nature aging

Supplementary information

https://doi.org/10.1038/s43587-026-01149-4

The blood metabolome of brain health in midlife and influences of genes, microbiome and exposome

In the format provided by the authors and unedited

Supplementary Materials

Supplementary Text

Comparison of explained variance estimates to previous studies

For genetic and gut microbiome determinants, we compared the estimates for the explained variance of metabolite levels from our study with the ones reported in Chen et al. (1) for the 133 metabolites measured in both studies (**Supplementary Table 8**). In Chen et al. and our study, 11 and 17 metabolites showed genetic contribution (EV > 0) in this set, respectively, with variance being explained through genetics for five metabolites in both studies: N(6)-methyllysine (our study: 61.6 % vs. Chen et al.: 15.8 %), N6,N6-dimethyllysine (41.4 % vs. 9.2 %), N-acetylornithine (23.1 % vs. 14.5 %), cis-4-decenoate (10:1n6) (2.8 % vs. 3.2 %), and 1-arachidonoyl-GPC (20:4n6) (1.9 % vs. 4.5 %). Notably, while the overlap seems moderate, the metabolites with EV>0 for genetics in only one of the studies have been identified (and replicated) as significant hits in previous studies on metabolite quantitative trait loci (mQTL) (e.g., (2)). For microbial features, the explained variance was more consistent across the two studies, with, for example, indolepropionate (our study: 15.8 % vs. Chen et al.: 12.2 %), p-cresol sulfate (31.9 % vs. 23.8 %), hippurate (16.1 % vs. 13.4 %), and phenol sulfate (11.0 % vs. 11.5 %) showing similar levels of variance explained.

Comparing EVs for genetics reported by Bar et al. (3) with those in our study, 44 metabolites with EV>10 % in Bar et al. (3) also showed evidence of genetic contribution (EV>0) in our study, with varying similarity in the actual estimates (e.g., 3-methylcitidine (our study: 16 % versus Bar et al.: 16%), imidazole lactate (4 % vs. 11 %)) (**Supplementary Table 8**). When contrasting EVs for gut-microbiota between the two studies, the number of metabolites with EV>0 were comparable (our study: n = 148 vs. Bar et al.: n = 182) with very similar estimates for EV percentages (e.g., cinnamoylglycine (our study: 24% versus Bar et al.: 24%), 3-phenylpropionate (31 % versus 26 %), p-cresol sulfate (32 % versus 42 %), and p-cresol glucuronide (22 % versus 41 %) (3).

Details of metabolomics data acquisition at Metabolon Inc.

Sample Accessioning: Following receipt, samples were immediately stored at -80^{o}C. Each sample received was accessioned into the Metabolon Laboratory Information Management System (LIMS) and was assigned a unique identifier. This identifier was used to track all sample handling, tasks, results, etc. The samples (and all derived aliquots) were tracked by the LIMS system.

Sample preparation: Samples were prepared using the automated MicroLab STAR® system from Hamilton. Several recovery standards were added prior to the first step in the extraction process for quality control (QC) purposes. To remove protein, dissociate small molecules bound to protein or trapped in the precipitated protein matrix, and to recover chemically diverse metabolites, proteins were precipitated with methanol under vigorous shaking for 2 min (Glen Mills GenoGrinder 2000) followed by centrifugation. The resulting extract was divided into five fractions: two for analysis by two separate reverse phase (RP)/UPLC-MS/MS

methods with positive ion mode electrospray ionization (ESI), one for analysis by RP/UPLC-MS/MS with negative ion mode ESI, one for analysis by HILIC/UPLC-MS/MS with negative ion mode ESI, and one sample was reserved for backup. Samples were placed briefly on a TurboVap® (Zymark) to remove the organic solvent. The sample extracts were stored overnight under nitrogen before preparation for analysis.

Quality assurance/quality control: Several types of controls were analyzed in concert with the experimental samples: a pooled matrix sample of well-characterized human plasma served as a technical replicate throughout the data set; extracted water samples served as process blanks; and a cocktail of QC standards that were carefully chosen not to interfere with the measurement of endogenous compounds were spiked into every analyzed sample, allowed instrument performance monitoring and aided chromatographic alignment. Instrument variability was determined by calculating the median relative standard deviation (RSD) for these standards. Overall process variability was determined by calculating the median RSD for all endogenous metabolites (i.e., non-instrument standards) present in 100% of the pooled matrix samples. Experimental samples were randomized across the platform run with QC samples spaced evenly among the injections.

Ultrahigh performance liquid chromatography-tandem mass spectroscopy (UPLC-MS/MS): All methods utilized a Waters ACQUITY ultra-performance liquid chromatography (UPLC) and a Thermo Scientific Q-Exactive high resolution/accurate mass spectrometer interfaced with a heated electrospray ionization (HESI-II) source and Orbitrap mass analyzer operated at 35,000 mass resolution. The sample extracts were dried then reconstituted in solvents compatible to each of the four methods. Each reconstitution solvent contained a series of standards at fixed concentrations to ensure injection and chromatographic consistency. One aliquot was analyzed using acidic positive ion conditions, chromatographically optimized for more hydrophilic compounds. In this method, the extract was gradient eluted from a C18 column (Waters UPLC BEH C18-2.1x100 mm, 1.7 µm) using water and methanol, containing 0.05% perfluoropentanoic acid (PFPA) and 0.1% formic acid (FA). Another aliquot was also analyzed using acidic positive ion conditions; however, it was chromatographically optimized for more hydrophobic compounds. In this method, the extract was gradient eluted from the same afore mentioned C18 column using methanol, acetonitrile, water, 0.05% PFPA and 0.01% FA and was operated at an overall higher organic content. Another aliquot was analyzed using basic negative ion optimized conditions using a separate dedicated C18 column. The basic extracts were gradient eluted from the column using methanol and water, however with 6.5mM Ammonium Bicarbonate at pH 8. The fourth aliquot was analyzed via negative ionization following elution from a HILIC column (Waters UPLC BEH Amide 2.1x150 mm, 1.7 µm) using a gradient consisting of water and acetonitrile with 10mM Ammonium Formate, pH 10.8. The MS analysis alternated between MS and data-dependent MS^n scans using dynamic exclusion. The scan range varied slightly between methods but covered 70-1000 m/z. Raw data files are archived and extracted as described below.

Bioinformatics: The informatics system consisted of four major components, the LIMS, the data extraction and peak-identification software, data processing tools for QC and compound identification, and a collection of information interpretation and visualization tools for use by

data analysts. The hardware and software foundations for these informatics components were the LAN backbone, and a database server running Oracle 10.2.0.1 Enterprise Edition.

Data extraction and compound identification: Raw data was extracted, peak-identified and QC processed using Metabolon's hardware and software. Compounds were identified by comparison to library entries of purified standards or recurrent unknown entities. Metabolon maintains a library based on authenticated standards that contains the retention time/index (RI), mass to charge ratio (m/z), and chromatographic data (including MS/MS spectral data) on all molecules present in the library. Furthermore, biochemical identifications are based on three criteria: retention index within a narrow RI window of the proposed identification, accurate mass match to the library +/- 10 ppm, and the MS/MS forward and reverse scores between the experimental data and authentic standards. The MS/MS scores are based on a comparison of the ions present in the experimental spectrum to the ions present in the library spectrum. While there may be similarities between these molecules based on one of these factors, the use of all three data points can be utilized to distinguish and differentiate biochemicals. More than 3300 commercially available purified standard compounds have been acquired and registered into LIMS for analysis on all platforms for determination of their analytical characteristics. Additional mass spectral entries have been created for structurally unnamed biochemicals, which have been identified by virtue of their recurrent nature (both chromatographic and mass spectral). These compounds have the potential to be identified by future acquisition of a matching purified standard or by classical structural analysis.

Curation: A variety of curation procedures were carried out to ensure accurate and consistent identification of true chemical entities, and to remove those representing system artifacts, mis-assignments, and background noise. Metabolon data analysts use proprietary visualization and interpretation software to confirm the consistency of peak identification among the various samples. Library matches for each compound were checked for each sample and corrected if necessary.

Metabolite quantification and data normalization: Peaks were quantified using area-under-the-curve. For studies spanning multiple days, a data normalization step was performed to correct variation resulting from instrument inter-day tuning differences. Essentially, each compound was corrected in run-day blocks by registering the medians to equal one (1.00) and normalizing each data point proportionately (termed the "block correction").

Additional details of preprocessing data from gut microbiome profiling

The 16S rRNA sequencing data from the samples of the RS-III cohort have been processed as follows.

Barcodes were separated from the reads and both ends were pasted together. These barcodes were then used to separate (demultiplex) the reads for each sample using an in-house developed script, allowing 1 error in each 12 nucleotide half of the 24 nucleotides long barcode. Reads were then cleaned of primer sequences and heterogeneity spacers using tag cleaner version 0.16 (4). Trimmed reads were then imported into the DADA2 R package version 1.18.0 (5). Samples without reads were removed and remaining reads were used as input for the filter step of DADA2. This filtering step removed reads with an expected error rate higher than 2 in both the forward and reverse reads, as well as any reads which had at

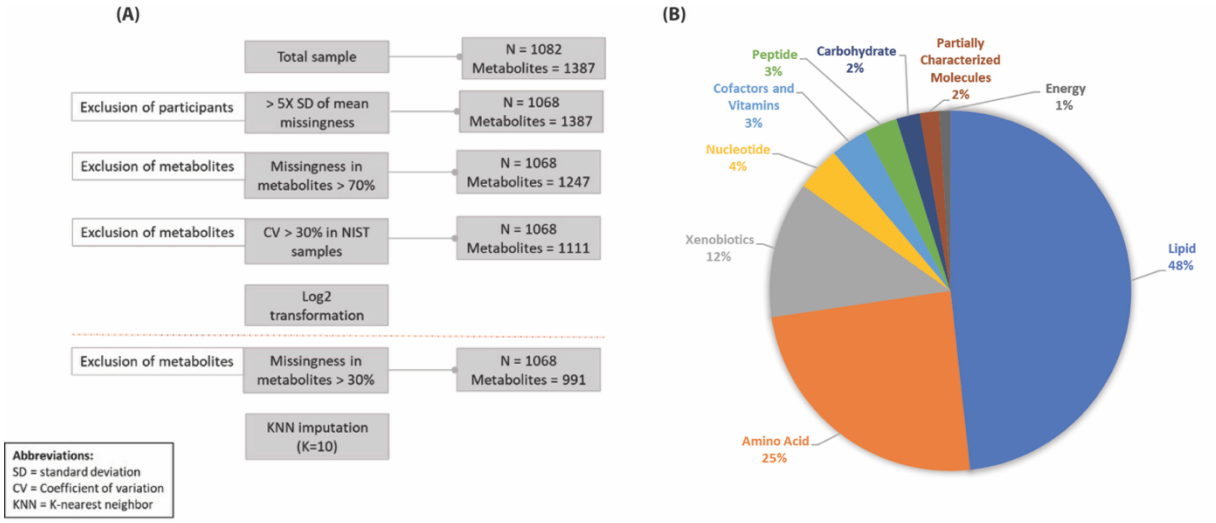
least one or more ambiguous bases in them ("N"). Additionally, reads were truncated if a low-quality base was encountered (Q-score =< 2). Next, reads were denoised by clustering them together based on similarity, starting with the most abundant read. All other reads with similarity within 10% were aligned against this cluster and included if the abundance p-value (likelihood of the read being too abundant to be explained by predicted errors in the main cluster sequence) was below the default threshold. The algorithm was then repeated with the remaining reads for the next abundant cluster, and so forth. Due to the nature of the algorithm requiring multiples of the same read, singletons are automatically excluded. The forward and reverse of the denoised reads were then merged if the overlap between them was a 100% match. Reads which were truncated in the filter step of DADA2 were also more likely to get removed in this step, due to possibly not having an overlap anymore. Next, the remaining chimeric sequences were removed using remove BimeraDenovo in consensus mode within DADA2. These cleaned, clustered reads (hereafter named Amplicon Sequence Variant, ASV) were then used as input for the RDP naive Bayesian classifier, which was trained on the SILVA version 138.1 microbial database (6, 7). Taxonomy for Kingdom through Genus was assigned if the bootstrap confidence score was above 50, meaning that a random part of the read could be assigned to those taxa at least 50% of the time. Species assignment was only reported if the ASV could be matched to a species for 100% and could not be mapped to other species either; otherwise "NA" was reported after the genus name. The ASV table, taxonomy table, and metadata were then combined into a phyloseq object for ease of analysis (8).

As described in **Methods**, further steps were taken to remove spurious and likely false-positive ASVs, filter samples, and for phylogenetic tree construction, which was added to the phyloseq object.

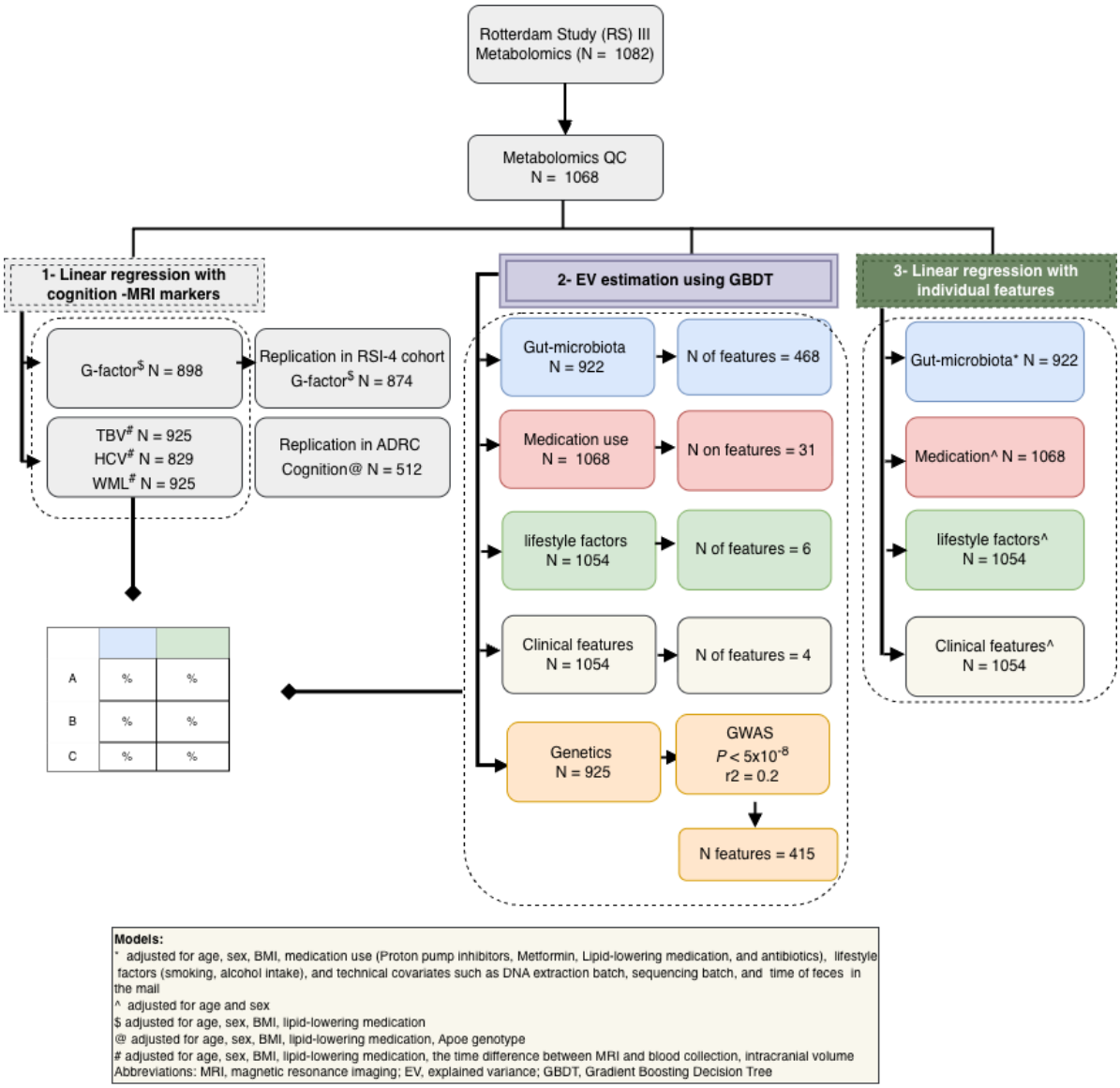
REFERENCES

1. Chen L, Zhernakova DV, Kurilshikov A, Andreu-Sanchez S, Wang D, Augustijn HE, et al. Influence of the microbiome, diet and genetics on inter-individual variation in the human plasma metabolome. Nat Med. 2022;28(11):2333-43.
2. Surendran P, Stewart ID, Au Yeung VPW, Pietzner M, Raffler J, Worheide MA, et al. Rare and common genetic determinants of metabolic individuality and their effects on human health. Nat Med. 2022;28(11):2321-32.
3. Bar N, Korem T, Weissbrod O, Zeevi D, Rothschild D, Leviatan S, et al. A reference map of potential determinants for the human serum metabolome. Nature. 2020;588(7836):135-40.
4. Schmieder R, Lim YW, Rohwer F, Edwards R. TagCleaner: Identification and removal of tag sequences from genomic and metagenomic datasets. BMC bioinformatics. 2010;11(1):1-14.
5. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nature methods. 2016;13(7):581-3.
6. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and environmental microbiology. 2007;73(16):5261-7.
7. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic acids research. 2012;41(D1):D590-D6.
8. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PloS one. 2013;8(4):e61217.

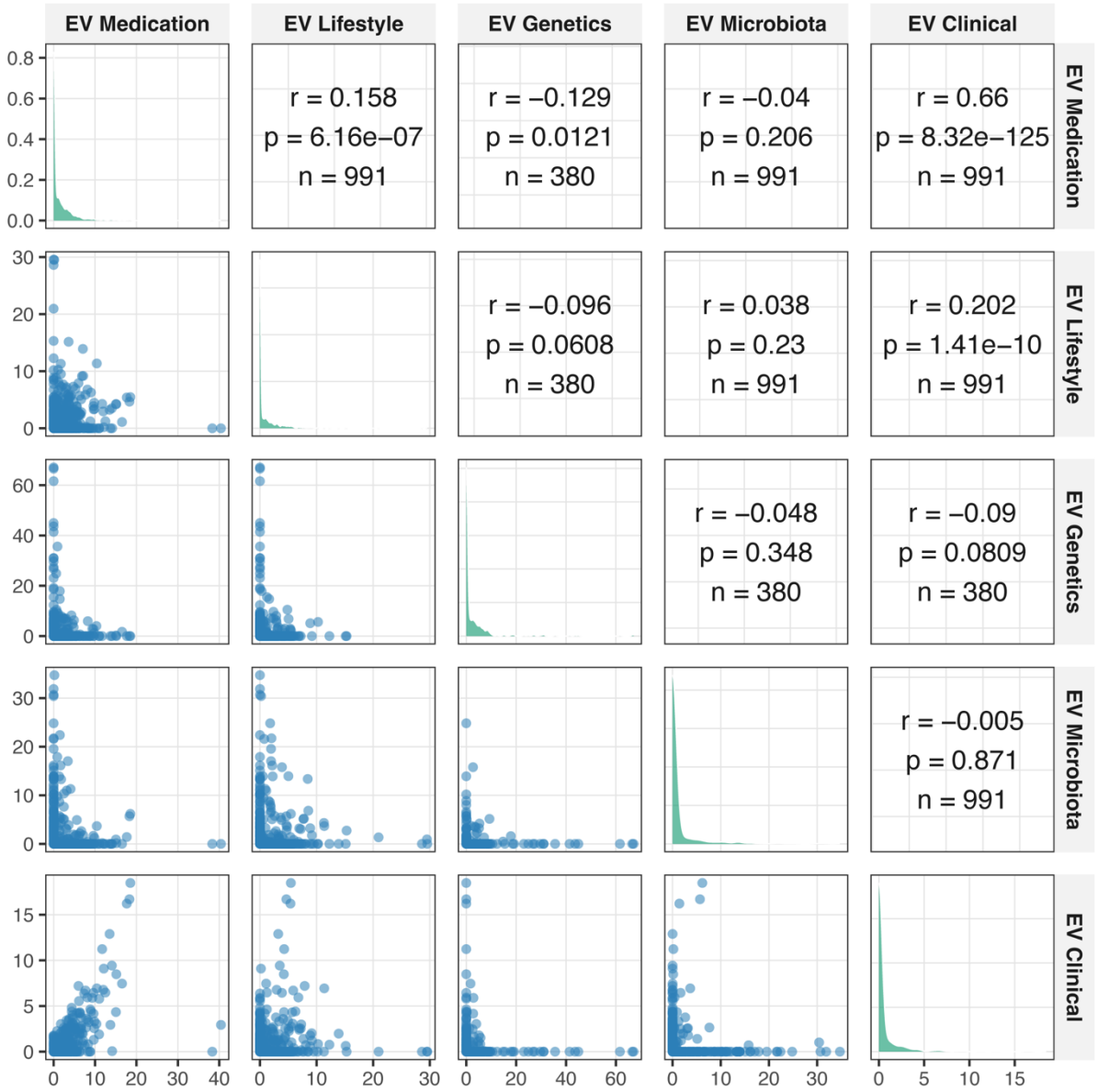
Supplementary Figures



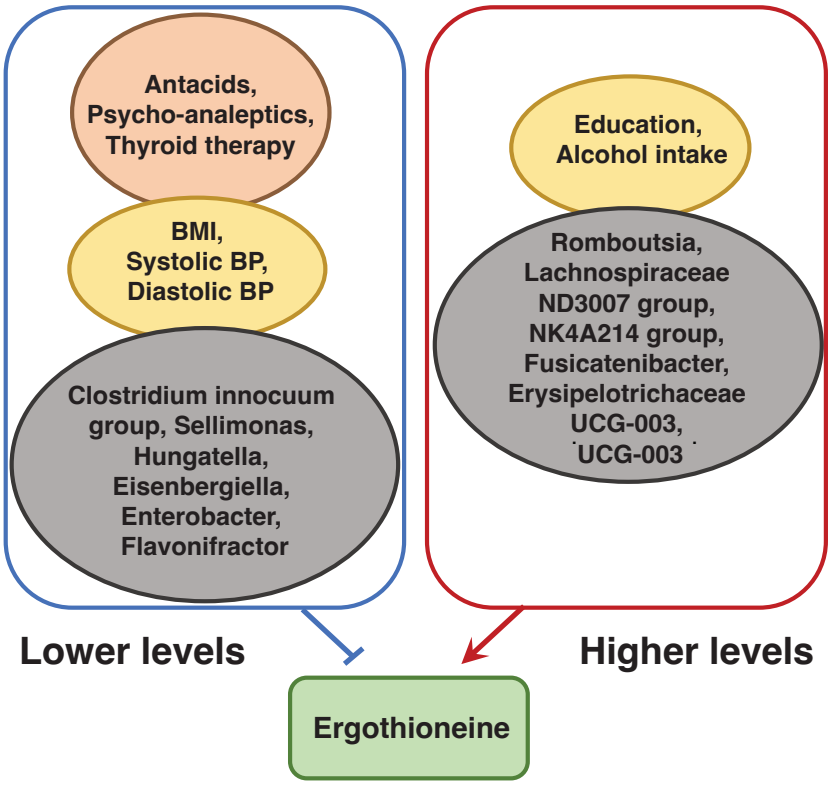
Supplementary Figure 1: Preprocessing of the metabolomics data and metabolite classes. (A) flowchart showing the steps performed during preprocessing of the metabolomics dataset. (B) Pi-chart showing the percentage of different metabolite classes covered in the metabolomics data set.



Supplementary Figure 2: Detailed flowchart of study design and analysis steps.



Supplementary Figure 3: Pairwise correlation between explained variance (EV) across different feature classes. For the correlation analysis, we only considered metabolites with explained variance (EV) > 0 and false discovery rate (*FDR*) of spearman correlation < 0.05 in the EV analysis. The Pearson correlation coefficients of the pairwise correlation between EVs, two-sided *P* values, number of observations in each test are provided. No multiple testing correction was performed for the correlation *P* values.



Supplementary Figure 4: Relationship of blood ergothioneine levels with microbiome, lifestyle and medication use. Overview of significant associations (false discovery rate (FDR) < 0.05) of ergothioneine blood levels with gut microbial, clinical, and lifestyle factors identified in this study.

The Alzheimer's Disease Metabolomics Consortium

Alexandra Kueider-Paisley, P. Murali Doraiswamy, Colette Blach, Arthur Moseley, Siamak Mahmoudiandehkhordi, Kathleen Welsh-Balmer, Brenda Plassman, Rima Kaddurah-Daouk.

Duke University, Durham, NC, USA

Kwangsik Nho, Shannon Risacher, Andrew Saykin.

Indiana University School of Medicine, Indianapolis, IN, USA

Matthias Arnold, Gabi Kastenmüller.

Helmholtz Zentrum München, Neuherberg, Germany

Xianlin Han.

University of Texas Health Science Center at San Antonio, San Antonio, TX, USA

Rebecca Baillie.

Rosa & Co., LLC, San Carlos, CA, USA

Pieter Dorrestein, James Brewer, Rob Knight.

University of California, San Diego, San Diego, CA, USA

Jennifer Labus, Pierre Baldi, Arpana Gupta, Emeran Mayer.

University of California, Los Angeles, Los Angeles, CA, USA

Oliver Fiehn.

West Coast Metabolomics Center, University of California, Davis, Davis, CA, USA

Dinesh Barupal.

Icahn School of Medicine at Mount Sinai, New York, NY, USA

Peter Meikle.

Baker Heart and Diabetes Institute, Melbourne, VIC, Australia

Sarkis Mazmanian.

California Institute of Technology, Pasadena, CA, USA

Leslie Shaw, Dan Rader.

University of Pennsylvania, Philadelphia, PA, USA

Najaf Amin, Alejo Nevado-Holgado, Cornelia van Duijn.

University of Oxford, Oxford, UK

Ranga Krishnan, Ali Keshavarzian, Robin Vogt, David Bennett.

Rush University, Chicago, IL, USA

Arfan Ikram.

Erasmus MC, Rotterdam, The Netherlands

Thomas Hankemeier.

Leiden University Metabolomics Center, Leiden, The Netherlands

Ines Thiele.

National University of Ireland – Galway, Galway, Ireland

Cory Funk.

Institute for Systems Biology, Seattle, WA, USA

Priyanka Baloni.

Purdue University, West Lafayette, IN, USA

Wei Jia.

University of Hawaii Cancer Center, Honolulu, HI, USA

David Wishart.

The Metabolomics Innovation Centre (TMIC), Edmonton, Alberta, Canada

Roberta Brinton.

University of Arizona, Tucson, AZ, USA

Lindsay Farrer, Wendy Qiu, Rhoda Au.

Boston University, Boston, MA, USA

Peter Würtz.

Nightingale Health, Helsinki, Finland

Therese Koal.

Biocrates Life Sciences AG, Innsbruck, Austria

Anna Greenwood.

Sage Bionetworks, Seattle, WA, USA

Jan Krumsiek.

Weill Medical College of Cornell, New York, NY, USA

Karsten Suhre.

Weill Cornell Medicine-Qatar, Doha, Qatar

John Newman.

USDA ARS, Washington, DC, USA

Ivan Hernandez.

SUNY Downstate, Brooklyn, NY, USA

Tatiana Foroud.

National Centralized Repository for Alzheimer's and other Dementias (NCRAD), Indianapolis, IN, USA

Frank Sacks.

Harvard School of Public Health (Harvard T.H. Chan School of Public Health), Boston, MA, USA