Title: Determinants of the plasma metabolome: cross-sectional and longitudinal associations over 6 years in the NESDA cohort

Authors: Dennis Klose[1,2*], Yuri Milaneschi[1,3], Laura K M Han[1,2,3], Nic J A van der Wee[4], Matthias Arnold[5,6], Christopher R Brydges[6], Siamak Mahmoudian Dehkordi[6], Gabi Kastenmüller[5], Rima Kaddurah Daouk[6,7,8], Brenda W J H Penninx[1,3]

Affiliations:

1 Amsterdam University Medical Center, Vrije Universiteit Amsterdam, Department of Psychiatry, Amsterdam, The Netherlands

2 Amsterdam University Medical Center, Amsterdam Public Health Research Institute, Personalized Medicine and Mental Health Programs, Amsterdam, The Netherlands

3 Amsterdam University Medical Center, Amsterdam Neuroscience Research Institute, Mood, Anxiety, Psychosis, Sleep & Stress Program and Complex Trait Genetics Program, Amsterdam, The Netherlands, The Netherlands

4 Leiden University Medical Center, Universiteit Leiden, Department of Psychiatry and Psychology, Leiden, The Netherlands

5 Institute of Computational Biology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany

6 Department of Psychiatry and Behavioral Sciences, Duke University, Durham, NC, USA

7 Duke Institute of Brain Sciences, Duke University, Durham, NC, USA

8 Department of Medicine, Duke University, Durham, NC, USA

* Corresponding author: d.klose@amsterdamumc.nl

Abstract

The plasma metabolome represents a valuable molecular readout of a person's physiological state, yet its relation to health, stress and lifestyle remains underexplored collectively. Here, we conducted an untargeted metabolomics analysis using 3804 paired samples from 1902 participants of the Netherlands Study of Depression and Anxiety at baseline and 6-year follow-up, quantifying 680 plasma metabolites. We characterised five metabolome profiles using principal component analysis, three with distinct biochemical enrichments related to transmembrane transport, sphingolipid and amino acid metabolism. Metabolite levels showed moderate intrapersonal correlation between baseline and 6-year follow-up (ICC_{median} = 0.482), and only 22% of metabolites showed standardized mean differences greater than 0.2, indicating minimal 6-year population-level change. Multivariate linear modelling on 18 determinants across demographics, psychosocial environment, lifestyle, somatic and mental health explained a maximum of 35% of baseline metabolome profile variance and 12.2% in 6-year changes. Demographic (e.g., sex, age), somatic health (e.g., BMI, medication) and lifestyle factors (e.g., smoking, alcohol intake) demonstrated strong associations, while psychosocial factors and mental health contributed minor explained variance. Altogether, our study provides novel insights into the cross-sectional and longitudinal implications of health, stress and lifestyle exposures for the plasma metabolome, contributing to the understanding of metabolomic signatures in population studies.

Introduction

The plasma metabolome represents the collection of exogenous chemicals and endogenous, small molecules originating from active metabolic processes, consisting of diverse compounds which facilitate energy uptake and storage, support cellular communication and maintain structural integrity (1). While several key influences on the blood metabolome composition have been identified in previous analyses using serum or plasma—including sex, age, genetics, diet, the gut microbiome, anthropometrics, circadian rhythm and diseases (2-6)—many sources of variation remain unexplored. Importantly, existing research has predominantly examined these influences in isolation, leaving a significant gap in our understanding of how multiple diverse factors contribute collectively. Metabolic alterations may play a fundamental role in the development of somatic and mental illnesses and can often be associated with environmental exposures such as medication, pollutants, lifestyle choices or psychological stressors (7-17). While the metabolome-wide influence of psychological stressors and mental health has been investigated in previous studies (3, 18), the role of more varied and detailed psychosocial factors has not been thoroughly investigated so far.

Here, we present the largest LC-MS-based study to date (n=1902) examining the associations of 18 determinants with 680 identified human plasma metabolites leveraging both cross-sectional and longitudinal data from 6-year follow-up assessments (FU6), overcoming limitations of previous studies pertaining to sample size, assessed determinants and timepoints. Our primary objective is to estimate the percentage of variance explained by determinants spanning five conceptual domains (demographics, psychosocial environment, lifestyle, somatic and mental health) in metabolite levels measured cross-sectionally or modelled as longitudinal change over 6 years. For this, we derive distinct metabolome profiles using principal component analysis and characterize them based on their potential enrichment in biochemical processes and subsequently use these to establish metabolome baseline scores

and metabolome change scores for linear modelling. Altogether our study gives insight into which characteristics and life experiences are embedded into the plasma metabolome on two timescales. These findings may provide the basis for future research to develop personalised lifestyle interventions targeting psychosocial wellbeing or lifestyle-related behaviours to optimise metabolic health conditions.

Results

Metabolome principal components differ in their biochemical composition - To characterise the metabolome profiles of our study participants, we first performed a principal component analysis (PCA) on 3804 paired samples from 1902 subjects and 680 metabolite levels (**Figure 1**, **Suppl. Table 1**). We observed a shift of the FU6 samples compared to the baseline samples in the dimensionality-reduced metabolome space (**Figure 2a**). To enhance the relevance and interpretability of our findings, we retained PC1 to PC5 because these together captured the substantial majority (27%) of total variance between samples (**Suppl. Figure 1a**). Inspection of the scree plot demonstrated a clear elbow at PC5, after which additional PCs only contributed minimally (< 3% per PC) to the total explained variance. Beyond PC5, incremental increases were negligible and did not meaningfully improve representation of the underlying data structure. To validate the underlying PC structure, we conducted a separate PCA on 910 held-out baseline samples for which no follow-up measurement was available (**Suppl. Table 2**). We correlated the 680 metabolite loading scores from the first PCA (discovery) with those from the second PCA (validation) and determined moderate to high Pearson's correlations for the respective loading scores (PC1: r = -0.94, PC2: r = 0.96, PC3: r = -0.94, PC4: r = -0.82, PC5: r = 0.58), confirming that the analysed dimensionality-reduced data captures meaningful and consistent information (**Suppl. Figure 1b**). Likewise, the explained variance per PC from the discovery PCA could be confirmed using the validation

samples (**Suppl. Figure 1a**). When considering change over 6 years, PC1 and PC3 showed the strongest change, although of small size, between the two timepoints with standardised mean differences (SMD) of -0.28 and 0.25 (**Suppl. Figure 1c**). After conducting a PCA test, 604 out of 680 metabolites (88%) significantly contributed to PC1 scores, whereas 257 (37%), 232 (34%), 235 (34%) and 180 (26%) metabolites significantly contributed to PC2,3,4 and 5 scores, respectively (**Suppl. Table 3**). Unique or overlapping metabolites from all five PCs are shown in **Suppl. Figure 1d**. In addition to the broad 6-year changes observable in PCA, we assessed the stability of metabolite levels between the two timepoints individually: Population-level changes higher than absolute SMD > 0.2 were not common, shown by only 22% of metabolites (**Figure 2b**), with moderate within-subject correlation (ICC_{median} = 0.482). Interestingly, we observed a strong reduction in per- and polyfluoroalkyl substances (PFAS) over 6 years and a striking increase in argininate, dibutyl sulfosuccinate and sarcosine. Dehydroepiandrosterone sulfate and pregnenolone sulfate (neurosteroids known to decrease with age) were also among the metabolites with the highest reduction over 6 years. For detailed metabolite level differences between timepoints see **Suppl. Figure 2** and **Suppl. Table 4**.

We next aimed to describe the significant metabolites per PC in terms of class and importance: While PC1 metabolites were a general reflection of the overall composition of measured metabolites, we further observed distinct metabolite proportions across the other components: PC2 was characterized by decreased lipids and cofactors/vitamins alongside increased amino acids and xenobiotics. PC3 showed an enrichment in lipids and xenobiotics but depletion in amino acids, peptides, and nucleotides. PC4 exhibited reduced amino acids yet elevated xenobiotics and cofactors/vitamins. Finally, PC5 displayed lower lipid content while showing higher levels of amino acids, peptides, and nucleotides. (**Figure 2c**). The top ten metabolites with the highest loading scores also varied per PC and are shown in **Figure**

2d. The highest loading metabolites were 2,3−dihydroxy−5−methylthio−4−pentenoate (DMTPA) for PC1, 5α-androstan-3α,17β-diol monosulfate for PC2, sphingomyelin (d17:1/14:0, d16:1/15:0) for PC3, hydroxy palmitoyl sphingomyelin (d18:1/16:0(OH)) for PC4, and 3-hydroxybutyroylglycine for PC5.

In summary, we reduced the dimensionality of each participant's metabolome to five principal component scores, which collectively explain 27% of the total population variance and represent specific biochemical profiles.

Metabolome PCs show distinct functional enrichments - Next, we investigated whether the described metabolome PCs represent concrete biological functions. Metabolite set enrichment analysis (MSEA) on the five sorted lists of metabolome PC loadings revealed that PC2 metabolites with high loadings were significantly enriched for “Transport of small molecules“ and “SLC-mediated transmembrane transport“ (protein-mediated transfer of molecules between the extra- and intracellular space) while PC3 and PC4 metabolites with high loadings respectively belonged to “Sphingolipid metabolism: integrated pathway” (involved in cell membrane formation, apoptosis and T-cell function) and “Metabolism of amino acids and derivatives“. MSEA revealed no significant results for PC1 or PC5 (**Figure 3a, Suppl. Figure 3a, Suppl. Table 3**). Since MSEA required use of PubChem IDs which were missing in 19-29% of cases per PC, we also used the sub-pathways assigned to each metabolite provided by Metabolon Inc. to retain all metabolites. We found high proportions (> 6%) of metabolites belonging to “Leucine, Isoleucine and Valine metabolism”, “Food component/plant”, and “Androgenic steroids” for PC2, “Sphingomyelins” and “Phosphatidylcholines” for PC3, and “Sphingomyelins”, “Food component/Plant” and “Benzoate metabolism” for PC4. Again, neither PC1 nor PC5 showed strong enrichments in any sub-pathway (**Suppl. Figure 3b**). We finally performed gene ontology analysis (GO) on all Human Metabolome Database (HMDB)

documented protein interaction partners of the metabolites per PC. Of note, 13-25% of metabolites could not be used for protein interaction partner retrieval due to missing HMDB IDs (**Suppl. Table 5**). Most PCs showed unique pathways not observed among the top 10 hits for the others (**Figure 3b, Suppl. Table 6**): PC2 metabolite protein interaction partners showed unique enrichments in "Arginine biosynthesis", "Biosynthesis of amino acids", "Biosynthesis of cofactors", "Lysine degradation" and "Pentose and glucuronate interconversions", while PC3, 4 and 5 showed unique enrichments for "Glutamatergic synapse", "Glycine, serine and threonine metabolism"/"Porphyrin metabolism" and "Glutathione metabolism", respectively.

Overall, this functional characterisation showed that metabolome profiles can be subdivided into principal components that capture metabolite sets representative of distinct biochemical processes. These components can therefore be used as informative outcomes for further linear modelling to assess determinants of the plasma metabolome.

Demographics and somatic health have the strongest associations with metabolome PCs, cross-sectionally and longitudinally - Having established the biological significance of metabolomic PC scores, we next aimed to identify baseline determinants of the baseline metabolome and of metabolomic changes over 6 years (**Table 1**). We extracted five metabolome baseline scores for PC1-5 and calculated five metabolome change scores termed ΔPC1-5, which were used as model outcomes.

First, we assessed whether any particular domain of determinants (demographics, psychosocial environment, lifestyle, somatic health and mental health) explained metabolome PCs or longitudinal changes better than the others. The demographic and somatic health domains explained most variance in PC1 and PC4, respectively (R^2 = 0.33 and 0.16, **Figure 4a**). The lifestyle domain explained most variance for PC1 (R^2 = 0.11). For the psychosocial

and mental health domain, we observed the highest R^2 of 0.08 and 0.06 also in PC1. The longitudinal analyses revealed generally weaker associations: We saw that the demographic domain also explained most additional variance for ΔPC1 scores (adj. R^2 = 0.11 in addition to the base model including only PC-baseline and shipment batch), while the somatic health domain explained most additional variance in ΔPC5 scores (additional adj. R^2 = 0.08). The lifestyle domain explained most additional variance for ΔPC4 scores (additional adj. R^2 = 0.01, respectively). For the psychosocial and the mental health domain, the highest additional R^2 of 0.012 and 0.002 were seen for ΔPC1/3 and ΔPC4, respectively. Overall, the highest explained variance was observed using the full model for PC1 (adj. R^2 = 0.35) and ΔPC1 (additional adj. R^2 = 0.12). An overview of all significant determinants per domain model and PC is shown in **Figure 4b**. In brief, all of the 18 chosen determinants showed significant associations across (Δ)PC1-5, with *anxiety diagnosis* being the only exception. An *anxiety diagnosis* showed nominally significant associations in the adjusted mental health models, but it did not withstand Benjamini-Hochberg (BH) correction.

As opposed to the separate domain models, the full models provide an overview of the independent associations of all 18 baseline determinants on PC-baseline or ΔPC scores (**Figure 5a,b**). In the following, the individual determinants' associations with metabolome PCs are listed in order of strong to weak effect sizes. *Sex* emerged as one of the most predictive determinants of metabolome PCs at baseline and for 6-year change, showing associations for PC1-4 and ΔPC1-3. Similarly, *age* was associated with PC1-5 and ΔPC1,2,3,5. The *number of medications* influenced PC2-5 scores and ΔPC2,4,5. *Alcohol intake* was reflected in PC1,3,4,5 and ΔPC4,5. *BMI* showed associations with PC1,4,5 and ΔPC4,5. *Current smoking* was associated with metabolites across PC1-2 and ΔPC5, while *former smoking* showed lasting influences on ΔPC5. *Years of education* were associated with PC4 and ΔPC1,5. *Daily hassles* were associated with PC2 and ΔPC1. *Degree of urbanisation*

was linked to PC4,5. *No high strain work* (versus *No work*) and *high strain work* showed associations with ΔPC2. *Leisure activity* and *reported childhood trauma* showed associations with PC4. *Number of somatic diseases* showed an association with ΔPC5. *Partner status*, *physical activity* and a *depression diagnosis* had only nominally significant associations with PC2, ΔPC4,5 and PC5/ΔPC3, respectively, which did not withstand BH correction (**Suppl. Table 7**). An *anxiety diagnosis* and *number of recent negative life events* had no nominally significant associations. All associations between the metabolome PCs, their functional enrichments and significant determinants from the cross-sectional and longitudinal modelling are listed as a summary in **Table 2**. Moreover, a PermANOVA of *age*, *sex*, *smoking status* and *BMI* on n=2803 baseline metabolome measurements revealed that those determinants alone explained 8.9% of variance in our sample ($p < 0.001$).

When we expanded our analysis from domain-specific to full models, we observed several changes. Most notably, for PC2, *daily hassles* remained significantly associated while other psychosocial factors became weaker associated or lost statistical significance. A similar attenuation was noted for *physical activity* and *number of recent negative life events* and all PC-baseline associations for *number of somatic diseases*.

Given the high influence of age, sex and BMI on metabolite levels as seen in the analysis above, additional domain model results adjusted individually for age, sex and BMI can be found in **Suppl. Table 7**. In general, subsequent results were highly similar to the full model results, indicating that the associations between certain determinants and metabolite PCs were attenuated by these important determinants.

Taken together, these results showed that baseline determinants explained a larger proportion of variance in PC-baseline scores as compared to ΔPC scores and that metabolome PC1-5 are differentially associated with the five domains. Across most metabolome PCs, the R^2 increased predominantly by adding information on somatic health and lifestyle beyond

demographic characteristics. We further observed that, while the main influences on the metabolome in our study pertained to sex, age, alcohol intake, BMI and medication, a wide variety of other factors contributed additional but subtler influences. These subtle influences were more abundant in domain models, when fewer mediators and confounders were present. Lastly, we observed that determinants linked to the baseline metabolome profile did not necessarily have significant associations with changes over 6 years and vice versa.

Discussion

In this analysis of the plasma metabolome, we examined five metabolome PCs derived from principal component analysis and their associations with participant characteristics across demographic, psychosocial, lifestyle, somatic and mental health domains. Principal component, metabolite set enrichment and Metabolon Inc. sub-pathway analysis revealed that metabolome PCs captured distinct biological processes. Most importantly, gene ontology analysis on metabolites' protein interaction partners reflected potential influences of specific exposures: PC2, primarily influenced by demographic factors (especially sex), showed enrichment in pentose and glucuronate interconversions linked to Uridine 5'-diphospho-glucuronosyltransferase (UGT) enzymes which mediate the inactivation of e.g. oestrogens via glucuronidation. In the domain model analysis, it was PC2 which had the most significant associations with psychosocial factors out of all five PCs, next to PC1. PC4, predominantly shaped by somatic health factors including medication use and especially BMI, was enriched for drug metabolism and porphyrin metabolism, an essential pathway for haemoglobin synthesis, and glycine-serine-and-threonine metabolism, a pathway which has been observed for maternal obesity in cord blood and general adiposity (19, 20). PC5 showed consistent associations with lifestyle factors and medication and was uniquely enriched for glutathione

metabolism, reflecting the antioxidant and detoxification demands associated with alcohol consumption, smoking and medication use (21).

We observed that our chosen determinants explained a limited amount of additional variance in metabolite change over time beyond that already captured by baseline metabolite levels. This is likely due to the restricted range of variability in 6-year change in our longitudinal analysis, which is consistent with previous studies, showing similar metabolomic stability over several years of follow-up (4, 22, 23). Strikingly, *age* influenced all metabolome change scores except for ΔPC4. Other studies also found that age had a strong influence on longitudinal metabolomic trajectories (24, 25). In fact, the robust association between *age* and the metabolome likely explains why *number of somatic diseases* and *number of recent negative life events*, both correlated with age though with different directionality, showed fewer significant independent associations in the full model. Notably, many other determinants in our study showed significant associations with longitudinal metabolomic change in our analysis, but these did not seem to follow a systematic pattern. *Education* is worth highlighting in this context, as it showed significant longitudinal associations with ΔPC5 in both the domain and full model. Due to its close link to socioeconomic status, this underlines the long-term potential of socioeconomic inequality to influence the metabolism (26, 27). When focusing on individual metabolites, we observed strong reductions in PFAS molecules over 6 years, potentially due to legislative efforts aimed at reducing the industrial use of these synthetic chemicals. Moreover, the neurosteroids dehydroepiandrosterone sulfate (DHEA-S) and the closely related pregnenolone sulfate (PS) were also among the most decreasing metabolites over 6 years. DHEA-S and PS have been investigated for their memory-enhancing and antidepressant effects and were shown to attenuate memory deficits in Alzheimer's disease induced by beta-amyloid β25-35 (28, 29). They have also been discussed

as potential therapeutics for several psychiatric conditions (30). Therefore, the observed reduction in DHEA-S and PS over 6 years in our cohort could be linked to detrimental aging phenotypes, although a prominent study has also revealed no beneficial effects of DHEA supplementation (31). Among the highest increasing compounds, we found argininate, dibutyl sulfosuccinate and sarcosine. The arginine metabolism is shown to be involved in cancer malignancy (32). Dibutyl (sodium) sulfosuccinate is a synthetic surfactant used in pharmaceuticals, detergents as well as personal care products (33). Contrastingly, sarcosine has previously been found to decrease with age in mice (34). Importantly, these results stem from a paired, but unadjusted comparison of baseline and FU6 data including all 1902 participants and might therefore differ if important covariates such as age, sex, BMI, smoking, medication use or health status are included.

Our findings further revealed that different metabolome PCs are differentially shaped by the assessed domains, reflecting the notion that exposures might leave specific rather than broad traces in the organism. The most pronounced effects were observable for *sex*, *age* (demographics), *medication*, *BMI* (somatic health), and lastly *alcohol consumption* and *smoking* (lifestyle). These associations align with previous studies and likely occur via interaction with established biological mechanisms (3-5), including the influence of sex hormones (35), aging and mortality-related metabolomic processes (25, 36, 37), microbial metabolite production (38-40) and xenobiotics exposure (41) which collectively shape the level of circulating metabolites. Our cross-sectional PermANOVA R^2 estimate of 8.9% for the combined effects of *age*, *sex*, *BMI*, and *smoking* on the metabolome exceeds previous reports by almost double, although this may partly be explained by differences in metabolite coverage and technical factors (4). Interestingly, *former smoking* demonstrated weaker associations with metabolome PCs than *current smoking*, suggesting the possibility of metabolic recovery

following smoking cessation. Furthermore, *leisure activity* had one significant association with PC4 even after adjusting for *alcohol intake*, *smoking*, *physical activity*, and *partner status*. This suggests an influence of spending free time outdoors, a finding that warrants structured follow-up analyses. While *leisure activity* and *physical activity* share some overlapping elements, leisure activity primarily reflects social activity, whereas physical activity mainly captures sports, exercise and movement. The consistent association of *degree of urbanisation* with PC4,5 in the demographic domain model and in the full model suggests that exposure to city environments might lead to a metabolic signature presumably shaped by airborne xenobiotics and toxins, but also lifestyle factors which differ between the rural and the urban population. Similar observations in plasma have been made by a recent Chinese study (42).

Psychological stress emerged as another contributor to metabolome profiles, with correlations observed across multiple stressors including *daily hassles*, *occupational stress*, and *childhood trauma*. These stress-related traces, while more subtle than those from other domains, may occur through multiple pathways including epigenetic modifications (43-46), altered microbial metabolite production (47), or behavioural adaptations affecting body composition and sleep patterns (48, 49). Influences of stress on the metabolome have been studied previously but none explicitly considered non-pathological occupational or daily stress in a large, representative cohort (12). Our findings on childhood trauma and its metabolomic signature recapitulate previous work from NESDA, which showed a dose-response relationship between childhood trauma and specific plasma metabolites (17). In line with previous studies (8, 50, 51) a depressive disorder diagnosis (MDD and/or dysthymia) was associated in the mental health domain model with metabolites captured in our study by PC5. This association was reduced and no longer statistically significant in the full model including

all determinants, which may suggest that the association between depression and metabolite levels could be largely explained by relevant factors included in the full model. Nevertheless, it is important to remark that estimates from the full model should be interpreted with caution due to potential overadjustment, as for each determinant all other factors may represent not only confounders to be taken into account but also mediators (e.g. unhealthy lifestyle as a consequence of depression) or colliders (e.g. chronic diseases determined by both depression and metabolic alterations) that should not be included in the model for proper interpretation.

This study represents the first large-scale investigation to simultaneously examine demographic, psychosocial, lifestyle, somatic and mental health determinants of the metabolome using both cross-sectional and longitudinal data. The comprehensive nature of the chosen determinants enables direct comparison of explained variance across domains and provides a basis for more targeted mechanistic investigations. However, several limitations warrant consideration. The reliance on PCA, while facilitating dimensional interpretation, necessarily involves loss of information compared to analyses on a metabolite-level. Additionally, HMDB protein interaction partners were used for gene ontology analysis, which also includes connections between proteins and metabolites that have not been validated. Furthermore, the ethnic, genetic and geographic homogeneity of our Dutch cohort may limit generalizability to other populations. Finally, untargeted metabolomic analyses are highly sensitive to laboratory protocols, which can introduce noise through anti-clotting agents, sample storage, shipment, and handling procedures (52). However, the large sample size and inclusion of shipment as a covariate should be sufficient to mitigate technical confounders in this study.

Future research should focus on disentangling the overlapping influences of the here presented determinants on the metabolome, which will help to better understand the

mechanisms that can either harm or protect metabolic health. Moreover, follow-up investigations should zoom in on individual metabolites to identify specific biomarkers underlying the associations presented here, especially exposure to medication/antidepressants, which usually have distinct metabolic effects which are hard to capture with principal components. Given its distinct metabolite class composition, PC2 metabolites emerged as particularly interesting due to their characteristic pathway enrichment, strong associations with sex and age and associations with psychosocial factors (including psychological stress) in the domain and full models, respectively. Comparative analysis of PC2 metabolites between both sexes could enhance our understanding of the molecular mechanisms underlying psychological stress exposures and inform targeted interventions with potential relevance for females, who are at higher risk of developing stress-related diseases such as depression (53, 54).

In conclusion, this study represents the largest epidemiological investigation to date examining a wide variety of determinants that influence metabolite levels in plasma cross-sectionally and longitudinally. Our approach extends previous research by investigating metabolome PCs rather than the metabolome as a whole, and by examining not only standard factors but also associations with psychosocial and lifestyle factors. Through a comprehensive analysis across five conceptual domains, we discovered that many previously unknown determinants beyond demographics shape the plasma metabolome. These findings open up new questions of how life exposures are embedded into the metabolome and how these could inform therapeutic interventions or prevention options for conditions involving metabolic dysregulation.

Methods

Study design and participants

Data are from the Netherlands Study of Depression and Anxiety (NESDA), an ongoing multi-centre, observational, naturalistic and longitudinal cohort study examining the course and consequences of depressive and anxiety disorders. A detailed description of the study rationale, design and methods is given elsewhere (55). Briefly, 2981 adults (18-65 years old) with or without a diagnosis of depression and/or anxiety disorder were recruited from community samples, primary care practices, and mental health organisations between 2004 and 2007. Detailed assessments were repeated after one, two, four and 6 years of follow-up. Exclusion criteria were being diagnosed with a current clinically overt psychiatric disorder other than depression and anxiety as well as lacking proficiency in Dutch. Ethical approval was obtained centrally from the Ethical Review Board of the Vrije Universiteit Medical Centre (reference number 2003/183) in Amsterdam, the Netherlands, as well as from the Ethics Review Boards of the other participating research centres. All participants provided written informed consent.

Metabolomic data generation and preprocessing

Plasma metabolite measurement and subsequent data pre-processing have been described elsewhere extensively (8). Data was analysed and pre-processed in line with protocols adopted by established international consortia (https://www.omicscience.org/), $\log_2$-transformed and winsorised. In brief, fasting EDTA plasma samples were collected in the morning at baseline and FU6 and were stored at -80°C. Samples were sent for analysis where metabolome profiles were assessed using the mass spectrometry-based untargeted HD4 platform from Metabolon Inc. (Durham, NC, USA), with paired samples being processed on the same plate within the same batch. Shipment was performed in two batches. In total, 4827

samples were measured, 2812 from baseline and 2015 from FU6. From that, 1902 study participants had available data across the two timepoints. From 820 measured metabolites in total, 139 unknown metabolites and the anti-clotting agent EDTA were excluded. Thus, 680 metabolites were available for analysis. Mutually exclusive labels provided by Metabolon Inc. were used to categorise metabolite classes: lipids, amino acids, xenobiotics, peptides, cofactors and vitamins, nucleotides, carbohydrates, partially characterised molecules and energy.

Clinical assessments and other measures at baseline

In total, 18 exposome factors (determinants) from five conceptual domains were selected based on their known general relevance for human metabolism, physiology and/or their influence on stress levels (42, 56-66). Of note, we did not observe concerning levels of multicollinearity across our chosen determinants (max. Pearson's r = 0.48 for N somatic diseases ~ N frequent medication), **Suppl. Figure 4**).

DEMOGRAPHIC: *Sex, age* and *years of education* were assessed via basic intake questionnaires at study begin. The *degree of urbanisation* was determined based on current zip code and population density information from the Central Bureau of Statistics of the Netherlands.

PSYCHOSOCIAL ENVIRONMENT: *Partner status* was assessed via a basic intake questionnaire at study begin. *Work related stress* was assessed using the Karasek questionnaire (67), with high strain work being defined as current work with self-reported high job demand scores (≥0.5) and low job control scores (≤0.5). Job demand refers to the expected pace of work, workload and work intensity whereas job control assesses self-perceived autonomy at the workplace. The *leisure activity score* is a summed score of five items from a questionnaire asking about

the current frequency of social free time activities such as going to a bar, museum, restaurant, association, amusement park, cinema, club or sports activity outside of home. The *daily hassles score* was derived using a questionnaire asking about daily life social stressors such as interpersonal difficulties or conflicts with friends, colleagues or family, high mental load, rejection, disappointment or financial struggles in the past month. The reported *childhood trauma score* was assessed using the Childhood Trauma Questionnaire from the NEMESIS study (68) and asked about traumatic experiences such as parental divorce, juvenile prison detention, foster family or child home placement, emotional neglect and psychological, physical or sexual abuse in the first 16 years of life. The *number of negative recent life events* was determined based on an abridged version of the Brugha questionnaire (69) which asks about the occurrence of adverse events in the past year, such as death or serious illness of loved ones, job loss, terminated relationship or friendship, contact with the police or court and other events.

LIFESTYLE: *Alcohol intake* was calculated based on data originating from the AUDIT Questionnaire and is expressed as drinks per week (70). *Smoking status* was determined based on data originating from the Fagerström Questionnaire and is expressed as never smoker, former smoker and current smoker (71). *Physical activity* was assessed using the International Physical Activity Questionnaire and is expressed as MET-minutes per week (72), capturing sports activity, movement and commuting.

SOMATIC HEALTH: We deduced the *number of somatic diseases* of participants based on a questionnaire asking about the following current, chronic conditions: asthma, heart disease, diabetes, stroke, arthritis, cancer, hypertension, intestinal disorders, liver disease, epilepsy, chronic fatigue, allergies, thyroid gland disease, injury in past year, head injury (ever), and

others. *Number of medications* was defined as drugs taken on ≥50% of the days in a current month for conditions falling under a WHO ATC classification. Dietary supplements and homeopathic medication were excluded. The top 10 most frequently used medications were paracetamol, ibuprofen, paroxetine, oxazepam, omeprazole, metoprolol, citalopram, venlafaxine, amiloride and simvastatin. For the frequency of ATC codes in our sample see **Suppl. Table 8**. *BMI* was determined as current body weight [kg] /height^2 [m].

MENTAL HEALTH: The "*depression diagnosis*" variable in this study encompasses CIDI diagnosed major depressive disorder or dysthymia in the past 6 months. The "*anxiety diagnosis*" encompasses CIDI diagnosed generalised anxiety disorder, agoraphobia, panic disorder or social phobia in the past 6 months.

Statistical analyses

If not stated otherwise, all analyses were conducted in R (version 4.3.3) in Microsoft VS Code. See **Suppl. Information** for further details on packages used for analysis and plotting (excl. dependencies).

INTRACLASS CORRELATION: We assessed intraclass correlations between baseline and FU6 for 680 metabolites to assess within-subject change using the ICC function from the R psych package. We reported ICC1.

PRINCIPAL COMPONENT ANALYSIS: We performed principal component analysis on a data matrix of 3804 samples from baseline and FU6 x 680 metabolites features using the prcomp function from base R. The scale and center arguments were set to TRUE. We calculated a

metabolome change score per PC for each sample as $\Delta PC = PC_{FU6} - PC_{Baseline}$. To perform a significance test for principal components themselves and the respective loadings for each principal component, we used the PCAtest function from the PCAtest R package (73). All PCs were significant, while only a subset of metabolites represented significant loadings for each PC. Significance was determined based on 100 permutations x 100 bootstraps of the original data matrix. For internal validation of the PCA results (scree plot and loadings) this was repeated for 2812 – 1902 = 910 additional baseline samples, including NESDA study participants and sibling samples.

METABOLITE IDS: For each metabolite, Metabolon Inc. provided the chemical name, its PubChem ID and its HMDB ID (74, 75). PubChem IDs had lower missingness than HMDB IDs and were thus used for metabolite set enrichment analysis (MSEA). For gene ontology analyses (GO), protein interaction partners of metabolites per PC1-5 were used as input and expressed as gene symbols. The HMDB API was used to derive those protein interaction partners, requiring HMDB IDs as queries. Therefore, different IDs were used compared to MSEA. In cases where multiple IDs were available per metabolite as provided by Metabolon Inc., the first one listed was chosen.

METABOLITE SET ENRICHMENT ANALYSIS: To identify enrichments of certain functional pathways along a metabolite list (ranked by metabolite loading score per PC from highest to lowest), we used the WebGestaltR function of the WebGestaltR package, setting enrichMethod to “GSEA“, organism to “hsapiens“, enrichDatabase to “pathway_RampDB_Metabolomics “, interestGeneType to “pubchem” and otherwise standard settings (76). All RampDB metabolites were considered as background. Higher

normalised enrichment scores indicate enrichment of metabolites closer to the top of the list, where the most important metabolites for the respective PC are found.

GENE ONTOLOGY ANALYSIS: To identify significant overlaps between metabolite protein interaction partners per PC and Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway sets, we used the ShinyGO v0.82 website https://bioinformatics.sdstate.edu/go/ (77) with "Use pathway DB for gene counts" set to TRUE, showing the top 10 enriched pathways with standard settings. KEGG was chosen to investigate general human biochemical pathways (78).

LINEAR REGRESSION: To model PC-baseline and ΔPC scores (outcomes) on baseline determinants (predictors), we used the lm function from base R. Shipment batch was included as covariate for both outcomes, the respective PC-baseline score was included as a covariate for all ΔPC models. Outcomes and all numerical predictors were scaled ("robust scaling" method, using the ratio of median and interquartile range). Degree of urbanisation was an ordinal variable. Smoking status and work-related stress were categorical variables. Sex, partner status, depression diagnosis and anxiety diagnosis were dichotomous variables. All other variables were numeric. To compare the models, R^2 coefficients of determination were adjusted for the number of predictors in the model. Sample sizes per model differed depending on missingness of data for some determinants. Sensitivity analyses restricted to participants with complete data (n=1648) yielded nearly identical results (correlations with original adj. R^2: Pearson's r = 0.999 for PC-baseline models, r = 0.995 for ΔPC models), confirming our findings' robustness. Nominal *P*-values were corrected using the BH method per linear model (79).

As part of the longitudinal analysis, we made use of a base model in addition to the five domain models, which included only the respective PC-baseline score and the shipment covariate. This ensures that we quantify the variance of metabolome change scores solely explained by baseline metabolome PCs and technical confounding to understand the additional explained variance coming from the determinants of interest in the other models. Therefore, "additional explained variance" refers to adj. R^2 from which the base model adj. R^2 has been subtracted.

PERMANOVA: To model the whole metabolite matrix on a set of determinants for comparison purposes to a previous publication, we used the adonis2 function from the vegan R package with 1000 permutations and the "euclidean" method. Since only baseline metabolomics data was assessed, we included samples without follow-up measurement, increasing sample size to n=2803 study participants of whom the respective data was collected.

(For methods used for Suppl. Figures see ***Suppl. Information****)*

Table 1 Cohort characteristics and determinant metrics.

Domain	Determinants			Missing
Demographic	Sex	Male	658 (34.6%)	0 (0%)
		Female	1244 (65.4%)	
	Age (years)		44 [18, 65]	0 (0%)
	Education (years)		12 [5, 18]	0 (0%)
	Degree of urbanisation	Not urbanised	104 (5.5%)	0 (0%)
		Hardly urbanised	143 (7.5%)	
		Moderately urbanised	287 (15.1%)	
		Strongly urbanised	245 (12.9%)	
		Extremely urbanised	1123 (59%)	
Psychosocial	Partner	No	570 (30%)	0 (0%)
		Yes	1332 (70%)	
	Work related stress	No work	430 (22.6%)	144 (7.6%)
		No high strain work	1121 (58.9%)	
		High strain work	207 (10.9%)	
	Leisure activity score		14 [5, 26]	24 (1.3%)
	Daily hassles score		30 [20, 64]	19 (1%)
	Childhood trauma score		0 [0, 8]	4 (0.2%)
	N recent negative life events		0 [0, 7]	0 (0%)
Lifestyle	Alcohol intake (drinks per week)		3.8 [0, 66]	17 (0.9%)
	Smoking status	Never smoker	561 (29.5%)	0 (0%)
		Former smoker	680 (35.8%)	
		Current smoker	661 (34.8%)	
	Physical activity (MET-minutes per week)		2856 [0, 19278]	111 (5.8%)
Somatic health	N somatic diseases		1 [0, 7]	0 (0%)
	N frequent medications		1 [0, 10]	0 (0%)
	BMI		24.5 [14.7, 55.8]	1 (0.1%)
Mental health	Depression	No diagnosis	1247 (65.6%)	0 (0%)
		Depression diagnosis	655 (34.4%)	
	Anxiety	No diagnosis	1166 (61.3%)	0 (0%)
		Anxiety diagnosis	736 (38.7%)	

Note: For categorical variables, values are N (%); for continuous variables, values are Median [Min, Max]

Table 2 Results summary.

Metabolome PC	**Metabolite class representation**	**MSEA results**	**GO results on protein interaction partners** (Top 3, ranked by fold enrichment)	**Best domain model for PC-baseline and ΔPC**	**Significant determinants** (Full model for PC-baseline score)	**Significant determinants** (Full model for ΔPC change score)
PC1 (604 metabolites)	General (similar to overall dataset composition)	n.s.	Alanine aspartate and glutamate metabolism Drug metabolism-cytochrome P450 Chemical carcinogenesis-DNA adducts	Demographic	Sex, Age Alcohol intake, Current smoking BMI	Sex, Age, Education Daily hassles
PC2 (257 metabolites)	Less lipids, more amino acids, more xenobiotics, less cofactors and vitamins than PC1	Transport of small molecules SLC-mediated transmembrane transport	Arginine biosynthesis Alanine aspartate and glutamate metabolism Ascorbate and aldarate metabolism	Demographic	Sex, Age Daily hassles Current smoking Medication	Sex, Age No high strain work, High strain work Medication
PC3 (232 metabolites)	More lipids, less amino acids, more xenobiotics, less peptides, less nucleotides than PC1	Sphingolipid metabolism: integrated pathway	Alanine aspartate and glutamate metabolism Sphingolipid metabolism Ether lipid metabolism	Demographic	Sex, Age Alcohol intake Medication	Sex, Age
PC4 (235 metabolites)	Less amino acids, more xenobiotics, more cofactors and vitamins than PC1	Metabolism of amino acids and derivatives	GPI-anchor biosynthesis Ascorbate and aldarate metabolism Ether lipid metabolism	Somatic health	Sex, Age, Urbanisation, Education Leisure activity, Childhood trauma Alcohol intake Medication, BMI	Alcohol intake Medication, BMI
PC5 (180 metabolites)	Less lipids, more amino acids, more peptides, more nucleotides than PC1	n.s.	Alanine aspartate and glutamate metabolism Linoleic acid metabolism GPI-anchor biosynthesis	Somatic health	Age, Urbanisation Alcohol intake Medication, BMI	Age, Education Alcohol intake, Former smoking, Current smoking Somatic diseases, Medication, BMI

Funding declaration

The infrastructure for the NESDA study (https://www.nesda.nl/) is funded through the Geestkracht program of the Netherlands Organisation for Health Research and Development (ZonMw, grant number 10-000-1002) and financial contributions by participating universities and mental health care organizations (VU University Medical Center, GGZ inGeest, Leiden University Medical Center, Leiden University, GGZ Rivierduinen, University Medical Center Groningen, University of Groningen, Lentis, GGZ Friesland, GGZ Drenthe, Rob Giel Onderzoekscentrum). This work was supported by the National Institute of Mental Health (Grant No. R01MH108348) and by a series of grants issued through National Institute on Aging (Grant Nos. U19AG063744 and U01AG061359 [to principal investigator, RKD]), which supported a large number of scientists working on metabolomics, neuropsychiatric disorders and acylcarnitines. RKD has also received funding from the NIA (Grant No. 1RF1AG058942, 1RF1AG057452, RF1AG051550, U01AG088562, RF1AG059093, and R01AG046171) and FNIH: #DAOU16AMPA. This funding enabled consortia that she leads including the Mood Disorder Precision Medicine Consortium (MDPMC), the Alzheimer's Disease Metabolomics Consortium (ADMC), and the Alzheimer Gut Microbiome Project (AGMP) that contributed to the metabolomic analyses of the NESDA sample used in this article and to acylcarnitine discoveries. A listing of AGMP Investigators can be found at https://alzheimergut.org/meet-the-team/. A listing of ADMC investigators can be found at: https://sites.duke.edu/adnimetab/team/. A listing of MDPMC investigators can be found at https://sites.duke.edu/mdpmc/files/2022/06/MDPMC-Members2022.pdf. The funders listed above had no role in the design or conduct of the study; the collection, management, analysis, or interpretation of the data; preparation, review, or approval of the article; or the decision to submit the article for publication. In addition, DK and BWJHP are supported by the research project Stress in Action – Innovation in the understanding and application of daily-life stress.

which is financially supported by the Dutch Research Council and the Dutch Ministry of Education, Culture and Science (NWO gravitation Grant No. 024.005.010). YM has received funding from Amsterdam UMC (Starter Grant Ronde 2), Amsterdam Neuroscience (PoC funding 2024–2026) and ImmunoMIND, funded by UK Research and Innovation as part of the UK national Mental Health Platform. LKMH was funded by the Veni award (Grant No. 09150162210201) from the Dutch Research Council (NWO).

Conflict of interest

Dr. Kaddurah-Daouk is an inventor on a series of patents on use of metabolomics for the diagnosis and treatment of central nervous system diseases and holds equity in Metabolon Inc., Chymia, and Metabosensor. Drs. Arnold and Kastenmüller are inventors of patents on applications of metabolomics in neurodegenerative/neuropsychiatric diseases and hold equity in Chymia LLC, which had no role in this work. The other authors declare no conflict of interest.

Acknowledgements

We thank all participants and their families for their time, data, and samples. Additionally, we thank the Amsterdam Psychiatry Department's HPC for technical support during the computational analyses and Rick Jansen for metabolomic data management. Generative AI (Anthropic Claude Sonnet 3.7/4.5) has been used to support coding.

Author contributions

DK: Conceptualization, Methodology, Formal analysis, Writing - Original Draft, Visualization; YM: Conceptualisation, Methodology, Writing – Review & Editing, Supervision; LKMH: Conceptualisation, Methodology, Writing – Review & Editing,

Supervision; NJAvdW: Writing – Review & Editing; MA: Writing – Review & Editing; CRB: Writing – Review & Editing; SM: Writing – Review & Editing; GK: Data generation, Writing – Review & Editing; RKD: Data generation, Writing – Review & Editing; BWJHP: Data collection and generation, Conceptualisation, Writing – Review & Editing, Supervision

Data and code availability

Code is available from Zenodo at https://doi.org/10.5281/zenodo.17869122.

The data used to support the findings of this study are available upon reasonable request from NESDA, Amsterdam: nesda@amsterdamumc.nl. Information on how to request the study data, including the data sharing policy, can be found at https://www.nesda.nl/nesda-english/.

References

1. Johnson CH, Ivanisevic J, Siuzdak G. Metabolomics: beyond biomarkers and towards mechanisms. Nature Reviews Molecular Cell Biology. 2016;17(7):451–9.
2. Asnicar F, Berry SE, Valdes AM, Nguyen LH, Piccinno G, Drew DA, et al. Microbiome connections with host metabolism and habitual diet from 1,098 deeply phenotyped individuals. Nature Medicine. 2021;27(2):321–32.
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. Chen L, Zhernakova DV, Kurilshikov A, Andreu-Sánchez S, Wang D, Augustijn HE, et al. Influence of the microbiome, diet and genetics on inter-individual variation in the human plasma metabolome. Nature Medicine. 2022;28(11):2333–43.
5. Navarro SL, Nagana Gowda GA, Bettcher LF, Pepin R, Nguyen N, Ellenberger M, et al. Demographic, Health and Lifestyle Factors Associated with the Metabolome in Older Women. Metabolites. 2023;13(4).
6. You J, Cui X-H, Chen Y-L, Wang Y-X, Li H-Y, Qiang Y-X, et al. Mapping the plasma metabolome to human health and disease in 274,241 adults. Nature Metabolism. 2025.
7. de Kluiver H, Jansen R, Penninx BWJH, Giltay EJ, Schoevers RA, Milaneschi Y. Metabolomics signatures of depression: the role of symptom profiles. Translational Psychiatry. 2023;13(1):198.
8. Jansen R, Milaneschi Y, Schranner D, Kastenmuller G, Arnold M, Han X, et al. The metabolome-wide signature of major depressive disorder. Molecular Psychiatry. 2024;29(12):3722–33.
9. Ginos BNR, de Crom TOE, Ghanbari M, Voortman T. Long-term air pollution exposure and the blood metabolome: The rotterdam study. Environmental Research. 2024;263:120131.
10. Liang D, Tang Z, Diver WR, Sarnat JA, Chow SS, Cheng H, et al. Metabolomics Signatures of Exposure to Ambient Air Pollution: A Large-Scale Metabolome-Wide Association Study in the Cancer Prevention Study-II Nutrition Cohort. Environ Sci Technol. 2025;59(1):212–23.
11. Altmaier E, Fobo G, Heier M, Thorand B, Meisinger C, Römisch-Margl W, et al. Metabolomics approach reveals effects of antihypertensives and lipid-lowering drugs on the human metabolism. Eur J Epidemiol. 2014;29(5):325–36.
12. Zhu Y, Jha SC, Shutta KH, Huang T, Balasubramanian R, Clish CB, et al. Psychological distress and metabolomic markers: A systematic review of posttraumatic stress disorder, anxiety, and subclinical distress. Neurosci Biobehav Rev. 2022;143:104954.

13. Zhang S, Wang Z, Wang Y, Zhu Y, Zhou Q, Jian X, et al. A metabolomic profile of biological aging in 250,341 individuals from the UK Biobank. Nat Commun. 2024;15(1):8081.
14. Tessier AJ, Wang F, Liang L, Wittenbecher C, Haslam DE, Eliassen AH, et al. Plasma metabolites of a healthy lifestyle in relation to mortality and longevity: Four prospective US cohort studies. Med. 2024;5(3):224–38.e5.
15. Pathmasiri W, Rushing BR, McRitchie S, Choudhari M, Du X, Smirnov A, et al. Untargeted metabolomics reveal signatures of a healthy lifestyle. Scientific Reports. 2024;14(1):13630.
16. Beuchel C, Becker S, Dittrich J, Kirsten H, Toenjes A, Stumvoll M, et al. Clinical and lifestyle related factors influencing whole blood metabolite levels – A comparative analysis of three large cohorts. Molecular Metabolism. 2019;29:76–85.
17. Souama C, Lamers F, Milaneschi Y, Jansen R, Vinkers CH, Giltay EJ, et al. The Metabolomic Signature of Childhood Trauma. Biol Psychiatry. 2025.
18. Zhu Y, Jha SC, Shutta KH, Huang T, Balasubramanian R, Clish CB, et al. Psychological distress and metabolomic markers: A systematic review of posttraumatic stress disorder, anxiety, and subclinical distress. Neuroscience & Biobehavioral Reviews. 2022;143:104954.
19. Palmnäs MSA, Kopciuk KA, Shaykhutdinov RA, Robson PJ, Mignault D, Rabasa-Lhoret R, et al. Serum Metabolomics of Activity Energy Expenditure and its Relation to Metabolic Syndrome and Obesity. Scientific Reports. 2018;8(1):3308.
20. Tao L, Li B, Du Y, Hung S, Garmire L. An ensemble method associates prepregnancy BMI and maternal ethnicity with key cord blood metabolomic changes in a multi-ethnic cohort from Hawaii. medRxiv. 2025:2025.08.14.25333702.
21. Meister A, Anderson ME. Glutathione. Annu Rev Biochem. 1983;52:711–60.
22. Lacruz ME, Kluttig A, Tiller D, Medenwald D, Giegling I, Rujescu D, et al. Instability of personal human metabotype is linked to all-cause mortality. Sci Rep. 2018;8(1):9810.
23. Yousri NA, Kastenmüller G, Gieger C, Shin SY, Erte I, Menni C, et al. Long term conservation of human metabolic phenotypes and link to heritability. Metabolomics. 2014;10(5):1005–17.
24. Darst BF, Koscik RL, Hogan KJ, Johnson SC, Engelman CD. Longitudinal plasma metabolomics of aging and sex. Aging (Albany NY). 2019;11(4):1262–82.
25. Mäkinen VP, Ala-Korpela M. Influence of age and sex on longitudinal metabolic profiles and body weight trajectories in the UK Biobank. Int J Epidemiol. 2024;53(3).
26. Bishehsari F, Drees M, Adnan D, Sharma D, Green S, Koshy J, et al. Multi-omics approach to socioeconomic disparity in metabolic syndrome reveals roles of diet and microbiome. Proteomics. 2023;23(19):e2300023.
27. Wang P, Gao X, Willett WC, Giovannucci EL. Socioeconomic Status, Diet, and Behavioral Factors and Cardiometabolic Diseases and Mortality. JAMA Netw Open. 2024;7(12):e2451837.
28. Maurice T, Su TP, Privat A. Sigma1 (sigma 1) receptor agonists and neurosteroids attenuate B25-35-amyloid peptide-induced amnesia in mice through a common mechanism. Neuroscience. 1998;83(2):413–28.
29. Reddy DS. Neurosteroids: endogenous role in the human brain and therapeutic potentials. Prog Brain Res. 2010;186:113–37.
30. Patterson R, Balan I, Morrow AL, Meltzer-Brody S. Novel neurosteroid therapeutics for post-partum depression: perspectives on clinical trials, program development, active research, and future directions. Neuropsychopharmacology. 2024;49(1):67–72.
31. Nair KS, Rizza RA, O'Brien P, Dhatariya K, Short KR, Nehra A, et al. DHEA in Elderly Women and DHEA or Testosterone in Elderly Men. New England Journal of Medicine. 2006;355(16):1647–59.
32. Al-Koussa H, El Mais N, Maalouf H, Abi-Habib R, El-Sibai M. Arginine deprivation: a potential therapeutic for cancer cell metastasis? A review. Cancer Cell International. 2020;20(1):150.
33. LookChem [Available from: https://www.lookchem.com/casno5144-51-4.html.
34. Walters RO, Arias E, Diaz A, Burgos ES, Guan F, Tiano S, et al. Sarcosine Is Uniquely Modulated by Aging and Dietary Restriction in Rodents and Humans. Cell Rep. 2018;25(3):663–76.e6.
35. Clayton GL, Borges MC, Lawlor DA. The impact of reproductive factors on the metabolic profile of females from menarche to menopause. Nat Commun. 2024;15(1):1103.
36. Bucaciuc Mracica T, Anghel A, Ion CF, Moraru CV, Tacutu R, Lazar GA. MetaboAge DB: a repository of known ageing-related changes in the human metabolome. Biogerontology. 2020;21(6):763–71.
37. Deelen J, Kettunen J, Fischer K, van der Spek A, Trompet S, Kastenmüller G, et al. A metabolic profile of all-cause mortality risk identified in an observational study of 44,168 individuals. Nat Commun. 2019;10(1):3346.

38. Hu X, Fan Y, Li H, Zhou R, Zhao X, Sun Y, et al. Impacts of Cigarette Smoking Status on Metabolomic and Gut Microbiota Profile in Male Patients With Coronary Artery Disease: A Multi-Omics Study. Front Cardiovasc Med. 2021;8:766739.
39. Koutromanos I, Legaki E, Gazouli M, Vasilopoulos E, Kouzoupis A, Tzavellas E. Gut microbiome in alcohol use disorder: Implications for health outcomes and therapeutic strategies-a literature review. World J Methodol. 2024;14(1):88519.
40. Lee HK, Kim NE, Shin CM, Oh TJ, Yoon H, Park YS, et al. Gut microbiome signature of metabolically healthy obese individuals according to anthropometric, metabolic and inflammatory parameters. Sci Rep. 2024;14(1):3449.
41. Gu F, Derkach A, Freedman ND, Landi MT, Albanes D, Weinstein SJ, et al. Cigarette smoking behaviour and blood metabolomics. International Journal of Epidemiology. 2015;45(5):1421–32.
42. Wang Y, Sha W, Wang H, Howard AG, Tsilimigras MCB, Zhang J, et al. Urbanization in China is associated with pronounced perturbation of plasma metabolites. Metabolomics. 2020;16(10):103.
43. Mellon SH, Bersani FS, Lindqvist D, Hammamieh R, Donohue D, Dean K, et al. Metabolomic analysis of male combat veterans with post traumatic stress disorder. PLoS One. 2019;14(3):e0213839.
44. Beaupere C, Liboz A, Fève B, Blondeau B, Guillemain G. Molecular Mechanisms of Glucocorticoid-Induced Insulin Resistance. Int J Mol Sci. 2021;22(2).
45. van Bodegom M, Homberg JR, Henckens M. Modulation of the Hypothalamic-Pituitary-Adrenal Axis by Early Life Stress Exposure. Front Cell Neurosci. 2017;11:87.
46. Yaribeygi H, Maleki M, Butler AE, Jamialahmadi T, Sahebkar A. Molecular mechanisms linking stress and insulin resistance. Excli j. 2022;21:317–34.
47. Ying L, Yuhao W, Yafang H, Jiao L, Lina D, Qinze S, et al. Chronic stress is associated with altered gut microbiota profile and relevant metabolites in adolescents. BMC Microbiology. 2025;25(1):423.
48. van den Berk-Clark C, Secrest S, Walls J, Hallberg E, Lustman PJ, Schneider FD, et al. Association between posttraumatic stress disorder and lack of exercise, poor diet, obesity, and co-occuring smoking: A systematic review and meta-analysis. Health Psychol. 2018;37(5):407–16.
49. Hirotsu C, Tufik S, Andersen ML. Interactions between sleep, stress, and metabolism: From physiological to pathological conditions. Sleep Sci. 2015;8(3):143–52.
50. Zacharias HU, Hertel J, Johar H, Pietzner M, Lukaschek K, Atasoy S, et al. A metabolome-wide association study in the general population reveals decreased levels of serum laurylcarnitine in people with depression. Molecular Psychiatry. 2021;26(12):7372–83.
51. Montanari S, Jansen R, Schranner D, Kastenmüller G, Arnold M, Janiri D, et al. Acylcarnitines metabolism in depression: association with diagnostic status, depression severity and symptom profile in the NESDA cohort. Translational Psychiatry. 2025;15(1):65.
52. Stevens VL, Hoover E, Wang Y, Zanetti KA. Pre-Analytical Factors that Affect Metabolite Stability in Human Urine, Plasma, and Serum: A Review. Metabolites. 2019;9(8).
53. Mengelkoch S, Slavich GM. Sex Differences in Stress Susceptibility as a Key Mechanism Underlying Depression Risk. Curr Psychiatry Rep. 2024;26(4):157–65.
54. Smith MD, Wesselbaum D. Global evidence on the prevalence of and risk factors associated with stress. Journal of Affective Disorders. 2025;374:179–83.
55. Penninx B, Eikelenboom M, Giltay EJ, van Hemert AM, Riese H, Schoevers RA, et al. Cohort profile of the longitudinal Netherlands Study of Depression and Anxiety (NESDA) on etiology, course and consequences of depressive and anxiety disorders. J Affect Disord. 2021;287:69–77.
56. Chandola T, Brunner E, Marmot M. Chronic stress at work and the metabolic syndrome: prospective study. Bmj. 2006;332(7540):521–5.
57. Kivimäki M, Bartolomucci A, Kawachi I. The multiple roles of life stress in metabolic disorders. Nat Rev Endocrinol. 2023;19(1):10–27.
58. Robinson O, Carter AR, Ala-Korpela M, Casas JP, Chaturvedi N, Engmann J, et al. Metabolic profiles of socio-economic position: a multi-cohort analysis. Int J Epidemiol. 2021;50(3):768–82.
59. Suzuki A, Poon L, Papadopoulos AS, Kumari V, Cleare AJ. Long term effects of childhood trauma on cortisol stress reactivity in adulthood and relationship to the occurrence of depression. Psychoneuroendocrinology. 2014;50:289–99.
60. Du D, Bruno R, Blizzard L, Venn A, Dwyer T, Smith KJ, et al. The metabolomic signatures of alcohol consumption in young adults. Eur J Prev Cardiol. 2020;27(8):840–9.
61. Xu T, Holzapfel C, Dong X, Bader E, Yu Z, Prehn C, et al. Effects of smoking and smoking cessation on human serum metabolite profile: results from the KORA cohort study. BMC Medicine. 2013;11(1):60.

62. Kelly RS, Kelly MP, Kelly P. Metabolomics, physical activity, exercise and health: A review of the current evidence. Biochim Biophys Acta Mol Basis Dis. 2020;1866(12):165936.
63. Jaagura M, Kronberg J, Reigo A, Aasmets O, Nikopensius T, Võsa U, et al. Comorbidities confound metabolomics studies of human disease. Scientific Reports. 2024;14(1):24810.
64. Vich Vila A, Collij V, Sanna S, Sinha T, Imhann F, Bourgonje AR, et al. Impact of commonly used drugs on the composition and metabolic function of the gut microbiota. Nat Commun. 2020;11(1):362.
65. Penninx BWJH, Lamers F, Jansen R, Berk M, Khandaker GM, De Picker L, et al. Immuno-metabolic depression: from concept to implementation. The Lancet Regional Health – Europe. 2025;48.
66. Humer E, Pieh C, Probst T. Metabolomic Biomarkers in Anxiety Disorders. Int J Mol Sci. 2020;21(13).
67. Karasek R. Job Content Questionnaire user's guide. Department of Work Environemnt. 1985.
68. de Graaf R, Bijl RV, Smit F, Vollebergh WA, Spijker J. Risk factors for 12-month comorbidity of mood, anxiety, and substance use disorders: findings from the Netherlands Mental Health Survey and Incidence Study. Am J Psychiatry. 2002;159(4):620–9.
69. Brugha T, Bebbington P, Tennant C, Hurry J. The List of Threatening Experiences: a subset of 12 life event categories with considerable long-term contextual threat. Psychol Med. 1985;15(1):189–94.
70. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption--II. Addiction. 1993;88(6):791–804.
71. Fagerström K-O. Measuring degree of physical dependence to tobacco smoking with reference to individualization of treatment. Addictive Behaviors. 1978;3(3):235–41.
72. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc. 2003;35(8):1381–95.
73. Camargo A. PCAtest: testing the statistical significance of Principal Component Analysis in R. PeerJ. 2022;10:e12967.
74. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem in 2021: new data content and improved web interfaces. Nucleic Acids Research. 2020;49(D1):D1388–D95.
75. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, et al. HMDB: the Human Metabolome Database. Nucleic Acids Research. 2007;35(suppl_1):D521–D6.
76. Elizarraras JM, Liao Y, Shi Z, Zhu Q, Pico AR, Zhang B. WebGestalt 2024: faster gene set analysis and new support for metabolomics and multi-omics. Nucleic Acids Res. 2024;52(W1):W415–w21.
77. Ge SX, Jung D, Yao R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics. 2019;36(8):2628–9.
78. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses and cellular organisms. Nucleic Acids Research. 2020;49(D1):D545–D51.
79. 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.

Figures

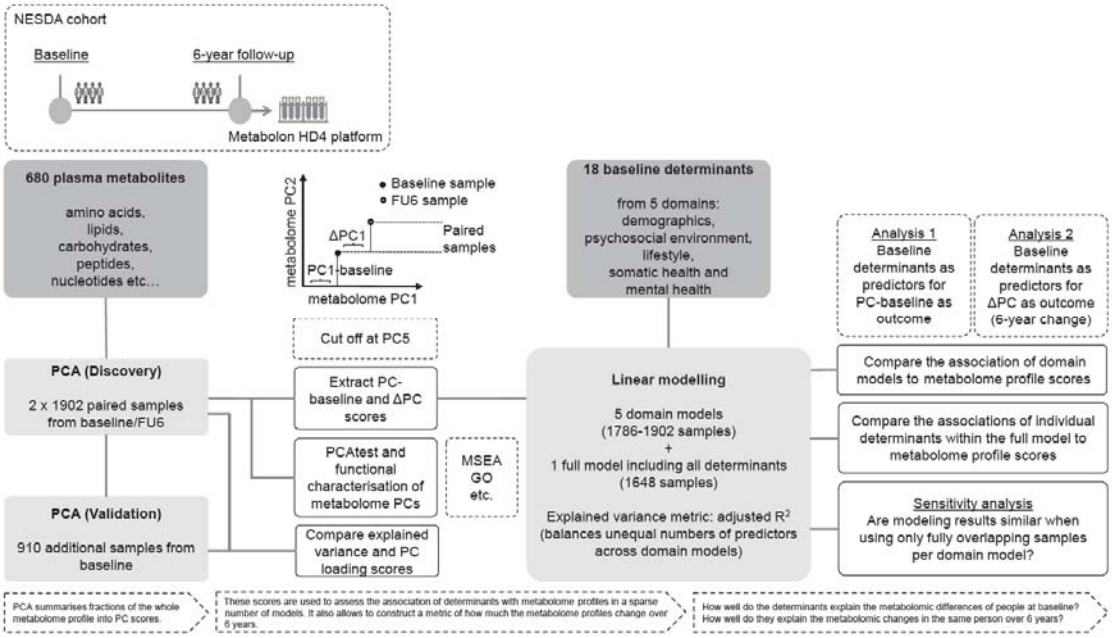


Figure legend 1 - Study design and workflow. PCA – Principal Component Analysis, FU6 – 6-year follow-up, MSEA – Metabolite Set Enrichment Analysis, GO – Gene ontology analysis.

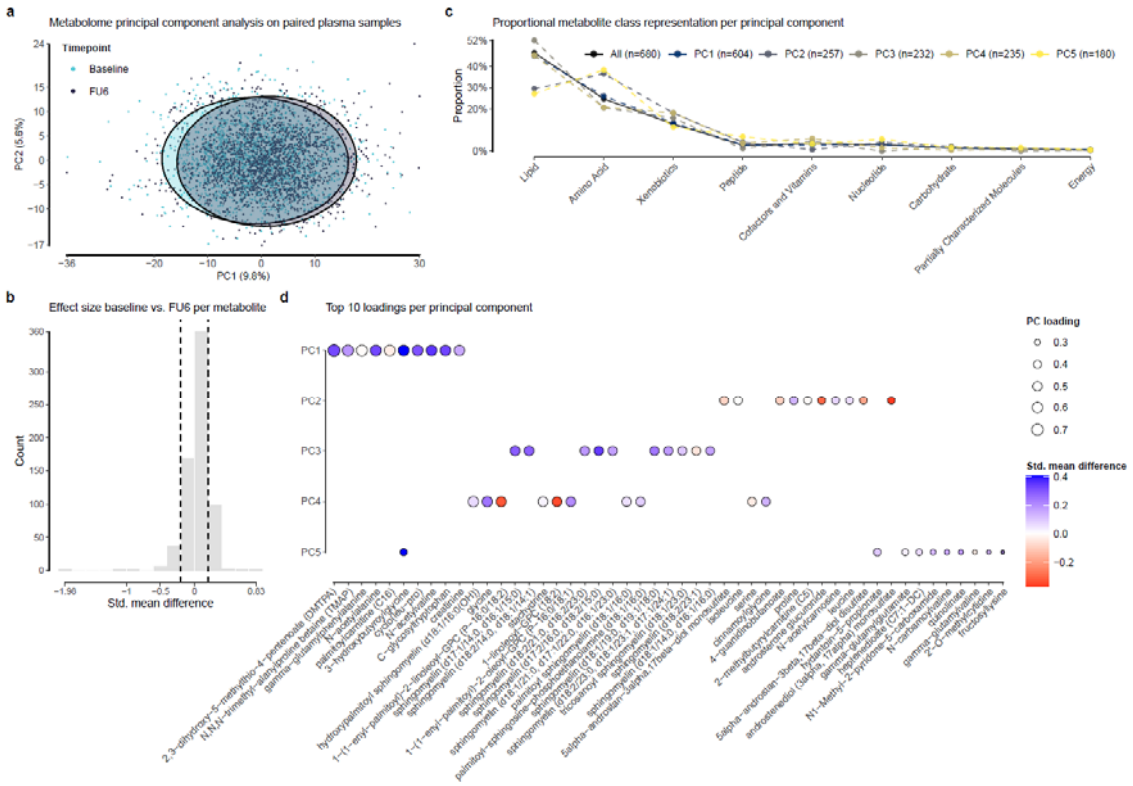


Figure legend 2 **-** Description of the first five metabolome PCs. **a** In a PCA plot using PC1 and PC2, the samples from baseline and 6-year follow-up (FU6) display a shifted metabolomic state (the two density circles do not fully overlap). **b** All 680 metabolites' standardized mean difference (SMD) for baseline versus FU6 shown as a histogram. Dashed lines mark SMD = ±0.2. Count refers to number of metabolites. **c** After a PCA test, the significant loadings (metabolites) per PC are shown categorised by metabolite class. While PC1 metabolites do not show different metabolite class proportions compared to all measured metabolites, PC2-5 display divergent proportions for several metabolite classes, especially lipids and amino acids. **d** Metabolites with highest loading scores per principal component. A higher loading score indicates more importance for the respective PC score.

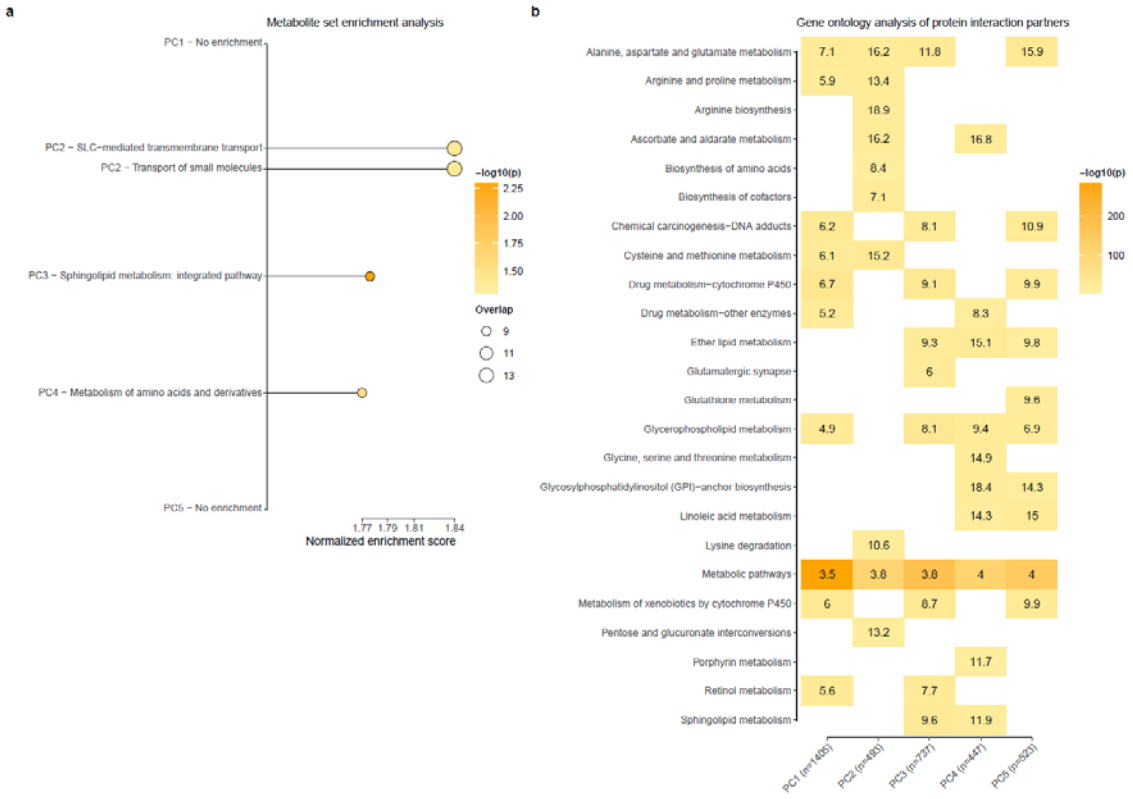


Figure legend 3 **-** Functional characterisation of the first five metabolome PCs. **a** Ranking all metabolites per PC by their loading score resulted in five lists used as input for metabolite set enrichment analysis (MSEA) in WebgestaltR. Here, all significant RampDB terms are shown. PC1 and 5 did not show any significant enrichment of metabolite pathways at the top of the ranked input list. The size of the circles refers to the number of overlapping metabolites from the input list and the respective RampDB set. **b** Results from GO analysis of all available HMDB-derived protein interaction partners of significant PC metabolites using ShinyGO. Shown are only the top 10 results per PC. Numbers in the grid signify fold enrichment scores (higher fold enrichment ~ higher coverage of KEGG set by protein input list). All reported *P*-values have been corrected for multiple testing by the used tools.

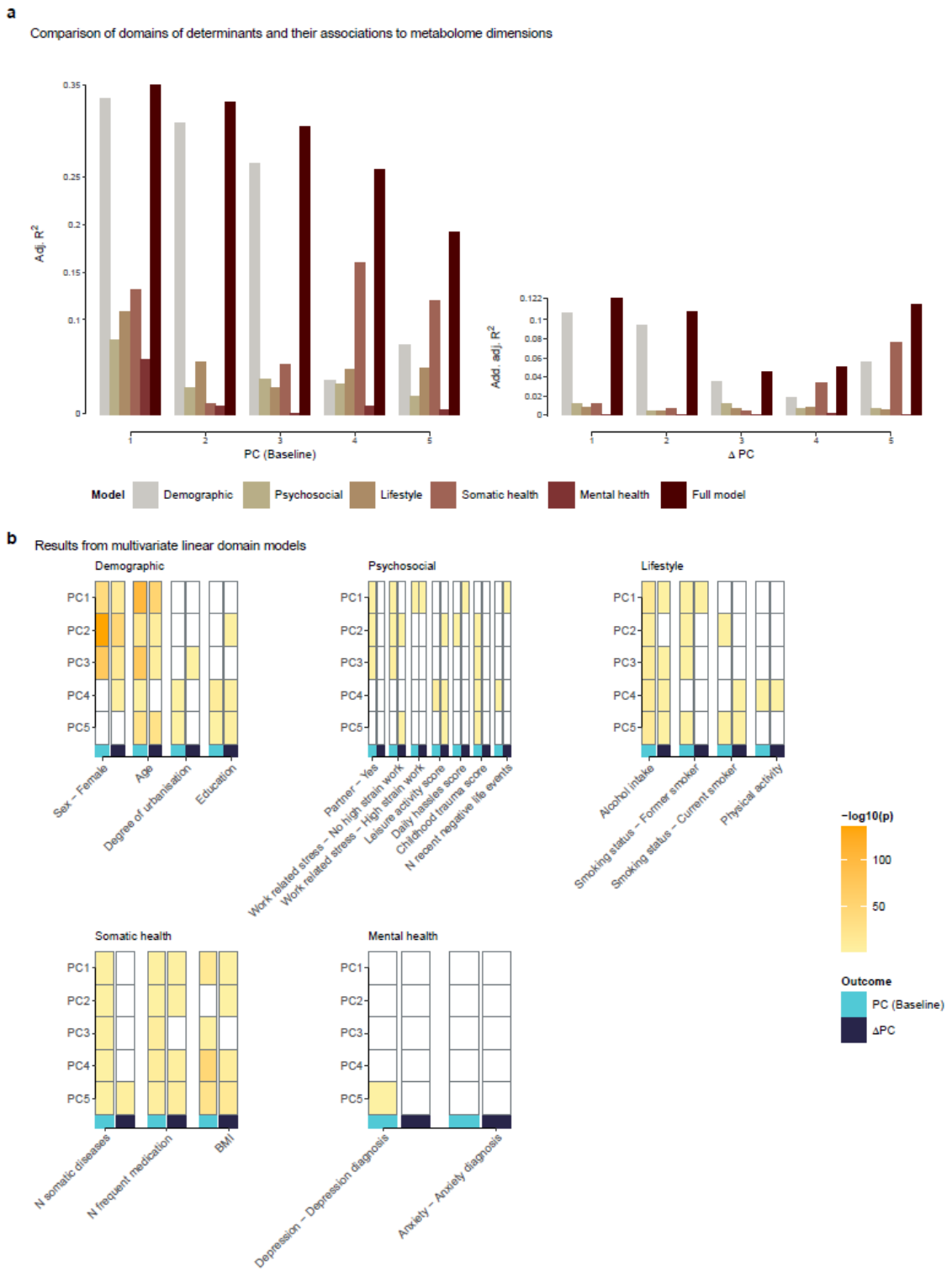


Figure legend 4 - Linear modelling results from the domain models. **a** Comparison of adjusted R^2 values from linear regression using the domain specific and full models with baseline determinants, with PC-baseline or ΔPC score as outcomes. **b** Visual overview of significant determinants per domain model. For detailed statistics see ***Suppl. Information***.

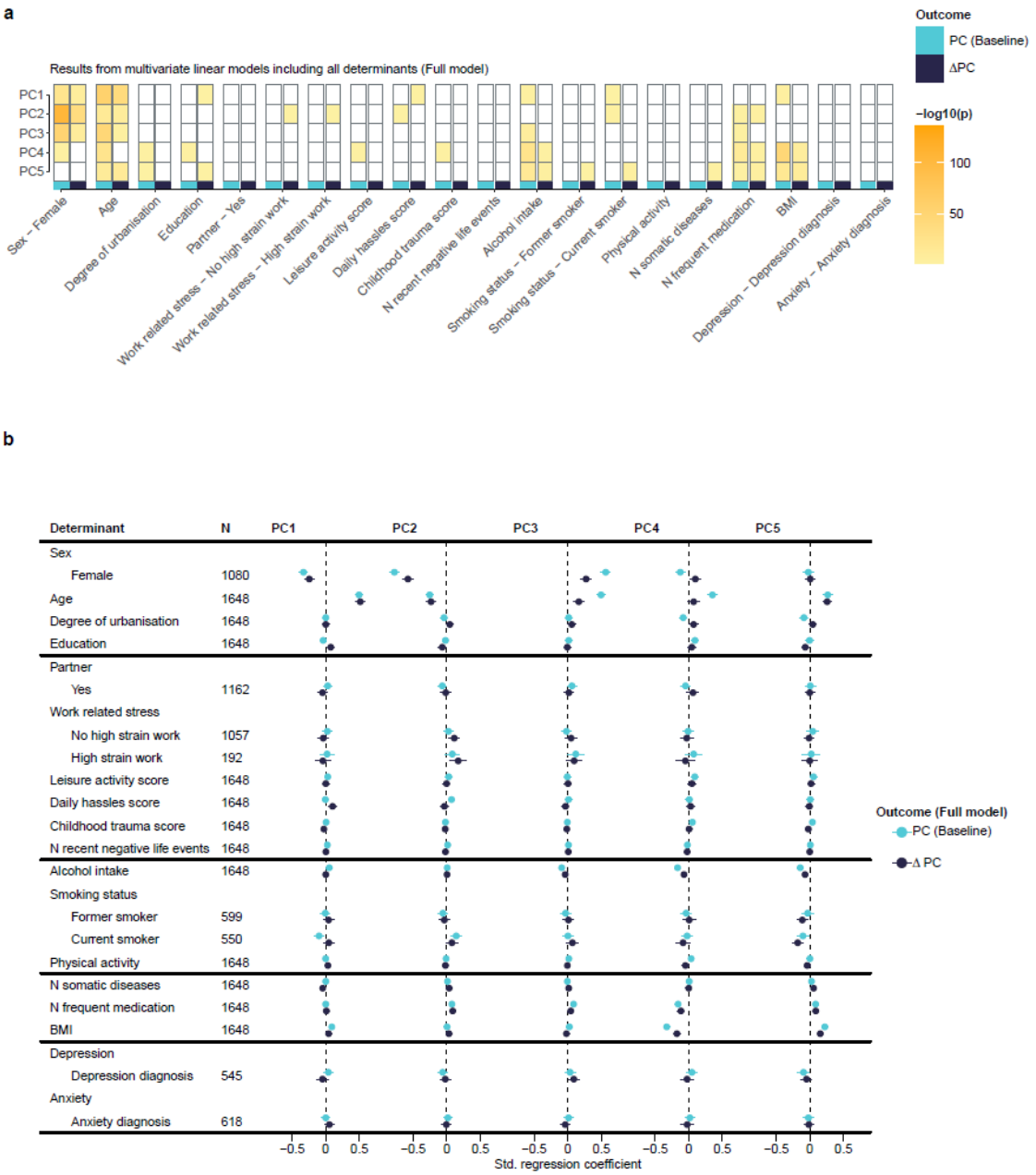
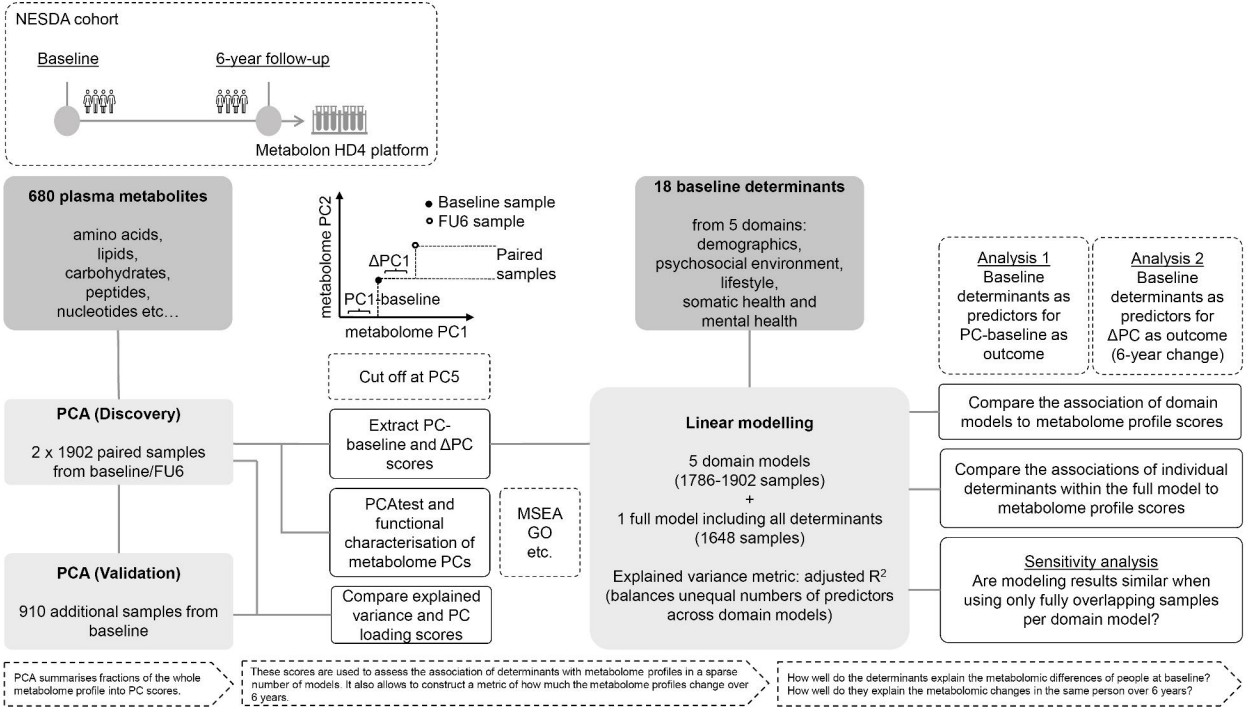
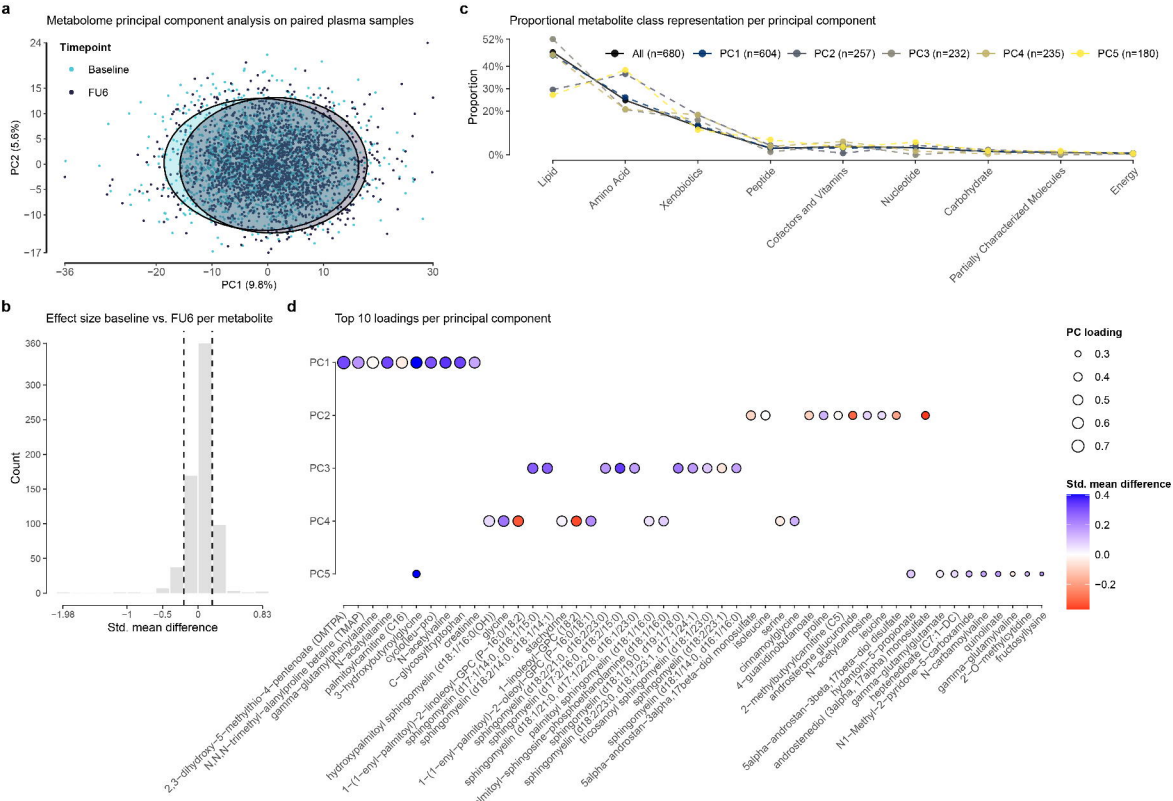
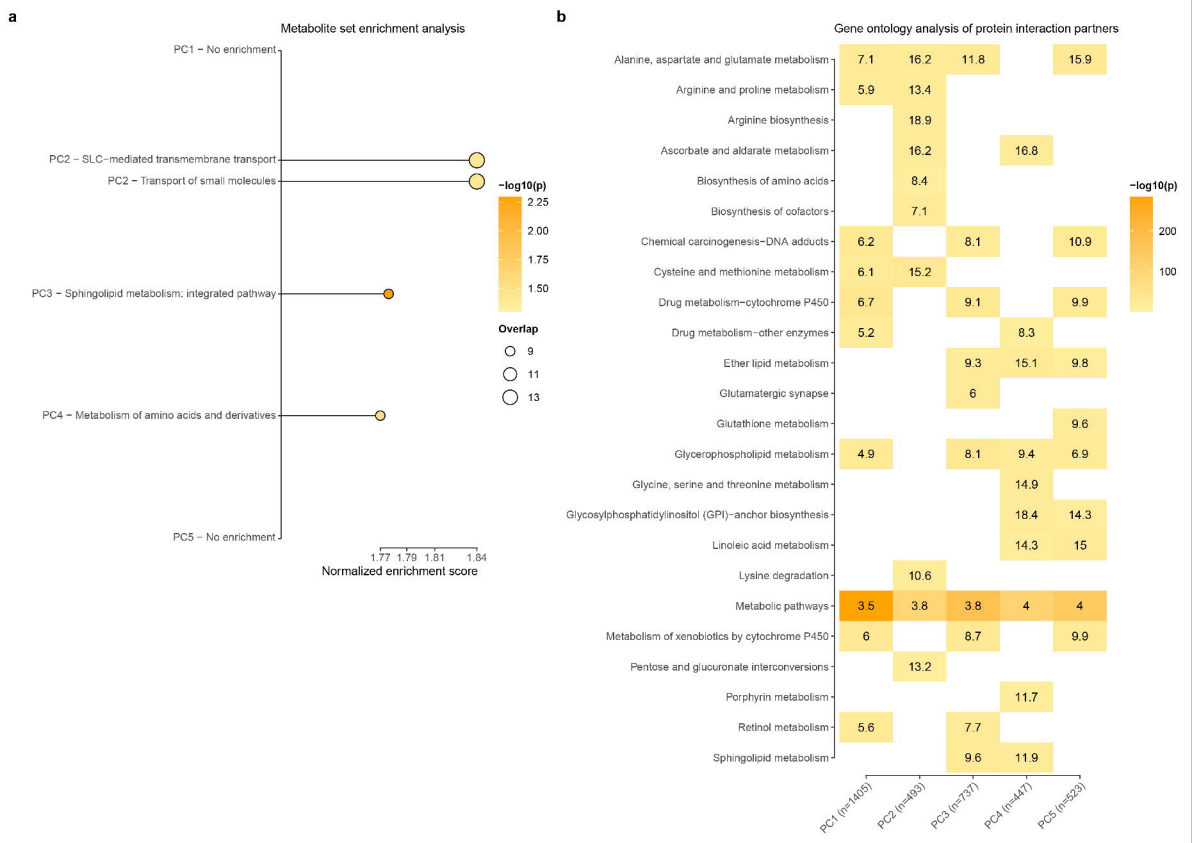


Figure legend 5 - Linear modelling results from the full models. **a** Visual overview of significant determinants. **b** Forest plot of standardized regression coefficients and 95% confidence intervals for all determinants in the full model. For detailed statistics see ***Suppl. Information***.

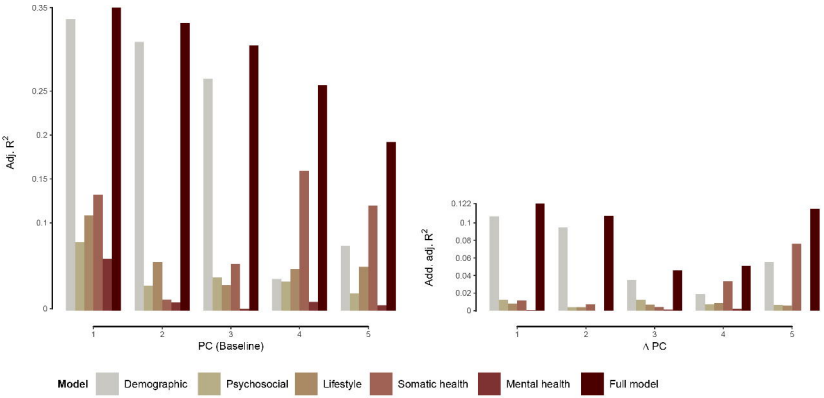

NESDA cohort
Baseline
6-year follow-up
Metabolon HD4 platform
680 plasma metabolites
amino acids,
lipids,
carbohydrates,
peptides,
nucleotides etc…
Baseline sample
FU6 sample
metabolome PC2
ΔPC1
Paired samples
PC1
baseline
metabolome PC1
18 baseline determinants
from 5 domains:
demographics,
psychosocial environment,
lifestyle,
somatic health and
mental health
Analysis 1
Baseline determinants as predictors for PC-baseline as outcome
Analysis 2
Baseline determinants as predictors for ΔPC as outcome (6-year change)
Cut off at PC5
PCA (Discovery)
2 x 1902 paired samples from baseline/FU6
Extract PC-baseline and ΔPC scores
Linear modelling
5 domain models
(1786-1902 samples)
+
1 full model including all determinants
(1648 samples)
Explained variance metric: adjusted R2
(balances unequal numbers of predictors across domain models)
Compare the association of domain models to metabolome profile scores
PCAtest and functional characterisation of metabolome PCs
MSEA
GO
etc.
Compare the associations of individual determinants within the full model to metabolome profile scores
PCA (Validation)
910 additional samples from baseline
Compare explained variance and PC loading scores
Sensitivity analysis
Are modeling results similar when using only fully overlapping samples per domain model?
PCA summarises fractions of the whole metabolome profile into PC scores.
These scores are used to assess the association of determinants with metabolome profiles in a sparse number of models. It also allows to construct a metric of how much the metabolome profiles change over 6 years.
How well do the determinants explain the metabolomic differences of people at baseline?
How well do they explain the metabolomic changes in the same person over 6 years?


a
Metabolome principal component analysis on paired plasma samples
Timepoint
Baseline
FU6
PC2 (5.6%)
PC1 (9.8%)
b
Effect size baseline vs. FU6 per metabolite
Count
Std. mean difference
c
Proportional metabolite class representation per principal component
All (n=680)
PC1 (n=604)
PC2 (n=257)
PC3 (n=232)
PC4 (n=235)
PC5 (n=180)
Proportion
Lipid
Amino Acid
Xenobiotics
Peptide
Cofactors and Vitamins
Nucleotide
Carbohydrate
Partially Characterized Molecules
Energy
d
Top 10 loadings per principal component
PC1
PC2
PC3
PC4
PC5
PC loading
Std. mean difference
2,3-dihydroxy-5-methylthio-4-pentenoate (DMTPA)
N,N,N-trimethyl-alanylproline betaine (TMAP)
gamma-glutamylphenylalanine
N-acetylalanine
palmitoylcarnitine (C16)
3-hydroxybutyroylglycine
cyclotelin-pro
N-acetylvaline
C-glycosyltryptophan
creatinine
hydroxypalmitoyl sphingomyelin (d18:1/16:0(OH))
glycine
1-(1-enyl-palmitoyl)-2-linoleoyl-GPC (P-16:0/18:2)
sphingomyelin (d17:1/14:0, d16:1/15:0)
sphingomyelin (d18:2/14:0, d18:1/14:1)
1-linoleoyl-GPC (P-16:0/18:2)
sphingomyelin (d18:2/21:0, d18:2/21:0)
sphingomyelin (d17:2/16:0, d18:2/15:0)
sphingomyelin (d18:1/21:0, d17:1/22:0, d16:1/23:0)
palmitoyl sphingomyelin (d18:1/16:0)
sphingomyelin (d18:1/18:0, d19:1/18:0)
palmitoyl-sphingosine-phosphoethanolamine (d18:1/16:0)
sphingomyelin (d18:2/23:0, d18:1/23:1)
sphingomyelin (d18:1/23:1)
sphingomyelin (d18:1/24:1)
sphingomyelin (d18:1/23:0)
sphingomyelin (d18:2/23:1)
sphingomyelin (d18:1/14:0, d16:1/16:0)
tricosanoyl sphingomyelin (d18:1/23:0)
sphingomyelin (d18:2/23:1)
sphingomyelin (d18:1/14:0, d16:1/16:0)
5alpha-androstan-3alpha,17beta-diol monosulfate
isoleucine
serine
cinnamoylglycine
4-guanidinobutanoate
proline
2-methylbutyrylcarnitine (C5)
androsterone glucuronide
N-acetylcarnosine
leucine
5alpha-androstan-3beta,17beta-diol disulfate
androstenediol (3alpha,17alpha) monosulfate
gamma-glutamylglutamate
heptenedioate (C7:1-DC)
N1-Methyl-2-pyridone-5-carboxamide
N-carbamoylvaline
glutamylvaline
gamma-glutamylvaline
2-O-methylcytidine
fructosylvaline


a
Metabolite set enrichment analysis
PC1 – No enrichment
PC2 – SLC–mediated transmembrane transport
PC2 – Transport of small molecules
PC3 – Sphingolipid metabolism: integrated pathway
PC4 – Metabolism of amino acids and derivatives
PC5 – No enrichment
1.77 1.79 1.81 1.84
Normalized enrichment score
−log10(p)
2.25
2.00
1.75
1.50
Overlap
9
11
13
b
Gene ontology analysis of protein interaction partners
Alanine, aspartate and glutamate metabolism
Arginine and proline metabolism
Arginine biosynthesis
Ascorbate and aldarate metabolism
Biosynthesis of amino acids
Biosynthesis of cofactors
Chemical carcinogenesis–DNA adducts
Cysteine and methionine metabolism
Drug metabolism–cytochrome P450
Drug metabolism–other enzymes
Ether lipid metabolism
Glutamatergic synapse
Glutathione metabolism
Glycerophospholipid metabolism
Glycine, serine and threonine metabolism
Glycosylphosphatidylinositol (GPI)–anchor biosynthesis
Linoleic acid metabolism
Lysine degradation
Metabolic pathways
Metabolism of xenobiotics by cytochrome P450
Pentose and glucuronate interconversions
Porphyrin metabolism
Retinol metabolism
Sphingolipid metabolism
PC1 (n=1405)
PC2 (n=493)
PC3 (n=737)
PC4 (n=447)
PC5 (n=523)
−log10(p)
200
100

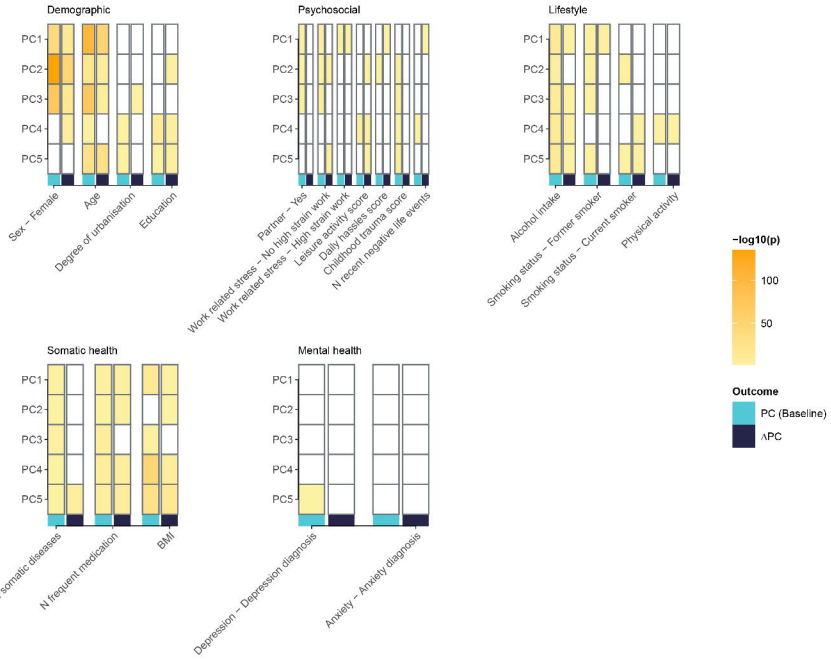
a

Comparison of domains of determinants and their associations to metabolome dimensions

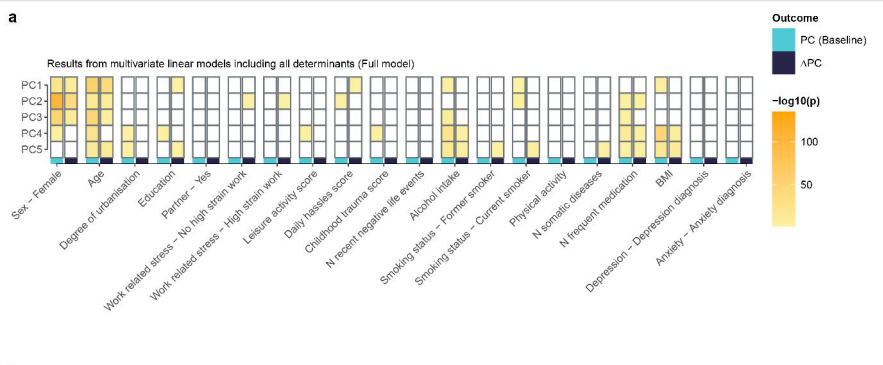


b

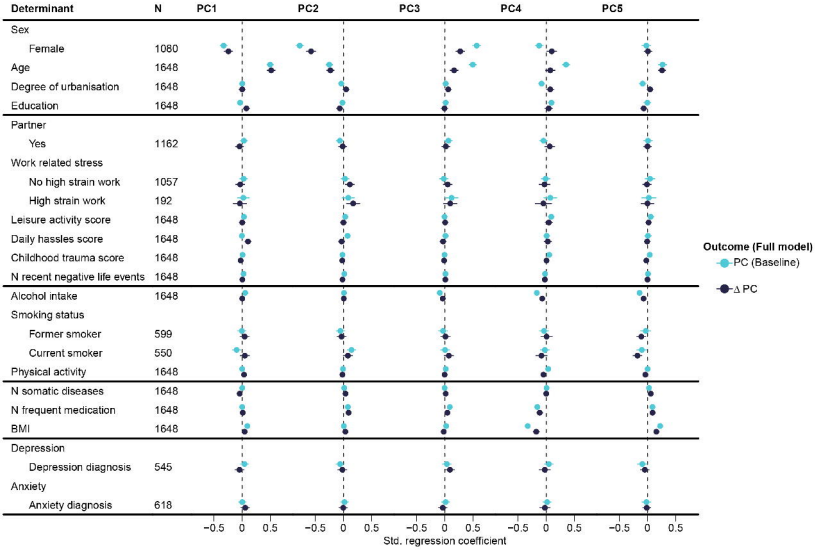
Results from multivariate linear domain models



a


Results from multivariate linear models including all determinants (Full model)
PC1
PC2
PC3
PC4
PC5
Sex - Female
Age
Degree of urbanisation
Education
Partner - Yes
Work related stress - No high strain work
Work related stress - High strain work
Leisure activity score
Daily hassles score
Childhood trauma score
N recent negative life events
Alcohol intake
Smoking status - Former smoker
Smoking status - Current smoker
Physical activity
N somatic diseases
N frequent medication
BMI
Depression - Depression diagnosis
Anxiety - Anxiety diagnosis
Outcome
PC (Baseline)
ΔPC
−log10(p)
100
50

b


Determinant
N
PC1
PC2
PC3
PC4
PC5
Sex
Female 1080
Age 1648
Degree of urbanisation 1648
Education 1648
Partner
Yes 1162
Work related stress
No high strain work 1057
High strain work 192
Leisure activity score 1648
Daily hassles score 1648
Childhood trauma score 1648
N recent negative life events 1648
Alcohol intake 1648
Smoking status
Former smoker 599
Current smoker 550
Physical activity 1648
N somatic diseases 1648
N frequent medication 1648
BMI 1648
Depression
Depression diagnosis 545
Anxiety
Anxiety diagnosis 618
−0.5 0 0.5
Std. regression coefficient
Outcome (Full model)
PC (Baseline)
Δ PC