



A metabolomic signature of the APOE2 allele

Paola Sebastiani · Zeyuan Song · Dylan Ellis · Qu Tian · Michaela Schwaiger-Haber · Ethan Stancliffe · Michael S. Lustgarten · Cory C. Funk · Priyanka Baloni · Cong-Hui Yao · Shakchhi Joshi · Megan M. Marron · Anastasia Gurinovich · Mengze Li · Anastasia Leshchik · Qingyan Xiang · Stacy L. Andersen · Mary F. Feitosa · Svetlana Ukraintseva · Mette Soerensen · Oliver Fiehn · Jose M. Ordovas · Marcia Haigis · Stefano Monti · Nir Barzilai · Sofiya Milman · Luigi Ferrucci · Noa Rappaport · Gary J. Patti · Thomas T. Perls

Received: 6 June 2022 / Accepted: 15 August 2022 / Published online: 23 August 2022
© The Author(s), under exclusive licence to American Aging Association 2022

Abstract With the goal of identifying metabolites that significantly correlate with the protective $\epsilon 2$ allele of the apolipoprotein E (*APOE*) gene, we established

a consortium of five studies of healthy aging and extreme human longevity with 3545 participants. This consortium includes the New England Centenarian Study, the Baltimore Longitudinal Study of Aging, the Arivale study, the Longevity Genes Project/LonGenity

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11357-022-00646-9>.

P. Sebastiani (✉) · A. Gurinovich
Institute for Clinical Research and Health Policy Studies,
Tufts Medical Center, 800 Washington Street, Boston,
MA 02111, USA
e-mail: psebastiani@tuftsmedicalcenter.org

Z. Song · Q. Xiang
Department of Biostatistics, Boston University, Boston,
MA, USA

D. Ellis · C. C. Funk · P. Baloni · N. Rappaport
Institute for Systems Biology, Seattle, WA, USA

Q. Tian · L. Ferrucci
Longitudinal Studies Section, Translational Gerontology
Branch, National Institute On Aging, Baltimore, MD, USA

M. Schwaiger-Haber · E. Stancliffe · G. J. Patti
Department of Chemistry, Department of Medicine,
Center for Metabolomics and Isotope Tracing, Washington
University in St. Louis, St. Louis, USA

M. S. Lustgarten
Nutrition, Exercise Physiology, and Sarcopenia
Laboratory, Jean Mayer USDA Human Nutrition Research
Center On Aging at Tufts University, Boston, MA, USA

C.-H. Yao · S. Joshi · M. Haigis
Department of Cell Biology at Harvard Medical School,
Boston, MA, USA

M. M. Marron
Department of Epidemiology, School of Public Health,
University of Pittsburgh, Pittsburgh, PA, USA

M. Li · A. Leshchik · S. Monti
Bioinformatics Program, Boston University, Boston, MA,
USA

S. L. Andersen · T. T. Perls
Department of Medicine, Boston University School
of Medicine, Boston, MA, USA

M. F. Feitosa
Division of Statistical Genomics, Department of Genetics,
Washington University School of Medicine, St Louis, MI,
USA

S. Ukraintseva
Biodemography of Aging Research Unit, Social Science
Research, Duke University, Durham, NC, USA

M. Soerensen
Epidemiology, Biostatistics and Biodemography,
Department of Public Health, University of Southern
Denmark, Odense, Denmark

O. Fiehn
West Coast Metabolomics Center, University of California,
Davis, CA, USA

studies, and the Long Life Family Study. We analyzed the association between *APOE* genotype groups E2 (e2e2 and e2e3 genotypes, $N=544$), E3 (e3e3 genotypes, $N=2299$), and E4 (e3e4 and e4e4 genotypes, $N=702$) with metabolite profiles in the five studies and used fixed effect meta-analysis to aggregate the results. Our meta-analysis identified a signature of 19 metabolites that are significantly associated with the E2 genotype group at $FDR < 10\%$. The group includes 10 glycerolipids and 4 glycerophospholipids that were all higher in E2 carriers compared to E3, with fold change ranging from 1.08 to 1.25. The organic acid 6-hydroxyindole sulfate, previously linked to changes in gut microbiome that were reflective of healthy aging and longevity, was also higher in E2 carriers compared to E3 carriers. Three sterol lipids and one sphingolipid species were significantly lower in carriers of the E2 genotype group. For some of these metabolites, the effect of the E2 genotype opposed the age effect. No metabolites reached a statistically significant association with the E4 group. This work confirms and expands previous results connecting the *APOE* gene to lipid regulation and suggests new links between the e2 allele, lipid metabolism, aging, and the gut-brain axis.

Keywords Apolipoprotein E · Longevity · Metabolomics · Lipid metabolism

Introduction

The apolipoprotein E (*APOE*) gene is associated with cognitive change and late-onset Alzheimer's disease [1]. The gene has three well-characterized alleles that are defined by a two-SNP haplotype (rs7412 and rs429358). The e3 allele is the most

common and it is considered the neutral allele. The e4 allele is associated with an increased risk for Alzheimer's disease (AD) and cognitive decline in whites, while the effect in other genetic backgrounds is less clear [2]. The e2 allele is the least common in whites with frequency ranging between 0.02 and 0.12 [3]. It is associated with increased human longevity [4] and neuroprotection and decreased risk for AD [5–7].

The biological mechanisms related to the various effects of the *APOE* e2 allele on cognitive function, resistance to AD and other dementias, and ultimately longevity are not well understood. *APOE* transports lipids across cells and tissues by binding to receptors of lipoproteins with affinity that is allele-dependent [8]. The gene is also a key regulator of plasma lipids, and a growing body of work has characterized lipid species that are associated with different *APOE* isoforms [9, 10]. For example, a genome-wide genetic study of 10,654 plasma metabolomic profiles identified a signature of 10 sphingolipids that were lower in e2 carriers compared to other *APOE* alleles [11]. Additionally, a large multi-cohort, multi-platform study identified 237 lipid species associated with the e2 allele [9].

To expand these results to cohorts of older individuals, we established a consortium of five studies of healthy aging and extreme human longevity that include, in aggregation, 544 carriers of e2, and metabolomic profiles of 3545 individuals with a total of 1453 chemicals measured in at least one study. This consortium includes the New England Centenarian Study [12], the Baltimore Longitudinal Study of Aging [13], the Arivale study [14], the Longevity Genes Project/LonGenity studies [15, 16], and the Long Life Family Study [17]. We harmonized the results across the studies that used different platforms for metabolomic profiling and used a fixed effect meta-analysis of the study-specific results to characterize a serum metabolomic signature of the e2 allele. The resulting signature includes glycerol, sterol and sphingolipids, and 6-hydroxyindole sulfate: a novel metabolite that is associated with gut microbiome patterns that are predictive of healthy aging and longevity [18]. By leveraging the wide range of ages of study participants, we were also able to investigate the effect of e2 on aging trends of some of these metabolites.

J. M. Ordovas
Nutrition and Genomics Team, Jean Mayer USDA
Human Nutrition Research Center On Aging and Gerald
J. and Dorothy R. Friedman School of Nutrition Science
and Policy, Tufts University, Boston, MA, USA

S. Monti
Section of Computational Biomedicine, Boston University
School of Medicine, Boston, MA, USA

N. Barzilai · S. Milman
Institute for Aging Research, Albert Einstein College
of Medicine, Bronx, NY, USA

Methods

Participant studies

New England Centenarian Study (NECS) A subset of 195 unrelated NECS participants comprising centenarians, centenarians' offspring, and unrelated controls were selected to be healthy at the time of blood draw (at least 1 year off from major clinical events), to be free of major medications, and to have survived at least 1 year after the blood draw. The age of NECS participants is carefully validated [12] and, in addition to an expansive data collection at enrollment, the participants are followed up annually to update their medical history and to assess their physical and cognitive function. Blood samples were collected at enrollment and transferred to the Boston University Molecular Genetic Core biobank for DNA, serum, and plasma extractions. Both serum and plasma samples were maintained at -80°C . The alleles of the *APOE* genes were determined from SNPs rs7412 and rs429358 (e2: rs7412=T rs429358=T; e3: rs7412=C rs429358=T; e4: rs7412=C rs429358=C) that were either genotyped using real-time PCR or imputed using IMPUTE2 in participants with genome-wide genotype data as described in [4]. We used the genetic data to enrich the selection of subjects with carriers of the allele e2 of *APOE* as described in [19]. A total of 1495 chemicals including 1213 compounds of known identity and 282 compounds of unknown structural identity were profiled in serum samples collected at enrollment at Metabolon, Inc. A detailed description for sample preparation and liquid chromatography-tandem mass spectrometry (LC-MS/MS) is in the Supplementary Material. Raw data was extracted, peak-identified, and QC-processed using Metabolon's hardware and software. Data were normalized to remove batch effects due to day-to-day variations. The compounds identified with the Metabolon bioinformatics pipeline were renamed using search tools from the Metabolomics workbench (Metabolomics Workbench: Tools: MS Searches), and standardized names from RefMet [20], Lipid Maps (LIPID MAPS), and the HMDB [21] were used when known. Supplementary Table 1 includes the mapping between the compound names provided by the Metabolon bioinformatics pipeline and the standardized names, as well as additional

annotations. The NECS protocol was approved by the Institutional Review Board of Boston University. Participants provided written informed consent at enrollment.

Arivale A subset of 634 Arivale wellness program participants with *APOE* genotype data, plasma metabolomic data, and other covariates data were included in this study. Arivale includes adults (age range 18–90+) who self-enrolled in a consumer facing wellness program (Arivale, Inc. 2015–2019) [14]. The participants were required to be over the age of 18 and not pregnant, with no additional screening of participants. The participants were provided wellness coaching, which targeted several clinical outcomes based on longitudinal omics measurements and detailed health history and behavioral questionnaires. Western International Review Board found that this study met the requirements for a waiver of consent under 45 CFR 46.116(d). *APOE* genotyping was based on whole genome sequencing of DNA extracted from whole blood using the Illumina TruSeq Nano Library prep kit and sequenced using Illumina HiSeq X, PE-150, target 30X coverage at a single CLIA approved sequencing laboratory. Metabolomics measurements of plasma extracted from blood samples collected from participants by certified phlebotomists were generated at Metabolon, Inc., using the Metabolon HD4 discovery platform and similar quality control pipelines used for NECS data. The study was reviewed and approved by the Western International Review Board. Participants consented to have their de-identified data being used for research purposes.

Baltimore Longitudinal Study of Aging (BLSA) A subset of 989 BLSA participants with *APOE* genotype data and plasma metabolites were selected for this study. The BLSA is a longitudinal study of community-dwelling adults selected to be free of cognitive impairment, functional limitations, chronic diseases, and cancer within the previous 10 years [13]. Blood samples were drawn from the antecubital vein between 7 and 8 am after an overnight fast. Participants were also requested not to smoke, exercise, or take medications before the blood sample collection. Plasma samples were prepared following the recommendations in [22]. *APOE* genotypes were determined using polymerase chain reaction (PCR) amplification

of leukocyte DNA [23] for earlier assays and using the TaqMan method for more recent assays [24]. Metabolomic analyses were conducted in plasma specimens via Biocrates p500 kit between April 2004 and December 2019. The BLSA protocol was approved by the Institutional Review Board of the National Institutes of Health. Participants provided written informed consent at each BLSA visit. Plasma metabolites were measured using LC–MS/MS. Metabolites were extracted and concentrations were measured using the MxP Quant 500 kit (Biocrates Life Sciences AG, Innsbruck, Austria) following the manufacturer’s protocol for a 5500 QTrap (Sciex, Framingham, MA, USA), as described in detail previously [25].

Long Life Family Study (LLFS) A subset of 1580 LLFS participants with *APOE* genotype data and data of 193 lipids that were profiled in plasma samples were included in this analysis. LLFS is a family-based study of healthy aging and longevity that enrolled 4981 family members from 552 families selected for familial longevity (6, 7). Participants were enrolled at three American field centers (Boston, Pittsburgh, and New York), and a Danish field center. Ages of all participants have been validated (8) and participants are followed up annually (9). *APOE* alleles were inferred from SNPs rs7412 and rs429358 haplotypes that were genotyped using whole genome sequencing [17]. All subjects provided informed consent and data are available via dbGaP (dbGaP Study Accession: phs000397.v1.p1). Experimental methods are described in detail in the Supplementary Material. In brief, lipid metabolites in plasma were extracted by using solid-phase extraction kits. Samples were subsequently separated by reversed-phase chromatography prior to being analyzed by an Agilent 6545 quadrupole time-of-flight mass spectrometer at Washington University St Louis. Data were processed with a combination of XCMS, DecoID, and Skyline to facilitate removal of background, annotation of adducts, and compound identification [26–28]. Lipid and polar metabolites were analyzed on a LC/MS platform using batches of approximately 90 samples each, and data were normalized by applying Combat [29]. From the metabolomic data, we focused on 193 lipid metabolites that were identified from plasma samples by matching the fragmentation patterns to metabolites databases. According to the Metabolomics Standard Initiative [30], these identifications are level 2 confidence.

Longevity Genes Project (LGP)/LonGenity The LGP and LonGenity cohorts were recruited at Albert Einstein College of Medicine to study the biological mechanisms of healthy longevity. LGP is a cross-sectional study of individuals aged 95 and older, their offspring, and controls without a history of parental longevity [15]. LonGenity is a longitudinal study of offspring of parents with exceptional longevity, defined as having at least one parent living to age 95 or older, and age- and sex-matched controls without parental longevity [16]. Both studies enroll individuals of Ashkenazi Jewish descent from Northeastern United States. Venous blood samples were collected at enrollment and serum was stored at -80°C until metabolomic analysis. Metabolomic analysis was performed on a subset of 150 samples from both studies, 50 samples from individuals aged 95 and older, 50 offspring, and 50 controls using LC/MS as in LLFS. By design, the subset was enriched for the oldest individuals whose offspring were also enrolled in the study. *APOE* genotype data was available for 147 individuals. The study protocols were approved by the Institutional Review Board at Albert Einstein College of Medicine. Participants provided written informed consent at enrollment.

Statistical analysis

NECS Raw data from the 1487 metabolites were inspected for residual batch effects but no evident bias was detected (Supplementary Fig. 2). The missing data pattern of each metabolite was analyzed using a logistic regression model of *APOE* genotypes, adjusting for age at blood draw, sex, batch indicator, and education, to identify metabolites that were not detected in carriers of specific *APOE* genotypes. This analysis did not identify any such metabolite. Only metabolites with less than 20% missing values were included in the final analysis. We used multiple imputation to fill-in missing metabolite values and repeated the imputation 5 times to maximize the power of the study. We imputed the non-detected metabolites using a uniform distribution defined between 0 and the minimum detected value of each metabolite, assuming that missing values were below the range of detection. In each imputed data set, each metabolite was transformed in natural log-scale, and correlated to *APOE* genotype group defined as $E2 = e2e2$ or $e2e3$; $E3 = e3e3$; and $E4 = e3e4$ or $e4e4$,

adjusting for age at blood draw, sex, education, and batch effect. The ambiguous e2e4 genotypes were removed from the data. The E3 group was used as the referent and the estimated effects of the E2 and E4 group on each metabolite in the 5 imputed data sets were aggregated using the formulas for repeated-imputation inference provided by Little and Rubin [31]. See Supplementary Fig. 2 for details.

Arivale This study provided data of 722 metabolites with less than 20% missing values in 634 study participants. Metabolite values were natural log-transformed to account for right skewness and outliers and the associations between each metabolite and the *APOE* genotype groups were estimated using linear regression, adjusted for age, sex, education, baseline BMI, and self-reported statin medication use. Missing metabolite values were ignored. Arivale analyses were performed in Python v3.7.6 with the statsmodels package v0.11.1

BLSA Metabolites with more than 30% missing values were excluded and missing values for the remaining metabolites were imputed using half of the minimum value. Metabolites were natural log-transformed and values equal to or greater than 4 standard deviation were considered outliers and excluded from the analysis. After pre-processing, 465 of the original 630 metabolites in 989 study participants were included in this analysis. The associations between metabolomics and *APOE* genotype groups were estimated using multivariable linear regression and were adjusted for age at blood draw, sex, batch effect, and years of education.

Long Life Family Study For the purposes of this work, we analyzed 193 lipid metabolites from 1580 plasma samples. Missing values were imputed using the half-minimum approach. The intensities of each compound were natural log-transformed and then correlated to the *APOE* genotype groups using a linear mixed effects model to account for genetic relationships. The analyses were adjusted for age at blood draw, sex, years of education, and the first four genome-wide principal components. The same model including only age at blood draw, sex, years of education, and the first four genome-wide principal components was used to characterize the distribution of metabolites at varying ages. The genetic relations,

genome-wide principal components analysis, and association analyses were conducted using the R package GENESIS.

LGP/LonGenity Data from 412 identified metabolites and less than 20% missing values in 147 serum samples were included in the analysis. Associations between metabolites and *APOE* genotype groups were estimated using linear regression adjusted for age at blood draw, sex, education, and batch effect. Missing values were ignored in this analysis. The association analyses were conducted using the R programming language (version 4.1.1).

Meta-analysis Results of the five studies were linked by the standardized name of metabolites, and the results of metabolites that were detected in at least 2 studies were aggregated by inverse-weighting, fixed effect meta-analysis using the rmeta package in the R software, V 4.1. Adjusted *p* values were calculated using the Benjamini–Hochberg correction [32] and results with a < 10% false discovery rate are included in Table 2. Complete meta-analysis results are in Supplementary Table 3.

Results

Study characteristics Table 1 summarizes the main characteristics of the five study populations. NECS participants were older, and with less years of education compared to the other studies. By design, the NECS participants were selected to be enriched for carriers of the E2 genotype group (24%), and had the smallest proportion of E4 carriers (10%). LLFS and LGP/LonGenity had a similar proportion of E2 carriers, higher than BLSA and Arivale but lower than NECS. The Arivale study included the youngest group of participants and the smallest proportion of E2 carriers.

Study-specific results The number and type of compounds being profiled by metabolomics are strongly dependent on methods. As an example, extracting samples with aqueous solvents is best suited for analysis of polar metabolites whereas extracting samples with organic solvents is better for examining lipid metabolites. The metabolomic data in all five studies described here were generated by using LC–MS, but

Table 1 Summary of patients' characteristic. Age is reported in median and range. E2=e2e2, e2e3; E3=e3e3; E4=e3e4, e4e4. Only categories of education were available in Arivale

	<i>N</i>	Age (years)	Sex (% F)	Education (years)	E2	E3	E4	Metabolomics
NECS	195	88 (56–115)	62%	14.4 (3.90)	47 (24%)	128 (66%)	20 (10%)	LC/MS with Metabolon Inc
LLFS	1580	83 (25–110)	52%	10.3 (3.99)	263 (17%)	1067 (68%)	250 (16%)	LC/MS
Arivale	634	49 (22–87)	61%	–	71 (11%)	415 (66%)	148 (23%)	LC/MS with Metabolon Inc
BLSA	989	67 (24–96)	52%	17.6 (2.7)	139 (14%)	583 (59%)	267 (27%)	LC/MS with Biocrates
LGP/LonGenity	147	80 (58–110)	69%	–	24 (16%)	106 (72%)	17 (12%)	LC–MS

different experimental and data-processing strategies were applied at Metabolon, Inc., Biocrates, Washington University in St. Louis, and Harvard Medical School (see “Methods”). As a result, each platform provided unique coverage of the metabolome. The lists of metabolites quantitated in each study are shown in the Supplementary Table 1. After reannotation of the metabolite names using RefMet, the five studies contributed data for a total of 1453 metabolites that were detected in at least one study. No metabolite was common to all studies; 21 were detected in four studies, 105 in three studies, and 427 in two studies (Supplement Fig. 3).

APOE-metabolite associations were initially conducted in each study, using regression analysis adapted to the specific study design and available meta-data of the individual cohorts, and the lists of study-specific results are included in Supplementary Table 2. In the NECS, no single *APOE*-metabolite association reached statistical significance after correction for multiple testing ($FDR < 10\%$), although we noticed a cluster of lipids including ceramides and glycerols showing strong, positive associations with the E2 genotype group. In the Arivale study, the metabolite 1-docosapentaenoyl-GPC (22:5n6) was significantly associated with the E2 genotype group (1.4 fold change (FC) comparing the E2 to the E3 group, adjusted p value = 0.0024). This metabolite did not pass the QC step in NECS, and it was not profiled in the other studies. In the BLSA, despite the substantially larger sample size, no single metabolite was significantly associated with any of the *APOE* genotype groups, after correcting for multiple testing ($FDR < 10\%$). However, we noticed a cluster of lipids mainly comprising triglycerides and sphingolipids that were nominally associated with the E2 genotype group (p value < 0.05). Several triglycerides

were higher in E2 carriers compared to E3 carriers, while sphingolipids were lower in E2 carriers compared to E3 carriers. In addition, putrescine was borderline significantly lower in carriers of the E4 genotype group (0.82 FC comparing the E4 to the E3 group, adjusted p value = 0.07174). In the LLFS, many lipids showed strong significant associations with the E2 genotype group. A few triglycerides and glycerolipids were significantly higher in carriers of the E2 genotype group compared to carriers of the E3 genotype group (largest effect: TG 56:1, FC = 1.45, p value = $3E - 10$), while a few sterol and sphingolipids were significantly lower in carriers of the E2 genotype group, compared to carriers of the E3 genotype group (largest effect: 22:6 cholesterol ester, CE 22:6, FC = 0.75, adjusted p value = $2E - 07$). Two metabolites reached statistically significant association in the LGP/LonGenity study (PC(O-40:6), FC = 1.26, p value = $3E - 05$; SM 46:3;O2, FC = 1.36, p value = $8E - 05$).

Results from meta-analysis Table 2 summarizes the 19 metabolites that were significantly associated with the E2 genotype group in the meta-analysis at $FDR < 10\%$. The group includes 10 glycerolipids that were all higher in E2 carriers compared to E3 carriers, with FC ranging from 1.14 (oleoyl-linoleoyl-glycerol, DG 18:1_18:2, adjusted p value = 0.001) to 1.23 (palmitoyl-oleoyl-glycerol, DG 16:0_18:1, adjusted p value = 0.0002) and 1.26 (1-oleoylglycerol, MG 18:1/0:0/0:0, adjusted p value = 0.03). Four glycerophospholipids were also significantly higher in E2 carriers compared to E3 carriers, with FC ranging from 1.08 (LPE 18:1, adjusted p value = 0.07) to 1.13 (PE 36:1, adjusted p value = 0.0054). Three sterol lipids were significantly lower in carriers of the

Table 2 List of metabolites that were significantly associated with APOE E2, FC, Est, SE, and Adj pvalue are fold change (FC) comparing E2 to E3 carriers, E2 effect relative to E3 in log-scale (Est), standard error of the effect (SE), and adjusted *p* values from the meta-analysis. NECS, Arivale, BLSA, LGP, and LLFS are the FC estimates obtained in the individual studies. Age effect is the effect of 1 year of age on log-transformed changes of the metabolites (adjusted by sex, education, genome-wide principal components) and values in parenthesis are the unadjusted *p* values to test the hypothesis of a significant age effect

Standardized name	Super class	FC	Est (SE)	Adj pvalue	FC NECS	FC Arivale	FC BLSA	FC LGP	FC LLFS	Age Effect	Metabolite
CE 22:6	Sterol lipid	0.78	-0.24(0.05)	0.00004			0.85		0.75	0.002 (0.08)	22:6 Cholesterol ester
DG 18:1_20:4	Glycerolipids	1.19	0.18(0.03)	0.00004	1.21	1.24			1.16	-0.008 (0)	Oleoyl-arachidonoyl-glycerol (18:1/20:4) [1]*
CE 20:4	Sterol lipid	0.84	-0.17(0.03)	0.00005			0.84		0.85	-0.004 (0)	20:4 Cholesterol ester
DG 16:0_18:2	Glycerolipids	1.19	0.17(0.03)	0.00007	1.23	1.18	1.23		1.17	-0.003 (0.026)	Palmitoyl-linoleoyl-glycerol (16:0/18:2) [1]*
DG 16:0_18:1	Glycerolipids	1.23	0.21(0.04)	0.00022	1.25	1.00	1.21		1.23	-0.004 (0.002)	Palmitoyl-oleoyl-glycerol (16:0/18:1) [1]*
DG 18:1_18:1	Glycerolipids	1.16	0.15(0.03)	0.00052	1.19	1.14	1.13		1.16	-0.001 (0.274)	Oleoyl-oleoyl-glycerol (18:1/18:1) [1]*
DG 18:1_18:2	Glycerolipids	1.15	0.14(0.03)	0.00105	1.21	1.14	1.16		1.12	-0.001 (0.633)	Oleoyl-linoleoyl-glycerol (18:1/18:2) [1]
PE 36:1	Glycerophospholipids	1.13	0.12(0.03)	0.00543				1.16	1.12	-0.006 (0)	
CE 22:5	Sterol lipid	0.87	-0.14(0.04)	0.00615			0.87		0.87	0.002 (0.046)	22:5 Cholesterol ester
DG 18:2_18:2	Glycerolipids	1.15	0.14(0.04)	0.01417	1.31	1.14	1.17		1.12	-0.002 (0.194)	linoleoyl-linoleoyl-glycerol (18:2/18:2) [1]*
MG 18:1/0/0/0	Glycerolipids	1.26	0.23(0.07)	0.03020	1.49	1.19					1-Oleoylglycerol (18:1)
SM 18:1/O2/24:0	Sphingolipids	0.93	-0.07(0.02)	0.03020	0.99	0.94	0.86		0.92	-0.01 (0)	Lignoceroyl sphingomyelin (d18:1/24:0)
LPE 18:2	Glycerophospholipids	1.08	0.08(0.02)	0.03521	1.04	1.05		1.19	1.09	0.001 (0.107)	1-Linoleoyl-GPE (18:2)*
PC P-16:0/20:4 or PC O-16:1/20:4	Glycerophospholipids	1.10	0.09(0.03)	0.06860	1.03	1.14					1-(1-Enyl-palmitoyl)-2-arachidonoyl-GPC (P-16:0/20:4)*
LPE 18:1	Glycerophospholipids	1.08	0.07(0.02)	0.07108	1.00	1.09		1.14	1.08	-0.002 (0.044)	1-Oleoyl-GPE (18:1)
6-Hydroxyindole sulfate	Organic acids	1.20	0.18(0.06)	0.07108	1.18	1.22					6-Hydroxyindole sulfate

Table 2 (continued)

Standardized name	Super class	FC	Est (SE)	Adj pvalue	FC NECS	FC Arivale	FC BLSA	FC LGP	FC LLFS	Age Effect	Metabolite
DG 16:0_18:1	Glycerolipids	1.19	0.18(0.06)	0.07108	1.25	1.16					Palmitoyl-oleoyl-glycerol (16:0/18:1) [2]*
DG 16:1_18:2	Glycerolipids	1.15	0.14(0.05)	0.08292	1.66	1.14	1.11	1.15		0 (0.998)	Palmitoleoyl-linoleoyl-glycerol (16:1/18:2) [1]*
DG 18:2_20:4	Glycerolipids	1.19	0.17(0.06)	0.08238	1.24	1.16					Linoleoyl-arachidonoyl-glycerol (18:2/20:4) [1]*

E2 genotype group (22:6 cholesterol ester, CE 22:6, FC=0.78, adjusted p value=0.00004; 20:4 cholesterol ester, CE 20:4, FC=0.84, adjusted p value=0.00005; 22:5 cholesterol ester, CE 22:5, FC=0.87, adjusted p value=0.0142). The sphingolipid lignoceroyl sphingomyelin SM 18:1;O2/24:0 was also significantly lower in carriers of the E2 genotype group (FC=0.93, adjusted p value=0.03). The study-specific results were very similar, as shown in the forest plots in Fig. 1 for the lipid species DG 18:1_20:4, and SM 18:1;O2/24:0, while additional examples are in Supplementary Fig. 4. In addition to lipids, the organic acid 6-hydroxyindole sulfate was higher in E2 carriers compared to E3 carriers (FC=1.20, adjusted p value=0.07). No metabolite was significantly associated with the E4 genotype group, after correcting for multiple testing.

Ageing trends To test whether the signature was enriched for metabolomic markers of aging, we examined the associations of these metabolites with age at enrollment in the LLFS that included the largest sample size and widest range of ages spanning 25 to 110 years (Table 1). Four lipid species were significantly associated with age after correction for multiple testing. Specifically, the oleoyl-arachidonoyl-glycerol (DG 18:1_20:4) showed lower values at older ages (p value=1.11E−12), and the PE 36:1 showed a decreasing trend with older age (p value=6.8E−13). Similarly, the 20:4 cholesterol ester (CE 20:4) and the lignoceroyl sphingomyelin (SM 18:1;O2/24:0) showed lower values at older ages (Table 2).

Discussion

Summary of finding We examined the association between *APOE* genotype groups and 1453 metabolites in 5 different studies comprising 3545 participants and 544 e2 carriers. After aggregating the study-specific results of 553 metabolites that were shared by at least 2 studies with a meta-analysis, 19 metabolites were significantly associated with the E2 genotype group at 10% FDR. The metabolomic signature of *APOE* E2 included 4 lipid classes (glycerolipids, glycerophospholipids, sphingolipids, and sterol lipids) and 6-hydroxyindole sulfate: an organic acid that may be a marker of gut microbiome health [18]. Four species of lipids from the 4 classes were also correlated with age. Interestingly, the effect of the E2 genotype group

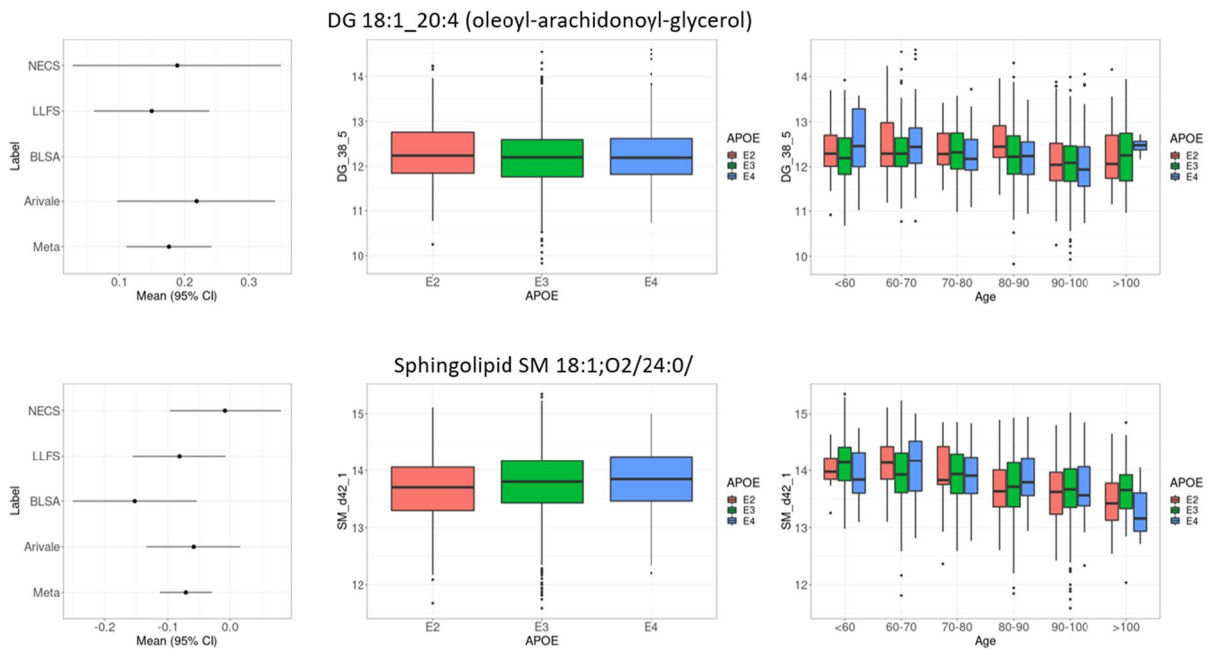


Fig. 1 Left panel: Forest plot of the study-specific effects of E2 = e2e2, e2e3 vs E3 = e3e3 and meta-analysis results for DG 18:1_20:4 (oleoyl-arachidonoyl-glycerol) and sphingolipid SM 18:1;O2/24:0/. The genetic effects on the *x*-axis are in log-

scale. The panels in the middle and the right display the distribution of the log-transformed levels of these two lipid species by APOE groups and age strata by APOE groups in LLFS participants

on the glycerolipid DG 18:1_20:4 and on the glycerophospholipid PE 36:1 was in the opposite direction to the age effect and carriers of the E2 genotype group appear to maintain, on average, higher values of these lipid species at any age. On the other hand, the effects of the E2 genotype group on the sphingolipid SM 18:1;O2/24:0 and the cholesterol ester CE 20:4 were concordant with the age effect, and carrying this genotype was associated with lower levels of both lipid species that were also lower at older ages. Last, it is noteworthy that many of the significant findings related to lipids with polyunsaturated fatty acyl chains (20:4, 22:5, 22:6) that were found to be linked to a variety of different lipid carriers. This finding supports the hypothesis that altered levels of lipid-bound polyunsaturated fatty acids may be part of the AD-protective and longevity-enhancing effects of APOE.

Discussion of key results and relevant literature The patterns of associations between *APOE* and the four classes of lipids in our study are consistent with those reported in recent studies and, particularly, the comprehensive analysis of Wang et al. [9] that identified 237

lipid species from more than 20 lipid classes that correlate with the e2 allele. Our results confirm the general trend of lower sphingo- and sterol-lipids and higher glycerolipids and glycerophospholipids in e2 carriers described in [9]. Lower levels of the three cholesterol esters CE 20:4, CE 22:5, and CE 22:6 in e2 carriers compared to e3 homozygotes have also been observed in older community-dwelling black men [33]. In addition, the same authors found a diacylglycerol, DG 38:5, that was similarly higher in e2 carriers vs. e3 homozygotes, though this result was only borderline significant.

In our analysis, we noticed that four of the lipid species that were associated with the e2 allele of *APOE* were also associated with age in the LLFS. In particular, levels of SM 18:1;O2/24:0 and CE 20:4 decreased with older age and were lower in e2 carriers, while levels of DG 18:1_20:4 (DG 38:5) and PE 36:1 decreased with older age but were higher in carriers of the e2 allele. We posit that since the e2 allele has a protective effect, the concordant associations of SM 18:1;O2/24:0 and CE 20:4 with both the e2 allele and older age suggest that decreasing levels of these two lipid species with older age may represent a protective mechanism that is enhanced in e2

carriers. Supporting this hypothesis, Wang et al. reported that elevated levels of some cholesteryl esters were associated with increased risk for AD [9], replicating data also provided in [34] and [10]. In addition, higher levels of cholesterol esters and sphingomyelins were associated with higher risk for cardiovascular disease and mortality [35]. It has been recently shown that an imbalance in sphingomyelin species is associated with AD pathogenesis and modulation of sphingosine-1-phosphate receptor activity could restore this imbalance [36]. Our results suggest that the e2 allele of APOE could have a role in maintaining protective levels of sphingomyelin species that warrant further investigation.

On the other hand, the discordant associations of the glycerol lipid DG 18:1_20:4 (DG 38:5) and the glycerophospholipid PE 36:1 with the E2 group and age (higher in E2 carriers and lower with older age) suggest that maintaining higher levels of this lipid species as people age might have a beneficial effect. Elevated levels of several diacylglycerols in E2 carriers have been previously reported in [9] and, while only lower levels of the species DG 16:0_22:6 (DG 38:6) have been correlated to AD by these investigators, lower levels of several DG species have been associated with higher risk for cardiovascular events and mortality in large cohorts [35]. Earlier reports that measured DG species using a different technology in a small number of participants reported significantly higher levels of DG 34:2 and DG 36:2 in participants with mild cognitive impairment versus individuals with mild and severe dementia but also when compared with controls [37], and similar patterns were noted by Wood et al. [38]. It has been postulated that glycerol lipids could be blood biomarkers of immune activation in the brain [39] and possibly modulate autophagy by induction of protein kinase [40, 41]. Declines of glycerophospholipids have been associated with older age in mice, and the decline appeared to be reduced after treatment with acarbose although the effect was significant only in males [42].

A novel finding in our analysis is the higher level of 6-hydroxyindole sulfate detected in e2 carriers. This metabolite is a member of the class of indoles that have several protective features, including anti-oxidant effects, and have been shown to increase health span and extend survival in a number of animal models [43], although elevated levels of this metabolite were noted in individuals with impaired kidney function [44]. This metabolite was associated with changes in gut microbiome that were

reflective of healthy aging and longevity in the Arivale cohort [18]. Interestingly, plasma levels of 6-hydroxyindole sulfate terms were positively correlated with prevalence of *Akkermansia* and *Lachnoclostridium* in the atlas of metabolomic signatures of the gut microbiome [45]. *Akkermansia* is considered a marker of gut health and is highly prevalent in centenarians [46]. *Lachnoclostridium* abundance was increased in e2 carriers, when compared with e3 carriers, in mice engineered to express *APOE* alleles [47]. This metabolite is structurally similar to 5-hydroxyindole, which is an inhibitor of monoamine oxidase in brain (MAO) [48], that is used as a therapeutic approach in Alzheimer's disease treatment [49]. Collectively, these data suggest a link between the e2 allele and the gut-brain axis.

No clear role for E4 Our study did not find any significant association between metabolites and the E4 genotype group, despite the collectively large number of e4 carriers. Wang et al. [9] also reported a smaller number of metabolites associated with e4 than e2. It has previously been noted that differences for e2 carriers when compared with non-e2 carriers may be easier to detect because of allele differences in apolipoprotein affinity [33]. For example, when compared with the e3 or e4 alleles, the e2 allele causes the resulting *APOE2* isoform to have a 50× lower binding affinity for the low-density lipoprotein receptor [50], potentially resulting in larger differences in circulating lipids that are easier to detect [33, 51]. In addition, serum and/or plasma may not be the best tissue. For example, very recently Miranda et al. [52] found lower levels of some diglycerol lipids in brain tissues of e4 carriers compared to e3 homozygotes.

Limitations Our analysis aggregated the results derived from five different studies that used a different study design, as well as different enrollment criteria, blood collection protocols, and approaches to metabolomics. Despite these differences, the metabolomic signature of the E2 genotype group is remarkably robust but probably limited to metabolites that were detected with multiple approaches. There certainly will be additional metabolites that correlate with APOE alleles and that were not detected in our analysis. In addition, inclusion of a larger number of normally aging individuals could unveil associations between the E4 genotype group and metabolites.

Conclusions

Our analysis discovered a robust metabolomics signature of *APOE* alleles that confirms an important role of isoforms of this gene in regulation of lipids and suggests a possible link between *APOE* and the gut-brain axis that warrants further investigation. These results expand our ongoing characterization of the role of *APOE2* in longevity [4] and preservation of good cognitive functions [5] through biological mechanisms [19]. They also add new evidence supporting a possible role of longevity genetic variants in slowing down the rate of molecular aging [53]. However, many questions remain to be answered. Future studies should replicate the results in larger, ethnically, and racially diverse groups, and validate the use of different technologies. It will also be important to quantify the portion of the protective effects of $\epsilon 2$ on longevity and neuroprotection that is mediated by the metabolomic signature to assess the value of this signature as target for therapeutics. Finally, as we enrich our ongoing studies of different molecular data such as serum proteomics, integrating the multi-omic associations with *APOE* genotype into one model will help decipher the molecular steps between genotype and the various paths to healthy aging and extreme human longevity.

Acknowledgements BLSA was supported by the Intramural Research Program of the National Institute on Aging. We acknowledge Mr. Brendan A. Mitchell for statistical assistance.

Funding NIA R01AG061844 (PS, TTP), NIA U19-AG023122 (to NR, TTP, PS); USDA 58–1950-4–003 (MSL), NHLBI T32-HL083825 (MMM). NIA R01AG057909 (NB, SM), NIA R01AG061155 (SM, NB). NIA U19AG063893 (GJP, TTP), NIA UH2AG064704 (TTP, PS, SM, NB). This project was supported in part by the USDA Agricultural Research Service Cooperative Agreement 58-8050-9-004. The content is the sole responsibility of the authors and does not necessarily represent the official views of the USDA.

The authors declare no competing interests. All study participants provided informed consent and the studies were approved by the Institutions' IRB as described in the "Methods."

References

- Ji Y, et al. Apolipoprotein Epsilon epsilon4 frequency is increased among Chinese patients with frontotemporal dementia and Alzheimer's disease. *Dement Geriatr Cogn Disord*. 2013;36(3–4):163–70.
- Farrer LA, et al. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis APOE and Alzheimer Disease Meta Analysis Consortium. *Jama*. 1997;278(16):1349–56.
- Corbo RM, R Scacchi, Apolipoprotein E (APOE) allele distribution in the world. Is APOE*4 a 'thrifty' allele? *Ann Hum Genet*, 1999; **63**(Pt 4) 301–10
- Sebastiani P, et al. APOE alleles and extreme human longevity. *J Gerontol A Biol Sci Med Sci*. 2019;74(1):44–51.
- Sweigart B, et al. APOE E2/E2 is associated with slower rate of cognitive decline with age. *J Alzheimers Dis*, 2021.
- Kim YJ, et al. Protective effects of APOE $\epsilon 2$ against disease progression in subcortical vascular mild cognitive impairment patients: a three-year longitudinal study. *Sci Rep*. 2017;7(1):1910.
- Reiman EM, et al. Exceptionally low likelihood of Alzheimer's dementia in APOE2 homozygotes from a 5,000-person neuropathological study. *Nat Commun*. 2020;11(1):667.
- Huang Y, Mahley RW. *Apolipoprotein E: structure and function in lipid metabolism, neurobiology, and Alzheimer's diseases*. *Neurobiology of Disease*. 2014;72:3–12.
- Wang T, et al. APOE epsilon2 resilience for Alzheimer's disease is mediated by plasma lipid species: analysis of three independent cohort studies. *Alzheimers Dement*, 2022.
- Liu Y, et al. Plasma lipidome is dysregulated in Alzheimer's disease and is associated with disease risk genes. *Transl Psychiatry*. 2021;11(1):344.
- Hysi PG, et al. Metabolome genome-wide association study identifies 74 novel genomic regions influencing plasma metabolites levels. *Metabolites*. 2022;12(1):61.
- Sebastiani P, Perls TT. The genetics of extreme longevity: lessons from the New England Centenarian Study. *Front Genet*. 2012;3:277.
- Ferrucci L. The Baltimore Longitudinal Study of Aging (BLSA): a 50-year-long journey and plans for the future. *J Gerontol A Biol Sci Med Sci*. 2008;63(12):1416–9.
- Earls JC, et al. Multi-omic biological age estimation and its correlation with wellness and disease phenotypes: a longitudinal study of 3,558 individuals. *J Gerontol A Biol Sci Med Sci*. 2019;74(Suppl_1):S52–s60.
- Barzilai N, et al. Unique lipoprotein phenotype and genotype associated with exceptional longevity. *JAMA*. 2003;290(15):2030–40.
- Gubbi S, et al. Effect of exceptional parental longevity and lifestyle factors on prevalence of cardiovascular disease in offspring. *Am J Cardiol*. 2017;120(12):2170–5.
- Wojczynski MK, et al. NIA Long Life Family Study: objectives, design, and heritability of cross sectional and longitudinal phenotypes. *J Gerontol A Biol Sci Med Sci*, 2021.
- Wilmanski T, et al. Gut microbiome pattern reflects healthy ageing and predicts survival in humans. *Nat Metab*. 2021;3(2):274–86.
- Sebastiani P, et al. A serum protein signature of APOE genotypes in centenarians. *Aging Cell*, 2019; e13023.
- Fahy E, Subramaniam S. RefMet: a reference nomenclature for metabolomics. *Nat Methods*. 2020;17(12):1173–4.
- Wishart DS, et al. HMDB 5.0: the Human Metabolome Database for 2022. *Nucleic Acids Res*. 2021;50(D1):D622–31.
- Tuck MK, et al. Standard operating procedures for serum and plasma collection: Early Detection Research Network

- Consensus Statement Standard Operating Procedure Integration Working Group. *J Proteome Res.* 2009;8(1):113–7.
23. Hixson JE, Vernier DT. Restriction isotyping of human apolipoprotein E by gene amplification and cleavage with HhaI. *J Lipid Res.* 1990;31(3):545–8.
 24. Koch W et al. TaqMan systems for genotyping of disease-related polymorphisms present in the gene encoding apolipoprotein E. 2002;40(11):1123–1131.
 25. Yamaguchi Y, et al. Plasma metabolites associated with chronic kidney disease and renal function in adults from the Baltimore Longitudinal Study of Aging. *Metabolomics.* 2021;17(1):9.
 26. Mahieu NG, et al. Defining and detecting complex peak relationships in mass spectral data: the Mz unity Algorithm. *Anal Chem.* 2016;88(18):9037–46.
 27. Cho K, et al. Targeting unique biological signals on the fly to improve MS/MS coverage and identification efficiency in metabolomics. *Anal Chim Acta.* 2021;1149:338210.
 28. Stancliffe E, et al. DecoID improves identification rates in metabolomics through database-assisted MS/MS deconvolution. *Nat Methods.* 2021;18(7):779–87.
 29. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics.* 2006;8(1):118–27.
 30. Sumner LW, et al. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics.* 2007;3(3):211–21.
 31. Little, R.J.A. and D.B. Rubin, *Statistical analysis with missing data.* 1987: Wiley.
 32. Benjamini Y, Hochberg Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J Roy Statist Soc Series B.* 1995;57(1):289–300.
 33. Marron MM et al. Using lipid profiling to better characterize metabolic differences in apolipoprotein E (APOE) genotype among community-dwelling older Black men. *Geroscience.* 2021.
 34. Proitsi P, et al. Plasma lipidomics analysis finds long chain cholesteryl esters to be associated with Alzheimer's disease. *Transl Psychiatry.* 2015;5(1):e494–e494.
 35. Mundra PA et al. Large-scale plasma lipidomic profiling identifies lipids that predict cardiovascular events in secondary prevention. *JCI Insight;* 2018. 3(17).
 36. Baloni P et al. Multi-omic analyses characterize the ceramide/sphingomyelin pathway as a therapeutic target in Alzheimer's disease. *medRxiv.* 2021 2021.07.16.21260601.
 37. Wood PL, et al. Targeted lipidomics distinguishes patient subgroups in mild cognitive impairment (MCI) and late onset Alzheimer's disease (LOAD). *BBA clinical.* 2015;5:25–8.
 38. Wood PL, et al. Targeted lipidomics of frontal 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 : JAD.* 2015;48(2):537–46.
 39. Wood PL, Cebak JE, Woltjer RL. Diacylglycerols as biomarkers of sustained immune activation in proteinopathies associated with dementia. *Clinica Chimica Acta.* 2018;476:107–10.
 40. Callender JA, Newton AC. Conventional protein kinase C in the brain: 40 years later. *Neuronal signaling.* 2017;1(2):NS0160005-NS20160005.
 41. Soto-Avellaneda A, Morrison BE. Signaling and other functions of lipids in autophagy: a review. *Lipids Health Dis.* 2020;19(1):214.
 42. Herrera JJ et al. Acarbose has sex-dependent and -independent effects on age-related physical function, cardiac health, and lipid biology. *JCI Insight.* 2020; 5(21).
 43. Sonowal R, et al. Indoles from commensal bacteria extend healthspan. *Proc Natl Acad Sci.* 2017;114(36):E7506–15.
 44. Masuo Y, et al. 6-Hydroxyindole is an endogenous long-lasting OATP1B1 inhibitor elevated in renal failure patients. *Drug Metab Pharmacokinet.* 2020;35(6):555–62.
 45. Dekkers KF et al. An online atlas of human plasma metabolite signatures of gut microbiome composition. *medRxiv.* 2021 2021.12.23.21268179.
 46. Badal VD, et al. The gut microbiome, aging, and longevity: a systematic review. *Nutrients.* 2020;12(12):3759.
 47. Zajac DJ, et al. APOE genetics influence murine gut microbiome. *Sci Rep.* 2022;12(1):1906.
 48. Crumeyrolle-Arias M, et al. Inhibition of brain mitochondrial monoamine oxidases by the endogenous compound 5-hydroxyoxindole. *Biochem Pharmacol.* 2004;67(5):977–9.
 49. Behl T et al. Role of monoamine oxidase activity in Alzheimer's disease: an insight into the therapeutic potential of inhibitors. *Molecules.* 2021; 26(12).
 50. Bu G. Apolipoprotein E and its receptors in Alzheimer's disease: pathways, pathogenesis and therapy. *Nat Rev Neurosci.* 2009;10(5):333–44.
 51. Wong MWK, et al. APOE genotype differentially modulates plasma lipids in healthy older individuals, with relevance to brain health. *J Alzheimers Dis.* 2019;72(3):703–16.
 52. Miranda AM, et al. Effects of APOE4 allelic dosage on lipidomic signatures in the entorhinal cortex of aged mice. *Transl Psychiatry.* 2022;12(1):129.
 53. Gurinovich A et al. Effect of longevity genetic variants on the molecular aging rate. *Geroscience.* 2021.

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

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.