

TITLE PAGE

Metabolic correlates of late midlife cognitive function: findings from the 1946

British Birth Cohort

Rebecca Green^{ab}, Jodie Lord^a, Jin Xu^{ac}, Jane Maddock^d, Min Kim^e, Richard Dobson^{ab},
Cristina Legido-Quigley^{ce}, Andrew Wong^d, Marcus Richards^{d*}, Petroula Proitsi^{a*}

*equally contributing and co-corresponding authors

^a*Institute of Psychology, Psychiatry & Neuroscience, King's College London, London, UK.*

^b*UK National Institute for Health Research (NIHR) Maudsley Biomedical Research Centre, South London and Maudsley Hospital, London.*

^c*Institute of Pharmaceutical Science, King's College London, London, UK.*

^d*MRC Unit for Lifelong Health & Ageing at UCL, University College London, London, UK.*

^e*Steno Diabetes Center Copenhagen, Gentofte, Denmark.*

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

Abstract

Background: Investigating associations between metabolites and late midlife cognitive function could reveal potential markers and mechanistic insights relevant to early dementia. Here, we aimed to identify the metabolic underpinnings of cognitive outcomes in late midlife by exploring and integrating associations of single metabolites, metabolic pathways and networks. We further aimed to untangle the influence of life course factors on these relationships; a previously unexplored avenue using a systems biology approach.

Methods and Findings: Levels of 1019 metabolites were detected by liquid chromatography-mass spectrometry (Metabolon Inc) and quantified at age 60-64 among participants of the British 1946 Birth Cohort (N=1740). Cognitive outcomes were assessed at the same age and 5-9 years later, and included short-term memory (age 60-64, 69 and change), delayed memory (age 60-64), processing speed (age 60-64, 69 and change) and Addenbrooke's Cognitive Examination III (age 69). Using a combination of linear regression analysis, quantitative pathway analysis and weighted gene correlation network analysis, we evaluated relationships between metabolite measures (single-metabolites, pathways and network modules) and cognitive outcomes. Single-metabolite and network analyses were sequentially adjusted for life course factors across four models, including: sex and blood clinic information (model 1); model 1 + BMI and lipid medication (model 2); model 2 + childhood cognition, education and socioeconomic position (model 3); model 3 + smoking, exercise, alcohol intake, blood pressure and diet (model 4). After correcting for multiple tests, we identified 155 metabolites, 10 pathways and 5 modules to show relationships with cognitive outcomes. Thirty-five metabolites were influential in their module and identified in single-metabolite analyses. Notably, we report

independent relationships between a module comprised of acylcarnitines and processing speed, revealing palmitoylcarnitine (C16) as a key driver of associations (model 4: $\beta = -0.10$, 95%CI = -0.15 to -0.052). Two modules demonstrated associations with several cognitive outcomes that were partly explained by life course factors: one enriched in modified nucleosides and amino acids (β range (model 1) = -0.12 to -0.09, attenuation (model 4) = 39.2 to 55.5%), and another in vitamin A and C metabolites (β range (model 1) = 0.11 to 0.23, attenuation (model 4) = 68.6 to 92.6%). Our other findings, including a module enriched in sphingolipid pathways (β range (model 1) = 0.085 to 0.10, attenuation (model 4) = 87.0 to 116%), were entirely explained by life course factors - particularly childhood cognition and education. The limitations of this study include those commonly seen with population-based cohorts, such as possible residual confounding and generalisability to other populations, as well as a lack of longitudinal metabolite data.

Conclusions: Using a large birth cohort study with information across the life course, we highlighted potential metabolic mechanisms underlying cognitive function in late midlife, suggesting marker candidates and life course relationships for further study.

Introduction

Cognitive function in late midlife is indicative of future cognitive trajectories and risk of dementia (1). As dementia is proposed to have a long prodrome, where pathology is accumulating but clinical criteria are not yet met, there presents a promising window to prevent or delay pathology (1). However, a lack of clinically significant symptoms impedes our ability to identify individuals for potential risk-reduction and treatment strategies. As such, our understanding of early disease mechanisms are not well established and no effective disease-modifying treatments are currently used in clinic (2).

Comprehensive longitudinal studies are required to detect early mechanisms and markers preceding diagnosis, for which studying metabolic correlates may be fruitful. Metabolites, such as fatty acids and amino acids, are low molecular weight compounds derived from cellular metabolism. Lying in closest proximity to the phenotype, they integrate upstream biological systems (e.g. genetics, transcriptomics, proteomics) as well as environmental and lifestyle influences, allowing for a holistic insight into the physiological status of an individual (3). Additionally, they are accessible and potentially modifiable, presenting as promising candidates for markers of pathology (4).

The biological relevance of metabolic alterations in cognitive function and dementia has been established. Contextually, genome-wide association studies have highlighted enrichment in lipid metabolism pathways in the genetic underpinnings of Alzheimer's disease (AD) (5). Further, many studies have linked metabolites to cognitive function and AD, consistently highlighting species such as sphingolipids, phospholipids, fatty acids,

cholesterol and amino acids (6–10), although replication of specific metabolites has proved challenging.

As metabolites are known to act in concert, investigating many metabolites together can enhance biological interpretation. Systems-level analyses, that are able to capture the complex interactions between metabolites, are required to guide our understanding and identify potential marker candidates for future studies. We recently employed a network approach to explore systems-level changes in AD phenotypes, whereby lipids and proteins were grouped into clusters based on their connectivity, highlighting important pathways in AD pathogenesis (11). A similar methodology was previously applied in the Bogalusa Heart Study to explore the metabolic signature of cognitive function in early midlife (10).

However, the molecular correlates of cognitive function relevant to this prodromal window remains unknown. Additionally, given that both metabolites and cognitive function can be influenced by life course factors, it is necessary to evaluate relationships from this context. To our knowledge, life course influences have not yet been explored using a systems approach; it is hoped that this could highlight independent associations as well as suggest relationships for further study - a potentially invaluable layer in untangling early pathology.

Using the Medical Research Council (MRC) National Survey of Health and Disease (NSHD) - the British 1946 Birth Cohort - we aimed to comprehensively investigate associations between metabolites and cognitive function in late midlife using a life course approach (Fig 1). Previously, levels of 233 metabolites and their association with cognitive function were explored in this cohort (6). Since that publication, we now have levels of >1000 metabolites quantified, as well as availability of the Addenbrooke's Cognitive Examination (ACE-III), a

111 comprehensive measure of cognitive state that is also able to screen for cognitive
 112 impairment and dementia (12). This provides a good opportunity to delineate pathways and
 113 networks associated with cognitive outcomes in late midlife. Integrating the depth and
 114 breadth of metabolite-level, pathway-level and network-level approaches, we aimed to
 115 identify functionally significant metabolites that may show merit as markers of early
 116 pathology. With lifelong information available, we explored the influence of life course
 117 factors to untangle these associations further.

MRC 1946 British Birth Cohort Study (n=1740)

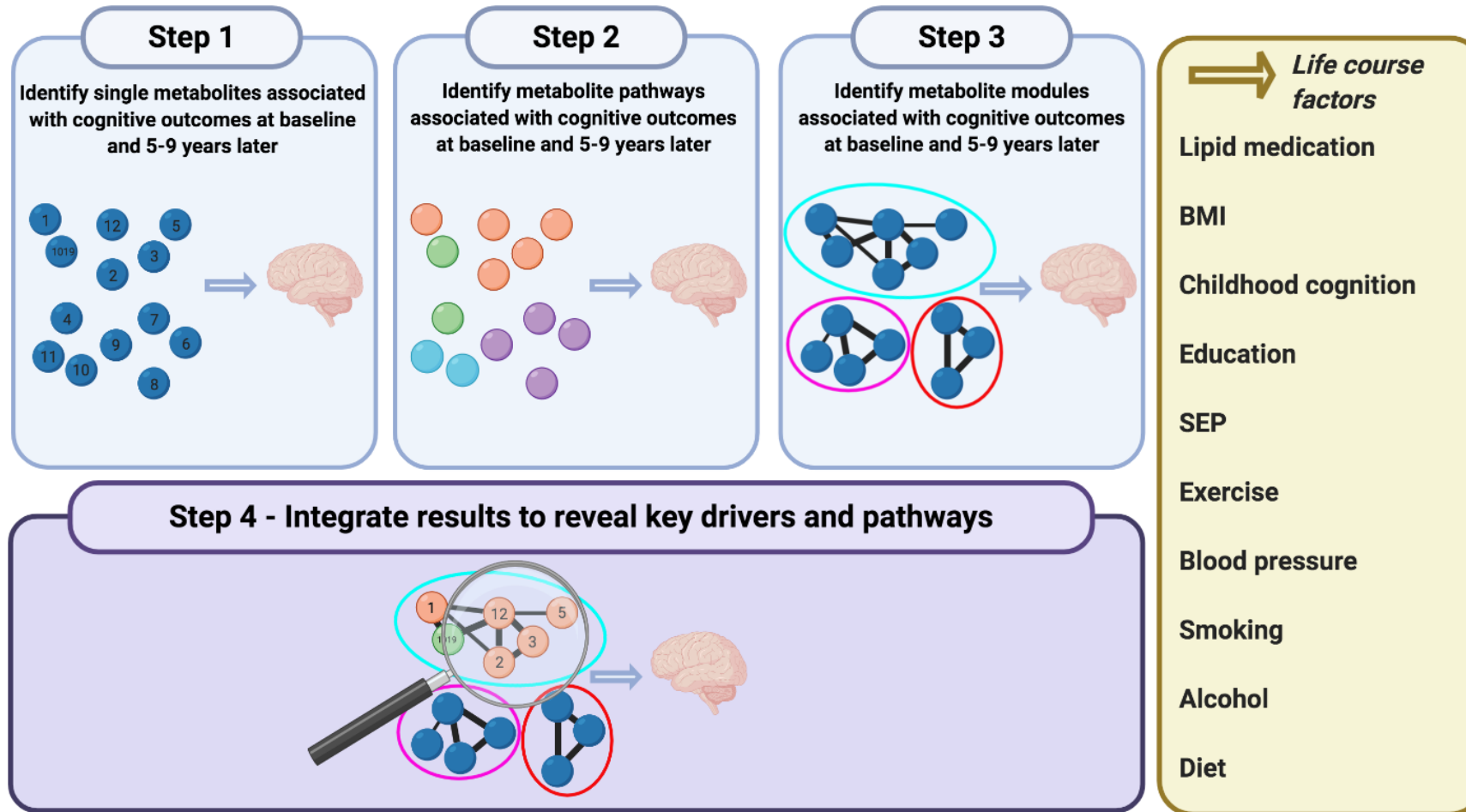


Fig 1. Overview of study workflow. Created with BioRender.com

2. Methods and Materials

2.1 Participants

The MRC NSHD is a nationally representative study, originally consisting of 5362 participants born in England, Scotland and Wales during one week of March in 1946 (13). Twenty-four waves of data have been collected since birth, with the most recent two follow-ups occurring at ages 60-64 (n=2229) and 68-69 years (n=2149). The study sample remains broadly representative of the British-born population at the same age (see (13,14) for further details on participant attrition and representativeness). Participants with full cognition, metabolite and blood clinic data at ages 60-64 were included for this analysis, giving a total sample of 1740 (see S1 Fig for further details on participant flow).

At age 60-64, ethical approval was obtained from the Greater Manchester Local Research Ethics Committee for the four English sites and the Scotland A Research Ethics Committee. Ethical approval for age 69 was obtained from Queen Square Research Ethics Committee (13/LO/1073) and Scotland A Research Ethics Committee (14/SS/1009). Research was conducted in accordance with the Declaration of Helsinki and all participants provided written informed consent at each wave.

2.2 Metabolomics

2.2.1. Metabolomics platform

Blood samples were collected during the nurse visit at ages 60-64 (96% fasted, 4% non-fasted). Samples were aliquoted and stored at -80°C.

Levels of 1401 plasma metabolites were profiled by Metabolon Inc (Durham, NC, USA) using Ultrahigh Performance Liquid Chromatography-Tandem Mass Spectrometry (UPLC-MS/MS). (see S1 Methods for further details). Metabolites were assigned to nine families: lipids (n=494), xenobiotics (n=256), amino acids (n=211), cofactors and vitamins (n=35), nucleotides (n=38), peptides (n=28), carbohydrates (n=22), partially characterised molecules (n=17) and energy (n=9). Additionally, the identity of 291 metabolites were unknown; these metabolites were allocated numbers prefixed by an “X” and were not assigned to any family. Metabolites were further organised into pathways by *Metabolon* based on their proposed biological function informed by the Kyoto Encyclopaedia of Genes and Genomes (KEGG) database.

2.2.2. Metabolomics data quality control (QC)

For full details on data QC, see S1 Fig. Briefly, metabolites with >20% of missing data were excluded, leaving 1019 for further analysis. Of the 1019, 193 were metabolites with an unknown identity. Remaining missing data were then taken forward for imputation using k-nearest-neighbours with k=10, as recommended for LC-MS data elsewhere (15). Imputation was carried out using the *impute* package in R (16). Imputed data were then log₁₀ transformed to achieve approximately normal distributions.

2.3 Cognitive outcomes

Cognitive outcome measures were recorded at two ages, 60-64 and 69. Four aspects of cognitive function were assessed:

Short-term memory (age 60-64 & 69)

Participants were asked to recall a 15-item word list, developed by the NSHD, after being presented with each word for two seconds. The task was repeated over three trials and the number of accurately recalled words was recorded (max score=45) (17).

Processing speed (age 60-64 & 69)

Participants were asked to cross out the letters P and W, randomly distributed on a page containing other letters. One minute was given to complete the task and participants were scored by the number and accuracy of the letters crossed out (max score=600) (17).

Delayed memory (age 60-64)

After the processing speed task, an uncued delayed free recall trial was administered (17).

Addenbrooke's Cognitive Examination-III (ACE-III) (age 69)

The ACE-III captures cognitive state, and is also a screening tool for cognitive impairment, comprised of five domains: attention and orientation, verbal fluency, memory, language and visuospatial function. Scores represent the total over all domains (max score=100), with lower scores indicating poorer cognitive function (12).

Cognitive change

For outcomes with available data at two time points - short-term memory and processing speed - we additionally investigated change in cognition, represented by the standardised residuals of a regression model fit between scores at age 60-64 and 69.

2.4 Covariables

As with previous analyses (6), covariables included the following: sex, blood clinic information (age at blood collection, clinic location and fasting status), body mass index (BMI), lipid medication, childhood cognition, educational attainment, childhood socioeconomic position (SEP), midlife SEP, lifetime smoking, alcohol intake, systolic blood pressure, physical activity and diet.

BMI was calculated using height and weight measures collected during the nurse visit at ages 60-64. At blood collection, the self-reported use of *lipid medication* was recorded and coded as a binary variable reflecting use in the previous 24 hours.

Childhood cognition was represented as a standardised composite score of four tests at age 15, including the Heim AH4 (measuring non verbal and verbal ability) (18), the Watts Vernon reading test (measuring reading comprehension) (19), and a test of mathematical ability (19). *Educational attainment* represented highest level of educational qualification by age 26, grouped into three categories: no qualifications, ordinary ('O') level secondary qualifications, or advanced ('A') level secondary and higher. *SEP* was represented in childhood and midlife, coded using the current or last known occupation of the father at age 11 and the study member at age 53, respectively. These categories corresponded to those specified in the UK Registrar General's classification: unskilled, partly skilled, skilled manual, skilled nonmanual, intermediate, or professional.

For lifestyle, *lifetime smoking* was represented by pack years per person between the ages of 20 and 60-64. *Physical activity* was coded as three categories depending on the self-

reported frequency of participation in sports, exercises or intense leisure activities in the month prior to the age 60-64 interview: none, 1-4 times per week, or >4 times per week. Where data were present for at least three of four timepoints, average daily *alcohol intake* during midlife was curated from 3-5 day diet diaries at age 36, 43, 53 and 60-64. This measure was then used to assign participants into three categories: no consumption, light-to-moderate consumption, and heavy consumption. *Systolic blood pressure* was represented by the second measurement (mmHg) taken at age 60-64. Finally, the *diet* variable represented adherence scores for the Dietary Approaches to Stop Hypertension (DASH) diet, estimated from 3-5 day diet diaries at age 60-64 (20). The DASH diet is based on a high intake of fruits, vegetables, low-fat dairy products and wholegrains, and a low intake of saturated fat and refined sugars (21). Participants were assigned to sex-specific quintiles, with lower quintiles indicating lower adherence, as described previously (20).

APOE genotype was determined from blood samples collected at age 53 or 69 and analysed as described previously (22). Participants with $\epsilon 2/\epsilon 4$ were excluded (n=46) and *APOE* genotype was coded as homozygous $\epsilon 4$ (n=46), heterozygous $\epsilon 4$ (n=361) or non $\epsilon 4$ (n=1068). Genotypes were treated as continuous variables.

2.5 Statistical analyses

To optimise power and minimise bias, missing covariate data were imputed using multiple imputation chained equations (*mice*) (23), resulting in fifty imputed datasets (see Table 1 for further details of missing data). All analyses were conducted in R and all results are reported in SD units.

2.5.1 Single-metabolite analyses

Predictors and outcomes were z-standardised prior to statistical analysis to allow for direct comparisons. Statistical models evaluated associations between metabolites (predictor) and cognitive test scores (outcome) using multiple linear regression. Regression analyses were performed on each imputed dataset, and estimates were pooled using Rubin's rules. To investigate associations in the context of life course influences, a series of statistical models with sequential adjustment were performed:

Model 1 (basic covariates): sex, blood clinic, age at blood clinic, fasting status

Model 2 (common metabolite confounders): model 1 + BMI, lipid medication

Model 3 (social factors and childhood cognition): model 2 + childhood cognition, attainment, SEP (childhood and midlife)

Model 4 (lifestyle influences): model 3 + blood pressure, physical activity, alcohol, smoking, diet

To account for the number of tests performed and correlations between metabolites, a Bonferroni-adjusted significance threshold was set as 0.05 divided by the number of principal components explaining >95% variance in the metabolite data (6); $p < 1.15 \times 10^{-4}$.

2.5.2 Pathway analyses

First, metabolon sub-pathway definitions were used to organise metabolites into pathways. Those containing <5 metabolites were excluded, resulting in 53 separate pathways. Metabolite pathway assignment can be found in S1 Table.

To investigate pathway-level associations, quantitative pathway analysis was performed using an approach reported previously (24). Briefly, we derived pathway z-scores for each participant, representing the standardised expression of metabolites in the pathway. To do this, metabolites were z-standardised, and the mean expression was computed for each pathway. Associations between pathway z-scores and outcomes were then evaluated using linear regression, adjusting for the basic (model 1) covariables listed above.

A Bonferroni-corrected significance threshold was set at 0.05/53 pathways; $p < 9.43 \times 10^{-4}$.

2.5.3 Network analyses

Network construction

To define metabolic networks, we applied weighted gene coexpression network analysis (WGCNA) to metabolite data, using the WGCNA package in R (25–27). WGCNA is a network analysis approach that organises data into densely connected modules based on pairwise correlations, whereby data in the same module will show high connectivity and those in differing modules, low. Subsequently, the first principal component of the module (module eigenvalue) can be derived and relationships between modules and outcomes can be explored. To infer biological function, overrepresentation analyses are commonly conducted to identify enriched pathways within the module. Additionally, the function of metabolites showing the greatest modular connectivity can be interrogated (25).

Metabolites were first adjusted for model 1 covariables and the standardised residuals were used for subsequent analysis. Next, the standardised connectivity (Z_k) for each sample was computed to identify outliers, resulting in the exclusion of ten individuals with a Z_k of < -4 .

We then derived a pairwise correlation matrix using biweight midcorrelations between all metabolites. From this, a weighted, signed adjacency matrix was constructed by raising correlations to a soft thresholding power of 9, chosen to meet a scale-free topology threshold of ≥ 0.85 while maximising mean connectivity (S3 Fig). Subsequently, the adjacency matrix was transformed into a topological overlap matrix (TOM), representing the network connectivity of metabolites. Metabolites were then hierarchically clustered into a dendrogram using an average linkage method based on their dissimilarity (1-TOM), and the dendrogram was cut using a dynamic hybrid tree cutting algorithm (28) (parameters - minModuleSize=20, deepSplit=4 and mergeHeight=0.3), resulting in 15 metabolite modules. Of these, the 'grey' module, comprised of metabolites that were not assigned to any particular module, was dropped from further analysis. Module eigenvalues were computed for the remaining 14 modules.

Overrepresentation analyses using the hypergeometric test were performed on modules to identify pathways expressed more than expected by chance. Pathway assignment is detailed in 2.5.2. For all module analyses, a Bonferroni-corrected significance threshold was set at $0.05/14$ modules; $p < 1.14 \times 10^{-3}$.

Regression analyses

To evaluate module-outcome relationships in the context of life course influences, modules were subject to the same series of regression models listed in 2.5.1, using z-standardised module eigenvalues as predictors. As modules were curated using the residuals of a regression model fit between metabolites and model 1 covariables, only models 2-4 were performed.

Identification of module hubs

Metabolites that have a high intramodular connectivity (“hub” metabolites) are highly influential in module structure and likely to play central roles in biological function. Highly connected metabolites that also associate with our outcomes thus present as relevant and functionally important candidates for further study.

To identify hubs, we evaluated associations between metabolites and their assigned module (module membership; kME) using correlations between metabolites and module eigenvalues. Metabolites with a kME of >0.65 were defined as hubs, and filtered for those identified in single-metabolite analyses.

2.5.4 Additional analyses

In our preliminary analysis, we investigated associations between all covariables and metabolites, and all covariables and outcomes (adjusting for model 1 covariables) (S2 Table). To further investigate whether particular covariables may be driving attenuations, we repeated single-metabolite and module regression analyses, adjusting for model 1 covariables and each additional covariable individually (S3 Table and S4 Table).

For any results significant at the Bonferroni threshold, analyses were rerun additionally adjusting for *APOE* genotype to investigate whether relationships were independent of *APOE* (S5 Table and S6 Table).

3. Results

3.1 Participant characteristics

Complete metabolite, cognition and blood clinic data at age 60-64 were available for 1740 study participants. Repeated measures at age 69 were present for 1482 (short-term memory) and 1496 (processing speed), and ACE-III scores were present for 1255. Characteristics of participants with complete data for each outcome are shown in Table 1.

Table 1. Clinical and demographic characteristics of participants of the MRC 1946 British Birth Cohort study

	Full cognition and metabolite data			
	Age 60-64 (N=1740)	Age 60-64 & processing speed age 69 (N=1496)	Age 60-64 & short-term memory age 69 (N=1482)	Age 60-64 & ACE-III age 69 (N=1255)
Sex				
Male	854 (49.1%)	732 (48.9%)	727 (49.1%)	607 (48.4%)
Female	886 (50.9%)	764 (51.1%)	755 (50.9%)	648 (51.6%)
Fasting blood sample				
Yes	1675 (96.3%)	1446 (96.7%)	1432 (96.6%)	1212 (96.6%)
No	65 (3.7%)	50 (3.3%)	50 (3.4%)	43 (3.4%)
Age				
Mean (SD)	63.20 (1.12)	63.21 (1.12)	63.21 (1.12)	63.22 (1.13)
Father's socioeconomic position, n (%)				
Missing (n)	86	75	75	60
Professional	128 (7.7%)	118 (8.3%)	118 (8.4%)	106 (8.9%)
Intermediate	375 (22.7%)	325 (22.9%)	323 (23.0%)	272 (22.8%)
Nonmanual skilled	293 (17.7%)	264 (18.6%)	261 (18.6%)	217 (18.2%)
Manual skilled	488 (29.5%)	421 (29.6%)	419 (29.8%)	362 (30.3%)
Partly skilled	282 (17.0%)	227 (16.0%)	225 (16.0%)	185 (15.5%)
Unskilled	88 (5.3%)	66 (4.6%)	61 (4.3%)	53 (4.4%)
Socioeconomic position (age 15-53), n (%)				
Missing (n)	10	5	5	3
Professional	139 (8.0%)	125 (8.4%)	124 (8.4%)	104 (8.3%)

Intermediate	724 (41.8%)	641 (43.0%)	638 (43.2%)	542 (43.3%)
Nonmanual skilled	404 (23.4%)	354 (23.7%)	349 (23.6%)	296 (23.6%)
Manual skilled	250 (14.5%)	200 (13.4%)	198 (13.4%)	164 (13.1%)
Partly skilled	165 (9.5%)	132 (8.9%)	130 (8.8%)	114 (9.1%)
Unskilled	48 (2.8%)	39 (2.6%)	38 (2.6%)	32 (2.6%)
Educational attainment (age 26), n (%)				
Missing (n)	90	71	68	55
No qualification	462 (28.0%)	376 (26.4%)	370 (26.2%)	308 (25.7%)
Up to GCSE	477 (28.9%)	409 (28.7%)	406 (28.7%)	354 (29.5%)
A-level or higher	711 (43.1%)	640 (44.9%)	638 (45.1%)	538 (44.8%)
Childhood cognition (age 15) (z-score)				
Missing (n)	252	217	215	177
Mean (SD)	0.00 (1.00)	0.05 (0.99)	0.06 (0.98)	0.07 (0.99)
Lipid medication (age 60-64), n (%)				
No	1318 (75.7%)	1135 (75.9%)	1124 (75.8%)	946 (75.4%)
Yes	422 (24.3%)	361 (24.1%)	358 (24.2%)	309 (24.6%)
BMI (age 60-64) (weight(kg)/height(m)²)				
Missing (n)	2	2	2	2
Mean (SD)	27.76 (4.70)	27.70 (4.56)	27.69 (4.58)	27.61 (4.45)
Physical activity (age 60-64), n (%)				
Missing (n)	48	40	37	32
None	1036 (61.2%)	862 (59.2%)	857 (59.3%)	714 (58.4%)
1-4 times a month	251 (14.8%)	229 (15.7%)	225 (15.6%)	192 (15.7%)
4+ times a month	405 (23.9%)	365 (25.1%)	363 (25.1%)	317 (25.9%)
Lifetime smoking (age 20-64), n (%)				
Missing (n)	347	291	285	243
Mean (SD)	10.79 (15.54)	10.40 (15.28)	10.26 (15.18)	10.51 (15.43)
Lifetime alcohol consumption (age 36-64), n (%)				
Missing (n)	285	226	221	192
No consumption	90 (6.2%)	77 (6.1%)	76 (6.0%)	63 (5.9%)
Light-moderate consumption	1134 (77.9%)	993 (78.2%)	987 (78.3%)	845 (79.5%)
Heavy consumption	231 (15.9%)	200 (15.7%)	198 (15.7%)	155 (14.6%)
Diet quintiles (age 60-64), n (%)				
Missing (n)	246	189	185	166
1	332 (22.2%)	276 (21.1%)	276 (21.3%)	231 (21.2%)
2	302 (20.2%)	266 (20.4%)	263 (20.3%)	209 (19.2%)
3	309 (20.7%)	266 (20.4%)	263 (20.3%)	230 (21.1%)
4	270 (18.1%)	244 (18.7%)	244 (18.8%)	199 (18.3%)

	5	281 (18.8%)	255 (19.5%)	251 (19.4%)	220 (20.2%)
Systolic blood pressure (age 60-64)					
Missing (n)		12	7	6	5
Mean (SD)		136.40 (18.30)	135.78 (18.18)	135.83 (18.20)	135.72 (18.28)
APOE genotype, n (%)					
Missing (n)		217	169	167	146
ε2/ε2		10 (0.7%)	9 (0.7%)	9 (0.7%)	5 (0.5%)
ε2/ε3		177 (11.6%)	152 (11.5%)	151 (11.5%)	119 (10.7%)
ε2/ε4		48 (3.2%)	39 (2.9%)	38 (2.9%)	31 (2.8%)
ε3/ε3		881 (57.8%)	771 (58.1%)	763 (58.0%)	650 (58.6%)
ε3/ε4		361 (23.7%)	316 (23.8%)	314 (23.9%)	270 (24.3%)
ε4/ε4		46 (3.0%)	40 (3.0%)	40 (3.0%)	34 (3.1%)

3.2 Single-metabolite analyses

Overall, we identified 155 metabolites to be associated with a least one cognitive outcome after adjusting for multiple testing: 65 metabolites showed positive relationships, with increased abundance reflecting better cognitive function, and 90 showed negative relationships. No metabolites were associated with cognitive change measures at the Bonferroni threshold; as such, these outcomes will not be further discussed in this section but are presented in S1 Table. Significant metabolites represented all metabolon families but were largely defined as lipids (n=46) and amino acids (n=32). The largest proportion of these metabolites were associated with short-term memory at age 60-64 only, and just over half were associated with more than one outcome (Fig 2). As expected, this indicates some molecular overlap between the cognitive outcomes, although we did observe some unique relationships for processing speed in particular (Fig 2). Full summary statistics are available in S1 Table and visualised in S2 Fig.

369

370

371

372

373

374

375

376

377

378

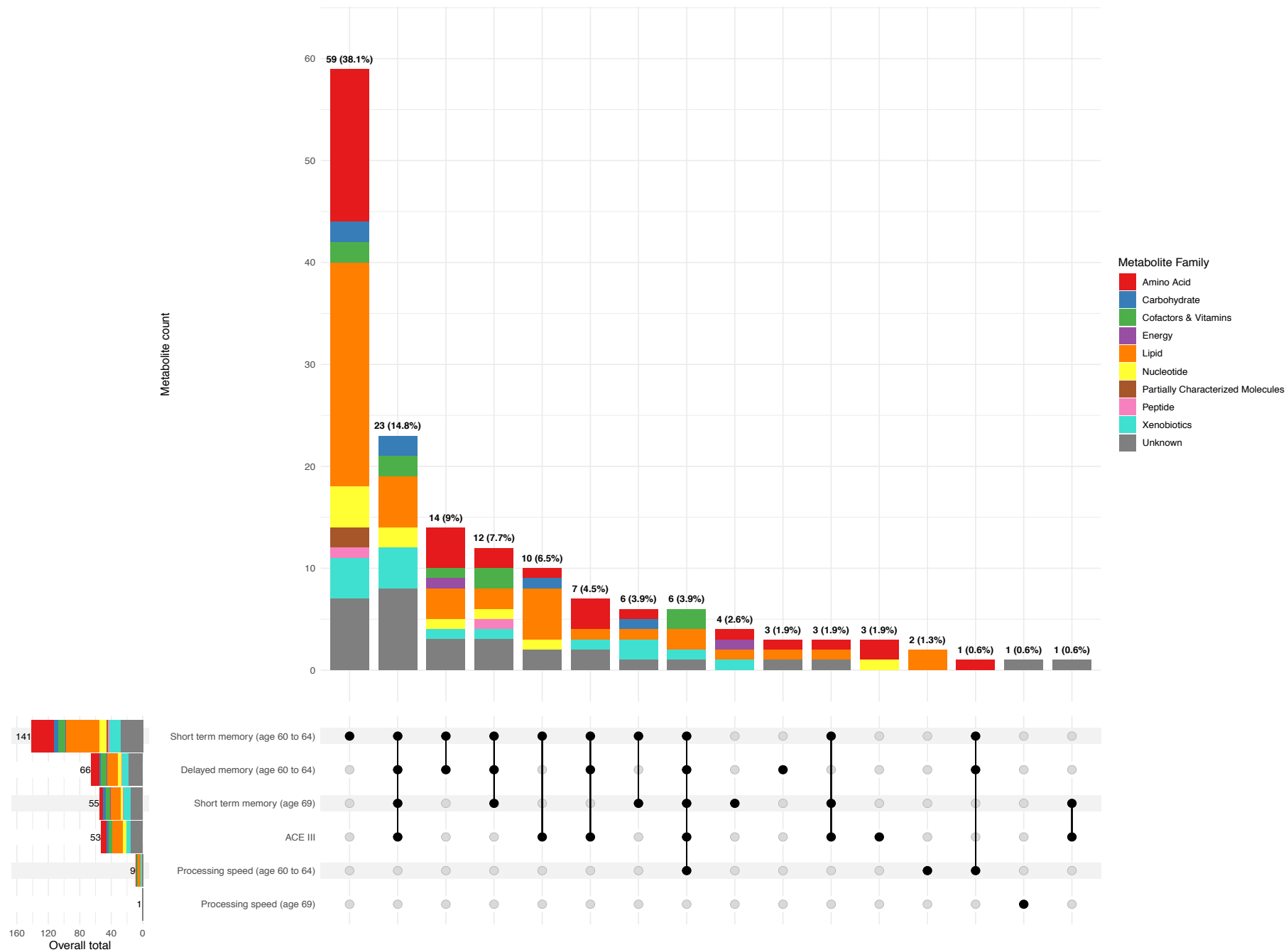
379

380

381

382

383



384 **Fig 2. (Main) Upset plot depicting the number of metabolites associated with each outcome and combination of outcomes at the adjusted**
385 **threshold, split by metabolite family. Outcomes are shown in the matrix below, with shaded circles demonstrating those represented by**
386 **the bar chart above. Where more than one outcome is indicated, lines further highlight these intersections. Bars are coloured by**
387 **metabolite family and total metabolite counts are displayed on top of each bar, alongside the corresponding percentage proportion of all**
388 **significant metabolites identified across all outcomes. As no metabolites were significant at the adjusted threshold for change outcomes,**
389 **these were not included. (Left) Barplot showing the number of metabolites identified overall for each outcome. Bars are coloured by**
390 **metabolite family and counts are displayed on the top of each bar. This plot was produced using (66). Underlying data are present in S1**
391 **Table. ACE-III = Addenbrooke's Cognitive Examination-III.**

Of the 155 metabolites identified across all models, 151 metabolites were identified in model 1. For all non-change outcomes, the metabolites showing the strongest positive effects were cofactors and vitamins involved in vitamin A metabolism; β -cryptoxanthin and carotene diol (1), although this relationship was not significant at the adjusted threshold for processing speed at age 69 ($\beta=0.08$, 95%CI=0.027 to 0.13, $p=3.02 \times 10^{-3}$). The strongest negative relationships were more distinct. Palmitoylcarnitine (C16), a lipid involved in fatty acid (acyl carnitine) metabolism, showed the largest negative association with processing speed at age 60-64. For the other outcomes, several unknown metabolites were revealed: X-24953 (short-term memory (both ages) and delayed memory), X-11372 (ACE-III) and X-21470 (processing speed at age 69).

After adjusting for BMI and lipid medication (model 2), average effect sizes reduced by 4.1% to 14.8%. However, the bulk of associations were attenuated after adjusting for model 3 (parental and own SEP, childhood cognition and own education). Model 3 attenuations were most pronounced for memory outcomes and ACE-III, ranging from 51.3% to 58.4%, and milder for processing speed (age 69 - 10.7%, age 60-64 - 31.5%). Subsequently, seven metabolites were significant at the adjusted threshold, including one previously nominal association – margaroylcarnitine (C17)*.

In the fully adjusted model (model 4), four metabolites remained significant at the adjusted threshold: X – 17676 (short-term memory and delayed memory at age 60-64), C16 (processing speed at age 60-64), C17 (processing speed at age 60-64) and imidazole propionate (short-term memory age 69) (see Table 2 for final results and overall change from models 1 to 4). Eighty metabolites remained nominally significant. For processing

speed at age 60-64, model 4 adjustments resulted in a modest decrease in effect magnitude (19.7%), with milder reductions for the other outcomes (3.2-7.9%). Of note, we did not identify any significant changes in our results following adjustment for APOE (S5 Table).

Table 2. Biological family, pathway and key results for significant metabolites in the final model (Model 4)

Metabolite	Family	Pathway	Model 1	Model 4	β change from Model 1 to 4 (%)
X – 17676	Unknown	Unknown	STM (69y) β = -0.14 95%CI = -0.19 to -0.089 p = 5.29×10^{-8} DM (60-64y) β = -0.11 95%CI = -0.15 to -0.061 p = 5.45×10^{-6}	STM (69y) β = -0.11 95%CI = -0.15 to -0.06 p = 4.84×10^{-6} DM (60-64y) β = -0.082 95%CI = -0.12 to -0.04 p = 1.11×10^{-4}	STM (69y) β change = -24.69% DM (60-64y) β change = -23.56%
Palmitoylcarnitine (C16)	Lipid	Fatty Acid Metabolism (Acyl Carnitine)	PS (60-64y) β = -0.11 95%CI = -0.16 to -0.062 p = 9.19×10^{-6}	PS (60-64y) β = -0.10 95%CI = -0.15 to -0.052 p = 5.99×10^{-5}	PS (60-64y) β change = -9.49%
Margaroylcarnitine (C17)*	Lipid	Fatty Acid Metabolism (Acyl Carnitine)	PS (60-64y) β = -0.067 95%CI = -0.11 to 0.020 p = 5.40×10^{-3}	PS (60-64y) β = -0.095 95%CI = -0.14 to -0.048 p = 8.33×10^{-5}	PS (60-64y) β change = +40.75%
Imidazole propionate	Amino Acid	Histidine Metabolism	STM (69y) β = -0.14 95%CI = -0.19 to -0.091 p = 4.98×10^{-8}	STM (69y) β = -0.094 95%CI = -0.14 to -0.048 p = 7.91×10^{-5}	STM (69y) β change = -33.48%

STM = short-term memory, DM = delayed memory, PS = processing speed.

3.3 Pathway analyses

Next, we characterised potential pathway involvement. Our results are presented in Fig 3 and S7 Table. No pathway was significant at the adjusted threshold for processing speed at age 69, nor for the cognitive change measures, although some nominal associations were observed.

Pathways significant at the adjusted threshold ($p < 9.43 \times 10^{-4}$) are discussed hereafter.

Briefly, the Vitamin A metabolism pathway was positively associated with all other outcomes. Ascorbate and aldarate metabolism, another pathway related to cofactors and vitamins metabolism, was also positively associated with short-term memory at both time points.

Negative relationships were observed between the purine metabolism (adenine containing) pathway and short-term memory, delayed memory and ACE-III. Short-term memory displayed some unique relationships at the adjusted threshold, although these pathways did show nominal relationships with at least two of the other outcomes. At age 60-64, five pathways (methionine; cysteine, SAM and taurine metabolism; urea cycle (arginine and proline metabolism); purine metabolism ((hypo)xanthine/inosine); fatty acid (monohydroxy and endocannabinoid) showed additional negative associations. At age 69, two pathways (pyrimidine metabolism (uracil); gamma-glutamyl amino acid) were additionally associated.

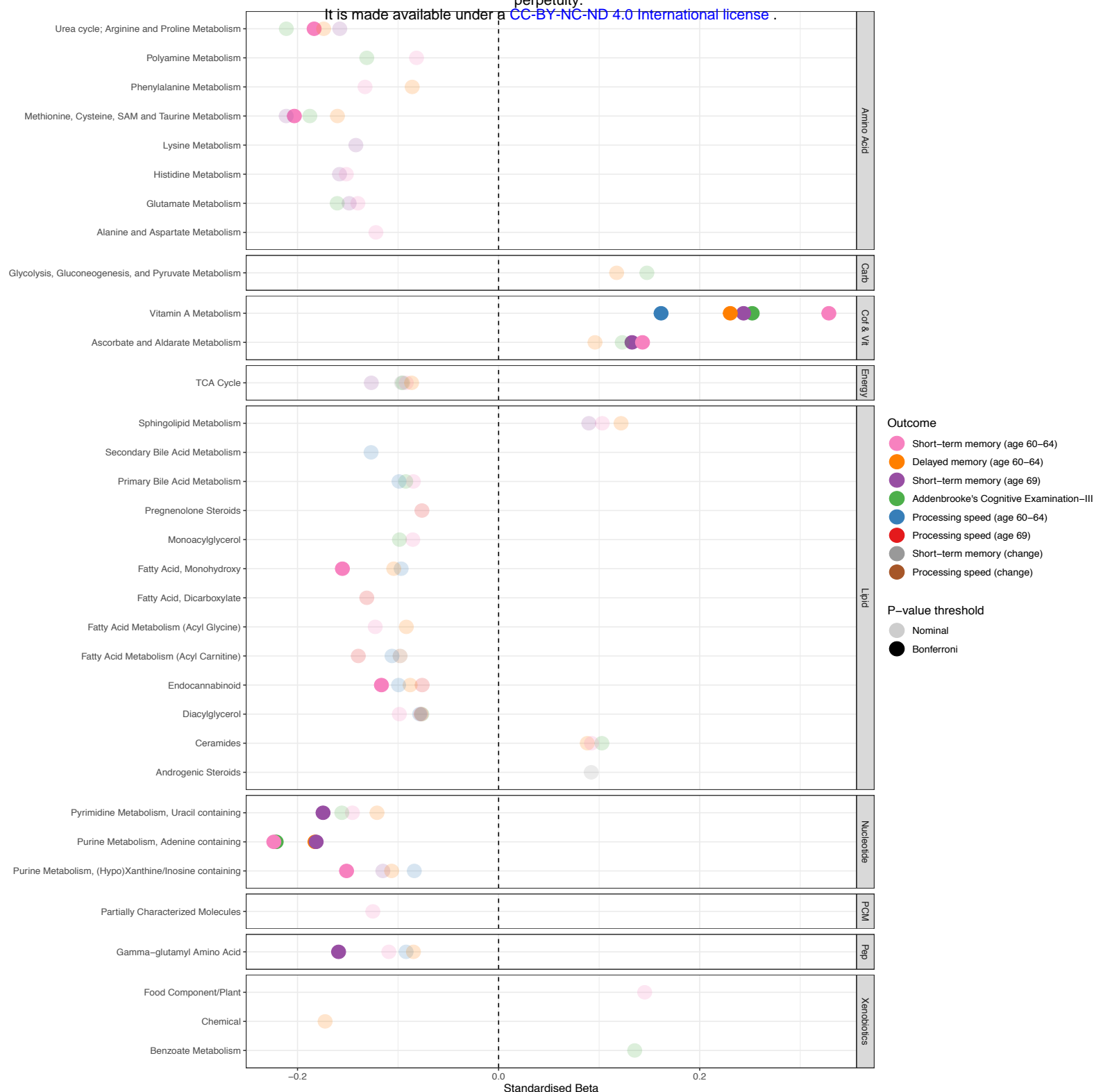


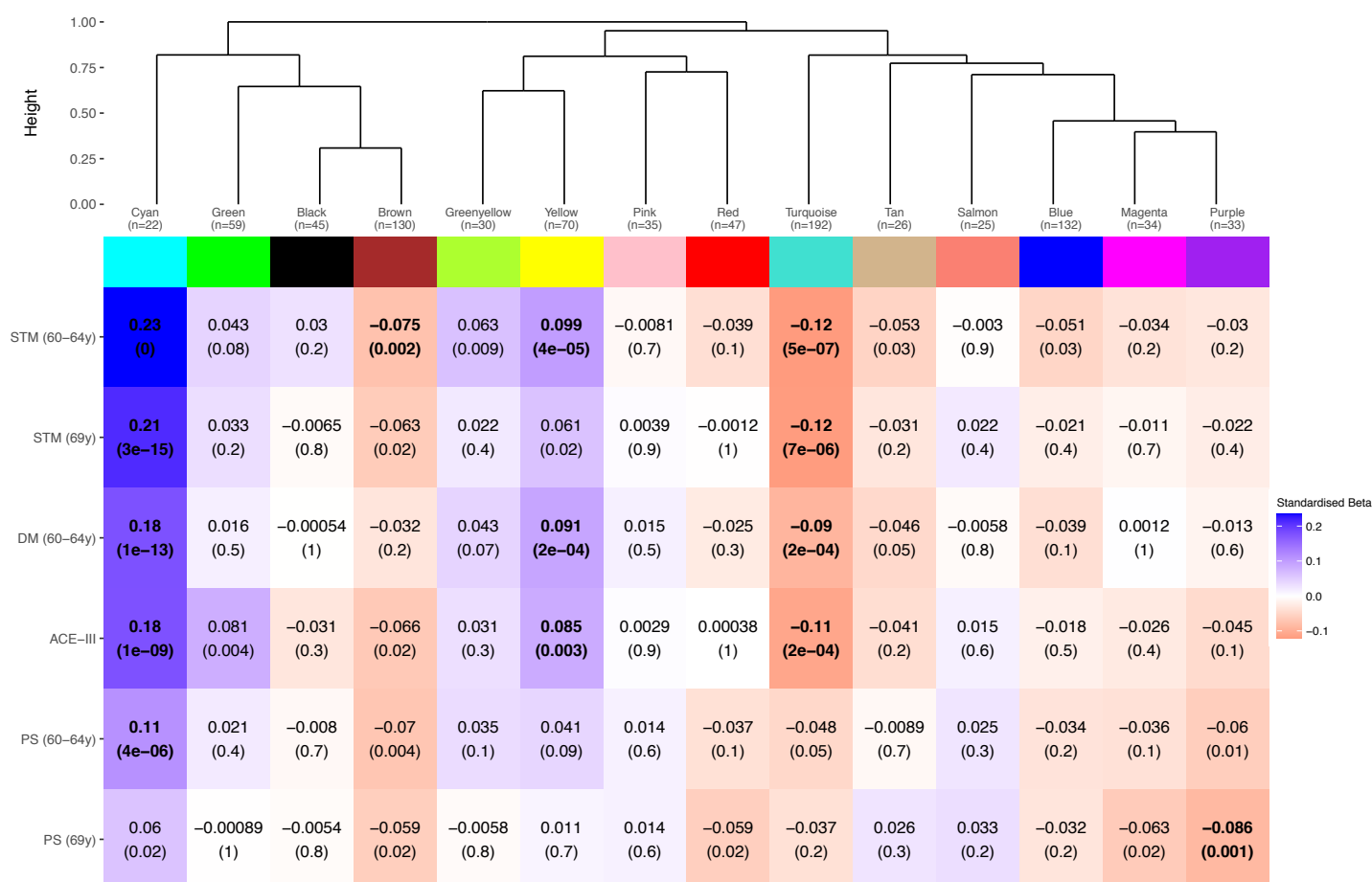
Fig 3. Plot representing pathway-outcome associations, organised by metabolite family.
Bonferroni-significant pathways ($p < 9.43 \times 10^{-4}$) are represented by a solid fill and nominal metabolites by a faint fill ($p < 0.05$. Underlying data are present in S7 Table.
Carb = carbohydrates, Cof & Vit = cofactors and vitamins, PCM = partially characterized molecules, Pep = peptides.

3.5 Network analyses

We next adopted a network approach, WGCNA, to explore relationships between groups of densely connected metabolites and cognitive outcomes. We identified 14 modules comprised of between 22 to 192 metabolites, with regression analyses revealing five of these to be associated with cognitive outcomes at the adjusted threshold ($p < 1.14 \times 10^{-3}$): two showing positive associations and three showing negative associations (Fig 4). All but one module were enriched in a biological pathway (Fig 4 and S8 Table), and no results were significant at the Bonferroni threshold for cognitive change measures. As with single-metabolite analyses, we further adjusted for life course factors, detailed below. Key results are presented in Fig 5 and full summary statistics are available in S9 Table.

480

A



B

Pathway overrepresentation analysis

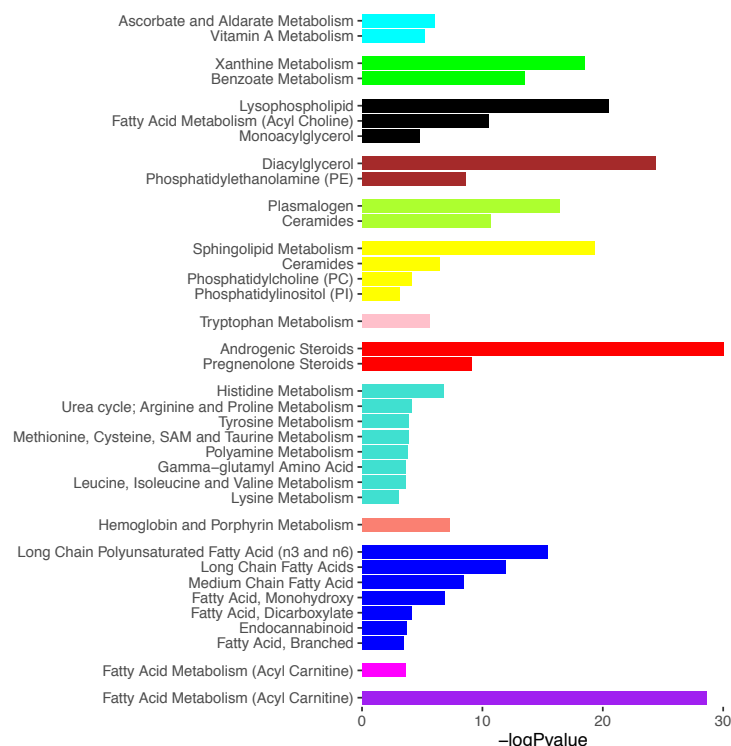


Fig 4. A) Module dendrogram and heatmap of module-outcome associations. Effect sizes and unadjusted p-values are presented in the tiles and significant associations ($p < 1.14 \times 10^{-3}$) are highlighted in bold. For clarity purposes, only outcomes demonstrating a Bonferroni-significant result are shown. B) Overrepresented pathways for each module. Only pathways significant at the adjusted threshold ($p < 9.43 \times 10^{-4}$) are shown. P-values are unadjusted. Underlying data are present in S8 Table and S9 Table.

STM = short-term memory, DM = delayed memory, ACE-III = Addenbrooke's Cognitive Examination-III, PS = processing speed.

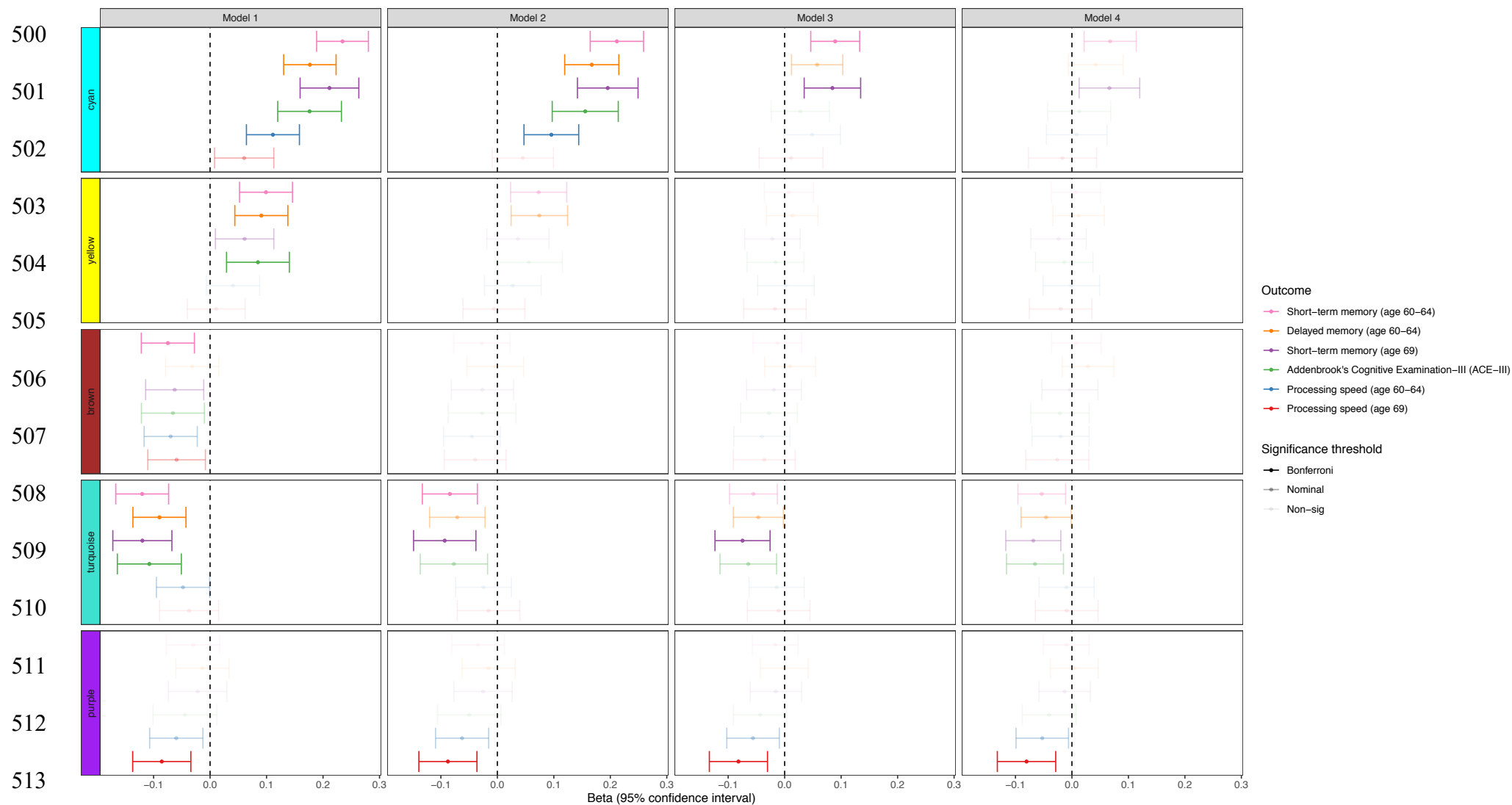


Fig 5. Forest plot showing associations between modules and outcomes in models 1-4. Bonferroni-significant modules ($p < 1.14 \times 10^{-3}$) are represented by a solid fill, nominal modules ($p < 0.05$) by a fainter fill and modules that are not significant at either threshold are represented by the faintest fill. For clarity purposes, only outcomes demonstrating a Bonferroni-significant result are shown. Underlying data are present in S9 Table.

3.5.1 Positive relationships were seen for yellow and cyan modules

The cyan module – enriched in ascorbate and aldarate metabolism, vitamin A metabolism and food/plant consumption – exhibited the strongest positive relationship with all but one of the cognitive outcomes. This effect was significant at the adjusted threshold for most outcomes, with the exception of processing speed at age 69 ($p=0.024$) and change measures (short-term memory – $p=8.03 \times 10^{-3}$, processing speed – $p=0.81$). Adjusting for BMI and lipid medication (model 2) slightly attenuated relationships (5.4-14.1%), with model 3 adjustments again showing the greatest reductions (42.2-72.5%). As a result, short-term memory at age 60-64 and 69 remained significant at the adjusted threshold, but all other associations were lost. Our additional analysis indicated that all model 3 covariables were able to reduce associations, with childhood cognition and education resulting in the biggest attenuations (S4 Table). After adjusting for lifestyle factors (model 4), effect sizes reduced by a further 36.1% for processing speed at age 60-64, with smaller reductions (8.4-9.3%) observed for the other outcomes (Fig 5). Subsequently, short-term memory was no longer significant at the adjusted threshold, but did remain nominally associated (age 60-64 - $\beta=0.068$, 95%CI =0.022 to 0.11, $p=4.10 \times 10^{-3}$; age 69 - $\beta=0.066$, 95%CI=0.013 to 0.12, $p=0.015$). Overall attenuations ranged from 68.6-92.6%.

We similarly found the yellow module – enriched in sphingolipid metabolism and other lipid pathways – to display positive associations with ACE-III, delayed memory and short-term memory at age 60-64 at the adjusted threshold. We saw modest to large effect size reductions (18.1-34.2%) in model 2, with all relationships lost after adjusting for multiple tests. Model 3 adjustments resulted in substantial further effect reductions (66.3-84.8%), again with childhood cognition and education resulting in the biggest attenuations (S4

Table). Minimal additional changes were seen in model 4, and overall attenuations ranged from 87.0-116%.

3.5.1 Negative relationships were seen for turquoise, purple and brown modules

The turquoise module - enriched in several amino acid metabolism pathways - displayed negative associations with ACE-III, short-term memory (both time points) and delayed memory. Module-outcome relationships were most sensitive to model 2 adjustments, decreasing by 20.9-30%. As such, associations were lost at the adjusted threshold for ACE-III and delayed memory, but remained nominally significant. A further attenuation of 11.7-27.1% was observed in model 3, before changing by a small margin (0.95-5.2%) in model 4. No associations passed multiple testing correction in the final model, but all relationships remained at the nominal threshold: standardised betas ranged from -0.068 to -0.046; and p values from 0.0060 to 0.043. Overall attenuations ranged from 39.2-55.5%.

Negative associations were also identified between the purple module - enriched in fatty acid (acyl carnitine) metabolism – and processing speed at age 69. This relationship remained significant and relatively stable across all stages of covariable adjustment, demonstrating an effect size reduction of 6% overall and a final effect size of $\beta = -0.080$ (95%CI=-0.13 to -0.029, $p = 2.33 \times 10^{-3}$). Similar effect directions were detected for both processing speed at age 60-64 and change in processing speed throughout, although this was nominally significant (model 4: age 60-64 - $\beta = -0.053$, 95%CI=-0.10 to -0.0061, $p = 0.027$; change - $\beta = -0.065$, 95%CI=-0.12 to -0.012, $p = 0.016$).

Finally, we identified the brown module - enriched in diacylglycerol and phosphatidylethanolamine (PE) pathways - to be negatively associated with short-term memory at age 60-64. In model 2, associations were largely reduced (63.3%) and no longer significant; this appeared to be mainly driven by BMI (S4 Table). Subsequent model adjustments resulted in full attenuation.

Results were largely unchanged after adjusting for APOE, with all previously significant modules remaining significant at the nominal or Bonferroni threshold in the basic model. In the final model we observed a similar pattern, with the exception of the turquoise module and delayed memory which was no longer significant at the nominal threshold (S6 Table).

3.5.3 Hub metabolites

To highlight possible key drivers, we focused in on metabolites showing a high intramodular connectivity ($kME > 0.65$) and integrated this with our single-metabolite results. Thirty-five of the 155 metabolites identified in single-metabolite analyses were revealed to be hubs (Fig 6 and S1 Table). Hub metabolites belonged to 10 modules and represented various pathways, including eight of the 10 identified in our pathway analyses at the adjusted threshold: gamma-glutamyl amino acid; methionine, cysteine, SAM and taurine metabolism; purine metabolism, (hypo)xanthine/inosine containing; purine metabolism, adenine containing; pyrimidine metabolism, uracil containing; ascorbate and aldarate metabolism; vitamin A metabolism; and fatty acid (monohydroxy). In model 4, one hub – C16 – was significant at the adjusted threshold and 17 were nominally significant.

It is made available under a [CC-BY-NC-ND 4.0 International license](#).

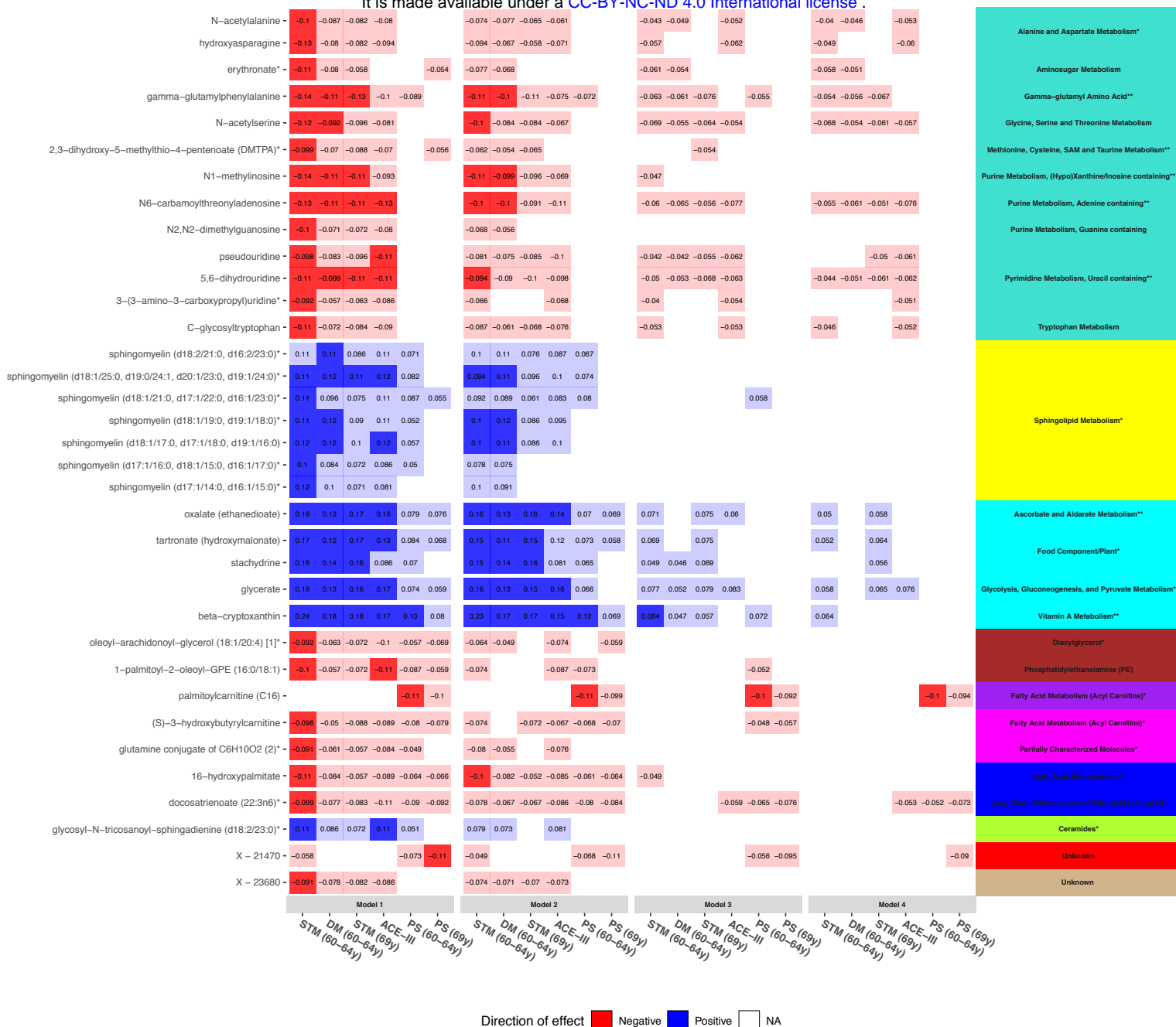


Fig 6. Heatmap showing trends of associations between hub metabolites and cognitive outcomes in models 1-4. Panels on the right indicate the pathways (text) and modules (colour) represented by metabolites. Pathways are suffixed with an asterisk if they were previously identified in our pathway analyses (= $p < 9.43 \times 10^{-4}$, * = $p < 0.05$). Bonferroni-significant metabolites ($p < 1.15 \times 10^{-4}$) are represented by a solid fill, nominal metabolites by a faint fill ($p < 0.05$), and non-significant metabolites by no fill ($p > 0.05$). Tiles are coloured by effect direction and effect sizes are noted in the centre. For clarity**

612 **purposes, only outcomes demonstrating a Bonferroni-significant result are shown.**

613 **Underlying data are present in S1 Table.**

614 *STM = short-term memory, PS = processing speed, DM = delayed memory, ACE-III = Addenbrooke's*

615 *Cognitive Examination-III.*

616

4. Discussion

Using the British 1946 Birth Cohort, we systematically evaluated the metabolic correlates of cognitive function in late midlife while untangling the influence of life course factors. We identified 155 metabolites, ten metabolic pathways and five modules of coexpressed metabolites to show associations with cognitive outcomes. A large proportion of our results converged using multiple analytical approaches, adding weight to their involvement in cognitive outcomes in late midlife. Integrating these, 35 hub metabolites were revealed to show potential as candidates for further study. Some relationships were independent of life course influences, providing insight into unique metabolic mechanisms underlying cognitive outcomes. However, consistent with our previous analyses in the MRC 1946 (6), as well as a previous lipidomics study in the Lothian Birth Cohort (29), many were sensitive to social factors and childhood cognition, suggesting important considerations for future studies.

Our most robust finding was for increased serum acylcarnitines and worse processing speed. The purple module - enriched in medium and long chain acylcarnitines - showed negative relationships that were independent of life course factors. Interestingly, associations were strongest for processing speed 5-9 years later (age 69), although similar patterns were observed for both baseline (age 60-64) and change at the nominal threshold. These relationships were specific to processing speed, indicating a possible mechanism unique to this outcome; our pathway analysis was largely in support of this. One metabolite, C16 (a long chain acylcarnitine) appeared to be a key driver in these associations, suggesting a potential candidate and pathway for further investigation.

Biologically, medium and long chain acylcarnitines are derivatives of fatty acid metabolism and known to be pivotal in mitochondrial fatty acid oxidation (30). Increased abundances in serum have thus been regarded as proxies for mitochondrial dysfunction and impairments in subsequent energy production (30). More specifically, C16 has also been linked to the induction and regulation of apoptotic events (31). Both apoptosis and mitochondrial dysfunction have been implicated in neurodegeneration, indicating a plausible biological mechanism behind our observations (32,33).

Perturbations in acylcarnitine levels have been reported in early cognitive impairments and AD (7,34,35). Serum levels have also been linked to other traits, such as insulin resistance(36,37), obesity (37) and cardiovascular disease (38), similarly demonstrating independent effects from known risk factors. Although these factors show influence on cognitive function, our associations remained following adjustment for several related factors such as blood pressure, BMI and lipid medication. Taken together, this presents the possibility that alterations may reflect metabolic dysfunction, conferring vulnerability to adverse health outcomes including decline in processing speed. Future studies will seek to further establish whether these changes lie on the causal pathway.

We also discovered a module comprised of nucleotides and amino acids, turquoise, to demonstrate negative associations with short-term memory, delayed memory and ACE-III. As the largest module, it was enriched in several pathways, including three which were brought to our attention in pathway analyses (histidine metabolism; methionine, cysteine, SAM and taurine metabolism; gamma-glutamyl amino acid metabolism). Purine and pyrimidine metabolism pathways were additionally represented by module hubs and

further corroborated by pathway analyses. Relationships remained nominally significant after adjusting for life course factors, reducing by 39.2-55.5% overall.

The top driver of the module was 2,3-dihydroxy-5-methylthio-4-pentenoic acid (DMTPA), belonging to the methionine, cysteine, SAM and taurine metabolism pathway. Additional module hubs were amino acids and nucleosides that were unified by the presence of modifications. Included in these were several markers of RNA turnover, a gamma-glutamyl amino acid and a marker of oxidative stress (39–41). The methionine, cysteine, SAM and taurine metabolism pathway has a key role in post-transcriptional and post-translational modifications, including the sourcing of methyl and aminocarboxypropyl groups demonstrated in some module hubs (42,43), indicating plausible relationships linking these metabolites.

With key roles in the expression, function and stability of molecules, RNA and amino acid modifications are able to regulate a multitude of biological and pathological processes, although precise mechanisms remain elusive (44,45). Aberrant patterns are thought to underly a host of chronic diseases (44), and are gaining attention in the field of neurodegeneration, with a recent paper reporting differential expression of RNA modifications in the post mortem brains of dementia patients (46). Interestingly, the many module hubs have been consistently reported together in adverse outcomes, including hypertension (47), chronic kidney disease (48,49), inflammation (50) and mortality (51). This signature has been previously shown to be relevant to cognitive function as well; in the Bogalusa heart study, a module was curated with similar top drivers, showing associations with cognitive outcomes in midlife (10). For the first time, we have reported such an

association in the context of late midlife cognitive outcomes, finding observational relationships with ACE-III, as well as with memory more generally.

The widespread associations of these module hubs together in adverse outcomes presents the possibility that they could reflect global disease burden or converging pathological mechanisms. A previous study hypothesised that increased levels of modified amino acids may represent an accelerated ageing phenotype and suggested modified nucleotides to be a marker of tissue breakdown and oxidative stress (52), which we know to be implicated in decline in cognitive function (53). Lending some support for this, life course factors related to these processes, such as BMI and smoking, were associated with the module. However, none were able to suitably explain relationships, suggesting that our findings could be capturing converging mechanisms that are independent of these factors.

We reported relationships between sphingolipids and improved cognitive function which were entirely explained by life course factors, particularly childhood cognition and education. Positive associations were observed between the yellow module - enriched in several pathways related to sphingolipid metabolism - and the ACE-III, as well as short-term and delayed memory at age 60-64. Nominal associations were also recorded for sphingolipid pathways and these outcomes. Module associations exhibited major reductions in model 3, and attenuated by 87-116% overall, with no hubs remaining associated with cognitive outcomes; this was surprising given the attention sphingolipid metabolism has received in the neurodegeneration field. We saw similar patterns for the brown module - enriched in diacylglycerol and PE – for which relationships attenuated in model 2, although effects were relatively small and only reported for short-term memory at age 60-64.

Previous research has implicated disturbances in sphingolipid balance in cognitive development (54,55), function (55), ageing (55,56) and AD (57). Sphingolipids are a lipid family comprised of sphingomyelins, ceramides and glycosphingolipids, and are present in large quantities in the CNS (57). Forming important components of cell membranes, they are highly dynamic and display crucial roles in cognitive development and function (55). In this regard, all yellow module hubs were sphingomyelin species - the most abundant sphingolipid with pivotal roles in myelin integrity and function (54). However, despite findings highlighting altered sphingolipid balance in cognitive decline and AD, replication of specific sphingolipids has proved challenging. This could be due to their delicate balance and dynamicity, or perhaps due to fluctuations as a function of pathology, with differing relationships observed throughout the disease course. Alternatively, inconsistencies in incorporating life course factors in analytical models could be a contributor.

Given observational findings linking sphingolipids and cognitive function at several stages of the life course, there exists several possibilities of the loss of relationships after model 3 adjustments, and particularly for childhood cognition and education. Attenuations following adjustment for childhood cognition suggests confounding through reverse causation. In the case of education, this may influence levels through other exposures, resulting in spurious associations between late midlife sphingolipid levels and late midlife cognitive outcomes. However, due to a lack of longitudinal metabolite data, we cannot rule out the possibility that later life sphingolipid levels may be a proxy for those of earlier life. In this regard, childhood cognition and education could be mediators of early life sphingolipid levels and later life cognitive function; or perhaps they may be linked by common genetic or environmental causes. Future studies aiming to unravel the precise nature of these life

course relationships will be crucial in understanding the importance of sphingolipids in cognitive outcomes in later life. Nevertheless, our findings suggest that these factors need to be carefully considered in future investigations into metabolic underpinnings of cognitive function and decline.

Finally, a module enriched in vitamin A and C metabolites, cyan, showed positive associations with most cognitive outcomes. Notably, our model 1 results suggested a ubiquitous positive role of vitamin A metabolites for cognitive outcomes, with the cyan module, vitamin A metabolism pathway and metabolites (carotenoids) contained within this pathway displaying the largest overall effects across all stages of analysis. Despite large effect sizes, associations were sensitive to adjustment for life course factors, namely childhood cognition and social factors. In the final model, relationships were largely explained for processing speed and ACE-III, displaying attenuations of >90%. Reductions were also seen for short-term memory at both time points (68.6 and 71.1%), but associations remained at the nominal threshold.

Oxidative stress is thought to be involved in the pathogenesis of neurodegenerative diseases and is characterised by an overabundance of reactive oxygen species, initiating a host of deleterious effects (58). Antioxidants, including vitamin C metabolites and carotenoids, may inhibit such processes through scavenging of these species (59). Due to this, their involvement in ageing, cognitive decline and AD has been discussed, and all five of the present hubs have previously shown apparent protective effects in small studies investigating cognitive impairments (60–62). Nevertheless, contributions are debated, with epidemiological studies showing conflicting results (63,64).

While generally acquired through food consumption, it is unclear whether our findings reflect increased intake or bioavailability, or accumulate consequentially as a marker of reduced antioxidant requirements. In a previous study exploring metabolite markers of healthy nutritional patterns, three of the module hubs (glycerate, stachydrine and oxalate) were highly correlated with the intake of foods rich in antioxidants and minerals (65). Although we identified strong relationships between the module and dietary scores, diet was not able to suitably explain observed relationships. As the cyan module exhibited significant associations with all covariables, it is plausible that it could be capturing a more pervasive mechanism such as oxidative stress and subsequent antioxidant requirements.

It was interesting to note that, following life course adjustment, relationships were largely explained for processing speed and ACE-III, but nominal residual relationships remained for short-term memory at both time points, with age 60-64 just shy of the adjusted threshold. Although these relationships still showed marked attenuations, this indicates that there may be some independent effects specific to short-term memory at these ages. As discussed with previously, attenuations suggest confounding by childhood cognition and social factors, or may reflect relationships we are not able to capture without longitudinal metabolite data. Further research is required to understand these relationships in greater depth.

Findings should be considered in light of several strengths and limitations. First, our study incorporated a variety of known confounders and our results converged through a variety of methodological techniques. Using a sequential analytical design further allowed for greater insight into influencing life course factors. However, our results may still be subject to residual confounding. We additionally do not have longitudinal metabolomics data,

precluding investigation into lifelong relationships and directionality. Our next step will be to interrogate the causal nature of associations highlighted in this study through methods such as Mendelian Randomisation, as well as to externally replicate our work. Next, cognitive change measures were curated from data collected at two time points within a small time window, which could explain the lack of relationships observed; it will be interesting to investigate this in future waves. Further, change measures were represented by residualised change scores which can be subject to bias and our findings should be interpreted with this in caution. The NSHD is a large nationally recruited sample with a narrow age range, thus overcoming confounding by age or recruitment strategy. With a large array of available measures, we were able to investigate prospective measures collected from birth as well as comprehensive measures of cognitive function, including the ACE-III. Looking at several domains of cognition further allowed us to identify more specific relationships that may otherwise have been missed. As seen with many cohort studies, individuals remaining in the study at this stage were generally of higher cognitive ability in childhood and more socially advantaged compared to the sample initially recruited at birth. Additionally, the study sample was ethnically homogenous. For generalisation, it is paramount to replicate this work in more diverse populations. Finally, we used a comprehensive metabolomics platform to acquire untargeted metabolite data, which provided unbiased and granular metabolite profiles for our analyses. Due to this, levels of metabolites with unknown identities were captured; module analysis yielded some insight into a broad identity and function, but precise characterisation is not yet available.

In summary, we conducted one of the largest LC-MS studies to date on cognitive outcomes in late midlife, and are the first to systematically evaluate associations in the context of life

808 course factors. We integrated metabolites, pathways and networks, offering biological
 809 interpretation while retaining granularity, illuminating important pathways and drivers of
 810 cognitive outcomes in late midlife. Our results illustrate the importance of incorporating life
 811 course influences, with many relationships largely explained by childhood cognition and
 812 education; our further studies will aim to unpick this. Finally, we identified several
 813 metabolites (e.g. C16) that were both pivotal to module function and associated with our
 814 outcomes, presenting as potential marker candidates for additional study.
 815

Acknowledgements

PP is funded by Alzheimer's Research UK and RG by the National Institute for Health Research (NIHR) Biomedical Research Centre.

This paper represents independent research part-funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. This work was further supported by the UK Medical Research Council which provides core funding for the MRC National Survey of Health and Development.

We thank NSHD study members for their lifelong participation and past and present members of the NSHD study team who helped to collect the data. We also thank ACE Mobile for providing a customised version of the ACE-III for NSHD.

Competing interests

The authors declare no competing interests.

Correspondence

Correspondence to Petroula Proitsi & Marcus Richards.

References

1. Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet*. 2020;396(10248):413–46.
2. Cummings J. Lessons Learned from Alzheimer Disease: Clinical Trials with Negative Outcomes. *Clin Transl Sci*. 2018;11(2):147–52.
3. Wishart DS. Emerging applications of metabolomics in drug discovery and precision medicine. *Nat Rev Drug Discov*. 2016;15(7):473.
4. Beebe K, Kennedy AD. Sharpening Precision Medicine by a Thorough Interrogation of Metabolic Individuality. *Comput Struct Biotechnol J*. 2016;14:97–105.
5. Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates A β , tau, immunity and lipid processing. *Nat Genet*. 2019;51(3):414–30.
6. Proitsi P, Kuh D, Wong A, Maddock J, Bendayan R, Wulaningsih W, et al. Lifetime cognition and late midlife blood metabolites: findings from a British birth cohort. *Transl Psychiatry*. 2018;8(1):203.
7. Toledo JB, Arnold M, Kastenmüller 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.
8. Jiang Y, Zhu Z, Shi J, An Y, Zhang K, Wang Y, et al. Metabolomics in the Development and Progression of Dementia: A Systematic Review. *Front Neurosci*. 2019;13:343.
9. 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. *Alzheimer's Dement*. 2018;14(6):707–22.

10. Shi M, Bazzano LA, He J, Gu X, Li C, Li S, et al. Novel serum metabolites associate with cognition phenotypes among Bogalusa Heart Study participants. *Aging (Albany NY)*. 2019;11(14):5124–39.
11. Xu J, Bankov G, Kim M, Wretling A, Lord J, Green R, et al. Integrated lipidomics and proteomics network analysis highlights lipid and immunity pathways associated with Alzheimer’s disease. *Transl Neurodegener*. 2020;9(1):36.
12. Hsieh S, Schubert S, Hoon C, Mioshi E, Hodges JR. Validation of the Addenbrooke’s Cognitive Examination III in Frontotemporal Dementia and Alzheimer’s Disease. *Dement Geriatr Cogn Disord*. 2013;36(3–4):242–50.
13. Stafford M, Black S, Shah I, Hardy R, Pierce M, Richards M, et al. Using a birth cohort to study ageing: representativeness and response rates in the National Survey of Health and Development. *Eur J Ageing*. 2013;10(2):145–57.
14. Kuh D, Wong A, Shah I, Moore A, Popham M, Curran P, et al. The MRC National Survey of Health and Development reaches age 70: maintaining participation at older ages in a birth cohort study. *Eur J Epidemiol*. 2016;31(11):1135–47.
15. Do KT, Wahl S, Raffler J, Molnos S, Laimighofer M, Adamski J, et al. Characterization of missing values in untargeted MS-based metabolomics data and evaluation of missing data handling strategies. *Metabolomic*. 2018;14(10):128.
16. Hastie T, Tibshirani R, Narasimhan B, Gilbert C. impute: Imputation for microarray data [software]. 2019.
17. James S-N, Davis D, O’Hare C, Sharma N, John A, Gaysina D, et al. Lifetime affective problems and later-life cognitive state: Over 50 years of follow-up in a British birth cohort study. *J Affect Disord*. 2018;241:348–55.
18. Heim AW. Manual for the AH4 Group Test of General Intelligence. Windsor: NFER.

- 1970.
19. Pigeon DA, Douglas JWB, Ross JM, Simpson HR. Details of the fifteen years tests. All our Future. London: Davies. 1968;11.
20. Maddock J, Ziauddeen N, Ambrosini GL, Wong A, Hardy R, Ray S. Adherence to a Dietary Approaches to Stop Hypertension (DASH)-type diet over the life course and associated vascular function: a study based on the MRC 1946 British birth cohort. Br J Nutr. 2018;119(5):581–9.
21. Appel LJ, Moore TJ, Obarzanek E, Vollmer WM, Svetkey LP, Sacks FM, et al. A clinical trial of the effects of dietary patterns on blood pressure. DASH Collaborative Research Group. N Engl J Med. 1997;336(16):1117–24.
22. Richards M, James S-N, Sizer A, Sharma N, Rawle M, Davis DHJ, et al. Identifying the lifetime cognitive and socioeconomic antecedents of cognitive state: seven decades of follow-up in a British birth cohort study. BMJ Open. 2019;9(4):e024404.
23. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. J Stat Software; 2011;45(3):1-67.
24. Krumsiek J, Mittelstrass K, Do KT, Stücker F, Ried J, Adamski J, et al. Gender-specific pathway differences in the human serum metabolome. Metabolomics. 2015;11(6):1815–33.
25. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol. 2005;4(1).
26. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9(1):559.
27. Langfelder P, Horvath S. Fast R functions for robust correlations and hierarchical clustering. J Stat Softw. 2012;46(11).

- 909 28. Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree:
910 the Dynamic Tree Cut package for R. *Bioinformatics*. 2008;24(5):719–20.
- 911 29. Corley J, Starr JM, Deary IJ. Serum cholesterol and cognitive functions: the Lothian
912 Birth Cohort 1936. *Int Psychogeriatrics*. 2015;27(3):439–53.
- 913 30. McCoin CS, Knotts TA, Adams SH. Acylcarnitines—old actors auditioning for new roles
914 in metabolic physiology. *Nat Rev Endocrinol*. 2015;11(10):617.
- 915 31. Mutomba M, Yuan H, Konyavko M, Adachi S, Yokoyama C, McGarry JD, et al.
916 Regulation of the activity of caspases by L-carnitine and palmitoylcarnitine. *FEBS Lett*.
917 2000;478:19–25.
- 918 32. Jones LL, McDonald DA, Borum PR. Acylcarnitines: Role in brain. *Prog Lipid Res*.
919 2010;49(1):61–75.
- 920 33. Shimohama S. Apoptosis in Alzheimer’s disease—an update. *Apoptosis*. 2000;5(1):9–
921 16.
- 922 34. van der Velpen V, Teav T, Gallart-Ayala H, Mehl F, Konz I, Clark C, et al. Systemic and
923 central nervous system metabolic alterations in Alzheimer’s disease. *Alzheimers Res*
924 *Ther*. 2019;11(1):93.
- 925 35. Ciavardelli D, Piras F, Consalvo A, Rossi C, Zucchelli M, Di Ilio C, et al. Medium-chain
926 plasma acylcarnitines, ketone levels, cognition, and gray matter volumes in healthy
927 elderly, mildly cognitively impaired, or Alzheimer’s disease subjects. *Neurobiol Aging*.
928 2016;43:1–12.
- 929 36. Guasch-Ferré M, Ruiz-Canela M, Li J, Zheng Y, Bulló M, Wang DD, et al. Plasma
930 Acylcarnitines and Risk of Type 2 Diabetes in a Mediterranean Population at High
931 Cardiovascular Risk. *J Clin Endocrinol Metab*. 2019;104(5):1508–19.
- 932 37. Mihalik SJ, Goodpaster BH, Kelley DE, Chace DH, Vockley J, Toledo FGS, et al.

- 933 Increased levels of plasma acylcarnitines in obesity and type 2 diabetes and
934 identification of a marker of glucolipotoxicity. Obesity (Silver Spring).
935 2010;18(9):1695–700.
- 936 38. Strand E, Pedersen ER, Svingen GFT, Olsen T, Bjørndal B, Karlsson T, et al. Serum
937 Acylcarnitines and Risk of Cardiovascular Death and Acute Myocardial Infarction in
938 Patients With Stable Angina Pectoris. J Am Heart Assoc. 2017 Feb;6(2): e003620.
- 939 39. Topp H, Dörrenhaus A, Bolt H, Schöch G, Föllmann W. Determination of Degradation
940 Rates of Transfer and Ribosomal Ribonucleic Acids in Cultured Rat Hepatocytes by
941 Measuring N⁶-Threoninocarbonyladenine, Dihydrouridine, and Pseudouridine in
942 Medium Using High-Performance Liquid Chromatography. Anal Biochem.
943 1997;254(2):200–7.
- 944 40. Nakano K, Nakao T, Schram KH, Hammargren WM, McClure TD, Katz M, et al. Urinary
945 excretion of modified nucleosides as biological marker of RNA turnover in patients
946 with cancer and AIDS. Clin Chim Acta. 1993;218(2):169–83.
- 947 41. Zierer J, Kastenmüller G, Suhre K, Gieger C, Codd V, Tsai P-C, et al. Metabolomics
948 profiling reveals novel markers for leukocyte telomere length. Aging (Albany NY).
949 2016;8(1):77–94.
- 950 42. Lieu EL, Nguyen T, Rhyne S, Kim J. Amino acids in cancer. Exp Mol Med.
951 2020;52(1):15–30.
- 952 43. Fontecave M, Atta M, Mulliez E. S-adenosylmethionine: nothing goes to waste.
953 Trends Biochem Sci. 2004;29(5):243–9.
- 954 44. Gillery P, Jaisson S. Post-translational modification derived products (PTMDPs): toxins
955 in chronic diseases? Clin Chem Lab Med. 2014;52(1):33–8.
- 956 45. Boo SH, Kim YK. The emerging role of RNA modifications in the regulation of mRNA

957 stability. *Exp Mol Med*. 2020;52(3):400–8.

958 46. Zhang X, Trebak F, Souza LAC, Shi J, Zhou T, Kehoe PG, et al. Small RNA modifications
959 in Alzheimer’s disease. *Neurobiol Dis*. 2020;145:105058.

960 47. Rhodes CJ, Ghataorhe P, Wharton J, Rue-Albrecht KC, Hadinnapola C, Watson G, et al.
961 Plasma Metabolomics Implicates Modified Transfer RNAs and Altered Bioenergetics
962 in the Outcomes of Pulmonary Arterial Hypertension. *Circulation*. 2017;135(5):460–
963 75.

964 48. Sekula P, Del Greco M F, Pattaro C, Köttgen A. Mendelian Randomization as an
965 Approach to Assess Causality Using Observational Data. *J Am Soc Nephrol*.
966 2016;27(11):3253–3265.

967 49. Niewczas MA, Mathew A V, Croall S, Byun J, Major M, Sabiseti VS, et al. Circulating
968 Modified Metabolites and a Risk of ESRD in Patients With Type 1 Diabetes and
969 Chronic Kidney Disease. *Diabetes Care*. 2017;40(3):383–90.

970 50. Lustgarten MS, Fielding RA. Metabolites Associated With Circulating Interleukin-6 in
971 Older Adults. *J Gerontol A Biol Sci Med Sci*. 2017;72(9):1277–83.

972 51. Huang J, Weinstein SJ, Moore SC, Derkach A, Hua X, Liao LM, et al. Serum
973 Metabolomic Profiling of All-Cause Mortality: A Prospective Analysis in the Alpha-
974 Tocopherol, Beta-Carotene Cancer Prevention (ATBC) Study Cohort. *Am J Epidemiol*.
975 2018;187(8):1721–32.

976 52. Nierenberg JL, He J, Li C, Gu X, Shi M, Razavi AC, et al. Serum metabolites associate
977 with physical performance among middle-aged adults: Evidence from the Bogalusa
978 Heart Study. *Aging (Albany NY)*. 2020;12(12):11914–41.

979 53. Kim GH, Kim JE, Rhie SJ, Yoon S. The Role of Oxidative Stress in Neurodegenerative
980 Diseases. *Exp Neurobiol*. 2015;24(4):325–40.

981 54. Schneider N, Hauser J, Oliveira M, Cazaubon E, Mottaz SC, O'Neill B V, et al.
982 Sphingomyelin in Brain and Cognitive Development: Preliminary Data. eNeuro.
983 2019;6(4):ENEURO.0421-18.2019.

984 55. Olsen ASB, Færgeman NJ. Sphingolipids: membrane microdomains in brain
985 development, function and neurological diseases. Open Biol. 2017;7(5):170069.

986 56. Montoliu I, Scherer M, Beguelin F, DaSilva L, Mari D, Salvioli S, et al. Serum profiling of
987 healthy aging identifies phospho- and sphingolipid species as markers of human
988 longevity. Aging (Albany NY). 2014;6(1):9–25.

989 57. van Echten-Deckert G, Walter J. Sphingolipids: Critical players in Alzheimer's disease.
990 Prog Lipid Res. 2012;51(4):378–93.

991 58. Engelhart MJ, Geerlings MI, Ruitenberg A, van Swieten JC, Hofman A, Witteman JCM,
992 et al. Dietary intake of antioxidants and risk of Alzheimer disease. Jama.
993 2002;287(24):3223–9.

994 59. Pitchumoni SS, Doraiswamy PM. Current Status of Antioxidant Therapy for
995 Alzheimer's Disease. J Am Geriatr Soc. 1998;46(12):1566–72.

996 60. Rinaldi P, Polidori MC, Metastasio A, Mariani E, Mattioli P, Cherubini A, et al. Plasma
997 antioxidants are similarly depleted in mild cognitive impairment and in Alzheimer's
998 disease. Neurobiol Aging. 2003;24(7):915–9.

999 61. Burté F, Houghton D, Lowes H, Pyle A, Nesbitt S, Yarnall A, et al. metabolic profiling of
1000 Parkinson's disease and mild cognitive impairment. Mov Disord. 2017;32(6):927–32.

1001 62. Li L, Sun L, Qiu Y, Zhu W, Hu K, Mao J. Protective Effect of Stachydrine Against
1002 Cerebral Ischemia-Reperfusion Injury by Reducing Inflammation and Apoptosis
1003 Through P65 and JAK2/STAT3 Signaling Pathway. Frontiers in Pharmacology. 2020;11.
1004 64.

1005 63. Crichton GE, Bryan J, Murphy KJ. Dietary antioxidants, cognitive function and
1006 dementia--a systematic review. *Plant Foods Hum Nutr.* 2013;68(3):279–92.

1007 64. Kamat CD, Gadgil S, Mhatre M, Williamson KS, Pye QN, Hensley K. Antioxidants in
1008 central nervous system diseases: preclinical promise and translational challenges. *J*
1009 *Alzheimers Dis.* 2008;15(3):473–93.

1010 65. Playdon MC, Moore SC, Derkach A, Reedy J, Subar AF, Sampson JN, et al. Identifying
1011 biomarkers of dietary patterns by using metabolomics. *Am J Clin Nutr.* 2016
1012 ;105(2):450–65.

1013 66. Krassowski M. ComplexUpset: Create Complex UpSet Plots Using ggplot2
1014 Components. R package Version 0.5.18 [software]. 2020.

1015

1016

1017 **Supporting Information Captions**

1018

1019 **S1 Table. Full single-metabolite analyses results (models 1-4), including module**

1020 **membership and hub status.**

1021

1022 **S2 Table. Linear regression analyses results for the association of each covariable against**

1023 **each predictor. All analyses are adjusted for sex, age at blood collection and blood**

1024 **collection centre.**

1025

1026 **S3 Table. Pathway analyses results.**

1027

1028 **S4 Table. Module overrepresentation analysis results.**

1029

1030 **S5 Table. Full module analyses results (models 1-4).**

1031

1032 **S6 Table. Linear regression analyses results for the associations between metabolites and**

1033 **outcomes, after adjusting for each covariable individually. All analyses are adjusted for**

1034 **sex, age at blood collection and blood collection centre.**

1035

1036 **S7 Table. Linear regression analyses results for the associations between metabolites and**

1037 **outcomes with and without adjusting for APOE. The maximum N with APOE genotype was**

1038 **used.**

1039

S8 Table. Linear regression analyses results for the associations between modules and outcomes, after adjusting for each covariable individually.

S9 Table. Linear regression analyses results for the associations between modules and outcomes with and without adjusting for APOE. The maximum N with APOE genotype was used.

S1 Fig. Flow chart depicting metabolomic data quality control.

S2 Fig. Heat map showing trends of associations between the 155 metabolites and cognitive outcomes in models 1-4, organised by metabolite family. Bonferroni-significant metabolites ($p < 1.15 \times 10^{-4}$) are represented by a solid fill, nominal metabolites by a faint fill ($p < 0.05$), and non-significant metabolites by no fill ($p > 0.05$).

Carb = carbohydrates, Cof & Vit = cofactors & vitamins, En = energy, PCM = partially characterised molecules, Pep = peptides, ACE-III = Addenbrooke's Cognitive Examination-III, DM = delayed memory, STM = short-term memory, PS = processing speed

S3 Fig. Scale free topology and mean connectivity plots for weighted gene coexpression network analysis.

S1 Methods. Supplementary methods