Investigating associations between blood metabolites, later life brain imaging measures, and genetic risk for Alzheimer's disease

Rebecca Green[ab], Jodie Lord[a], Marzia A. Scelsi[c], Jin Xu[ad], Andrew Wong[e], Sarah Naomi-James[ef], Alex Handy[g], Lachlan Gilchrist[a] , Dylan M. Williams[eh], Thomas D. Parker[fij], Christopher A. Lane[f], Ian B. Malone[f], Carole H. Sudre[cefk], William Coath[f], David L. Thomas[fl], Sarah Keuss[f], Richard Dobson[abgmn], Cristina Legido-Quigley[do], Nick C. Fox[fp], Jonathan M. Schott[f*], Marcus Richards[e*], Petroula Proitsi[a*], The Insight 46 study team

*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 Trust, London, UK.

[c]Department of Medical Physics and Biomedical Engineering, Centre for Medical Image Computing (CMIC), University College London (UCL), London, UK.

[d]Institute of Pharmaceutical Science, King's College London, London, UK.

[e]MRC Unit for Lifelong Health & Ageing at UCL, University College London, London, UK.

[f]Dementia Research Centre, UCL Queen Square Institute of Neurology, London, UK.

[g]University College London, Institute of Health Informatics, London, UK.

[h]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.

[i]Department of Brain Sciences, Imperial College London, London W12 0NN, UK.

[j]UK DRI Centre for Care Research and Technology, Imperial College London, London W12 0BZ, UK.

[k]School of Biomedical Engineering and Imaging Sciences, King's College London, London, UK.

[l]Neuroradiological Academic Unit, Department of Brain Repair and Rehabilitation, UCL Queen Square Institute of Neurology, University College London, London, UK.

[m]Health Data Research UK London, University College London, London, UK.

[n]NIHR Biomedical Research Centre at University College London Hospitals NHS Foundation Trust, London, UK.

[o]Steno Diabetes Center Copenhagen, Gentofte, Denmark.

[p]UK Dementia Research Institute at University College London, London, UK.

Co-corresponding authors:

1. Dr Petroula Proitsi (petroula.proitsi@kcl.ac.uk)

Maurice Wohl Clinical Neuroscience Institute, 5 Cutcombe Road, London, SE5 9RT, UK.

2. Professor Marcus Richards (m.richards@ucl.ac.uk)

Floor 5, MRC LHA at UCL, 1 – 19, Torrington Place, London, WC1E 7HB, UK.

3. Professor Jonathan M. Schott (j.schott@ucl.ac.uk)

Dementia Research Centre, UCL Queen Square Institute of Neurology, 8-11 Queen Square, London, WC1N 3BG, UK.

ABSTRACT

Identifying blood-based signatures of brain health and preclinical pathology may offer insights into early disease mechanisms and highlight avenues for intervention. Here, we systematically profiled associations between blood metabolites and whole-brain volume, hippocampal volume, and amyloid-β status among participants of Insight 46 – the neuroscience sub-study of the National Survey of Health and Development (NSHD). We additionally explored whether key metabolites were associated with polygenic risk for Alzheimer's disease (AD).

Following quality control, concentrations of 1019 metabolites – detected with liquid chromatography-mass spectrometry – were available for 1740 participants at age 60-64. Metabolite data were subsequently clustered into modules of co-expressed metabolites using weighted coexpression network analysis. Accompanying MRI and amyloid-PET imaging data were present for 437 participants (age 69-71). Regression analyses tested relationships between metabolite measures – modules and hub metabolites – and imaging outcomes. Hub metabolites were defined as metabolites that were highly connected within significant ($p_{FDR}<0.05$) modules or previously identified as a hub for cognition in the same cohort. Regression models included adjustments for age, sex, *APOE* genotype, lipid medication use, childhood cognition and social factors. Finally, AD polygenic risk scores (PRS), including and excluding the *APOE* region, were tested for relationships with metabolites and modules that associated ($p_{FDR}<0.05$) with an imaging outcome (N=1638).

In the fully adjusted model, three lipid modules were associated with a brain volume measure ($p_{FDR}<0.05$): one enriched in sphingolipids (hippocampal volume: ß=0.14, 95%CI=[0.055,0.23]), one in several fatty acid pathways (whole-brain volume: ß=-0.072, 95%CI=[-0.12,-0.026]), and another in diacylglycerols and phosphatidylethanolamines (whole-brain volume: ß=-0.066, 95%CI=[-0.11,-0.020]). Twenty-two hub metabolites were associated ($p_{FDR}<0.05$) with an

imaging outcome (whole-brain volume: 22; hippocampal volume: 4). Some nominal associations were reported for amyloid-β, and with an AD PRS in our genetic analysis, but none survived multiple testing correction.

Our findings highlight key metabolites, with functions in membrane integrity and cell signalling, that associated with structural brain measures in later life. Future research should focus on replicating this work and interrogating causality.

ABBREVIATIONS

Aβ = amyloid-beta

AD = Alzheimer's disease

BMI = body mass index

DAG = diacylglycerol

FDR = false discovery rate

FLAIR = weighted fluid-attenuated inversion recovery

KEGG = Kyoto Encyclopaedia of Genes and Genomes

kME = module membership

LCMS = liquid chromatography-mass spectrometry

MR = Mendelian randomisation

n3 and n6 = Omega 3 and 6

NSHD = National Survey of Health and Development

PC = phosphatidylcholine

PE = phosphatidylethanolamine

PRS = polygenic risk score

PUFA = polyunsaturated fatty acid

QC = quality control

SEP = socioeconomic position

SUVR = standardised uptake volume ratio

WGCNA = weighted gene coexpression network analysis

1. INTRODUCTION

Brain changes accompanying ageing are varied and can include pathologies that lead to cognitive impairment, the commonest of which is Alzheimer's disease (AD). Identifying non-invasive and scalable markers of brain health and pathology in later life, including but not limited to those associated with AD, would be valuable for research and therapeutic trials. This has led to large efforts in detecting blood-based markers, with candidates such as neurofilament light and phosphorylated-tau showing particular promise (1). Blood metabolites – the products of chemical reactions occurring in the body – may also present as potential candidates for this goal. Due to their proximity to core biological processes, they are uniquely placed to capture real-time physiological changes and may allow insights into the processes associated with emergence of disease (2). Additionally, since they are potentially modifiable (3,4), they could represent possible targets for intervention.

Existing research has identified associations between several metabolite classes and imaging markers related to neurodegeneration, including particular lipids and amino acids (5–9). However, these studies have been directed towards clinical cohorts and thus may not be representative of the general preclinical population. Additionally, little is known about the involvement of groups of interrelated metabolites; utilising systems-level approaches could offer an improved understanding of these complex relationships and facilitate the identification of candidate markers. We previously employed a systems-level approach to explore the metabolic correlates of late midlife cognition in the Medical Research Council National Survey of Health and Development (NSHD; the British 1946 birth cohort) (10). We identified groups of highly coexpressed metabolites that associated with cognitive outcomes and key metabolites within these to explore further, including acylcarnitines, modified nucleotides and amino acids, vitamins, and sphingolipids; although many associations with late midlife cognition were explained by social factors and childhood cognition (10).

Here, we aimed to investigate the metabolic correlates of later life brain imaging measures relevant to AD and neurodegeneration – Aβ pathology, whole-brain volume and hippocampal volume. To provide a deeper understanding on the nature of potential relationships and how they may contribute to AD risk, we explored whether any key metabolites are additionally associated with polygenic risk for AD.

2. METHODS

2.1 Participants

The NSHD is a broadly representative birth cohort study, originally following 5362 individuals since their birth in mainland Britain during one week in March 1946 (11). At age 69-71, 502 participants enrolled in Insight 46, the neuroscience sub-study. At a University College London clinic they underwent comprehensive clinical and cognitive tests, MRI, and ^{18}F-florbetapir positron emission tomography (PET) imaging (12,13). Compared to the full NSHD cohort, participants of Insight 46 were of slightly higher cognitive ability, more socially advantaged, and of better overall health (13). Further details on participant eligibility and recruitment can be found elsewhere (12).

Ethical approval was obtained from the National Research Ethics Service Committee London (14/LO/1173). All participants provided written informed consent.

2.2 Materials

2.2.1 Metabolite quality control

At age 60-64, blood samples were collected by trained research nurses. Samples were stored at -80 degrees.

Using Ultrahigh Performance Liquid Chromatography-Tandem Mass Spectrometry (UPLC-MS/MS), concentrations of 1401 metabolites were detected and measured by Metabolon Inc (Durham, NC, USA) among 1814 NSHD participants. Metabolites were assigned to nine families (lipids, amino acids, xenobiotics, peptides, nucleotides, cofactors and vitamins, carbohydrates, energy and partially characterised molecules) and further organised into pathways based on their proposed biological function informed by the Kyoto Encyclopaedia of Genes and Genomes (KEGG) database (Supplementary Table 1). Unknown metabolites were assigned to an "Unknown" family and pathway and denoted by a number prefixed by an "X"; these were included in all analyses.

Metabolite data underwent strict quality control (QC), as detailed in (10), resulting in 1019 metabolites (Supplementary Notes).

2.2.2 Genetic quality control

Initial QC and imputation were performed centrally by the NSHD study team (Supplementary Notes). For this analysis, we removed variants that were rare (MAF <5%), with a low call rate (<98%), or that deviated from Hardy-Weinberg equilibrium ($p < 1x10^{-5}$). We additionally removed participants with a low call rate (<98%), mismatching biological and reported sex, or that were related (PIHAT <0.1). All QC was performed using PLINK v1.9 (https://www.cog-genomics.org/plink2) (14). Following QC, genetic and metabolomic data were available for 1638 participants.

2.2.3 Scanning procedure

The scanning procedure and data processing were undertaken by the Insight 46 team. Aβ-PET and MRI were acquired contemporaneously using a single Biograph mMR 3 Tesla PET/MRI scanner (Siemens Healthcare) (12). Aβ burden was quantified over 10 minutes, approximately 50

minutes after intravenous injection with ^{18}F-florbetapir (370 mBq. Standardised uptake value ratios (SUVRs) were derived using a grey matter cortical composite and an eroded subcortical white matter reference region. A cut-off of >0.6104 was used to define Aβ positivity being the 99th percentile of the lower (Aβ-negative) Gaussian distribution (15). Participants below this threshold were defined as Aβ-negative. Data were processed using an in-house pipeline, including attenuation correction using pseudo-CT (12). For volumetric T1-weighted MRI images, visual QC was performed as detailed in (12), and processed using the following: MAPS (16) for whole-brain volume (with manual editing if needed); STEPS (17) for left and right hippocampal volumes (with manual editing if needed); and SPM12 (fil.ion.ucl.ac.uk/spm) (18) for total intracranial volume (TIV). hippocampal volume was calculated as the mean volume of the left and right hippocampi.

2.2.4 Covariables

In line with our previous analysis in the full NSHD (10,19), covariables were: sex, blood collection information (clinic location, age, fasting status), age at scan, *APOE* genotype (ε4 carrier/non carrier; blood samples at age 53 or 69y), BMI (60-64y nurse visit), lipid medication (yes/no; self-reported use in 24 hours preceding blood collection at 60-64y), childhood cognition (15y), highest level of educational attainment (no qualifications/'O level'/'A level' or higher; 26y), childhood socioeconomic position (SEP) (Father's current or last known occupation categorised according to the UK Registrar General; 11y), and midlife SEP (own occupation categorised as for childhood SEP; 53y).

2.3 Statistical analysis

We previously imputed missing covariable data using multiple imputation chained equations (100 iterations and 50 imputations) (20) in the full NSHD metabolomics dataset. Further details of missing data can be found in Table 1. Unless otherwise specified, we carried out all analyses in R

version 3.6.0 (details of all software and packages used can be found in Supplementary Notes). A visual summary of our analytical pipeline can be found in Figure 1.

2.3.1 Network analysis

To explore associations of clusters (termed "modules") of co-expressed metabolites, we applied weighted gene coexpression network analysis (WGCNA) (21–23) to metabolite data, as detailed previously (10). First, metabolite data were adjusted for sex and blood clinic information and the standardised residuals were used for WGCNA. Fourteen modules of highly connected metabolites were then identified, and the first principal component of each module (termed "module eigenvalue") was derived to allow for relationships between modules and outcomes to be examined. Overrepresentation analyses were conducted, using the hypergeometric test, to identify enriched pathways within the module and provide insight into potential biological function (10). Modules were allocated an arbitrary colour name using the WGCNA package for ease of discussion.

2.3.2 Hub metabolite analysis

Metabolites that are highly connected to their module (termed "hub metabolites") are likely to be functionally important and thus present as valuable marker candidates (24). We previously identified associations between 35 hubs, defined using correlations between metabolites and module eigenvalues exceeding r=0.65 (termed "module membership"; kME), and cognitive outcomes in the NSHD (10). As these metabolites were selected on the premise of showing associations with cognition, we additionally looked for hubs that may be important in brain imaging outcomes. To do this, we extracted any additional metabolites exceeding the 0.65 threshold (10) in modules showing significant ($p_{FDR}<0.05$) associations in the present analysis.

2.3.3 Regression analysis

To allow for direct comparisons, we standardised all continuous predictors and outcomes to a mean of 0 and standard deviation of 1. We then tested relationships between a) modules, and b) hub metabolites using linear regression (for whole-brain volume and hippocampal volume) and logistic regression (for Aβ status). Model 1 adjusted for basic covariables: sex, blood collection information, age at scan, *APOE* genotype, and total intracranial volume (for whole-brain volume and hippocampal volume only). As modules were already adjusted for sex and blood clinic information, these covariables were not additionally included for module analyses. Model 2 additionally adjusted for lipid-related factors: BMI and lipid medication use. Finally, model 3 further adjusted for childhood cognition, educational attainment, and SEP (parental and midlife).

Analyses were conducted on each imputed dataset and pooled using Rubin's rules (25). Regression assumptions were checked by examination of the residuals. We applied False Discovery Rate (FDR) correction using the Benjamini-Hochberg procedure (26) with an alpha=0.05. FDR correction was applied separately to each analysis (module and hub) and each outcome.

2.3.4 Polygenic score analysis

To explore whether levels of key metabolites were influenced by genetic risk for AD, we investigated associations between AD polygenic risk scores (PRS) – a weighted sum of genetic variants associated with a trait or disease – and significant ($p_{FDR}<0.05$) hub metabolites and modules. We obtained genome-wide association study summary statistics from Kunkle et al. (27) (N=63,926; 21,982 AD clinically-ascertained cases, 41,944 controls), which were used as the base data for PRS analyses. Using PRSice-2 (28), we computed PRS in the NSHD, both including and excluding SNPs in the *APOE* region (chr 19, GRCh37 coordinates 44912079 to 45912079) (29). Two p-value thresholds (P_T) – previously identified to be optimal for PRS including and excluding the *APOE* region – were used for SNP selection: $5x10^{-8}$ (suggested for *APOE* region

included) and 0.1 (suggested for *APOE* region excluded) (30), resulting in four PRS. SNPs in linkage disequilibrium (r^2>0.001 within a 250kb window) were clumped and the SNP with the lowest p-value was retained.

We first standardised predictors and outcomes to a mean of 0 and standard deviation of 1. Then, we regressed PRS on key metabolites and modules, adjusting for sex, age, blood collection details, and seven principal components (to control for population stratification). We applied FDR correction to each analysis – module and hub metabolite – using the Benjamini-Hochberg procedure (26) with an alpha=0.05.

2.3.5 Additional analysis

We conducted several additional analyses to test the robustness of our findings (see Supplementary Notes for full details). In brief, we first investigated whether WGCNA modules, which were curated in the full NSHD, were preserved in the Insight 46 subset. Then, for our main analyses, we applied a more conservative Bonferroni correction to our findings (module: $p<3.57x10^{-3}$; hub metabolite: $p<1.06x10^{-3}$). Finally, we explored whether lifestyle and related factors (lifetime smoking, diet, exercise, blood pressure, and alcohol intake) explained any of our results.

2.4 Data availability

Anonymised data are available upon request to bona fide researchers (https://skylark.ucl.ac.uk/NSHD/doku.php).

3. RESULTS

3.1 Participant characteristics

Insight 46 participants who completed the scanning procedure, were dementia-free, and with full metabolite data were included for module and hub metabolite analysis (N=437; 47.6% female, 18.9% Aβ-positive). For PRS analysis, NSHD participants with metabolomics and genetic data were included (N=1638; 50.4% female). Participant characteristics can be found in Table 1 (see Supplementary Notes for characteristics split by amyloid status).

3.2 Metabolite network modules

Overall, we identified three modules that showed associations with brain volume outcomes ($p_{FDR}<0.05$) and none ($p>0.05$) with Aβ status. Full results can be found in Supplementary Table 2 and are visualised in Figure 2. Results of the fully adjusted model are discussed hereafter.

We report associations between higher expression of two lipid modules and smaller whole-brain volumes: the brown module, enriched in diacylglcerol (DAG) and phosphatidylethanolamine (PE) pathways (ß=-0.066, 95%CI=[-0.11,-0.019], p=0.006, p_{FDR}=0.044) and the blue module, enriched in various fatty acids pathways (ß =-0.072, 95%CI=[-0.12,-0.026], p=0.0021, p_{FDR}=0.035). Higher expression of the yellow module, enriched in sphingolipid metabolism and related pathways, was associated with a larger hippocampal volume (ß =0.14, 95%CI=[0.055,0.23], p=0.0017, p_{FDR}=0.035).

3.3 Hub metabolites

We explored associations between 81 metabolites that were highly connected (kME>0.65) in significant modules identified in 3.2, and 35 that were identified to be hubs in our previous study on cognition. Across all models, we report 30 key metabolites after FDR correction, of which 13 were previously associated with cognitive outcomes (Figure 3). No associations were detected for any metabolite and Aβ status after FDR correction, although some nominal associations were observed (3 cyan and 1 yellow metabolite; Supplementary Table 1 & Figure 3).

In the fully adjusted model, 22 metabolites were associated (p_{FDR}<0.05) with an imaging outcome (whole-brain volume: 22; hippocampal volume only: 4). Of the 22, 10 metabolites belonged to the yellow module (pathways: sphingolipid metabolism) and were positively associated with larger brain volumes. Twelve metabolites were negatively associated with brain volumes: six belonging to the blue module (pathways: fatty acid (monohydroxy; dicarboxylate), long chain PUFA (n3 and n6), glycerolipid metabolism), four to the brown module (pathways: PE, phosphatidylcholine (PC), DAG), and two to additional modules identified in our previous analysis of cognition (pathways: fatty acid metabolism (acyl carnitine); methionine, cysteine, SAM and taurine metabolism).

3.4 Polygenic score analysis

We investigated whether modules and hub metabolites that associated with an imaging measure (p_{FDR}<0.05) were also associated with polygenic risk for AD (*APOE* region included and excluded). We observed no significant relationships following FDR correction. At the nominal threshold, we report relationships between higher AD PRS and decreased levels of five hub metabolites (Figure 3 and Supplementary Table 3). Three DAG and phosphoethanolamine hub metabolites were associated with the *APOE* AD PRS only (P_T=5x10^{-8}; ß range=-0.061 to -0.050, p range=0.012 to 0.04, p_{FDR}>0.05). Two PUFA were associated with the non-*APOE* AD PRS (P_T=0.1; DPA: ß =-0.077, 95%CI=[0.13,-0.029], p=0.0018; linolenate: ß =-0.054, 95%CI=[-0.10,-0.0056], p=0.029, p_{FDR}>0.05); these associations weakened with the *APOE* region additionally included (P_T=0.1; DPA: ß =-0.058, 95%CI=[-0.11,-0.0097], p=0.019, linolenate: ß =-0.040, 95%CI=[-0.088,0.0088], p=0.11). No relationships were observed for metabolite modules at either the nominal or adjusted level of significance.

3.5 Additional analysis

Full results from our additional analyses are discussed in Supplementary Notes. Briefly, modules showed moderate to large preservation in the Insight 46 subset of the NSHD (Supplementary Figure 1). Following Bonferroni correction (module: $p<3.57x10^{-3}$; hub metabolite: $p<1.06x10^{-3}$) all modules and nine metabolites remained associated with an outcome, with 21 no longer reaching the adjusted level of significance. Further adjustment for life course factors resulted in minimal changes (Supplementary Table 4 & 5).

4. DISCUSSION

In a population-based cohort, we identified three modules of coexpressed lipids (phospholipids and DAGs, fatty acids, and sphingolipids) that were associated with whole-brain or hippocampal volume, and 22 highly connected metabolites within these that present as potential markers for additional study. We report no significant metabolic associations for Aβ status following multiple testing correction, nor for AD polygenic risk in our genetic analyses, although some relationships were seen at the nominal level. Taken together, this suggests that these findings may be relevant to later life brain structure but non-specific to AD.

Fatty acids in whole-brain volume

First, we found that higher expression of the blue module – enriched in several fatty acid pathways – was associated with smaller whole-brain volumes. We identified eight hub metabolites in this module, belonging to fatty acid pathways (long chain PUFA (n3 and n6); monohydroxy; dicarboxylate; medium chain), as well as glycerol from the glycerolipid metabolism pathway. These pathways are tightly linked: fatty acids and glycerol constitute phospholipids and triglycerides, and dicarboxylate and monohydroxy fatty acids are oxidative products of PUFA and other fatty acids.

Upregulation of this module and hub metabolites may thus represent changes in lipid metabolism, including enhanced lipid breakdown, accumulation of free fatty acids and glycerol in the blood, and alterations in fatty acid oxidation – all of which have been linked to neurodegeneration and AD (31–33). Worth noting however, is that while n6 PUFA have been typically linked to risk of neurodegenerative disease, n3 have been linked to decreased risk (34), although both are an area of contention (35). Here, we identified one hub (docosapentaenoate; DPA, 22:5n3) to be an n3 fatty acid, contrasting with the general consensus among the literature. We additionally observed some suggestive evidence of an association between AD genetic risk and PUFA, particularly with the *APOE* region removed; a higher AD PRS was associated with decreased levels of two long chain PUFA hubs (linolenate (18:3n3 or 3n6) and DPA) at the nominal level. These effect directions align with those reported in the literature, but were in the unexpected direction based on our findings in our imaging analysis. Nevertheless, we note caution in interpreting these findings until they are replicated in a larger, independent sample.

We further identified two module hubs (16-hydroxypalmitate and hexadecanedioate) which are products of microsomal omega-oxidation – a minor oxidation pathway for fatty acids. Accumulation thus points to defects in mitochondrial β-oxidation pathways, perhaps induced by an overload of free fatty acids or vice versa (36,37). Notably, defective β-oxidation and compensatory omega-oxidation pathways are thought to induce oxidative stress (38,39) – a key mechanism linked to neurodegeneration, which is marked among other things by reduction in brain volume (40). Nevertheless, to our knowledge, these metabolites have not been linked to brain health and neurodegeneration previously, although they have been found to play a role in other phenotypes, such as mortality and blood pressure (36).

Phospholipids and DAGs in whole-brain volume

Similar to the module of fatty acids, higher expression of the brown module – enriched in PEs and DAGs – associated with smaller whole-brain volumes. These pathways have key roles in membrane structure and cell signalling: PEs are a class of phospholipid which form important components of cellular membranes (41), and are a precursor to DAGs, which are components of cellular membranes and secondary messengers (42). We similarly reported associations between this module and short-term memory in our previous analysis of cognition in the NSHD; these associations were in the same direction, although they were mostly explained by BMI and lipid medication (10). Here, our results were independent of these factors with minimal attenuations overall.

Within the module, we identified four hub metabolites – two phospholipids and two DAGs. Both subclasses have been previously linked to AD and neurodegeneration (43,44), and increased serum concentrations have been hypothesised to represent alterations in membrane integrity and subsequent degradation (45,46), although some associations in the opposite direction have also been reported (32,47). Three module hubs were additionally associated with an AD PRS at the nominal level; this relationship appeared to be driven predominantly by *APOE* and may thus reflect pleiotropic pathways. Again, these were in the opposite direction to our findings for whole-brain volume and did not survive multiple testing correction.

Sphingolipids in hippocampal volume and whole-brain volume

We highlighted associations between higher levels of sphingolipids, a lipid class that contain important constituents of cellular membranes (48), and larger hippocampal and whole-brain volumes. Our findings were observed for sphingomyelins in particular, a subclass that are especially abundant in the CNS, where they form pivotal components of neuronal membranes and play key roles in signal transduction (49). Given their biological role, it is unsurprising that sphingolipids have been previously linked to brain health and pathology (50,51). Here, we report

associations between a module enriched in sphingolipids and hippocampal volume, as well as several sphingomyelin hub metabolites and whole-brain volume.

We previously reported similar findings for sphingolipids and several cognitive outcomes; however, after we adjusted for childhood cognition and social factors – particularly childhood cognition and education – these were entirely attenuated (10). Interestingly, we did not observe the same pattern here, with relationships showing negligible attenuations overall, although it is worth noting that childhood cognition and education show much weaker associations with structural brain measures and thus less likely to confound associations. We hypothesised that attenuations could represent earlier relationships we are unable to capture without longitudinal metabolite data, shared genetic or environmental underpinnings, or confounding by reverse causation, i.e. increased sphingolipid levels may be consequential to a higher level of cognition in early life, and education may be capturing shared components. With no earlier measure of brain volume, the latter could extend to our present findings, although previous research has linked sphingomyelins to longitudinal markers of pathology (50,52,53). Another possibility is that sphingolipids may have different involvements in cognition and later life brain volume, or may be particularly important during sensitive age periods, for example in cognitive development as well as during vulnerable periods in later life with regard to structural integrity (48). Expanding this to longitudinal data, alongside interrogating relationships using MR, will allow for a greater insight into our findings.

Limited relationships were seen for Aβ status and AD polygenic risk

Interestingly, we saw limited metabolic relationships for Aβ status; no module showed associations ($p>0.05$) and a handful of metabolites were associated at the nominal threshold only. Possibly there are no robust associations between these metabolites and Aβ, or independent of *APOE*, although other work has reported differently (5,7). Alternatively, this

may reflect power, particularly given the relatively young age of Insight 46 participants and the fact that they have evidence for Aβ accumulation but not frank dementia. Expanding our analyses to further sweeps will be necessary to address this, by enrolling both a larger sample size and of individuals with detectable Aβ burden, as well as longer term follow-up. Furthermore, five metabolites were associated with an AD PRS at the nominal level, but none survived multiple testing correction and no other metabolic associations were seen at either threshold. As modules and metabolites were selected for PRS analyses based on showing significant associations in our imaging analyses, which were observed for whole-brain or hippocampal volume, and not Aβ, this suggests that these metabolites may not be specific to AD. Alternatively, it may also reflect power; expanding our work in larger samples is warranted.

Strengths and limitations

Strengths of the study include age-matched cohort with information on a large range of confounders across the life course, including the rarely available measure of childhood cognition, as well as data on both imaging and subdomains of cognition. Moreover, the metabolomics data in this study represent a far more comprehensive proportion of the metabolome than in past clinical metabolomic studies of neuroimaging parameters (7,53). Nevertheless, the results of this study should be interpreted in the context of the following limitations. First, our findings may not extend to the general population. Study participants are all white and, compared to the full NSHD cohort, participants enrolled in Insight 46 were on average of slightly better self-rated health, cognitive ability, and SEP. Next, our findings may be subject to residual confounding; further study is needed to disentangle causal relationships. Finally, as there are currently few cohorts with genetic, serum LC-MS, and brain imaging data, we are not yet able to replicate our analyses elsewhere.

Conclusion

Our findings highlight relationships between groups of lipids and structural brain measures, as well as key metabolites within these that are likely to be driving associations. Future work should be directed towards understanding if these metabolites associate with longitudinal changes in brain volumes, and whether relationships are causal; this could advance our understanding of brain health and neurodegeneration, and reveal possible targets of intervention.

ACKNOWLEDGEMENTS

We are very grateful to those study members who helped in the design of the study through focus groups, and to the participants both for their contributions to Insight 46 and for their commitments to research over the last seven decades. We are grateful to the radiographers and nuclear medicine physicians at the UCL Institute of Nuclear Medicine, and to the staff at the Leonard Wolfson Experimental Neurology Centre at UCL. We are particularly indebted to the support of the late Dr Chris Clark of Avid Radiopharmaceuticals who championed this study from its outset.

We thank the International Genomics of Alzheimer's Project (IGAP) for providing summary results data for these analyses. The investigators within IGAP contributed to the design and implementation of IGAP and/or provided data but did not participate in analysis or writing of this report.

We acknowledge use of the research computing facility at King's College London, *Rosalind* (https://rosalind.kcl.ac.uk), which is delivered in partnership with the National Institute for Health Research (NIHR) Biomedical Research Centres at South London & Maudsley and Guy's & St. Thomas' NHS Foundation Trusts, and part-funded by capital equipment grants from the Maudsley Charity (award 980) and Guy's & St. Thomas' Charity (TR130505). We further acknowledge the use of BioRender.com in the creation of Figure 1.

FUNDING

This study is principally funded by grants from Alzheimer's Research UK (ARUK-PG2014-1946, ARUK-PG2017-1946), the Medical Research Council Dementias Platform UK (CSUB19166), the Wolfson Foundation (PR/ylr/18575), and the Alzheimer's Association (SG-666374-UK). The genetic analyses are funded by the Brain Research Trust (UCC14191). Florbetapir amyloid tracer was kindly provided by Avid Radiopharmaceuticals (a wholly owned subsidiary of Eli Lilly. The NSHD is funded by the Medical Research Council (MC_UU_00019/1, MC_UU_00019/3).

R.G. is supported by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. M.A.S. was supported by the EPSRC-funded UCL Centre for Doctoral Training in Medical Imaging (EP/L016478/1). T.D.P. was supported by a Wellcome Trust Clinical Research Fellowship (200109/Z/15/Z) and is supported by a NIHR Clinical Lectureship. C.H.S. is supported by an Alzheimer's Society Junior Fellowship (AS-JF-17-011). R.D. is supported by the following: (1) NIHR Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London, London, UK; (2) Health Data Research UK, which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation and Wellcome Trust; (3) The BigData@Heart Consortium, funded by the Innovative Medicines Initiative-2 Joint Undertaking under grant agreement No. 116074. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA; it is

chaired by DE Grobbee and SD Anker, partnering with 20 academic and industry partners and ESC; (4) the National Institute for Health Research University College London Hospitals Biomedical Research Centre; (5) the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London; (6) the UK Research and Innovation London Medical Imaging & Artificial Intelligence Centre for Value Based Healthcare; (7) the National Institute for Health Research (NIHR) Applied Research Collaboration South London (NIHR ARC South London) at King's College Hospital NHS Foundation Trust. J.M.S. acknowledges the support of the National Institute for Health Research University College London Hospitals Biomedical Research Centre. M.R. is supported by the MRC (MC_UU_00019/3). P.P. is supported by Alzheimer's Research UK.

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.

COMPETING INTERESTS

MAS is currently a full-time employee of F. Hoffman-La Roche AG. However, her contributions to this research precede in time and are unrelated to her present employment. J.M.S. has received research funding from Avid Radiopharmaceuticals (a wholly owned subsidiary of Eli Lilly), has consulted for Roche Pharmaceuticals, Biogen, Merck, and Eli Lilly, given educational lectures sponsored by GE Healthcare, Eli Lilly, and Biogen. He is Clinical Advisor to UK DRI and Chief Medical Officer for ARUK.

CODE AVAILABILITY

Code used in this manuscript will be made available here

(https://github.com/beckigreen/imaging-metabolites).

REFERENCES

1. Alawode DOT, Heslegrave AJ, Ashton NJ, Karikari TK, Simrén J, Montoliu-Gaya L, et al. Transitioning from cerebrospinal fluid to blood tests to facilitate diagnosis and disease monitoring in Alzheimer's disease. J Intern Med. 2021 Sep;290(3):583–601.

2. Beebe K, Kennedy AD. Sharpening Precision Medicine by a Thorough Interrogation of Metabolic Individuality. Comput Struct Biotechnol J. 2016 Jan 21;14:97–105.

3. Wishart DS. Emerging applications of metabolomics in drug discovery and precision medicine. Nat Rev Drug Discov. 2016 Jul;15(7):473–84.

4. Voyle N, Kim M, Proitsi P, Ashton NJ, Baird AL, Bazenet C, et al. Blood metabolite markers of neocortical amyloid-β burden: discovery and enrichment using candidate proteins. Transl Psychiatry. 2016 Jan;6(1):e719–e719.

5. Kim M, Snowden S, Suvitaival T, Ali A, Merkler DJ, Ahmad T, et al. Primary fatty amides in plasma associated with brain amyloid burden, hippocampal volume, and memory in the European Medical Information Framework for Alzheimer's Disease biomarker discovery cohort. Alzheimer's & Dementia. 2019 Jun 1;15(6):817–27.

6. Leeuw FA de, Karamujić-Čomić H, Tijms BM, Peeters CFW, Kester MI, Scheltens P, et al. Circulating metabolites are associated with brain atrophy and white matter hyperintensities. Alzheimer's & Dementia. 2021;17(2):205–14.

7. Nho K, Kueider-Paisley A, Arnold M, MahmoudianDehkordi S, Risacher SL, Louie G, et al. Serum metabolites associated with brain amyloid beta deposition, cognition and dementia progression. Brain Communications. 2021 Jul 1;3(3).

8. Proitsi P, Kim M, Whiley L, Simmons A, Sattlecker M, Velayudhan L, et al. Association of blood lipids with Alzheimer's disease: A comprehensive lipidomics analysis. Alzheimer's & Dementia. 2017 Feb 1;13(2):140–51.

9. Xu J, Bankov G, Kim M, Wretlind A, Lord J, Green R, et al. Integrated lipidomics and proteomics network analysis highlights lipid and immunity pathways associated with Alzheimer's disease. Translational Neurodegeneration. 2020 Sep 21;9(1):36.

10. Green R, Lord J, Xu J, Maddock J, Kim M, Dobson R, et al. Metabolic underpinnings of late midlife cognitive outcomes: findings from the 1946 British Birth Cohort. Brain Communications. 2021 Dec 15;fcab291.

11. 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 Jun;10(2):145–57.

12. Lane CA, Parker TD, Cash DM, Macpherson K, Donnachie E, Murray-Smith H, et al. Study protocol: Insight 46 – a neuroscience sub-study of the MRC National Survey of Health and Development. BMC Neurology. 2017 Apr 18;17(1):75.

13. James S-N, Lane CA, Parker TD, Lu K, Collins JD, Murray-Smith H, et al. Using a birth cohort to study brain health and preclinical dementia: recruitment and participation rates in Insight 46. BMC Research Notes. 2018 Dec 13;11(1):885.

14. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am J Hum Genet. 2007 Sep;81(3):559–75.

15. Lu K, Nicholas JM, Collins JD, James S-N, Parker TD, Lane CA, et al. Cognition at age 70: Life course predictors and associations with brain pathologies. Neurology. 2019 Dec 3;93(23):e2144–56.

16. Leung KK, Barnes J, Modat M, Ridgway GR, Bartlett JW, Fox NC, et al. Brain MAPS: An automated, accurate and robust brain extraction technique using a template library. NeuroImage. 2011 Apr 1;55(3):1091–108.

17. Jorge Cardoso M, Leung K, Modat M, Keihaninejad S, Cash D, Barnes J, et al. STEPS: Similarity and Truth Estimation for Propagated Segmentations and its application to hippocampal segmentation and brain parcelation. Medical Image Analysis. 2013 Aug 1;17(6):671–84.

18. Malone IB, Leung KK, Clegg S, Barnes J, Whitwell JL, Ashburner J, et al. Accurate automatic estimation of total intracranial volume: A nuisance variable with less nuisance. NeuroImage. 2015 Jan 1;104:366–72.

19. 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 Sep 26;8(1):1–11.

20. Buuren S van, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software. 2011 Dec 12;45(1):1–67.

21. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008 Dec 29;9(1):559.

22. Langfelder P, Horvath S. Fast R Functions for Robust Correlations and Hierarchical Clustering. J Stat Softw. 2012 Mar;46(11):11.

23. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol. 2005;4: 17.

24. Horvath S, Dong J. Geometric Interpretation of Gene Coexpression Network Analysis. PLoS Comput Biol. 2008 Aug 15;4(8):e1000117.

25. Rubin DB. Multiple Imputation for Nonresponse in Surveys. John Wiley & Sons; 2004. 326p.

26. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological). 1995;57(1):289–300.

27. 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 Aβ, tau, immunity and lipid processing. Nat Genet. 2019 Mar;51(3):414–30.

28. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. Bioinformatics. 2015 May 1;31(9):1466–8.

29. Zettergren A, Lord J, Ashton NJ, Benedet AL, Karikari TK, Lantero Rodriguez J, et al. Association between polygenic risk score of Alzheimer's disease and plasma phosphorylated tau in individuals from the Alzheimer's Disease Neuroimaging Initiative. Alzheimer's Research & Therapy. 2021 Jan 8;13(1):17.

30. Leonenko G, Baker E, Stevenson-Hoare J, Sierksma A, Fiers M, Williams J, et al. Identifying individuals with high risk of Alzheimer's disease using polygenic risk scores. Nat Commun. 2021 Jul 23;12(1):4506.

31. Chew H, Solomon VA, Fonteh AN. Involvement of Lipids in Alzheimer's Disease Pathology and Potential Therapies. Frontiers in Physiology. 2020;11:598.

32. Kao Y-C, Ho P-C, Tu Y-K, Jou I-M, Tsai K-J. Lipids and Alzheimer's Disease. Int J Mol Sci. 2020 Feb 22;21(4):1505.

33. Fonteh AN, Cipolla M, Chiang AJ, Edminster SP, Arakaki X, Harrington MG. Polyunsaturated Fatty Acid Composition of Cerebrospinal Fluid Fractions Shows Their Contribution to Cognitive Resilience of a Pre-symptomatic Alzheimer's Disease Cohort. Front Physiol. 2020 Feb 14;11:83.

34. Dyall SC. Long-chain omega-3 fatty acids and the brain: a review of the independent and shared effects of EPA, DPA and DHA. Front Aging Neurosci. 2015 Apr 21;7:52.

35. González-Domínguez R, García-Barrera T, Gómez-Ariza JL. Using direct infusion mass spectrometry for serum metabolomics in Alzheimer's disease. Anal Bioanal Chem. 2014 Nov 1;406(28):7137–48.

36. Menni C, Graham D, Kastenmüller G, Alharbi NHJ, Alsanosi SM, McBride M, et al. Metabolomic Identification of a Novel Pathway of Blood Pressure Regulation Involving Hexadecanedioate. Hypertension. 2015 Aug 1;66(2):422–9.

37. Wanders RJA, Komen J, Kemp S. Fatty acid omega-oxidation as a rescue pathway for fatty acid oxidation disorders in humans. The FEBS Journal. 2011;278(2):182–94.

38. Natarajan SK, Ibdah JA. Role of 3-Hydroxy Fatty Acid-Induced Hepatic Lipotoxicity in Acute Fatty Liver of Pregnancy. Int J Mol Sci. 2018 Jan 22;19(1):322.

39. Gusdon AM, Song K, Qu S. Nonalcoholic Fatty Liver Disease: Pathogenesis and Therapeutics from a Mitochondria-Centric Perspective. Oxidative Medicine and Cellular Longevity. 2014 Oct 13;2014:e637027.

40. Klein JA, Ackerman SL. Oxidative stress, cell cycle, and neurodegeneration. J Clin Invest. 2003 Mar 15;111(6):785–93.

41. Fullerton MD, Hakimuddin F, Bonen A, Bakovic M. The Development of a Metabolic Disease Phenotype in CTP:Phosphoethanolamine Cytidylyltransferase-deficient Mice. J Biol Chem. 2009 Sep 18;284(38):25704–13.

42. Carrasco S, Mérida I. Diacylglycerol, when simplicity becomes complex. Trends in Biochemical Sciences. 2007 Jan 1;32(1):27–36.

43. 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. Alzheimer's & Dementia. 2017;13(9):965–84.

44. Wood PL, Medicherla S, Sheikh N, Terry B, Phillipps A, Kaye JA, et al. Targeted Lipidomics of Fontal Cortex and Plasma Diacylglycerols (DAG) in Mild Cognitive Impairment and Alzheimer's Disease: Validation of DAG Accumulation Early in the Pathophysiology of Alzheimer's Disease. Journal of Alzheimer's Disease. 2015 Jan 1;48(2):537–46.

45. González-Domínguez R, García-Barrera T, Gómez-Ariza JL. Application of a novel metabolomic approach based on atmospheric pressure photoionization mass spectrometry using flow injection analysis for the study of Alzheimer's disease. Talanta. 2015 Jan 1;131:480–9.

46. Calzada E, Onguka O, Claypool SM. Phosphatidylethanolamine Metabolism in Health and Disease. Int Rev Cell Mol Biol. 2016;321:29–88.

47. Huo Z, Rana BK, Elman JA, Dong R, Engelman CD, Johnson SC, et al. Metabolic Profiling of Cognitive Aging in Midlife. Frontiers in Aging Neuroscience. 2020;12:397.

48. Olsen ASB, Færgeman NJ. Sphingolipids: membrane microdomains in brain development, function and neurological diseases. Open Biology. 7(5):170069.

49. Schneider N, Hauser J, Oliveira M, Cazaubon E, Mottaz SC, O'Neill BV, et al. Sphingomyelin in Brain and Cognitive Development: Preliminary Data. eNeuro. 2019 Aug 5;6(4):ENEURO.0421-18.2019.

50. Mielke MM, Haughey NJ. Could plasma sphingolipids be diagnostic or prognostic biomarkers for Alzheimer's disease? Clin Lipidol. 2012 Oct;7(5):525–36.

51. Sliz E, Shin J, Ahmad S, Williams DM, Frenzel S, Gauß F, et al. Circulating Metabolome and White Matter Hyperintensities in Females and Males. Circulation. 2022 Jan 20; Epub ahead of print.

52. Mielke MM, Haughey NJ, Bandaru VVR, Weinberg DD, Darby E, Zaidi N, et al. Plasma sphingomyelins are associated with cognitive progression in Alzheimer's Disease. J Alzheimers Dis. 2011 Jan 1;27(2):259–69.

53. 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 Medicine. 2018 Jan 25;15(1):e1002482.

FIGURE LEGENDS

Figure 1. Analysis pipeline.

Figure 2. Module results.

A) Heatmap showing relationships between modules and brain imaging outcomes in the basic model and final model. Tiles are coloured by effect direction (blue = associated with better outcomes, red = associated with worse outcomes). Effect sizes are presented inside the tiles. Associations significant after multiple testing correction are represented by a solid fill and nominal ($p<0.05$) by a fainter fill. Module names in bold were additionally associated with a cognitive outcome in our previous analysis. Source data are present in Supplementary Table 2.

B) Table presenting enriched pathways in each module, with the most highly enriched pathways presented first. No pathways were enriched for the tan module. Source data are available in (7).

Figure 3. Hub metabolite results. Heatmap showing relationships between key hub metabolites ($p_{FDR}<0.05$) and brain imaging outcomes, as well as relationships between these metabolites and Alzheimer's disease polygenic risk scores (best threshold shown). Tiles are coloured by effect direction (blue = associated with better outcomes, red = associated with worse outcomes). Effect sizes are presented inside the tiles. Associations significant after multiple testing correction are represented by a solid fill and nominal ($p<0.05$) by a fainter fill. Metabolites are organised by both module (indicated via the colour panel on the right) and pathway (specified next to the module colour panel). Metabolite names in bold were additionally associated with a cognitive outcome in our cognition analysis. Source data are present in Supplementary Table 1 & 3).

AD = Alzheimer's disease, DMTPA = 2,3 Dihydroxy-5-methylthio-4-pentenoic acid, NSHD = National Survey of Health and Development, PRS = polygenic risk score.

SUPPLEMENTARY LEGENDS

Table 1. Full hub metabolite results (models 1-3)

Table 2. Full module analysis results (models 1-3)

Table 3. PRS results

Table 4. Associations between metabolites, modules, outcomes and lifestyle factors

Table 5. Associations between metabolites and modules and whole-brain volume following adjustment for smoking and alcohol intake

Supplementary Figure 1. A. Plot showing the Zsummary preservation statistics for each module against the module size. Thresholds for moderate and strong evidence of module preservation are indicated. B. Plot showing the median rank for module preservation. Modules are ranked from most preserved (lowest number) to least preserved (highest number).

Supplementary Notes. Supplementary Notes (metabolomics acquisition, metabolomic and genomic quality control, polygenic risk scores, software and packages used, additional analyses).

		Participants with genetic and metabolite data		Participants with imaging and metabolite data	
		Overall	Missing (%)	Overall	Missing (%)
n		1638		437	
Sex, N (%)	Male	812 (49.6)	0	229 (52.4)	0
	Female	826 (50.4)		208 (47.6)	
Age at scan (years), mean (SD)				70.7 (0.7)	0
APOE4 carrier, N (%)	Non-carrier	1044 (69.6)	8.5	306 (70.3)	0.5
	Carrier	455 (30.4)		129 (29.7)	
Amyloid status, N (%)	Negative			348 (81.1)	1.8
	Positive			81 (18.9)	
Hippocampal volume in ml, mean (SD)				3.1 (0.3)	0.5
Brain volume in ml, mean (SD)				1101.8 (99.3)	0.5
Total intracranial volume in ml, mean (SD)				1435.9 (132.2)	0.5
Age at blood collection (years), mean (SD)		63.2 (1.1)	0	63.3 (1.1)	0
Fasting at blood collection , N (%)	No	63 (3.8)	0	6 (1.4)	0
	Yes	1575 (96.2)		431 (98.6)	
Lipid medication use (age 60-64), N (%)	No	1232 (75.2)	0	335 (76.7)	0
	Yes	406 (24.8)		102 (23.3)	
Body mass index (age 60-64) in kg/m2, mean (SD)		27.8 (4.7)	0.1	27.4 (4.0)	0
Childhood cognitive ability (age 15)*, z-score, mean (SD)		0.2 (0.8)	14.3	0.5 (0.7)	8.2
Childhood socioeconomic position (age 11), N (%)	Unskilled	80 (5.1)	5	18 (4.1)	0.5

	Partly skilled	274 (17.6)		64 (14.7)	
	Manual skilled	462 (29.7)		106 (24.4)	
	Nonmanual skilled	276 (17.7)		90 (20.7)	
	Intermediate	346 (22.2)		110 (25.3)	
	Professional	118 (7.6)		47 (10.8)	
Adult socioeconomic position (age 53), N (%)	Unskilled	46 (2.8)	0.4	4 (0.9)	0
	Partly skilled	159 (9.7)		22 (5.0)	
	Manual skilled	238 (14.6)		40 (9.2)	
	Nonmanual skilled	380 (23.3)		91 (20.8)	
	Intermediate	680 (41.7)		225 (51.5)	
	Professional	129 (7.9)		55 (12.6)	
Highest educational attainment (age 26), N (%)	No qualification	446 (28.6)	4.9	65 (15.3)	3
	Up to GCSE	446 (28.6)		125 (29.5)	
	A-level or higher	665 (42.7)		234 (55.2)	

TABLE 1. Characteristics of participants included in this analysis.
*Childhood cognition Z-scores were calculated in the full National Survey of Health and Development cohort (N=5362).

Investigating associations between blood metabolites, later life brain imaging measures, and genetic risk for Alzheimer's disease - Figures

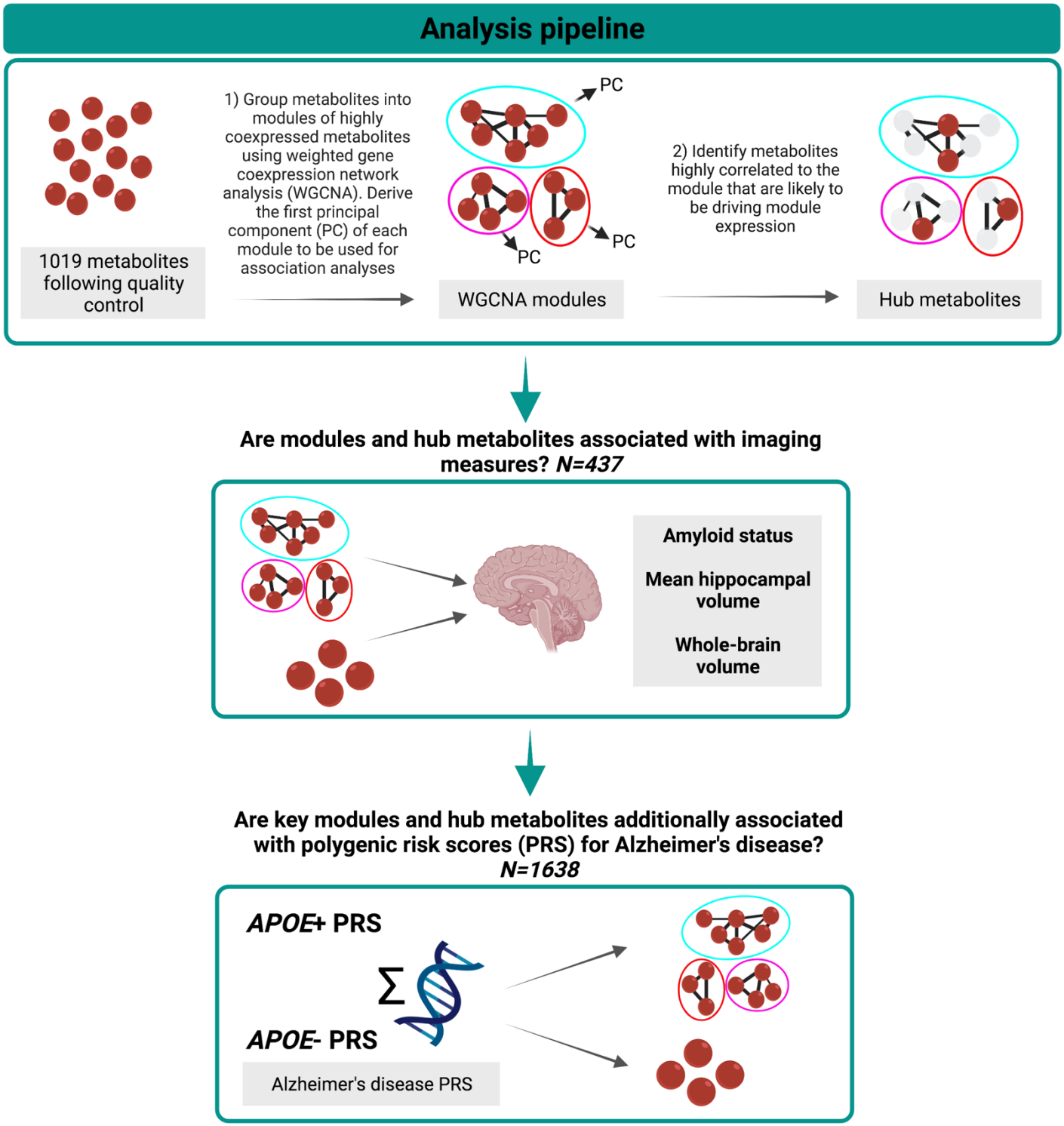
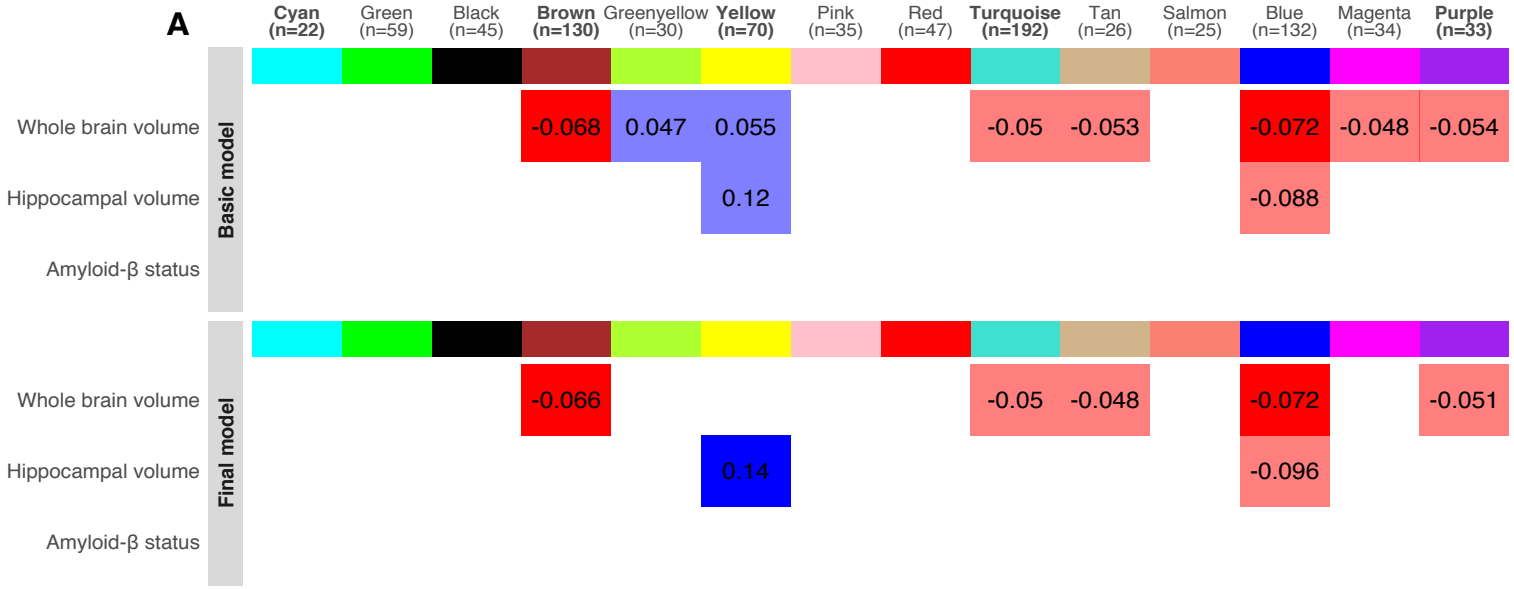


Figure 1. Analysis pipeline

A



B

Module	Pathway
cyan	Ascorbate and Aldarate Metabolism; Vitamin A Metabolism
green	Xanthine Metabolism; Benzoate Metabolism
black	Lysophospholipid; Fatty Acid Metabolism (Acyl Choline); Monoacylglycerol
brown	Diacylglycerol; Phosphatidylethanolamine (PE)
greenyellow	Plasmalogen; Ceramides
yellow	Sphingolipid Metabolism; Ceramides; Phosphatidylcholine (PC); Phosphatidylinositol (PI)
pink	Tryptophan Metabolism
red	Androgenic Steroids; Pregnenolone Steroids
turquoise	Histidine Metabolism; Urea cycle; Arginine and Proline Metabolism; Methionine, Cysteine, SAM and Taurine Metabolism; Tyrosine Metabolism; Polyamine Metabolism; Gamma-glutamyl Amino Acid; Leucine, Isoleucine and Valine Metabolism; Lysine Metabolism
salmon	Hemoglobin and Porphyrin Metabolism
blue	Long Chain Polyunsaturated Fatty Acid (n3 and n6); Long Chain Fatty Acids; Medium Chain Fatty Acid; Fatty Acid, Monohydroxy; Fatty Acid, Dicarboxylate; Endocannabinoid; Fatty Acid, Branched
magenta	Fatty Acid Metabolism (Acyl Carnitine)
purple	Fatty Acid Metabolism (Acyl Carnitine)

Figure 2.

A) Heatmap showing relationships between modules and brain imaging outcomes in the basic model and final model. Tiles are coloured by effect direction (blue = associated with better outcomes, red = associated with worse outcomes). Effect sizes are presented inside the tiles. Associations significant after multiple testing correction are represented by a solid fill and nominal ($p<0.05$) by a fainter fill. Module names in bold were additionally associated with a cognitive outcome in our previous analysis. Source data are present in Supplementary Table 2.

B) Table presenting enriched pathways in each module, with the most highly enriched pathways presented first. No pathways were enriched for the tan module. Source data are available in (7).

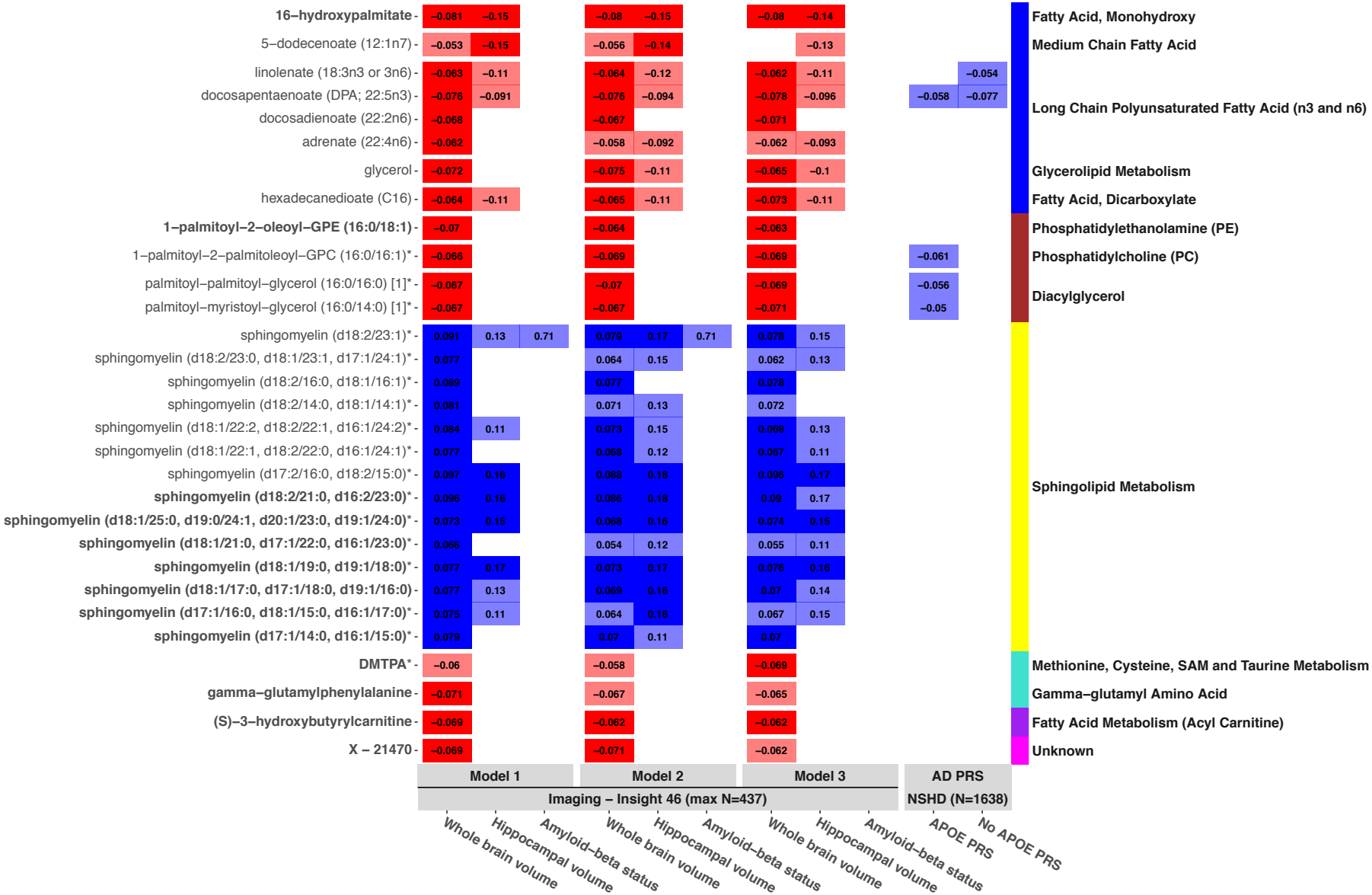


Figure 3. Hub metabolite results. Heatmap showing relationships between key hub metabolites (pFDR<0.05) and brain imaging outcomes, as well as relationships between these metabolites and Alzheimer's disease polygenic risk scores (best threshold shown). Tiles are coloured by effect direction (blue = associated with better outcomes, red = associated with worse outcomes). Effect sizes are presented inside the tiles. Associations significant after multiple testing correction are represented by a solid fill and nominal (p<0.05) by a fainter fill. Metabolites are organised by both module (indicated via the colour panel on the right) and pathway (specified next to the module colour panel). Metabolite names in bold were additionally associated with a cognitive outcome in our cognition analysis. Source data are present in Supplementary Table 1 & 3. *AD = Alzheimer's disease, DMTPA = 2,3 Dihydroxy-5-methylthio-4-pentenoic acid, NSHD = National Survey of Health and Development, PRS = polygenic risk score.*