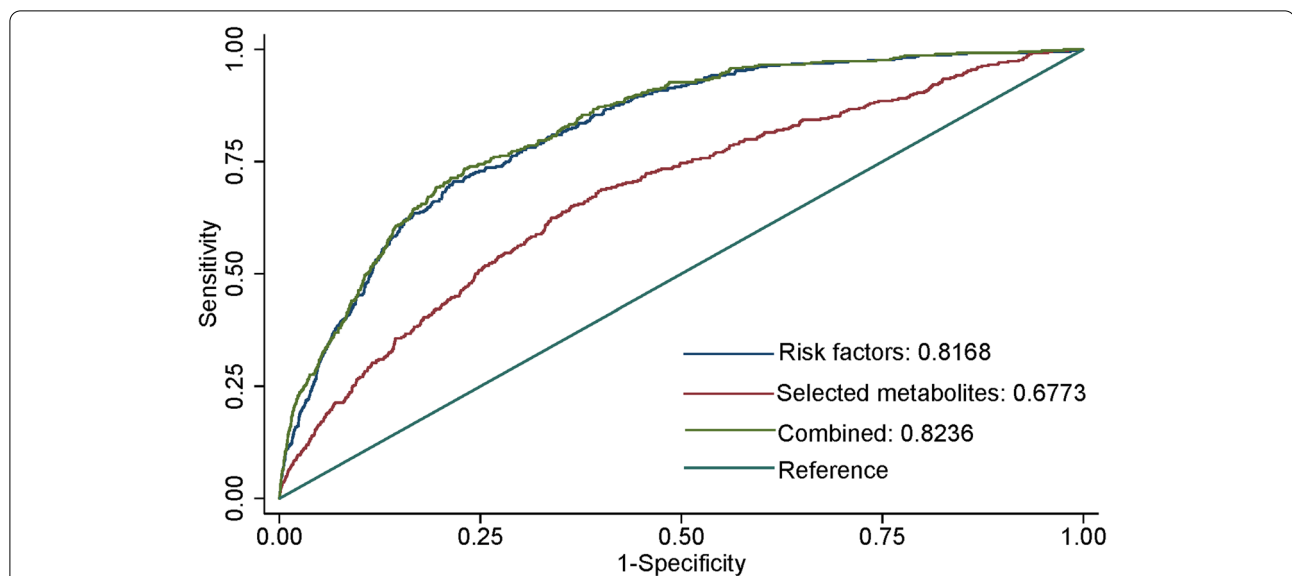


Fig. 3 ROC and AUC analysis of incident dementia prediction model development and predictive value comparison. An elastic net regression model based on lasso penalty was used for dementia prediction. After 10-fold cross-validation, 24 of 249 metabolites were selected for the dementia prediction model. Xb1 curve used conventional risk factors as input signals, while the Xb2 curve was for 24 selected metabolites and Xb3 was for conventional risk factors and 24 selected metabolites. There was no clinically significant difference ($P = 0.042$) found between the AUC of Xb1 and Xb3. Conventional risk factors included age, gender, education level, systolic pressure, anti-hypertension treatment, diabetes mellitus, smoking status, history of stroke, history of coronary heart disease, and APOE $\epsilon 4$ allele. ROC, receiver operating characteristic; AUC, area under curve

Table 2 Net reclassification improvement of adding 24 metabolites into a conventional risk prediction model

	Conventional prediction model	Updated model	
Incident dementia in 10 years	Risk (%)	<5	≥ 5
Yes	<5	668	48
	≥ 5	6	76
No	<5	101,185	654
	≥ 5	353	823

Values are *n*. Patients in the upper-right cell were rightfully up-reclassified ($n=48$, and 6 patients were mistakenly down-reclassified), indicating an improved sensitivity. Patients in the lower-left cell were rightfully down-reclassified ($n=353$, and 654 patients mistakenly up-reclassified), indicating an undermined specificity. NRI (SE) = 0.0497 (0.00921), P value <0.001.

in low carbohydrate diets, disrupted glucose uptake disorders like diabetes, or during long periods of fasting. Ketone bodies have also been implicated in cognitive impairment and Alzheimer’s dementia (AD) brains which were exposed to long-term glucose uptake and utilization disruption [33, 34]. This association is further explained by recent evidence suggesting the utilization of ketone bodies in the brain largely relies on their plasma concentration [35] and transportation capacity across the BBB [36]. Our association with peripheral citrate is concordant with FA Leeuw et al. whom observed higher levels of

plasma citrate were associated with brain and hippocampal atrophy, and white matter hyperintensity, which are known neurological changes associated with AD and dementia [37]. Furthermore, these metabolic disturbances are replicated in animal AD model brains [33, 38, 39]. Although our findings are consistent with previous studies, the exact mechanisms underlying these associations are lacking. We suggest future research to concentrate on laboratory and clinical studies to further define and elucidate the roles of ketones and citrate in dementia.

Strong negative correlations were observed between small high-density lipoproteins (HDL), its lipid constituents (including total lipids, cholesteryl esters, cholesterol, free cholesterol, and phospholipids) and incident dementia in the present analysis. Similar findings have been replicated by previous studies although an adequate explanation on these observations is yet to be fully realized [37, 40, 41]. Of note, Martinez et al. recently showed small HDL was the only lipoprotein that could pass the BBB, which implies it plays a role in the balance and distribution of fats within the brain [42]. Potentially, HDL exerts its neuroprotective effects through the redistribution of lipids which may affect neuronal membrane composition. This may have reverberating effects on β deposition and p-tau accumulation, and may protect neurons against oxidation and inflammation, thereby preserving vascular and synaptic function [40, 42, 43].

Total lipids, cholesteryl esters, cholesterol, triglycerides in VLDL and LDL subclasses, and polyunsaturated fatty acids (PUFA) were inversely associated with future risk of dementia, while free cholesterol concentration in very large VLDL was associated with an increased risk of dementia. These findings are aligned with some previous studies [1, 40, 44, 45], although a body of evidence has suggested conflicting results [46–48]. Of note, LDL-C is a conventional risk factor in cardiovascular and metabolic diseases, although its role in dementia has been historically conflicting [48, 49]. Furthermore, the different roles of lipoprotein subclasses may be attributed to their ability to cross the BBB, which makes the lipid environment in the brain very different from the periphery [37]. Nonetheless, several hypotheses may explain our findings, including the presumed role of triglycerides for interfering with peripheral A β transportation and facilitating PUFA absorption [50]. PUFAs are not only an anti-oxidative and anti-inflammatory regulator through modulating pro-inflammatory cytokines and decreasing microglial inflammatory activation of A β [50], but also are an essential component of neuronal membranes and can pass the BBB to participate in brain development [51], so theoretically their presence in serum should indicate their active synthesis and transportation which would be beneficial for the brain. The links between VLDL and LDL with dementia are less straightforward, as these lipoproteins cannot pass the BBB. Further studies are needed to investigate their mechanisms and tie in their association with dementia pathogenesis.

Furthermore, we successfully identified candidate metabolites, whose addition to the conventional model could significantly improve the accuracy of dementia prediction and reclassification of risk group, thus might improve the sensitivity of identifying patients in their prodromal phase. Such findings suggest the potential clinical use of metabolomic biomarkers as complementary information for early and population-based detection of dementia. However, though our model showed statistically significant improvement through the addition of metabolites, caution is still required when interpreting clinical implications as the absolute increase was not substantial. This was possible because conventional risk factors for dementia, e.g., diabetes and hypertension, already accounted for some of the metabolic changes in dementia pathogenesis.

Our study had several strengths, including the use of over 100,000 samples over a 14-year follow-up duration, the adjustment for confounding factors, and a homogeneous platform to analyse metabolites using the NMR technique [52]. Some limitations must also be acknowledged. Firstly, the participants were largely Caucasian and from developed countries with good socioeconomic standing [53] which may limit the generalizability of our results to other ethnicities and geographic backgrounds. Secondly, our longitudinal

associations do not imply causality, hence more research is needed to confirm our findings. Thirdly, though the identification of clinical dementia from the UK electronic healthcare dataset guaranteed the specificity of our model, the omission of subclinical dementia may undermine its sensitivity [54, 55]. Furthermore, the paper did not define the effect of metabolites on predicting various subtypes of dementia. Fourthly, as we used peripheral metabolomic data rather than CNS metabolites, these results may not reflect the brain microenvironment and should be interpreted with caution. Lastly, we cannot exclude residual confounders.

Conclusions

By use of a novel study design to investigate a prospective cohort of 110,655 participants over 14 years, several metabolomic biomarkers were found significantly associated with dementia incidence. The metabolites identified in this study may supplement future hypotheses explaining the complex interplay between energy and lipid metabolism, and the development of dementia. These metabolites improved risk stratification when added to conventional risk factor models, and this study provides an opportunity for considering the improvement of screening tools to better identify at-risk populations. Our findings also further discuss biological mechanisms underlying dementia and potentially facilitate the prediction, prevention, and treatment of dementia. Molecular and genetic studies were needed to determine the exact pathways mediating our observed associations, and clinical studies are needed to prove these metabolites can improve screening and prediction of dementia.

Abbreviations

AUC: Area under the receiver operating characteristic curves; AD: Alzheimer's dementia; BBB: Blood-brain barrier; BCAA: Branched-chain amino acids; CI: Coefficient interval; HDL: High-density lipoproteins; HR: Hazard ratio; ICD: International Classification of Diseases; IDL: Intermediate-density lipoproteins; IQR: Interquartile range; LASSO: Least absolute shrinkage and selection operator; LDL: Low-density lipoproteins; NMR: Nuclear magnetic resonance; NRI: Net reclassification improvement; PUFA: Polyunsaturated fatty acids; ROC: Receiver operating characteristic; SD: Standard deviation; UKB: The United Kingdom Biobank; VLDL: Very-low-density lipoproteins.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12916-022-02449-3>.

Additional file 1: Table S1. Baseline characteristics of study participants stratified by discovery or replication dataset. **Table S2.** β coefficient of adjusted HR (95% CI) and *P* values of Incident Dementia for All 249 Metabolites in Cox Proportional Hazards Model. **Table S3.** Coefficients of selected 24 metabolites in elastic net regularized logistic regression model after 10-fold cross validation, their coefficients in e-net regression model in training dataset, and their VIFs in testing dataset.

Additional file 2: Table S4. β coefficients (95% CI) and *P* values of conventional risk factors and metabolites in the multi-variable logistic model for 10-year incident dementia prediction.

Acknowledgements

Not Applicable.

Authors' contributions

ZTZ and WW conceptualized and designed the study. All authors contributed to the acquisition, analysis, or interpretation of the study. XYZ and YYW drafted the manuscript. ZTZ, MGH, XHY, KVK, and GB revised the manuscript for important intellectual content. ZTZ, WYH, HL, XYZ, XWS, YH, XLZ, SLT, and YJH did the statistical analyses. MGH, XHY, and HHY obtained funding. ZTZ, MGH and XHY provided administrative, technical, and material support. ZTZ, MGH, XHY, and HHY supervised the whole study. All authors read and approved the final manuscript.

Funding

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The present work was supported by Fundamental Research Funds of the State Key Laboratory of Ophthalmology, National Natural Science Foundation of China (82000901, 82101173, 81870663, 82171075), Outstanding Young Talent Trainee Program of Guangdong Provincial People's Hospital (KJ012019087), Guangdong Provincial People's Hospital Scientific Research Funds for Leading Medical Talents and Distinguished Young Scholars in Guangdong Province (KJ012019457), Talent Introduction Fund of Guangdong Provincial People's Hospital (Y012018145), Science and Technology Program of Guangzhou, China (20202020049), Project of Special Research on Cardiovascular Diseases (2020XXG007), and Research Foundation of Medical Science and Technology of Guangdong Province (B2021237). Prof. Mingguang He receives support from the University of Melbourne at Research Accelerator Program and the CERA Foundation. The Centre for Eye Research Australia receives Operational Infrastructure Support from the Victorian State Government. The sponsor or funding organization had no role in the design or conduct of this research. Mingguang He received the following grant: High-level Talent Flexible Introduction Fund of Guangdong Provincial People's Hospital (No. KJ012019530).

Availability of data and materials

The dataset analysed during the present study is available in the UK Biobank (<https://www.ukbiobank.ac.uk/>). These data are not publicly available because they were analysed under licence, but they are available from the corresponding author on reasonable request and with permission of UK Biobank.

Declarations**Ethics approval and consent to participate**

Northwest Multi-Centre Research Ethics Committee (11/NW/0382) gave their ethical approval for the original UK Biobank study. Informed consent was obtained from each participant. Our study was conducted in accordance to the Declarations of Helsinki. Our access to data from the UK Biobank cohort was approved by the UKB Ethics Advisory Committee (application ID: 62443).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Ophthalmology, Guangdong Academy of Medical Sciences, Guangdong Provincial People's Hospital, Guangzhou, China. ²Department of Ophthalmology, Shanghai General Hospital, Shanghai, China. ³State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, China. ⁴Neural Regeneration Group, Institute of Reconstructive Neurobiology, University of Bonn, Bonn, Germany. ⁵Centre for Eye Research, University of Melbourne, East Melbourne, Victoria, Australia. ⁶Aier Institute of Refractive Surgery, Refractive Surgery Center, Guangzhou Aier Eye Hospital, Guangzhou, China. ⁷Aier School of Ophthalmology, Central South University, Changsha, China.

Received: 13 February 2022 Accepted: 21 June 2022

Published online: 15 August 2022

References

1. Tynkkynen J, Chouraki V, van der Lee SJ, Hernesniemi J, Yang Q, Li S, et al. Association of branched-chain amino acids and other circulating metabolites with risk of incident dementia and Alzheimer's disease: a prospective study in eight cohorts. *Alzheimers Dement*. 2018;14(6):723–33.
2. Prince MJ, Wimo A, Guerchet MM, Ali GC, Wu Y-T, Prina M. World Alzheimer Report 2015 - The Global Impact of Dementia. London: Alzheimer's Disease International; 2015.
3. Shah H, Albanese E, Duggan C, Rudan I, Langa KM, Carrillo MC, et al. Research priorities to reduce the global burden of dementia by 2025. *Lancet Neurol*. 2016;15(12):1285–94.
4. Varma VR, Oommen AM, Varma S, Casanova R, An Y, Andrews RM, et al. Brain and blood metabolite signatures of pathology and progression in Alzheimer disease: a targeted metabolomics study. *PLoS Med*. 2018;15(1):e1002482.
5. Huo Z, Yu L, Yang J, Zhu Y, Bennett DA, Zhao J. Brain and blood metabolome for Alzheimer's dementia: findings from a targeted metabolomics analysis. *Neurobiol Aging*. 2020;86:123–33.
6. Saji N, Murotani K, Hisada T, Kunihiro T, Tsuduki T, Sugimoto T, et al. Relationship between dementia and gut microbiome-associated metabolites: a cross-sectional study in Japan. *Sci Rep*. 2020;10(1):8088.
7. Oresic M, Hyotylainen T, Herukka SK, Sysi-Aho M, Mattila I, Seppanan-Laakso T, et al. Metabolome in progression to Alzheimer's disease. *Transl Psychiatry*. 2011;1:e57.
8. Mapstone M, Cheema AK, Fiandaca MS, Zhong X, Mhyre TR, MacArthur LH, et al. Plasma phospholipids identify antecedent memory impairment in older adults. *Nat Med*. 2014;20(4):415–8.
9. Mousavi M, Jonsson P, Antti H, Adolfsson R, Nordin A, Bergdahl J, et al. Serum metabolomic biomarkers of dementia. *Dement Geriatr Cogn Dis Extra*. 2014;4(2):252–62.
10. Graham SF, Chevallier OP, Elliott CT, Holscher C, Johnston J, McGuinness B, et al. Untargeted metabolomic analysis of human plasma indicates differentially affected polyamine and L-arginine metabolism in mild cognitive impairment subjects converting to Alzheimer's disease. *PLoS One*. 2015;10(3):e0119452.
11. Casanova R, Varma S, Simpson B, Kim M, An Y, Saldana S, et al. Blood metabolite markers of preclinical Alzheimer's disease in two longitudinally followed cohorts of older individuals. *Alzheimers Dement*. 2016;12(7):815–22.
12. Abdullah L, Evans JE, Emmerich T, Crynen G, Shackleton B, Keegan AP, et al. APOE epsilon4 specific imbalance of arachidonic acid and docosahexaenoic acid in serum phospholipids identifies individuals with preclinical Mild Cognitive Impairment/Alzheimer's Disease. *Aging (Albany NY)*. 2017;9(3):964–85.
13. Li D, Misialek JR, Boerwinkle E, Gottesman RF, Sharrett AR, Mosley TH, et al. Prospective associations of plasma phospholipids and mild cognitive impairment/dementia among African Americans in the ARIC Neurocognitive Study. *Alzheimers Dement (Amst)*. 2017;6:1–10.
14. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12(3):e1001779. <https://doi.org/10.1371/journal.pmed.1001779>.
15. Soininen P, Kangas AJ, Wurtz P, Suna T, Ala-Korpela M. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circ Cardiovasc Genet*. 2015;8(1):192–206.
16. Wurtz P, Kangas AJ, Soininen P, Lawlor DA, Davey Smith G, Ala-Korpela M. Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Large-Scale Epidemiology: A Primer on -Omics Technologies. *Am J Epidemiol*. 2017;186(9):1084–96.
17. Soininen P, Kangas AJ, Wurtz P, Tukiainen T, Tynkkynen T, Laatikainen R, et al. High-throughput serum NMR metabolomics for cost-effective holistic studies on systemic metabolism. *Analyst*. 2009;134(9):1781–5.
18. Plassman BL, Langa KM, Fisher GG, Heeringa SG, Weir DR, Ofstedal MB, et al. Prevalence of dementia in the United States: the aging, demographics, and memory study. *Neuroepidemiology*. 2007;29(1–2):125–32.

19. Ou Y-N, Tan C-C, Shen X-N, Xu W, Hou X-H, Dong Q, et al. Blood pressure and risks of cognitive impairment and dementia: a systematic review and meta-analysis of 209 prospective studies. *Hypertension*. 2020;76(1):217–25.
20. Walker KA, Power MC, Gottesman RF. Defining the relationship between hypertension, cognitive decline, and dementia: a review. *Curr Hypertens Rep*. 2017;19(3):24.
21. Biessels GJ, Despa F. Cognitive decline and dementia in diabetes mellitus: mechanisms and clinical implications. *Nat Rev Endocrinol*. 2018;14(10):591–604.
22. Durazzo TC, Mattsson N, Weiner MW. Initiative AsDN: Smoking and increased Alzheimer's disease risk: a review of potential mechanisms. *Alzheimers Dement*. 2014;10:S122–45.
23. Hachinski V, Einhäupl K, Ganten D, Alladi S, Brayne C, Stephan BC, et al. Preventing dementia by preventing stroke: the Berlin Manifesto. *Alzheimers Dement*. 2019;15(7):961–84.
24. Gorelick PB, Scuteri A, Black SE, DeCarli C, Greenberg SM, Iadecola C, et al. Vascular contributions to cognitive impairment and dementia: a statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*. 2011;42(9):2672–713.
25. Uddin MS, Kabir MT, Al Mamun A, Abdel-Daim MM, Barreto GE, Ashraf GM. APOE and Alzheimer's disease: evidence mounts that targeting APOE4 may combat Alzheimer's pathogenesis. *Mol Neurobiol*. 2019;56(4):2450–65.
26. Gao X, Starmer J, Martin ER. A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genet Epidemiol*. 2008;32(4):361–9.
27. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Series B Stat Methodology*. 2005;67(2):301–20.
28. Ahrens A, Hansen CB, Schaffer ME. lasso-pack: Model selection and prediction with regularized regression in Stata. *Stata J*. 2020;20(1):176–235.
29. Windham BG, Parker SB, Zhu X, Gabriel KP, Palta P, Sullivan KJ, et al. Endurance and gait speed relationships with mild cognitive impairment and dementia. *Alzheimers Dement (Amst)*. 2022;14(1):e12281.
30. Toledo JB, Arnold M, Kastenmuller G, Chang R, Baillie RA, Han X, et al. Metabolic network failures in Alzheimer's disease: A biochemical road map. *Alzheimers Dement*. 2017;13(9):965–84.
31. Polis B, Samson AO. Role of the metabolism of branched-chain amino acids in the development of Alzheimer's disease and other metabolic disorders. *Neural Regen Res*. 2020;15(8):1460–70.
32. Suzuki H, Yamashiro D, Ogawa S, Kobayashi M, Cho D, Iizuka A, et al. Intake of Seven Essential Amino Acids Improves Cognitive Function and Psychological and Social Function in Middle-Aged and Older Adults: A Double-Blind, Randomized, Placebo-Controlled Trial. *Front Nutr*. 2020;7:586166.
33. Butterfield DA, Halliwell B. Oxidative stress, dysfunctional glucose metabolism and Alzheimer disease. *Nat Rev Neurosci*. 2019;20(3):148–60.
34. Yao J, Brinton RD. Estrogen regulation of mitochondrial bioenergetics: implications for prevention of Alzheimer's disease. *Adv Pharmacol*. 2012;64:327–71.
35. Jensen NJ, Wodschow HZ, Nilsson M, Rungby J. Effects of Ketone Bodies on Brain Metabolism and Function in Neurodegenerative Diseases. *Int J Mol Sci*. 2020;21(22):8767.
36. Domingues R, Pereira C, Cruz MT, Silva A. Therapies for Alzheimer's disease: a metabolic perspective. *Mol Genet Metab*. 2021;132(3):162–72.
37. de Leeuw FA, Karamujic-Comic H, Tijms BM, Peeters CFW, Kester MI, Scheltens P, et al. Circulating metabolites are associated with brain atrophy and white matter hyperintensities. *Alzheimers Dement*. 2021;17(2):205–14.
38. Salminen A, Haapasalo A, Kauppinen A, Kaarniranta K, Soininen H, Hiltunen M. Impaired mitochondrial energy metabolism in Alzheimer's disease: Impact on pathogenesis via disturbed epigenetic regulation of chromatin landscape. *Prog Neurobiol*. 2015;131:1–20.
39. Teo E, Ravi S, Barardo D, Kim HS, Fong S, Cazenave-Gassiot A, et al. Metabolic stress is a primary pathogenic event in transgenic *Caenorhabditis elegans* expressing pan-neuronal human amyloid beta. *Elife*. 2019;8:e50069.
40. van der Lee SJ, Teunissen CE, Pool R, Shipley MJ, Teumer A, Chouraki V, et al. Circulating metabolites and general cognitive ability and dementia: Evidence from 11 cohort studies. *Alzheimers Dement*. 2018;14(6):707–22.
41. Pedrini S, Hone E, Gupta VB, James I, Teimouri E, Bush AI, et al. Plasma High Density Lipoprotein Small Subclass is Reduced in Alzheimer's Disease Patients and Correlates with Cognitive Performance. *J Alzheimers Dis*. 2020;77(2):733–44.
42. Martinez AE, Weissberger G, Kuklenyik Z, He X, Meuret C, Parekh T, et al. The small HDL particle hypothesis of Alzheimer's disease. *Alzheimers Dement*. 2022;10.1002/alz.12649.
43. Chernick D, Zhong R, Li L. The Role of HDL and HDL Mimetic Peptides as Potential Therapeutics for Alzheimer's Disease. *Biomolecules*. 2020;10(9):1276.
44. Proitsi P, Kim M, Whitley L, Simmons A, Sattler M, Velayudhan L, et al. Association of blood lipids with Alzheimer's disease: A comprehensive lipidomics analysis. *Alzheimers Dement*. 2017;13(2):140–51.
45. Hosseini M, Poljak A, Braidy N, Crawford J, Sachdev P. Blood fatty acids in Alzheimer's disease and mild cognitive impairment: A meta-analysis and systematic review. *Ageing Res Rev*. 2020;60:101043.
46. Schilling S, Tzourio C, Soumare A, Kaffashian S, Dartigues JF, Ancelin ML, et al. Differential associations of plasma lipids with incident dementia and dementia subtypes in the 3C Study: A longitudinal, population-based prospective cohort study. *PLoS Med*. 2017;14(3):e1002265.
47. Ancelin ML, Ripoch E, Dupuy AM, Barberger-Gateau P, Auriacombe S, Rouaud O, et al. Sex differences in the associations between lipid levels and incident dementia. *J Alzheimers Dis*. 2013;34(2):519–28.
48. Zhu Y, Liu X, Zhu R, Zhao J, Wang Q. Lipid levels and the risk of dementia: A dose-response meta-analysis of prospective cohort studies. *Ann Clin Transl Neurol*. 2022;9(3):296–311.
49. Benn M, Nordestgaard BG, Frikke-Schmidt R, Tybjaerg-Hansen A. Low LDL cholesterol, PCSK9 and HMGCR genetic variation, and risk of Alzheimer's disease and Parkinson's disease: Mendelian randomisation study. *BMJ*. 2017;357:j1648.
50. Bernath MM, Bhattacharyya S, Nho K, Barupal DK, Fiehn O, Baillie R, et al. Serum triglycerides in Alzheimer disease: Relation to neuroimaging and CSF biomarkers. *Neurology*. 2020;94(20):e2088–98.
51. Janssen CI, Kiliaan AJ. Long-chain polyunsaturated fatty acids (LCPUFA) from genesis to senescence: the influence of LCPUFA on neural development, aging, and neurodegeneration. *Prog Lipid Res*. 2014;53:1–17.
52. Julkunen H, Cichonska A, Slagboom PE, Wurtz P. Nightingale Health UKB: Metabolic biomarker profiling for identification of susceptibility to severe pneumonia and COVID-19 in the general population. *Elife*. 2021;10:e63033.
53. Petermann-Rocha F, Lyall DM, Gray SR, Esteban-Cornejo I, Quinn TJ, Ho FK, et al. Associations between physical frailty and dementia incidence: a prospective study from UK Biobank. *Lancet Healthy Longevity*. 2020;1(2):e58–68.
54. Wilkinson T, Schnier C, Bush K, Rannikmae K, Henshall DE, Lerpiniere C, et al. Identifying dementia outcomes in UK Biobank: a validation study of primary care, hospital admissions and mortality data. *Eur J Epidemiol*. 2019;34(6):557–65.
55. Zhang H, Greenwood DC, Risch HA, Bunce D, Hardie LJ, Cade JE. Meat consumption and risk of incident dementia: cohort study of 493,888 UK Biobank participants. *Am J Clin Nutr*. 2021;114(1):175–84.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.