1	Longitudinal analysis of genetic and environmental interplay in
2	human metabolic profiles and the implication for metabolic
3	health
4	Jing Wang ^{1#} , Alberto Zenere ^{1#} , Xingyue Wang ^{1#} , Göran Bergström ² , Fredrik Edfors ³ , Mathias Uhlén ³ ,
5	Wen Zhong ^{1*}
6	
7	¹ Science for Life Laboratory, Department of Biomedical and Clinical Sciences (BKV), Linköping
8	University, Linköping, Sweden.
9	² Department of Molecular and Clinical Medicine, Sahlgrenska Academy, University of Gothenburg,
10	and Clinical Physiology, Sahlgrenska University Hospital, Gothenburg, Sweden
11	³ Department of Protein Science, KTH - Royal Institute of Technology, Stockholm, Sweden
12	
13	
14	* To whom correspondence should be addressed: Wen Zhong
15	Address: Department of Biomedical and Clinical Sciences (BKV), Linköping University, SE-581 83,
16	Linköping
17	Email: wen.zhong@liu.se
18	Phone: + 46 729419808
19	# Equal contribution; these authors share first authorship.
20	

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

21 Abstract

22 Background

23 Understanding how genetics and environmental factors shape human metabolic profiles is crucial for

24 advancing metabolic health. Variability in metabolic profiles, influenced by genetic makeup, lifestyle,

and environmental exposures, plays a critical role in disease susceptibility and progression.

26 Methods

We conducted a two-year longitudinal study involving 101 clinically healthy individuals aged 50 to 65, integrating genomics, metabolomics, lipidomics, proteomics, clinical measurements, and lifestyle questionnaire data from repeat sampling. We evaluated the influence of both external and internal factors, including genetic predispositions, lifestyle factors, and physiological conditions, on individual metabolic profiles. Additionally, we developed an integrative metabolite-protein network to analyze protein-metabolite associations under both genetic and environmental regulations.

33 Results

Our findings highlighted the significant role of genetics in determining metabolic variability, identifying 34 35 22 plasma metabolites as genetically predetermined. Environmental factors such as seasonal variation, 36 weight management, smoking, and stress also significantly influenced metabolite levels. The integrative 37 metabolite-protein network comprised 5,649 significant protein-metabolite pairs and identified 87 38 causal metabolite-protein associations under genetic regulation, validated by showing a high replication 39 rate in an independent cohort. This network revealed stable and unique protein-metabolite profiles for 40 each individual, emphasizing metabolic individuality. Notably, our results demonstrated the importance 41 of plasma proteins in capturing individualized metabolic variabilities. Key proteins representing 42 individual metabolic profiles were identified and validated in the UK Biobank, showing great potential 43 for predicting metabolic diseases and metabolic risk assessment.

44 Conclusions

Our study provides longitudinal insights into how genetic and environmental factors shape human
 metabolic profiles, revealing unique and stable individual metabolic profiles. Plasma proteins emerged

as key indicators for capturing the variability in human metabolism and assessing metabolic risks. These
findings offer valuable tools for personalized medicine and the development of diagnostics for
metabolic diseases.

50 Keywords

51 Human metabolism, genetics, proteomics, metabolomics, lifestyle, environment, metabolic risk

52

53 Background

54 The human metabolome is dynamic, and the variability in human metabolic profiles across individuals is shaped by each person's unique genetic makeup, lifestyle, and environmental exposures^{1,2}. These 55 56 factors play critical roles in disease susceptibility and progression, including obesity, diabetes, hypertension, cardiovascular disease and other metabolic abnormalities³⁻⁶. Despite significant 57 58 advancements in metabolomics, the determinants of individual metabolic variability remain 59 incompletely understood. Twin studies have revealed a broad range of heritability for metabolite levels in human plasma^{7,8}, and genome-wide association studies (GWAS) have identified numerous genetic 60 61 variants influencing metabolite levels (mQTL, metabolite quantitative trait loci)^{9–13}. This demonstrates 62 the important role of genetics in human metabolism for the understanding of individual metabolic 63 diversity. Beyond genetics, human metabolic profiles are also influenced by various factors such as obesity^{14–16}, lifestyle ¹⁷, diet^{2,18,19}, microbiome^{2,18,19}, medications²⁰, and other environmental exposures. 64 For instance, by analyzing fasting plasma profiles of 1,183 metabolites in 1,679 samples from 1,368 65 individuals. Chen et al.²¹ found that diet, genetics and the gut microbiome could explain 9.3%, 3.3% 66 67 and 12.8%, respectively, of the inter-individual variations in plasma metabolomics.

However, these GWAS and association studies often overlook the temporal dynamics and environmental interactions that continuously influence the metabolome. Seasonal variations, for instance, introduce another layer of complexity in metabolite levels^{22–24}, reflecting changes in environmental conditions, dietary habits, physical activity, and other lifestyle factors^{25,26}. The heterogeneity in metabolite levels among individuals further underscores their multifaceted roles in

various biological processes^{27,28}. To accurately contextualize this variability in human metabolism, 73 74 incorporating longitudinal molecular data from the same individuals is crucial. This approach allows 75 for the monitoring of dynamics in metabolite levels in response to external influences and internal 76 physiological changes, providing a more comprehensive understanding of individual metabolite 77 variability^{5,29,30}. In addition, longitudinal data can reveal temporal metabolic regulation patterns that are 78 not apparent in cross-sectional studies³¹, offering deeper insights into the dynamic interplay between 79 genetic predispositions and environmental factors and advancing precision medicine for metabolic 80 health.

Here, we conducted a detailed longitudinal multi-omics analysis involving 101 individuals aged 50-65 over two years to explore the dynamics and individual differences in metabolic profiles. The influences of genetic predispositions, lifestyle factors, and physiological conditions on individual metabolic profiles have been investigated. Additionally, we established an integrative metabolite-protein network and identified key proteins and metabolites associated with human metabolic risk. This work contributes to a more comprehensive understanding of individual metabolite variability and advances our knowledge in more personalized approaches to monitoring metabolic health.

88 Methods

89 The wellness profiling study

90 The Swedish SciLifeLab SCAPIS Wellness Profiling (S3WP) study is an observational study aimed at 91 gathering longitudinal clinical and molecular data from a community-based cohort. This study derived 92 from the Swedish CArdioPulmonary bioImage Study (SCAPIS), which is a prospective observational 93 study with 30,154 individuals aged 50 to 65 years at enrollment, randomly selected from the general Swedish population between 2015 and 2018³². In SCAPIS, no exclusion criteria are applied except the 94 95 inability to understand written and spoken Swedish for informed consent. In the S3WP study, we 96 enrolled 101 healthy individuals with the following exclusion criteria: (1) previously received health 97 care for myocardial infarction, stroke, peripheral artery disease, or diabetes, (2) presence of any 98 clinically significant disease which, in the opinion of the investigator, may interfere with the results or

99 the subject's ability to participate in the study, (3) any major surgical procedure or trauma within 4 100 weeks of the first study visit, or (4) medication for hypertension or hyperlipidemia. Before enrolling in the S3WP study, all subjects had been extensively assessed by SCAPIS ³². Throughout the duration of 101 102 the S3WP study, follow-up visits are conducted every third month (± 2 weeks) in the first year and 103 approximately a 6-month interval in the second year. All subjects were fasting overnight (at least 8 h) 104 before the visits. Lifestyle questionnaires, anthropometric measurements, clinical measurements, and 105 plasma proteome profiling, metabolome profiling and lipidome profiling were examined at each of the 106 follow-up visit. Whole genome sequencing data were detected at the baseline (Fig. 1a, Additional file 107 2: Table S1). The study has been approved by the Ethical Review Board of Göteborg, Sweden 108 (registration number 407-15), and all participants provided written informed consent. The study 109 protocol adheres to the ethical guidelines outlined in the 1975 Declaration of Helsinki.

110 Self-reported questionnaires

Self-reported questionnaires, comprised 140 questions, were used to gather detailed information covering health, family history, medication, occupational and environmental exposure, lifestyle, psychosocial well-being, socioeconomic status, and other social determinants. These questionnaires had been administered in the SCAPIS trial. During each visit in the S3WP program, a selection of these questions was repeated to update the information from the initial SCAPIS questionnaire. Additionally, participants were asked about any changes in lifestyle factors between visits, such as infections, disease, medication, perceived health, and exercise level.

118 Anthropometric and clinical measurements

Height was measured to the nearest centimeter without shoes, with participants wearing indoor clothing. Weight was recorded on a calibrated digital scale under the same conditions. BMI was calculated by dividing weight (kg) by the square of height (m). Waist circumference was measured midway between the iliac crest and the lowest rib margin in the left and right mid-axillary lines. Hip circumference was measured at the widest part of the buttocks. Bioimpedance was assessed using the Tanita MC-780MA following the manufacturer's instructions. Systolic and diastolic blood pressure were measured in the supine position after a 5-minute rest using the Omron P10 automatic device. Blood pressure was initially

measured in both arms during the first visit, and subsequently in the arm that showed the highest reading.
Clinical chemistry and hematology assessments included capillary glucose (Hemocue), plasma glucose,
HbA1c, triglycerides, total cholesterol, LDL, HDL, ApoA1, ApoB, ApoA1/B ratio, creatinine, highsensitive C-reactive protein (hsCRP), ALAT, GGT, urate, cystatin C, vitamin D, TNT, NTproBNP,
hemoglobin, and blood cell count. The complete list of the clinical variables can be found in Additional
file 2: Table S1.

132 Whole genome sequencing

133 The whole genome-sequencing procedure has been detailed previously by Zhong et al. ^{33,34}. Briefly, 134 Genomic DNA was sequenced to average 30X coverage on the HiSeq X system (Illumina, paired-end 2×150 bp). The alignment was performed using BWAmem using GRCh38.p7 as reference genome. 135 Single-nucleotide and insertion/deletion variants were called following the GATK pipeline 136 (https://software.broadinstitute.org/gatk/best-practices; GATK v3.6). BCFtools³⁵ and PLINK 2.0³⁶ were 137 138 used to perform quality control (QC). The exclusion criteria included: (1) removing variants which did 139 not receive the "PASS" tag from GATK; (2) removing variants with minQUAL <30; (3) removing variants/samples that with a genotyping rate < 0.05; (5) removing variants with a low minor allele 140 141 frequency (MAF) (<5%); (6) removing variants that failed the Hardy–Weinberg equilibrium (HWE) 142 test ($P < 1 \times 10^{-6}$). In total, 6,691,390 high-quality variants were identified in all samples. Functional 143 annotation of variants was performed using Ensembl Variant Effect Predictor (VEP) v111³⁷.

144 Plasma metabolomics and lipidomics profiling

Plasma metabolites and lipids profiling were obtained by gas chromatography-mass spectrometry (GC-145 MS) and liquid chromatography-mass spectrometry (LC-MS)²⁹. Briefly, the metabolites were extracted 146 by using Agilent 1290 Infinity UHPLC-system (reverse phase chromatography) combined with an 147 148 Agilent 6550 Q-TOF mass spectrometer equipped with an electrospray Jetstream ion source operating 149 in positive and negative ion mode. The m/z range was 70-1700, and data were collected in centroid 150 mode with an acquisition rate of 4 scans/s. The mass spectrometry files were processed using Profinder 151 B.08.01 (Agilent Technologies Inc., Santa Clara, CA, USA). Peak detection was performed using mass 152 feature extraction. The lipids were extracted following a modified Folch protocol ³⁸. The data were

processed using Batch Targeted Feature Extraction algorithm within MassHunter[™] ProFinder version
B.08.00 (Agilent Technologies Inc., Santa Clara, CA, USA). In-house databases with exact mass and
experimental retention times of lipids were used for identification. The detailed method can be found
in Tebani et al.²⁹. In total, 456 metabolites were identified in 101 subjects and 173 lipids were measured
in 50 subjects.

All metabolite concentrations were log2-transformed to approximate a normal distribution. Metabolites with any of the following conditions have been removed: (1) metabolites that failed detection in at least 30% of samples; (2) the ratio of maximum and minimum interquartile range (IQR) of metabolite concentrations across four visits > 2; (3) duplicated metabolites. In addition, 5 subjects were removed because they participated in < 4 follow-ups. The filtering process retained a total of 527 unique metabolites for 96 subjects (357 metabolites based on 96 subjects and 170 lipids based on 50 subjects) for the downstream analysis of the study.

Metabolite annotation was performed using resources from the Human Metabolomics Database (HMDB) (version 5.0) ³⁹ and relevant literature ^{4,7–10}. A total of 527 metabolites were categorized into 9 main classes and 63 subclasses. Additionally, 11 lipid subclasses were further subdivided into 23 secondary lipid subclasses, resulting in the creation of 75 custom terms. These customized terms, along with metabolic pathways curated in the Small Molecule Pathway Database (SMPDB), were used for the enrichment analysis of metabolites. The complete list of the customized terms can be found in **Additional file 2: Table S3.**

172 Plasma protein profiling

We used a multiplex proximity extension assay (Olink Bioscience, Uppsala Sweden) to measure the relative concentrations of 794 plasma proteins in eleven Olink panels. To minimize inter-run and intrarun variation, the samples were randomized across plates and normalized using both an internal control (extension control) and an inter-plate control; then a pre-determined correction factor was applied to transform the data. The pre-processed data were provided in the arbitrary unit Normalized Protein eXpression (NPX) on a log2 scale. QC procedures were performed at both sample and protein level. Briefly, samples were flagged (did not pass QC) if the incubation control deviated more than a pre-

180 determined value (+/-0.3) from the median value of all samples on the plate (www.olink.com). To 181 reduce the batch effect between samples run at different times, bridging reference samples from 182 different visits were also run on plates from the different batches. Reference sample normalization based on bridging samples was conducted to minimize technical variation between batches (www.olink.com). 183 184 After QC, a total of 794 unique proteins for 90 subjects and 6 visits (540 samples) were retained for analysis. The detailed information about plasma protein profiling can be found in previous papers^{34,33}. 185 186 Proteins were annotated according to their molecular function, following the Human Protein Atlas v23 187 (https://www.proteinatlas.org/).

188 Co-expression analysis of plasma metabolites

189 Before performing co-expression analysis, metabolomics and lipidomics data for 527 metabolites were 190 scaled to zero mean and unit variance. UMAP was applied as an unsupervised clustering modeling 191 method for dimensionality reduction, providing an overview of clustering patterns among samples with 192 similar data profiles. The K Nearest Neighbour Search (KNN) algorithm was implemented to calculate the adjacency matrix using the 'nn2' function from the R package RANN v2.6.140, setting the maximum 193 194 number of nearest neighbors to 20. To calculate the number of shared nearest neighbors (SNN), we employed the 'sNN' function from the R package dbscan v1.1.11⁴¹, considering 5 neighbors for the 195 calculation. R package igraph v1.5.042 was used to build the adjacency matrix based on the nearest 196 197 neighbor results for each metabolite and to identify communities using the 'cluster louvain' function. 198 In total, 19 clusters were identified from the initial clustering. The mean scaled abundance of 199 metabolites within each of these 19 clusters was then used for a second step of hierarchical clustering. 200 Euclidean distances between the 19 clusters were computed, and similarities were assessed using the 201 ward.D2 method. The hierarchical clustering dendrogram was subsequently divided into 8 distinct 202 groups, resulting in the identification of 8 unique metabolite clusters (Additional file 2: Table S2).

203 Seasonal variation analysis

The seasonal patterns of both metabolites and proteins were analyzed by calculating the amplitude of their temporal expressions. The amplitude was defined as the square root of the sum of the squares of the coefficients of the sine and cosine terms of sampling month from the fitted seasonal model:

207
$$molecular abundance \sim \sin\left(\frac{2\pi \cdot month}{12}\right) + \cos\left(\frac{2\pi \cdot month}{12}\right) + Sex + BMI + Age + (1|individual) (1)$$

To identify temporally co-expressed metabolites and proteins, Euclidean distances were calculated for those showing significant results associated with sampling month (FDR-adjusted P < 0.05) based on their scaled mean intensity at each month. Clustering analysis was then performed using the 'ward.D2' method in the 'hclust' function from the R 'stats' package. The resulting hierarchical clustering tree was divided into distinct seasonal groups, revealing unique patterns of seasonal variation (Additional file

213 **2: Table S5 and Additional file 2: Table S8**).

214 Genome-wide association analysis of plasma metabolites

For each plasma metabolite, we calculated the coefficients of metabolite intensity for each subject using 215 a linear regression model, with visit included as a covariate. These coefficients served as adjusted 216 metabolite levels for subsequent GWAS analysis. GWAS was conducted using the PLINK v2.0³⁶, 217 218 employing a linear regression model that included body mass index (BMI), sex, and age as covariates. To identify independent mOTLs, linkage disequilibrium (LD) $r^2 > 0.1$ with window size 1 Mb was first 219 220 used to exclude the correlated variants. Given the high correlation among metabolites, we utilized an 221 eigendecomposition method to estimate the effective number of independent metabolites⁴. Out of the 222 527 metabolites, the estimated number of independent metabolites was 23, and this number was used to calculate the Bonferroni threshold for multiple hypothesis testing $(5 \times 10^{-8} / 23 = 2.17 \times 10^{-9})$. For 223 224 metabolites associated with multiple mQTLs, conditional analysis was performed by re-calculating 225 genetic associations using the lead single-nucleotide polymorphism (SNP) as a covariate. Only 226 associations with a conditional *P*-value < 0.01 were considered to be independent mQTLs. A total of 66 227 significant associations between genetic variants and individual blood metabolite concentrations were 228 identified. Among them, 19 independent metabolite quantitative trait loci (mQTLs) (Linkage Disequilibrium, LD, $r^2 < 0.1$, conditional P < 0.01) involving 22 metabolites were identified 229 230 (Additional file 2: Table S11).

231 Meta-analysis of plasma metabolite GWAS

232 For the meta-analysis of plasma metabolite GWAS, we included six out of 22 identified genetic-related 233 metabolites with available published GWAS summary results. The GWAS summary data from published studies^{4,7} was retrieved via the GWAS Catalog (www.ebi.ac.uk/gwas/, accession date: 234 2023.10), including 3 cohorts: Cooperative Health Research in the Region of Augsburg (KORA, N = 235 236 1768), TwinsUK (N = 1052) and Canadian Longitudinal Study of Aging (CLSA, N = 8,299). All these 237 three cohorts comprised relatively healthy European individuals. To ensure consistency across datasets, we used LiftOver⁴³ to convert the genome coordinates to the GRCh38 reference genome. The meta-238 239 analysis was performed on these three cohorts along with our study, involving a total of 5.8 million 240 SNPs, using an inverse-variance fixed-effect model using GWAMA v2.2.2⁴⁴. The genome-wide significance threshold for the meta-analysis was set at $P < 8.33 \times 10^{-9}$, accounting for multiple testing 241 correction $(5 \times 10^{-8} / 6)$. 242

243 Experimental validation

244 ACADS short interfering RNA (siRNA) and negative control siRNA were transfected into 293T cells. Following a 48-hour incubation period, the cells were collected and divided into two equal portions. 245 246 One portion was lysed and quantified using the Bicinchoninic Acid (BCA) Protein Assay Kit. Equal amounts of protein lysate from each sample were used for immunoblotting to assess the expression 247 248 levels of the ACADS protein. The immunoblotting results were analyzed to determine the efficacy of 249 ACADS knockdown. The remaining portion of the cells was processed for metabolite analysis. 250 Specifically, the relative content of butyrylcarnitine was measured using a liquid chromatographyquadrupole time-of-flight mass spectrometer (LC-QTOF MS, Agilent #1290-6546). The mass 251 252 spectrometer was equipped with Agilent Jet-stream source operating in negative and positive ion mode 253 with source parameters set as follow: Nebulizer gas, 45psi; Sheath gas temperature, 325 °C; Sheath gas 254 flow, 10L/min; Dry gas temperature, 280 °C; Dry gas flow, 8L/min; Capillary voltage, 3500v for two 255 ion modes and nozzle voltage, 500v for positive and 1000v for negative mode. The QTOF scan parameters were set as follows: Scan speed, 2 scan/s. Separation of metabolites was achieved in a Waters 256 ACQUITY UPLC BEH Amide column(2.1 mm \times 100 mm \times 1.7 µm) and guard column (2.1 mm \times 5 257

 $mm \times 1.7 \ \mu m$). Agilent Masshunter profinder 10.0 was used to extract its characteristic m/z 232.1547 ion from the total ion chromatogram. Butyrylcarnitine was quantified by peak area. Data analysis was conducted using Agilent MassHunter profinder 10.0 software. The detailed experimental setup are

261 provided in **Supplementary Information**.

262 Protein-metabolite association analysis, Mendelian randomization, and network analysis

Linear mixed modeling (LMM) was conducted to identify associations between 527 metabolites and 263 794 proteins, with sex, age, and BMI as fixed effects, and subject and visit as random effects. The 264 analysis was performed using the 'ImerTest' package⁴⁵, and the Kenward-Roger approximation was 265 applied to calculate P-value using the R package 'pbkrtest' v0.5.246. In addition, one-sample MR 266 267 analysis was performed to test the causal relationships between protein-metabolite pairs identified 268 through LMM (FDR-adjusted $P \le 0.05$). This analysis utilized instrumental variable (IV) regression by 269 two-stage least squares (2SLS) using the 'ivreg' function from the R package AER 1.2-1047, which is 270 based on Sex, age and BMI were included in the regression models as covariates, while the independent pQTLs (genome-wide significance, $P < 5 \times 10^{-8}$) associated with the protein were used as IVs. SNP 271 272 would be removed from the IVs if it had more than 5 association across the proteins¹³. The full list of 273 IVs was provided in Additional file 2: Table S15.

The protein-metabolite association network was constructed by combining both the LMM and MR results. Betweenness centrality score for each node in the network was calculated using the function 'betweenness' from the package 'igraph' v1.5.0⁴². Proteins and metabolites with a score > MAD above the respective median were classified as Tier 2, while proteins and metabolites with a score >2×MAD the median were categorized as Tier 1. All significant (FDR-adjusted P < 0.05) protein-metabolite pairs identified from the LMM were used to perform a UMAP clustering based on samples without missing values.

281 Metabolic risk stratification of participants in the study

To evaluate the metabolic risk levels of participants at various study visits, five classical metabolic risk indicators were utilized: body mass index (BMI), high-density lipoprotein (HDL), systolic blood pressure (SBP), fasting glucose (Gluc), and triglycerides (TG). Each measurement was normalized by

adjusting for sex-specific effects and then dividing by the standard deviation across all samples. Kmeans clustering was performed to stratify the samples into two groups (high risk and low risk) using the function 'kmeans' from the R package 'stats' ⁴⁸. Additionally, other clinical biochemical markers, including alanine aminotransferase (ALAT), gamma-glutamyltransferase (GGT), urate, troponin T (TNT), and white blood cells (WBC), were normalized in a same way. These markers were then used to test for differential expression levels between the two stratified metabolic risk groups.

291 Individuals with metabolic diseases in the UK Biobank

292 The UK Biobank is a large-scale biomedical database and research resource that includes genetic, 293 lifestyle, and health information from half a million UK participants aged 40-69 at baseline^{49,50}. 294 Participants were enrolled from 2006 to 2010 in 22 recruitment centers across the UK, with follow-up 295 data continuously collected. Proteomic profiling was conducted on blood plasma samples collected 296 during baseline recruitment from a randomized subset of UK Biobank participants. For this study, we 297 focused on Normalized Protein Expression (NPX) data for Tier 1 proteins and extracted metabolic 298 disease diagnoses information (field ID: 41202, ICD 10: E10, E11, E66, E03, E05, E78, M10, E88). 299 Participants with available proteome data for Tier 1 proteins were selected. To validate the predictive 300 power of Tier 1 proteins for various metabolic diseases, we focused on diseases with sufficient case 301 numbers for proteome analysis. Only data from participants diagnosed before baseline recruitment were 302 included, and individuals with multiple diagnoses were excluded. This resulted in the inclusion of type 303 2 diabetes (N = 144), obesity (N = 25), hyperthyroidism (N = 21), and gout (N = 27) for predictive 304 analysis.

For robust statistical analysis, NPX values were rank-inverse normal-transformed. Remaining missing values were imputed using the average value calculated across all individuals. Ultimately, the dataset comprised 242 participants in the metabolic disease group and 5,511 in the healthy group.

308 Identification of individuals who develop obesity in the UK Biobank

To identify individuals at risk of developing obesity in the future, we utilized BMI records (field ID:21001) from the UK Biobank, which included baseline recruitment data collected between 2006 and

311 2010 and three follow-up visits in 2012-2013, 2014+, and 2019+. Participants with a baseline BMI 312 lower than 25 kg/m² were included in the analysis. These individuals were further stratified into two groups: the 'future obese' group, consisting of individuals who recorded a BMI higher than 30 kg/m² in 313 314 at least one of the three follow-up visits, and the 'control' group, comprising individuals who maintained 315 a BMI lower than 30 kg/m² across all subsequent visits. To ensure data quality and relevance, we filtered the samples based on the availability of proteome data for Tier 1 proteins. This filtering process resulted 316 317 in the classification of fifteen individuals into the 'future obesity' group and 3,185 individuals into the 318 control group.

319 Machine learning for predictive tasks

To develop and validate prediction models, the samples were randomly split into training and validation 320 321 datasets using a 7:3 ratio. For analyses performed on the S3WP data, samples from the same individuals 322 were kept together in either the training or validation group. For the UK Biobank data, balanced datasets 323 were constructed by randomly selecting an equal number of control samples for predicting metabolic 324 diseases. All data were scaled before the prediction analysis. The prediction models were built using the XGBClassifier from the Python library xgboost v2.0.3⁵¹ with "binary:logistic" used as the objective. 325 To address class imbalance, the subsample parameter was set to 0.5, and the scale pos weight 326 327 parameter was adjusted to the ratio of negative and positive samples. The model training and validation 328 procedure was repeated 100 times, each time with a different random split of the data, to account for 329 stochastic variability in the selection of training and validation data. Receiver operating characteristic (ROC) curves were generated and visualized using the R package 'verification' v1.42⁵². 330

331 Statistical analysis and visualization

Uniform Manifold Approximation and Projection (UMAP) was performed by using the R package umap⁵³. Canonical correspondence analysis (CCA) was performed including lifestyle factors, anthropometric and clinical measurements, and visit as constraining variables using the R package vegan v2.6.4⁵⁴. The association between each variable and CCA1 and CCA2 was quantified by the estimated coefficient obtained from univariate linear regression. The mixed model expressed by equation (1) was used to determine the associations between sex, BMI, age, seasonality and metabolite

338 abundance, with metabolite abundance as response, sex, BMI, age, and month of sampling as fixed 339 effects and subject as random effect. All of mixed model analyses were performed using the ImerTest package⁴⁵ and Kenward-Roger approximation was applied to calculate *P*-value using the R package 340 341 pbkrtest v0.5.2⁴⁶. Variance analysis of metabolites and UMAP components was conducted using a linear 342 regression model including significantly associated mQTLs (and pQTLs for UMAP components), clinical measurements, lifestyle parameters, sex, age, and visit as variables. The fraction of explained 343 variability was measured as the Sum of Squares Explained (SSE) and p-value was determined using 344 345 Analysis of Variance (ANOVA) by the built-in R function aov. Pearson correlation was used to estimate 346 the correlation between metabolites and proteins. Kruskal-Walis test and t-test were used to compare 347 differences in the levels of metabolites or proteins between groups. Fisher's exact test was used for enrichment analysis of metabolites. FDR was calculated for multiple testing correction by using the 348 349 function p.adjust, selecting the Benjamini-Hochberg method. FDR-adjusted P < 0.05 was considered 350 significant in the analysis.

The Sankey plot was generated using the function sankeyNetwork of R package networkD3 v0.4⁵⁵. The heatmap was generated using the R package pheatmap v1.0.12⁵⁶. Networks were created using the R package igraph v1.5.0⁴². Chord diagrams were generated using the R packages circlize⁵⁷. Spider plots were generated using the package ggradar v0.2⁸⁴. The package ggalluvial v0.12.5⁵⁸ was used to show the group of the samples from the same individual across the six visits. All the other visualizations were performed using the ggplot2⁵⁹ R package. All of the data analysis and visualization was performed using the R project⁶⁰.

358 **Results**

359 Longitudinal analysis of human metabolic profiles in a wellness study

To systematically investigate the human metabolic profiles over time, we performed a comprehensive integrative multi-omics analysis of 101 participants in the Swedish SciLifeLab SCAPIS Wellness Profiling (S3WP) program over two years ²⁹. Whole-genome sequencing was performed at baseline for all participants. Extensive phenotyping of the individuals was conducted every three months in the first

364 year and at approximately a 6-month interval in the second year, which included plasma metabolome 365 and lipidome profiling, proteome profiling, clinical measurements, and detailed lifestyle questionnaires covering physical activity, mental health, substance intake, and other environmental factors, alongside 366 blood sample collection, to capture seasonal fluctuations and provide a robust temporal perspective (Fig. 367 368 1a and Additional file 2: Table S1). Using a combination of GC-MS and LC-MS technologies, we 369 identified a total of 527 metabolites in the study. These metabolites were classified into nine main 370 classes, covering a wide range of lipids (n = 335, 63.6%), amino acids (n = 77, 14.6%), xenobiotics (n 371 = 37, 7.0%), peptides (n = 17, 3.2%), carbohydrates (n = 20, 3.8%), cofactors and vitamins (n = 16, 3.0%), nucleotides (n = 14, 2.7%), energy (n =7, 1.3%), and other metabolites (n = 4, 0.8%). The 372 metabolites were further categorized into 63 subclasses based on their functions and biochemical 373

374 characteristics (Fig. 1a and Additional file 2: Table S2, see Methods for more details).

375 To explore the co-expression patterns of these identified metabolites, we conducted a clustering analysis applying K-nearest neighbor (KNN), shared nearest neighbor (SNN) and Louvain algorithms and 376 377 revealed that the 527 measured metabolites could be classified into eight major clusters (Fig. 1b-c, 378 Additional file 1: Fig. 1 and Additional file 2: Table S2). Functional enrichment analysis based on 379 customized class and pathway terms was further performed for each cluster to identify shared pathways 380 among groups of co-expressed metabolites (Fig. 1d and Additional file 2: Table S3). Specifically, cluster 1 showed a co-regulation of pathways involved in fatty acid metabolism, including steroids and 381 382 their derivatives, as well as monoglycerides. Cluster 2 mainly included metabolites related to the 383 metabolism of dietary components such as caffeine and benzoates. Cluster 3 included pathways central 384 to the urea cycle, specifically the arginine and proline metabolism. It also involved the metabolism of 385 amino acids like methionine, cysteine, glycine, serine, and theronine, which are crucial for nitrogen balance and protein turnover⁶¹. Cluster 4 exhibited a co-expression similar to Cluster 3, featured by 386 387 metabolites like dipeptides as well as the metabolisms of arginine, aspartate, and gamma-glutamyl 388 amino acids, which are mainly involved in protein synthesis and amino acid recycling⁶². Cluster 5 389 mainly consisted of metabolites essential for the cell membranes composition and signaling, including sphingomyelin, ceramide, and phosphatidylcholine⁶³. Cluster 6 showed the interconnected pathways of 390

391 glycolysis, gluconeogenesis, and sugar metabolisms, including the metabolism of fructose, mannose, 392 and galactose. Cluster 7 included metabolites related to energy storage and mobilization, such as diglycerides⁶⁴. Cluster 393 triglycerides and 8 comprises lyso-phosphatidylcholine, lyso-394 phosphatidylethanolamine, and phosphatidylcholine, key components involved in cell membrane 395 remodeling⁶⁵.

396 Individual and seasonal variations in plasma metabolome profiles

397 We assessed the variability in the expression of each metabolite over time by analyzing both inter-individual 398 and intra-individual variations, calculated using the coefficient of variance (CV; Fig. 2a and Additional file 399 2: Table S4). Notably, our analysis revealed that the variability between individuals for each measured 400 metabolite was greater than the variability observed within the same individual, with ratios of inter-401 individual to intra-individual CV ranging from 1.09 to 8.86. Among these metabolites, pyrrole-2-carboxylic 402 acid and picolinic acid showed the most significant differences between individuals (Fig. 2a). This suggests 403 that, despite the fluctuations in each person's metabolic profile over time due to various environmental 404 factors, each individual maintained a distinct metabolomic signature.

405 Interestingly, some metabolites showed seasonal patterns which partially contributed to the intra-individual 406 variations. To explore the seasonal influences on metabolite variability and identify potential seasonal 407 patterns, we performed an association analysis between metabolite levels and the month of sampling, with 408 sex, age and BMI considered as covariates. A total of 121 metabolites showed clear associations with the 409 month of sampling, with amplitudes ranging from 0.016 to 0.541 (Fig. 2b and Additional file 2: Table S5). 410 Hierarchical clustering of these metabolites further showed four distinct seasonal patterns: extremely low 411 levels from January to March (Cluster M1, n = 18); relatively low levels during the winter months (Cluster M2, n = 18); relatively high levels during winter (Cluster M3, n = 49); extremely high levels in December 412 413 (Cluster M4, n = 36) (Fig. 2c-g and Additional file 2: Table S6). The metabolites in Cluster M1, exhibiting 414 the lowest levels from January to March, were enriched in carbohydrate metabolism pathways (Fig. 2h and 415 Additional file 2: Table S7). As an example, fructose reached its lowest levels in January and February 416 (Additional file 1: Fig. 2a). This could reflect a post-holiday shift to baseline dietary habits or a reduction 417 in the consumption of sugars. Conversely, the elevation of energy-related pathways in Cluster M4 during 418 December may indicate an increased caloric intake during the festive period or an adaptive physiological

419 response to colder temperatures. Metabolites in Cluster M2, with relatively low levels in winter, were 420 enriched in pathways associated with NAD⁺ signaling, nicotinate and nicotinamide metabolism, purine 421 metabolism, and phosphatidylethanolamine biosynthesis (Fig. 2h). This could reflect a reduced requirement 422 for or availability of these pathways' end products during the winter, possibly due to changes in diet, reduced 423 exposure to sunlight affecting vitamin D synthesis, which in turn affects NAD⁺ synthesis ⁶⁶. Interestingly, 424 the seasonal pattern of vitamin D closely correlated with the metabolites in Cluster M2 (Pearson P = 0.04827, 425 Fig. 2i). In contrast, Cluster M3 metabolites, which were relatively high during winter, were associated with 426 amino acid metabolism, specifically glycine and serine metabolism, as well as pathways involved in 427 ammonia recycling and the metabolism of methionine and homocysteine (Fig. 2h). For instance, the highest 428 level of cysteine was observed during winter (Additional file 1: Fig. 2b). These increased levels of amino 429 acid metabolism could suggest a shift in fuel utilization towards amino acid catabolism for energy production 430 and could also be associated with lower levels of physical activity as reported by Pedersen EB, et al^{67} . 431 Interestingly, the energy metabolism-related fatty acid succinvlcarnitine in Cluster M3, which is involved in 432 carnitine synthesis and utilization pathway, also showed a significantly high inter-individual variation, 433 indicating varying levels of energy metabolism among individuals (Fig. 2j).

434 We further analyzed the seasonal variations of plasma proteins and found that some plasma proteins also 435 exhibited seasonal patterns, with two opposite seasonal clusters observed: Cluster P1, with low levels during 436 the colder season (Cluster P1, n = 51), and Cluster P2, with relatively low expression levels during the 437 warmer months (Cluster P2, n = 312) (Additional file 1: Fig. 2c-e and Additional file 2: Table S8). As an 438 example, 46 proteins in Cluster P2 were found to closely correlate with the seasonal patterns of metabolites 439 in Cluster M3. These proteins primarily function in amino acid, glycan, and fatty acid metabolism. 440 Specifically, angiopoietin-like protein 4 (ANGPTL4), which regulates lipoprotein lipase and has been found to be influenced by dietary fatty acids in both human muscle⁶⁸ and mice heart⁶⁹, showed associations with a 441 442 group of fatty acids, such as cis-4-decenoylcarnitine and laurylcarnitine (Additional file 1: Fig. 2f and 2g). 443 Along with these fatty acids, ANGPTL4 exhibited relatively lower expression levels from June to September 444 (Additional file 1: Fig. 2h). The associations between proteins and metabolites will be described in more 445 depth below.

446 Sex- and BMI-related divergences in plasma metabolite levels

447 The associations between plasma metabolite concentrations, lifestyle factors and clinical measurements were 448 analyzed using canonical correspondence analysis (CCA) that incorporated all 527 metabolites, 13 lifestyle 449 variables, and 43 clinical chemistry and anthropometric variables across visits. Regression analysis of the 450 two CCA components showed that CCA1 was mainly associated with BMI and lipid profiles, whereas CCA2 451 was more closely related to sex, body composition (muscle and fat percentage), hemoglobin, and urate. 452 Notably, a clear divergence between male and female samples was observed, emphasizing sex as a significant 453 factor influencing plasma metabolomic levels (Fig. 3a). The associations between metabolites and sex were 454 visualized in a volcano plot (Fig. 3b and Additional file 2: Table S9). Among the 119 metabolites that 455 showed significant sex differences, 35 were found to be elevated in females and 84 were more abundant in males (Fig. 3b). In particular, the peptide gamma-glutamylleucine exhibited higher levels in males (Fig. 3c) 456 and has been found to be associated with elevated cardio-metabolic risks⁷⁰. Additionally, 86 BMI-related 457 458 metabolites were identified (Fig. 3d and Additional file 2: Table S10), with glutamic acid showing the 459 strongest association (Fig. 3e), which aligns with previous reports⁷¹. Furthermore, the ratio of glutamic acid 460 to other amino acids, such as lysine, ornithine, and hydroxyproline, have been reported as promising biomarkers for identifying metabolically healthy obese individuals⁷¹. Interestingly, lifestyle factors such as 461 462 physical activity, stress, and smoking were found to correlate with metabolite levels and show collinearity with sex. In general, the stress levels in females were higher than in males, which was also significantly 463 associated with the elevation of the stress hormone corticosteroid⁷². On the other hands, a higher incidence 464 465 of smoking among males was associated with the alterations in several metabolites, such as glutamine, which 466 plays a central role in cellular metabolism and function, highlighting the influence of lifestyle factors on 467 metabolic health.

468 Genome-wide association analysis of the plasma metabolite profile

To investigate the genetic influence on inter-individual differences in plasma metabolite concentration, a GWAS was performed based on individual variation coefficients for 527 plasma metabolites and 6.7 million common genetic variants (minor allele frequency, MAF > 0.05) identified through wholegenome sequencing. A total of 66 significant associations between genetic variants and individual blood metabolite concentrations were identified ($P < 2.17 \times 10^{-9}$, conventional *P* of 5×10^{-8} adjusted for the

474 number of independent metabolites ⁴). Among them, 19 independent metabolite quantitative trait loci 475 (mQTLs) (Linkage Disequilibrium, LD, $r^2 < 0.1$, conditional P < 0.01) involving 22 metabolites were identified (Fig. 4a and Additional file 2: Table S11). Of 19 mOTLs, 4 were pleiotropic genetic variants 476 associated with multiple metabolites. Of these metabolites, 45% were lipids (n = 10) (Fig. 4a-b). Out 477 478 of the 22 genetically associated metabolites in our study, six have been previously analyzed in other 479 GWAS studies^{4,7}. To validate the associations between these metabolites and genetic variants, a meta-GWAS analysis was conducted for these six metabolites. Most of the genetic loci (8 out of 11) identified 480 481 from meta-analysis showed the same direction of effects as in our study (Additional file 2: Table S12). 482 Among these, the association between the genetic variant (rs34673751) from Acyl-CoA dehydrogenase short chain (ACADS) and butyrylcarnitine was the most significant in the meta-analysis. The ACADS 483 gene encodes the enzyme short-chain acyl-CoA dehydrogenase (SCAD), which is essential for 484 485 mitochondrial fatty acid beta-oxidation, while butyrylcarnitine is a short-chain acylcarnitine involved in fatty acid transport and energy metabolism. Our longitudinal analysis further demonstrated that 486 487 individuals carrying A allele at rs34673751 exhibited stable higher blood butyrylcarnitine levels. 488 Moreover, heterozygous individuals for the protein variant presented intermediate levels of blood 489 butyrylcarnitine compared to the homozygous groups (Fig. 4c-e). To experimentally validate the 490 association between ACADS and butyrylcarnitine, we knocked down the expression of ACADS in 293T 491 cell lines using siRNA. Subsequent metabolite analysis revealed that silencing the ACADS gene 492 increased the levels of butyrylcarnitine in these cells (Fig. 4f-g), providing direct evidence of ACADS's 493 role in the regulation of butyrylcarnitine levels.

Another notable example of the identified mQTLs is the association between metabolite 4-androsten-3alpha,17alpha-diol monosulfate (2), a sulfated steroid and a derivative of androstenediol, and the gene cytochrome P450 family 3 subfamily A member 7 (*CYP3A7*). (**Fig. 4h**). The highest association was found for the variant rs45446698, located upstream of the *CYP3A7* gene. Individuals carrying a TT homozygote had higher and more stable levels of 4-androsten-3alpha,17alpha-diol monosulfate (2) than individuals carrying a TG heterozygote (**Fig. 4i-i**).

500 Quantification of genetic and non-genetic effects on plasma metabolite levels

501 To quantify the influence of genetics, lifestyle, and physiological conditions on metabolite variability, 502 we applied a linear multivariate regression model to each metabolite. This model included all 19 mQTLs, 503 13 lifestyle factors, 43 anthropometric and clinical chemistry parameters, and visit. In the analysis, the 504 genetic variants were combined as "genetic component", all the lifestyle-related factors were combined 505 as "lifestyle component", and all anthropometric and clinical chemistry variables were categorized into 506 13 clinical classes. A summary of the analysis across all 527 analyzed plasma metabolites (Fig. 5a and 507 Additional file 2: Table S13) showed that the influence of genetics, lifestyle, and physiological 508 conditions on plasma metabolite level varied considerably. Genetics emerged as one of the important 509 factors; out of the 22 metabolites with at least one significant genetic association, 5 metabolites had a 510 genetic contribution greater than 20% (Fig. 5b). The metabolite most affected by genetics was 1,3-511 Dimethylurate, which is formed from caffeine and can be used as an indicator of caffeine metabolism 512 ⁷³, with 30% of its blood level variance explained by genetic variants. Besides genetics, 469 metabolites 513 were influenced by various clinical factors, with a total contribution greater than 20%. Consistent with 514 the CCA analysis (Fig. 3a), body composition and lipid profiles showed the most significant 515 associations with plasma metabolite levels, with 69 metabolites associated with each of them. In 516 addition, 54 metabolites were associated with urate levels, 48 with kidney function, 36 with glucose homeostasis, 18 with liver function, 17 with heart function, 15 with leukocytes, 6 with the acute phase 517 518 response, and 28 with other clinical parameters. The top 25 most significant metabolites associated with 519 clinical components were highlighted in Fig. 5c. As an example, a significant association was observed 520 between body composition and pyroglutamylvaline, aligning with previous study that found pyroglutamylvaline to be positively associated with leg muscle⁷⁴. Additionally, multiple metabolites 521 522 have been identified as being associated with immune cell and red cell populations, as well as 523 inflammatory biomarkers like C-reactive protein (CRP) (Fig. 5c and Additional file 2: Table S13). 524 These results indicated the intricate crosstalk between metabolism, immune response, and 525 hematopoiesis.

526 Furthermore, we identified 39 metabolites associated with various lifestyle factors, with 9 showing an 527 influence from lifestyle factors greater than 20% (Additional file 2: Table S13). These lifestyle-related metabolites included 15 associated with smoking, 8 with changes in housing, 4 with physical activity, 528 4 with stress, and 9 with other lifestyle factors. The top 25 most significant metabolites associated with 529 530 lifestyle factors are listed in Fig. 5d. Notably, smoking had the most prominent influence on blood 531 metabolite levels. Among the smoking associated metabolites, glutamine, the most abundant amino acid in the body and considered conditionally essential for critical illness and injury^{75,76}, was negatively 532 533 associated with smoking, along with factors such as 3-pyridinol, an intermediate product in nicotine degradation⁷⁷ (Additional file 2: Table S13). 534

535 Individual metabolite-protein profiles in human plasma

To investigate the co-expression patterns of plasma metabolites and proteins, we applied linear mixed 536 537 modeling (LMM) to 527 metabolites and 794 proteins, adjusting for cofounders including subject, visit, 538 sex, age, and BMI. The analysis revealed 5,649 significant protein-metabolite pairs, involving 479 539 metabolites and 625 proteins, each characterized by a correlation with a false discovery rate (FDR) 540 adjusted P of less than 0.05 (Additional file 2: Table S14). Among these significant associations, 459 541 involving 121 metabolites and 257 proteins (48.93% of the overlapping metabolite-protein associations) 542 aligned with previous findings from Benson MD, et al.⁷⁸, despite differences in molecular measurement 543 platforms. In Fig. 6a we present the 200 most significant protein-metabolite associations to illustrate 544 the complex interplay within the protein-metabolite network. Multiple important hormones, enzymes, 545 receptors and cytokines, such as glucagon (GCG), lipoprotein lipase (LPL), natriuretic peptide (NPPC), 546 and interleukin 10 (IL10), have been identified as hub proteins in the network, highlighting their broad 547 regulatory roles in human metabolism. Most of these protein-metabolite associations were connected to lifestyle and physiological conditions. As an example, we found a co-regulation between leptin (LEP) 548 549 and aceturic acid (Fig. 6b). LEP, a hormone secreted by adipose tissue, plays an important role in 550 regulating hunger and energy balance⁷⁹, with higher concentrations observed in females⁸⁰. Aceturic acid, 551 also known as N-acetylglycine, is a derivative of the amino acid glycine and has been reported to

modulate weight and associated adipose tissue immunity ⁸¹. Our findings suggested a significant
 interaction between LEP and aceturic acid with gender stratification.

554 Subsequent Mendelian Randomization (MR) analyses were conducted to investigate potential genetic 555 drivers of causality between plasma proteins and metabolites. A total of 87 putative causal associations 556 were identified between 38 proteins and 61 metabolites (FDR-adjusted P < 0.05, Additional file 1: Fig. 557 3a, Additional file 2: Table S15). As an example, we detected a significant MR association between 558 MDGA1, a cell surface glycoprotein involved in cell adhesion, migration, axon guidance, and neurodevelopment⁸²⁻⁸⁴, and 1-Arachidonoylglycerol (1-AG) (Fig. 6c), a stable regioisomer of the 559 560 endocannabinoid 2-AG engaged in physiological functions such as emotion, cognition, and 561 neuroinflammation^{85,86}. Our analyses showed that the cis instrumental variable (rs9349050, Fig. 6d) for 562 the MDGA1 protein consistently differentiated both MDGA1 and 1-AG levels in individuals carrying 563 different genotypes (Additional file 1: Fig. 3b-c). This stable and positive association between 564 MDGA1 and 1-AG indicated a potential genetic basis for the co-regulation of proteins and metabolites 565 in the nervous system.

566 Using Uniform Manifold Approximation and Projection (UMAP), we clustered the individual 567 molecular profiles based on the integrated metabolite-protein network. We noted that each individual 568 possessed a unique and stable protein-metabolic profile (Fig. 6e). Regression analysis of the two UMAP 569 components (Fig. 6f-g) revealed that UMAP1 was most significantly associated with genetic factors, 570 followed by kidney function, lipid profile, body composition and erythrocytes. UMAP2 showed 571 significant associations with body composition, blood pressure, genetics, and erythrocytes. Additionally, 572 immune response, vitamin D, urate levels, liver functions, glucose homeostasis were moderately 573 associated with the metabolite-protein profiles. Lifestyle factors such as stress, drug intake, smoking, and physical activity also exhibited minor influences on these profiles. 574

To identify key proteins and metabolites with the highest connectivity of the protein-metabolite network, we quantified their importance using the centrality betweenness score and categorized them into three tiers based on their deviation from the median level, measured in median absolute deviations (MAD). These categories include Tier 1, high importance (beyond two MADs); Tier 2, moderate importance

579 (beyond one MAD); and Other: low importance (within one MAD). (Fig 6 h-i). Notably, ANGPTL4 580 and LEP were the top two proteins with the highest scores (Fig. 6j). ANGPTL4 has been reported as a 581 key regulator in lipid metabolism, primarily by inhibiting the activity of lipoprotein lipase (LPL)⁸⁷, and 582 also involved in angiogenesis, vascular permeability, and inflammation processes. In our analysis, we 583 revealed that ANGPTL4 was associated with a broad spectrum of metabolites, involving those related 584 to lipid and glucose metabolism, tissue functions, and immune responses (Additional file 1: Fig. 3d). LEP, which is a well-known hormone, has been shown in recent studies to reflect systemic alterations 585 of the human metabolome^{88,89}. In our analysis, we observed that LEP was associated with a spectrum 586 587 of metabolites related to blood pressure and glucose homeostasis, as well as stress (Additional file 1: 588 Fig. 3d). On the other hand, N-Acetylaspartate (NAA), one of the most abundant metabolites in the 589 mammalian brain, and 3-Hydroxyhippurate, a microbial aromatic acid metabolite derived from dietary 590 polyphenols and flavonoids found in normal human urine, were identified as high-importance 591 metabolites through network analysis and found to be associated with kidney function in the study (Fig. 592 6k, Additional file 2: Table S13).

593 Variability of protein-metabolite profiles and the associations with metabolic health

594 To investigate the associations between protein-metabolite profile variability and metabolic health, we 595 stratified the analyzed samples into two risk groups (high-risk group and low-risk group) based on the 596 clustering patterns of five key indicators of metabolic syndrome: high density lipoprotein (HDL), body mass index (BMI), systolic blood pressure (SBP), triglycerides (TG) and fasting glucose (Gluc)⁹⁰. The 597 598 high-risk group was characterized by increased levels of BMI, SBP, TG, and Gluc, alongside decreased 599 levels of HDL, in comparison with the low-risk group (Fig. 7a). We then explored the differences in 600 clinical chemistry measurements between these two groups. We found significant higher levels of two 601 liver function biomarkers (alanine aminotransferase, ALAT; gamma glutamyltransferase, GGT), one kidney function biomarker (urate), one heart function biomarker (troponin T, TNT), and one immune 602 603 biomarker (white blood cells count, WBC) in the high-risk group, indicating an increased susceptibility 604 to cardiometabolic diseases (Fig. 7b).

605 Predictive modeling of the metabolic risk of the analyzed samples was conducted using proteins and 606 metabolites. Interestingly, we found that the predictive power for risk groups, based on a combination 607 of all proteins and metabolites, to be comparable to that based on Tier 1 proteins (n = 13) and metabolites (n = 14) alone, as well as models based on only proteins (Fig. 7c,d). This highlighted the potential 608 609 importance of Tier 1 proteins and metabolites for metabolic risk assessments. In addition, proteins 610 contributed more significantly to metabolic risk prediction than metabolites, with metabolites providing 611 only marginal additional predictive value. Remarkably, the predictive power of Tier 1 proteins alone 612 approached that of the full protein model (Fig. 7c,d), while metabolites demonstrated considerably 613 lower predictive power compared to proteins (Fig. 7c,d). The importance of Tier 1 proteins and 614 metabolites in relation to metabolic risk was quantified using a linear regression model to estimate the 615 impact coefficients (Fig. 7e,f). We further examined the expressions levels of the 13 Tier 1 proteins in 616 individuals with metabolic diseases, including obesity, type 2 diabetes (T2D), gout or hyperthyroidism, compared to healthy individuals using data from the UK Biobank⁴⁹. Our analysis revealed that the 617 618 expression levels of all Tier 1 proteins were significantly different (P < 0.05) between healthy individuals 619 and those diagnosed with any of these metabolic diseases (Fig. 7g, Additional file 1: Fig. 4). 620 Furthermore, the expression patterns of these 13 Tier 1 proteins varied significantly across different 621 metabolic diseases, indicating their potential use as diagnostic biomarkers for various metabolic 622 conditions. Moreover, we assessed the predictive power of Tier 1 proteins for these metabolic diseases, 623 resulting in average AUCs ranging from 0.643 and 0.878 across different diseases (Fig. 7h).

Next, we examined the stability of the risk levels of participants throughout the two-year study. In 624 625 general, the metabolic health status of most participants (70 out of 101) remained consistent, with 39 626 categorized within the high-risk group and 31 within the low-risk group at every visit (Fig. 7i). A notable 627 exceptional case was participant W0010, who experienced a significant weight reduction from 120 kg 628 during the third visit to 104.7 kg by the fourth visit (Fig. 7i). This remarkable change caused a shift 629 from the high-risk group in the initial three visits to the low-risk group in the subsequent three 630 assessments. Focusing on the Tier 1 proteins, we observed a significant decrease in their levels from 631 the third to the fourth visit (Fig.7k). However, between the fourth and sixth visits, protein levels

632 reversed despite no significant change in BMI (Fig. 7k), suggesting alterations in the individual's 633 underlying metabolism at the molecular level. Additionally, we found that nearly half of the Tier 1 proteins (6/13) displayed significantly different levels between lean (BMI < 25) and obese (BMI > 30) 634 groups (Additional file 1: Fig. 5). We further predicted future obesity in individuals with a baseline 635 636 BMI lower than 25 in the UK Biobank using the 13 Tier 1 protein expressions. The results showed that participants who later developed obesity (BMI > 30) already exhibited a distinct profile of Tier 1 637 638 proteins at baseline compared to those who maintained a BMI lower than 30 throughout the study period 639 (Additional file 1: Fig. 6a, b). Predictive models demonstrated that baseline expression levels of Tier 640 1 proteins could predict future obesity with an average AUC of 0.773 with a 95% confidence interval 641 of 0.739-0.806 (Fig. 7l).

642 **Discussion**

643 In this study, we have performed a longitudinal multi-omics analysis on a group of clinically healthy 644 participants aged 55 to 65 over two years to explore how genetics, lifestyle and physiological conditions 645 affect individual metabolic profiles. We systematically examined the abundances and dynamics of 527 blood metabolites, which were categorized into nine major classes: lipids, amino acids, xenobiotics, 646 647 peptides, carbohydrates, cofactors and vitamins, nucleotides, energy, and others. These metabolites 648 were analyzed alongside their co-regulations with 794 proteins. By integrating whole-genome 649 sequencing and extensive phenotyping data collected concurrently at each sampling time point, our results revealed an intricate interplay between genetic predispositions and environmental factors in 650 651 shaping individual metabolic profiles.

One unique aspect of our study is the focus on the temporal dynamics of individual metabolic profiles. We identified eight co-expressed metabolite clusters based on longitudinal data and linked them to various metabolic pathways that are susceptible to both internal and external modulators, such as energy mobilization and dietary intake. By analyzing intra-individual metabolite variation over time, we observed four distinct seasonal patterns of plasma metabolites, with more fluctuations or extreme concentrations during the summer and winter. These seasonal fluctuations of metabolites might reflect changes in physical activity ⁶⁷, diet ⁹¹, and metabolic rate ⁹², which are closely related to the season and

might be regulated by seasonal gene expressions in multiple tissue types, including adipose tissue, brain,
 and gonadal tissue ^{93,94}. Similar seasonal variations were observed in the proteomics profiling,
 suggesting a coordinated seasonal influence on both metabolite and protein levels.

662 In general, our analysis revealed higher inter-individual variations than intra-individual variations in 663 plasma metabolite levels. By integrating whole-genome sequencing, we identified significant 664 associations between genetic variants and blood metabolite levels (mQTLs). Numerous mQTLs have been identified within various populations^{8-13,95}, and in this study, we used individual coefficients 665 666 obtained from longitudinal data to better associate metabolite levels with genetic factors. Considering the significant seasonal variations observed in metabolite levels, this approach allows for more accurate 667 668 quantification of individual metabolite levels for mOTL identification. Our findings indicated that many 669 plasma metabolite levels during adult life were genetically predetermined at birth and remained stable 670 within a range of dynamics under healthy conditions. Additionally, our results showed that lifestyle and 671 clinical factors also significantly contributed to the variability in blood metabolite levels. Notably, 89% of (469 out of 527) metabolites exhibited at least a 20% variability attributable to lifestyle or clinical 672 673 factors, indicating a substantial influence from lifestyle and physiological conditions on human 674 metabolism. Among these, body composition and smoking were the top physiological and lifestyle factors influencing blood metabolite levels. 675

676 Proteins, as essential regulators and executors of metabolic processes, are critical components of human 677 metabolic profiles, making their analysis essential for a comprehensive understanding of human 678 metabolism. By integrating longitudinal metabolomics and proteomics data from the same individuals, 679 we established a protein-metabolite co-expression network and found that each individual possessed a 680 unique and stable protein-metabolic profile over time. This stability of individuals' protein-metabolite profiles suggested that, despite the changes in environment, there is an inherent individuality to 681 metabolic regulation. Our analysis showed that genetic factors were the most significant contributors to 682 683 these profiles, with additional moderate influences from physiological conditions such as kidney 684 function, blood pressure, and body compositions. Moreover, lifestyle factors like stress, smoking, and 685 drug intake also had considerable impacts.

686 Furthermore, we identified key signature proteins and metabolites that characterize individual protein-687 metabolic profiles and assessed the predictive power of the most important Tier 1 proteins and metabolites for identifying individuals at high metabolic risk. Remarkably, the predictive performance 688 689 of these Tier 1 biomarkers was comparable to that of the full set of proteins and metabolites, 690 underscoring their potential utility in metabolic health monitoring. This finding also revealed a superior 691 predictive power of proteins over metabolites. This is likely because proteins are directly involved in 692 essential metabolic processes, function as signaling molecules and regulators, and generally exhibit 693 greater stability and consistency over time. In our previous report, we also showed that plasma proteome exhibited most of the effects from clinical data^{34,33}. We further validated this panel of 13 Tier 1 proteins 694 695 in large-scale population data from the UK Biobank for the prediction of metabolic diseases. These 696 biomarkers showed significant potential in diagnosing metabolic diseases. Although, all of the 13 697 proteins were elevated in various metabolic diseases, they exhibited different patterns in different 698 metabolic abnormalities. Interestingly, we observed that the levels of the 13 proteins in an obese 699 individual in the S3WP study initially decreased following weight loss but subsequently started to revert 700 to the initial levels, despite a stable BMI during the study period. In the UK Biobank, we also noted that 701 individuals who eventually became obese already exhibited an altered profile of Tier 1 proteins at a 702 normal BMI and at baseline. These observations suggest that fluctuations in the proteome may precede 703 BMI changes, highlighting their potential value in weight management strategies. However, it should 704 be noted that more comprehensive analyses in large disease cohorts of these plasma proteins and 705 metabolites must be performed to validate their use as clinical biomarkers. Another limitation of this 706 study is the relatively small cohort size, which may not capture the full spectrum of individual metabolic 707 variability across diverse populations. Thus, further validation and generalization of the findings in larger, more diverse cohorts and in clinical settings are necessary. 708

709 Conclusions

710 In summary, our study provides a comprehensive longitudinal analysis on how genetics, lifestyle, and 711 physiological factors influence the human metabolic profiles. Our results demonstrated the dynamic 712 interplay between genetic predispositions and environmental factors, including seasonal variations,

- 713 lifestyle, and physiological changes. By integrating proteomics and metabolomics data, we established
- a detailed protein-metabolite network and identified key molecular signatures that enhanced metabolic
- 715 disease diagnostics and risk assessments. These findings offer promising avenues for improving
- 716 metabolic health monitoring and developing future targeted interventions based on an individual's
- 717 unique metabolic profile.

718 Abbreviations

- 719 1-AG: 1-Arachidonoylglycerol
- 720 ACADS: Acyl-CoA dehydrogenase short chain
- 721 ALAT: Alanine aminotransferase
- 722 ANGPTL4: Angiopoietin-like protein 4
- 723 ANOVA: Analysis of Variance
- 724 ApoA1: Apolipoprotein A1
- 725 ApoB: Apolipoprotein B
- 726 AUC: Area under curve
- 727 **BMI:** Body mass index
- 728 CCA: Canonical correspondence analysis
- 729 CRP: C-reactive protein
- 730 **CV:** Coefficient of variance
- 731 **CYP3A7:** Cytochrome P450 family 3 subfamily A member 7
- 732 **FDR:** False discovery rate
- 733 GCG: Glucagon
- 734 GC-MS: Gas chromatography-mass spectrometry
- 735 **GGT:** Gamma glutamyltransferase
- 736 Gluc: Fasting glucose
- 737 **GWAS:** Genome-wide association studies
- 738 HDL: High density lipoprotein
- 739 HMDB: Human Metabolomics Database

- 740 **IL10:** Interleukin 10
- 741 IV: Instrumental variable
- 742 KNN: K-nearest neighbor
- 743 LC-MS: Liquid chromatography-mass spectrometry
- 744 LD: Linkage Disequilibrium
- 745 LDL: Low density lipoprotein
- 746 LEP: Leptin
- 747 LMM: Linear mixed modeling
- 748 LPL: Lipoprotein lipase
- 749 MAD: Median absolute deviation
- 750 MAF: Minor allele frequency
- 751 MDGA1: MAM domain containing glycosylphosphatidylinositol anchor 1
- 752 **mQTL:** Metabolite quantitative trait loci
- 753 MR: Mendelian randomization
- 754 NAD+: Nicotinamide adenine dinucleotide+
- 755 NPPC: Natriuretic peptide
- 756 NPX: Normalized Protein eXpression
- 757 NTproBNP: N-Terminal pro-brain natriuretic peptide
- 758 **pQTL:** Protein quantitative trait locus
- 759 **QC:** Quality control
- 760 **Q-TOF:** Quadrupole Time-of-Flight
- 761 **ROC:** Receiver operating characteristic
- 762 S3WP: Swedish SciLifeLab SCAPIS Wellness Profiling
- 763 SBP: Systolic blood pressure
- 764 SCAD: Short-chain acyl-CoA dehydrogenase
- 765 SCAPIS: Swedish CArdioPulmonary bioImage Study
- 766 siRNA: short interfering RNA
- 767 SMPDB: Small Molecule Pathway Database

- 768 SNN: Shared nearest neighbor
- 769 **SNP:** Single-nucleotide polymorphism
- 770 **T2D:** Type 2 diabetes
- 771 TG: Triglyceride
- 772 **TNT:** Troponin T
- 773 UMAP: Uniform manifold approximation and projection
- 774 **WBC:** White blood cells count
- 775

776 **Declarations**

777 Ethics approval and consent to participate

- 778 The Swedish SciLifeLab SCAPIS Wellness Profiling (S3WP) study has been approved by the Ethical
- Review Board of Göteborg, Sweden (registration number 407-15), and all participants provided written
- informed consent. The study protocol adheres to the ethical guidelines outlined in the 1975 Declaration
- 781 of Helsinki.

782 **Consent for publication**

783 Not applicable.

784 Availability of data and materials

785 This study utilized participant-level datasets that have been securely deposited at the Swedish National 786 Data Service (SND), a repository certified by the Core Trust Seal (URL: https://snd.gu.se/). The GC 787 and LC-MS datasets are available as part of the full SW3P Wellness multi-omics dataset with the doi 788 10.5878/rdys-mz27. In compliance with patient consent and confidentiality agreements, these datasets 789 are accessible for validation purposes only. Requests for access can be directed to SND via email 790 (snd@snd.gu.se). The evaluation of such requests will be conducted in accordance with relevant 791 Swedish legislation. Additionally, for inquiries specifically on research within the scope of the S3WP 792 program, interested parties are encouraged to contact the corresponding author directly. The code can 793 be made available upon request to the corresponding author.

794 Competing interests

- 795 The authors declare no competing interests.
- 796 Funding

797 This work was supported by the SciLifeLab & Wallenberg Data Driven Life Science Program (grant:

798 KAW 2020.0239) and the Swedish Research Council (#2022-01562).

799 Authors' contributions

- 800 W.Z. conceived and designed the study. G.B. supplied clinical material. M.U., G.B., W.Z. and F.E.,
- 801 collected and contributed data to the study. W.Z., J.W., A.Z., and X.W. performed the data analysis.
- 802 W.Z., J.W., A.Z., and X.W. drafted the manuscript. All authors read and approved the final manuscript.

803 Acknowledgements

- Original human sensitive data was processed under Project sens2018115. The integrative multi-omics analysis of the processed data was performed using computational resources provided by SNIC through the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under Project Berzelius-2024-200. Part of this research was conducted using the UK Biobank Resource under Application 99914. We thank the Plasma Profiling Facility at SciLifeLab in Stockholm for conducting the Olink analyses and Dr. Linn Fagerberg for data collection and data analysis in the S3WP program.
- 810 We also thank the Proteomics and Metabolomics Platform at Guangzhou Laboratory for conducting the
- 811 cell-line validation experiment using liquid chromatography-time-of-flight mass spectrometry (Q-TOF).
- 812 We also express our gratitude to Yanfen Hong for her assistance with data acquisition and analysis.

813 Authors' information

- 814 Science for Life Laboratory, Department of Biomedical and Clinical Sciences (BKV), Linköping
- 815 University, Linköping, Sweden.
- 816 Jing Wang, Alberto Zenere, Xingyue Wang and Wen Zhong
- 817
- 818 Department of Molecular and Clinical Medicine, Sahlgrenska Academy, University of Gothenburg, and
- 819 Clinical Physiology, Sahlgrenska University Hospital, Gothenburg, Sweden

820 Göran Bergström

821

- 822 Department of Protein Science, KTH Royal Institute of Technology, Stockholm, Sweden
- 823 Fredrik Edfors and Mathias Uhlén

824 **References**

- 1. Gieger, C. *et al.* Genetics Meets Metabolomics: A Genome-Wide Association Study of Metabolite
- 826 Profiles in Human Serum. *PLOS Genet.* **4**, e1000282 (2008).
- 827 2. Chen, L. et al. Influence of the microbiome, diet and genetics on inter-individual variation in the
- human plasma metabolome. *Nat. Med.* **28**, 2333–2343 (2022).
- 829 3. Robinson, J. L. et al. An Atlas of Human Metabolism. Sci. Signal. 13, eaaz1482 (2020).
- 4. Chen, Y. *et al.* Genomic atlas of the plasma metabolome prioritizes metabolites implicated in human
 diseases. *Nat. Genet.* 55, 44–53 (2023).
- 5. Hornburg, D. *et al.* Dynamic lipidome alterations associated with human health, disease and ageing. *Nat. Metab.* 5, 1578–1594 (2023).
- 6. Merino, J., Udler, M. S., Leong, A. & Meigs, J. B. A Decade of Genetic and Metabolomic
 Contributions to Type 2 Diabetes Risk Prediction. *Curr. Diab. Rep.* 17, 135 (2017).
- 836 7. Shin, S.-Y. *et al.* An atlas of genetic influences on human blood metabolites. *Nat. Genet.* 46, 543–
 837 550 (2014).
- 838 8. Long, T. *et al.* Whole-genome sequencing identifies common-to-rare variants associated with human
- 839 blood metabolites. *Nat. Genet.* **49**, 568–578 (2017).
- 840 9. Surendran, P. *et al.* Rare and common genetic determinants of metabolic individuality and their
 841 effects on human health. *Nat. Med.* 28, 2321–2332 (2022).
- 842 10. Lotta, L. A. *et al.* A cross-platform approach identifies genetic regulators of human metabolism
 843 and health. *Nat. Genet.* 53, 54–64 (2021).
- 844 11. Illig, T. *et al.* A genome-wide perspective of genetic variation in human metabolism. *Nat. Genet.*845 42, 137–141 (2010).

- 846 12. Bomba, L. et al. Whole-exome sequencing identifies rare genetic variants associated with
- 847 human plasma metabolites. *Am. J. Hum. Genet.* **109**, 1038–1054 (2022).
- 848 13. Karjalainen, M. K. et al. Genome-wide characterization of circulating metabolic biomarkers.
- 849 *Nature* 1–9 (2024) doi:10.1038/s41586-024-07148-y.
- 850 14. Cirulli, E. T. et al. Profound Perturbation of the Metabolome in Obesity Is Associated with
- 851 Health Risk. *Cell Metab.* **29**, 488-500.e2 (2019).
- Watanabe, K. *et al.* Multiomic signatures of body mass index identify heterogeneous health
 phenotypes and responses to a lifestyle intervention. *Nat. Med.* 29, 996–1008 (2023).
- 16. Ottosson, F. et al. Metabolome-Defined Obesity and the Risk of Future Type 2 Diabetes and
- 855 Mortality. *Diabetes Care* **45**, 1260–1267 (2022).
- Xu, T. *et al.* Effects of smoking and smoking cessation on human serum metabolite profile:
 results from the KORA cohort study. *BMC Med.* 11, 60 (2013).
- Bar, N. *et al.* A reference map of potential determinants for the human serum metabolome. *Nature* 588, 135–140 (2020).
- 860 19. Asnicar, F. *et al.* Microbiome connections with host metabolism and habitual diet from 1,098
 861 deeply phenotyped individuals. *Nat. Med.* 27, 321–332 (2021).
- 862 20. Altmaier, E. *et al.* Metabolomics approach reveals effects of antihypertensives and lipid863 lowering drugs on the human metabolism. *Eur. J. Epidemiol.* 29, 325–336 (2014).
- Chen, L. *et al.* Influence of the microbiome, diet and genetics on inter-individual variation in
 the human plasma metabolome. *Nat. Med.* 28, 2333–2343 (2022).
- 866 22. Stamp, T. C. B. & Round, J. M. Seasonal Changes in Human Plasma Levels of 25867 Hydroxyvitamin D. *Nature* 247, 563–565 (1974).
- 868 23. Brewerton, T. D., Putnam, K. T., Lewine, R. R. J. & Risch, S. C. Seasonality of cerebrospinal
- 869 fluid monoamine metabolite concentrations and their associations with meteorological variables in
- 870 humans. J. Psychiatr. Res. 99, 76–82 (2018).
- 871 24. Martynova, E. et al. Seasonal Changes in Serum Metabolites in Multiple Sclerosis Relapse. Int.
- 872 J. Mol. Sci. 24, 3542 (2023).

- 873 25. Ma, Y. *et al.* Seasonal variation in food intake, physical activity, and body weight in a
 874 predominantly overweight population. *Eur. J. Clin. Nutr.* **60**, 519–528 (2006).
- 875 26. Küçükerdönmez, Ö. & Rakicioğlu, N. The effect of seasonal variations on food consumption,
- dietary habits, anthropometric measurements and serum vitamin levels of university students. *Prog. Nutr.* 20, (2018).
- 878 27. Chaleckis, R., Murakami, I., Takada, J., Kondoh, H. & Yanagida, M. Individual variability in
- human blood metabolites identifies age-related differences. *Proc. Natl. Acad. Sci.* 113, 4252–4259
 (2016).
- 28. Chen, Z.-Z. *et al.* Metabolite Profiles of Incident Diabetes and Heterogeneity of Treatment
 Effect in the Diabetes Prevention Program. *Diabetes* 68, 2337–2349 (2019).
- 883 29. Tebani, A. *et al.* Integration of molecular profiles in a longitudinal wellness profiling cohort.
 884 *Nat. Commun.* 11, 4487 (2020).
- 885 30. Kim, K. *et al.* Mealtime, temporal, and daily variability of the human urinary and plasma 886 metabolomes in a tightly controlled environment. *PloS One* **9**, e86223 (2014).
- 31. Ala-Korpela, M. *et al.* Cross-sectionally Calculated Metabolic Aging Does Not Relate to
 Longitudinal Metabolic Changes-Support for Stratified Aging Models. *J. Clin. Endocrinol. Metab.*108, 2099–2104 (2023).
- 890 32. Bergström, G. *et al.* The Swedish CArdioPulmonary BioImage Study: objectives and design. *J.*891 *Intern. Med.* 278, 645–659 (2015).
- 33. Zhong, W. *et al.* Next generation plasma proteome profiling to monitor health and disease. *Nat. Commun.* 12, 2493 (2021).
- 894 34. Zhong, W. *et al.* Whole-genome sequence association analysis of blood proteins in a
 895 longitudinal wellness cohort. *Genome Med.* 12, 53 (2020).
- Bioinformatics 33, 2037–2039 (2017).
- 898 36. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer
 899 datasets. *GigaScience* 4, 7 (2015).

- 900 37. McLaren, W. et al. Deriving the consequences of genomic variants with the Ensembl API and
- 901 SNP Effect Predictor. *Bioinformatics* **26**, 2069–2070 (2010).
- 902 38. Diab, J. et al. Lipidomics in Ulcerative Colitis Reveal Alteration in Mucosal Lipid Composition
- Associated With the Disease State. *Inflamm. Bowel Dis.* **25**, 1780–1787 (2019).
- 904 39. Wishart, D. S. et al. HMDB 5.0: the Human Metabolome Database for 2022. Nucleic Acids Res.
- 905 **50**, D622–D631 (2022).
- 40. Arya, S., Mount, D., Kemp, S. E. & Jefferis, G. RANN: Fast Nearest Neighbour Search (Wraps
 ANN Library) Using L2 Metric. (2019).
- 41. Hahsler, M., Piekenbrock, M. & Doran, D. dbscan: Fast Density-Based Clustering with R. J.
- 909 Stat. Softw. **91**, 1–30 (2019).
- 910 42. Csárdi, G. et al. igraph: Network Analysis and Visualization. (2024).
- 43. Hinrichs, A. S. *et al.* The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.*
- 912 **34**, D590–D598 (2006).

918

- 44. Mägi, R. & Morris, A. P. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* 11, 288 (2010).
- 915 45. Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. ImerTest Package: Tests in Linear
- 916 Mixed Effects Models. J. Stat. Softw. 82, 1–26 (2017).
- 917 46. Halekoh, U. & Højsgaard, S. A Kenward-Roger Approximation and Parametric Bootstrap

Methods for Tests in Linear Mixed Models – The R Package pbkrtest. J. Stat. Softw. 59, 1–32 (2014).

- 919 47. Kleiber, C. & Zeileis, A. AER: Applied Econometrics with R. Springer-Verlag (2008).
- 920 48. R Core Team and contributors worldwide. The R Stats Package.
- 49. Sun, B. B. *et al.* Plasma proteomic associations with genetics and health in the UK Biobank.
 Nature 622, 329–338 (2023).
- 923 50. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature*924 562, 203–209 (2018).
- 925 51. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. in Proceedings of the
- 926 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 785–794
- 927 (Association for Computing Machinery, New York, NY, USA, 2016). doi:10.1145/2939672.2939785.

- 928 52. Laboratory, N.-R. A. verification: Weather Forecast Verification Utilities. (2015).
- 929 53. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation
- and Projection. J. Open Source Softw. 3, 861 (2018).
- 931 54. Oksanen, J. et al. vegan: Community Ecology Package. (2022).
- 932 55. Allaire, J. J. et al. networkD3: D3 JavaScript Network Graphs from R. (2017).
- 933 56. Kolde, R. pheatmap: Pretty Heatmaps. (2019).
- 934 57. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize Implements and enhances circular
- 935 visualization in R. *Bioinforma. Oxf. Engl.* **30**, 2811–2812 (2014).
- 936 58. Read, Q. D. & Brunson, J. C. ggalluvial: Alluvial Plots in 'ggplot2'. (2023).
- 937 59. Wickham, H. ggplot2. WIREs Comput. Stat. 3, 180–185 (2011).
- 938 60. Team, R. R: A language and environment for statistical computing. MSOR Connect. (2014).
- 939 61. Wu, G. Amino acids: metabolism, functions, and nutrition. *Amino Acids* **37**, 1–17 (2009).
- 940 62. DeBerardinis, R. J. et al. Beyond aerobic glycolysis: Transformed cells can engage in glutamine
- 941 metabolism that exceeds the requirement for protein and nucleotide synthesis. *Proc. Natl. Acad. Sci.*
- **104**, 19345–19350 (2007).
- 843 63. Rodriguez-Cuenca, S., Pellegrinelli, V., Campbell, M., Oresic, M. & Vidal-Puig, A.
 844 Sphingolipids and glycerophospholipids The 'ying and yang' of lipotoxicity in metabolic diseases.
- 945 Prog. Lipid Res. 66, 14–29 (2017).
- 64. Grabner, G. F., Xie, H., Schweiger, M. & Zechner, R. Lipolysis: cellular mechanisms for lipid
 mobilization from fat stores. *Nat. Metab.* 3, 1445–1465 (2021).
- McMahon, H. T. & Gallop, J. L. Membrane curvature and mechanisms of dynamic cell
 membrane remodelling. *Nature* 438, 590–596 (2005).
- 66. Chang, E. & Kim, Y. Vitamin D decreases adipocyte lipid storage and increases NAD-SIRT1
 pathway in 3T3-L1 adipocytes. *Nutr. Burbank Los Angel. Cty. Calif* 32, 702–708 (2016).
- pathway In 515-11 addpocytes. Ivan. Barbank Los Iniger. Ciy. Carl <math>52, 702 -700 (2010).
- 952 67. Pedersen, E. B. et al. Plasma amino acids in Greenlanders and Danes: influence of seasons,
- residence, ethnicity, and diet. Am. J. Hum. Biol. Off. J. Hum. Biol. Counc. 18, 99–111 (2006).
- 68. Catoire, M. et al. Fatty acid-inducible ANGPTL4 governs lipid metabolic response to exercise.
- 955 Proc. Natl. Acad. Sci. U. S. A. 111, E1043-1052 (2014).

- 956 69. A, G. *et al.* Induction of cardiac Angptl4 by dietary fatty acids is mediated by peroxisome
 957 proliferator-activated receptor beta/delta and protects against fatty acid-induced oxidative stress.
 958 *Circ. Res.* 106, (2010).
- 959 70. Wu, Q. *et al.* Gamma-glutamyl-leucine levels are causally associated with elevated cardio960 metabolic risks. *Front. Nutr.* 9, 936220 (2022).
- 961 71. Badoud, F. *et al.* Serum and adipose tissue amino acid homeostasis in the metabolically healthy
 962 obese. *J. Proteome Res.* 13, 3455–3466 (2014).
- 963 72. Teo, C. H., Wong, A. C. H., Sivakumaran, R. N., Parhar, I. & Soga, T. Gender Differences in
- 964 Cortisol and Cortisol Receptors in Depression: A Narrative Review. *Int. J. Mol. Sci.* 24, 7129 (2023).
- 965 73. Aranda, J. V. *et al.* Metabolism of Theophylline to Caffeine in Human Fetal Liver. *Science* 206,
 966 1319–1321 (1979).
- 967 74. Diamanti, K. *et al.* Integration of whole-body [18F]FDG PET/MRI with non-targeted
 968 metabolomics can provide new insights on tissue-specific insulin resistance in type 2 diabetes. *Sci.*969 *Rep.* 10, 8343 (2020).
- 970 75. Oudemans-van Straaten, H. M., Bosman, R. J., Treskes, M., van der Spoel, H. J. & Zandstra,
- 971 D. F. Plasma glutamine depletion and patient outcome in acute ICU admissions. *Intensive Care Med.*972 27, 84–90 (2001).
- 973 76. Lacey, J. M. & Wilmore, D. W. Is Glutamine a Conditionally Essential Amino Acid? *Nutr. Rev.*974 48, 297–309 (1990).
- 975 77. Wen, Z. *et al.* Time-resolved online analysis of the gas- and particulate-phase of cigarette smoke
 976 generated by a heated tobacco product using vacuum ultraviolet photoionization mass spectrometry.
 977 *Talanta* 238, 123062 (2022).
- 978 78. Benson, M. D. *et al.* Protein-metabolite association studies identify novel proteomic
 979 determinants of metabolite levels in human plasma. *Cell Metab.* 35, 1646-1660.e3 (2023).
- 980 79. Van Dijk, G. The Role of Leptin in the Regulation of Energy Balance and Adiposity. J.
 981 Neuroendocrinol. 13, 913–921 (2001).

- 982 80. de Castro, M. A., Baltar, V. T., Marchioni, D. M. L. & Fisberg, R. M. Sex differences in serum
- leptin and its relation to markers of cardiometabolic risk in middle-aged adults: Evidence from a
 population-based study. *Nutrition* 31, 491–497 (2015).
- 81. Fluhr, L. *et al.* Gut microbiota modulates weight gain in mice after discontinued smoke
 exposure. *Nature* 600, 713–719 (2021).
- 82. Lewis, C. M. *et al.* Genome scan meta-analysis of schizophrenia and bipolar disorder, Part II:
 schizophrenia. *Am. J. Hum. Genet.* 73, 34–48 (2003).
- 83. Litwack, E. D., Babey, R., Buser, R., Gesemann, M. & O'Leary, D. D. M. Identification and

990 characterization of two novel brain-derived immunoglobulin superfamily members with a unique

991 structural organization. *Mol. Cell. Neurosci.* **25**, 263–274 (2004).

- 84. Takeuchi, A. & O'Leary, D. D. M. Radial Migration of Superficial Layer Cortical Neurons
 Controlled by Novel Ig Cell Adhesion Molecule MDGA1. *J. Neurosci.* 26, 4460–4464 (2006).
- 85. Dócs, K. *et al.* The Ratio of 2-AG to Its Isomer 1-AG as an Intrinsic Fine Tuning Mechanism
 of CB1 Receptor Activation. *Front. Cell. Neurosci.* 11, (2017).
- 86. Viveros, M. P., Marco, E. M. & File, S. E. Endocannabinoid system and stress and anxiety
 responses. *Pharmacol. Biochem. Behav.* 81, 331–342 (2005).
- 87. Kersten, S. Role and mechanism of the action of angiopoietin-like protein ANGPTL4 in plasma
 lipid metabolism. *J. Lipid Res.* 62, 100150 (2021).
- 1000 88. Kumar, A. A. *et al.* Plasma leptin level mirrors metabolome alterations in young adults.
 1001 *Metabolomics Off. J. Metabolomic Soc.* 16, 87 (2020).
- 1002 89. Lawler, K. *et al.* Leptin-Mediated Changes in the Human Metabolome. *J. Clin. Endocrinol.*1003 *Metab.* 105, 2541–2552 (2020).
- 1004 90. Alberti, K. G. M., Zimmet, P. & Shaw, J. The metabolic syndrome—a new worldwide definition.
 1005 *The Lancet* 366, 1059–1062 (2005).
- 1006 91. Stelmach-Mardas, M. *et al.* Seasonality of food groups and total energy intake: a systematic
 1007 review and meta-analysis. *Eur. J. Clin. Nutr.* **70**, 700–708 (2016).

- 1008 92. van Ooijen, A. M. J., van Marken Lichtenbelt, W. D., van Steenhoven, A. A. & Westerterp, K.
- 1009 R. Seasonal changes in metabolic and temperature responses to cold air in humans. *Physiol. Behav.*

1010 **82**, 545–553 (2004).

- 1011 93. Dopico, X. C. et al. Widespread seasonal gene expression reveals annual differences in human
- 1012 immunity and physiology. *Nat. Commun.* **6**, 7000 (2015).
- 1013 94. Wucher, V., Sodaei, R., Amador, R., Irimia, M. & Guigó, R. Day-night and seasonal variation
- 1014 of human gene expression across tissues. *PLoS Biol.* **21**, e3001986 (2023).
- 1015 95. Suhre, K. et al. Human metabolic individuality in biomedical and pharmaceutical research.
- 1016 *Nature* **477**, 54–60 (2011).

1017

1018 Figures

1019 Fig. 1



1020

Fig. 1 Longitudinal multi-omics profiling and co-expression analysis of human plasma
metabolites in 101 healthy individuals.

a Overview of the study design for the S3WP project, detailing the longitudinal data collection strategy
 over six visits in two years (created with biorender.com). This included whole-genome sequencing,

1025 proteomics, metabolomics, lipidomics, clinical measurements, and lifestyle questionnaires from 101 1026 individuals. A total of 527 identified metabolites were categorized into 9 major classes. **b** UMAP 1027 clustering of the 527 metabolites showing their co-expression patterns, color-coded by their classes and 1028 grouped by co-expression clusters. **c** Sankey plot showing the distribution of metabolic classes across 1029 the clusters. **d** Bar plot showing the functional enrichment results of the metabolite clusters using 1030 Fisher's exact test.

1031









1035 **a** The inter-individual and intra-individual variations of plasma metabolite levels, calculated as the 1036 coefficient of variation (CV) for each metabolite within each visit and across all participants, color-1037 coded by metabolite classes. b Seasonal variation analysis of plasma metabolites using amplitude analysis, color-coded by metabolic classes. Y-axis showing the adjusted *p*-values with multiple test 1038 1039 corrections using Benjamini and Hochberg method. c Heatmap showing the scaled levels of 121 1040 metabolites with significant seasonal variations across 12 months. d-g Plasma metabolite levels 1041 throughout the year for metabolites in Cluster M1-M4. h Pathway enrichment analysis of metabolites 1042 within each seasonal cluster. Significantly enriched pathways were defined with adjusted P-values < 1043 0.05 (Fisher's exact test with multiple test corrections using using Benjamini and Hochberg method). i. 1044 Vitamin D concentration levels across 12 months during the study period. j, Succinylcarnitine levels 1045 over 12 months as an example of metabolites in Cluster M3. Each line represents an individual; red 1046 lines indicate females and blue lines indicate males. Regression lines are added using trigonometric 1047 functions.





1049

1050 Fig. 3 Influence of clinical measurements and lifestyle factors on metabolite levels.

1051 a Canonical correspondence analysis (CCA) showing correlations between plasma metabolite levels, 1052 clinical measurements and lifestyle variables. The upper and left bar plots show the estimated linear 1053 regression coefficients for clinical and lifestyle variables with respect to CCA1 and CCA2. b, d Volcano 1054 plots showing the impacts of sex (b) and BMI (d) on plasma metabolite levels (Kenward-Roger 1055 approximation with Benjamini and Hochberg correction). c Concentration of gamma-glutamylleucine 1056 across four study visits, shown as an example of a sex-associated metabolite. Each line represents an 1057 individual; red lines indicate females and blue lines indicate males. e Scatter plot showing the 1058 correlation between glutamic acid concentration and BMI, color-coded by BMI.

1059 Fig. 4



1060



a Manhattan plot showing the identified mQTLs in the study. Significant loci are annotated based on the closest gene, with colors indicating the class of the corresponding metabolite. **b** Chord diagram of loci shared ($r^2 > 0.2$) among metabolites in GWAS study. Line thickness is proportionate to the -Log10(*P*). **c** Manhattan plot of butyrylcarnitine showing the genome locations of all associated mQTLs.

1066 **d** Boxplot showing the association between plasma levels of butyrylcarnitine and the genotype of 1067 rs34673751, color-coded by the genotype of rs34673751. e Plasma levels of butyrylcarnitine across 1068 study visits; each individual is represented by a line; color-coded by the genotype of rs34673751. f-g 1069 Western blot showing increased butyrylcarnitine levels in the ACADS know-down 293T cell lines 1070 compared to the control group. h Manhattan plot for 4-androsten-3alpha,17alpha-diol monosulfate (2), showing the associated genetic loci. i Boxplot showing the association between plasma levels of 4-1071 1072 androsten-3alpha,17alpha-diol monosulfate (2) and the genotype of rs45446698. j Plasma levels of 1073 butyrylcarnitine across study visits; each individual is represented by a line; color-coded by the 1074 genotype of rs34673751.





1076

1077 Fig. 5 Influence of genetic, clinical and lifestyle factors on plasma metabolite level variability.

1078 a Overview of influence of genetic, clinical and lifestyle factors on the plasma metabolite level 1079 variability. b Barplot showing the variance explanation fraction for each component across all 22

- 1080 genetic-related metabolites, color-coded by the variable classes. c Barplot showing the top 25 1081 metabolites most strongly associated with clinical components. d Barplot showing the top 25 1082 metabolites most strongly associated with lifestyle components.
- 1083 Fig. 6



1084

1085 Fig. 6 Characterization of protein-metabolite network

1086 a Network presenting the top 200 significant protein-metabolite pairs identified by the linear mixed 1087 model (LMM) (FDR-adjusted P < 0.05). Solid circles represent proteins in the inner ring, color-coded 1088 by protein annotation. Squares represent metabolites in the outer ring, color-coded by different 1089 metabolite-related influence factors. Pairs of related proteins and metabolites are connected by dashed 1090 lines (indicating correlations supported by LMM results) and solid lines (indicating correlations 1091 supported by both the LMM and Mendelian randomization (MR) analysis). Green lines indicate positive 1092 correlations between proteins and metabolites in the LMM, while red lines indicate negative 1093 correlations. b Scatter plot showing the correlation between aceturic acid concentration and LEP, color-1094 coded by sex. c Scatter plot showing the correlation between 1-Arachidonovlglycerol (1-AG) 1095 concentration and MDGA1, color-coded by different genotypes of rs9349050. d Manhattan plot for 1096 MDGA1 protein, showing the associated genetic variants with plasma levels of MDGA1. One of the 1097 most significant SNP (rs9349050) was used as an instrumental variable in the MR analysis. e UMAP 1098 clustering of the protein-metabolite profiles of the analyzed samples, color-coded by individual with 1099 lines connecting the visits for each individual. \mathbf{f}, \mathbf{g} Bar plots showing the variance explanation fraction 1100 of different genetic, clinical and lifestyle factors, calculated from linear mixed modeling, for UMAP1 1101 (f) and UMAP2 (g). h, i Distribution of betweenness score of proteins (h) and metabolites (i) in three 1102 tiers. **j**, **k** Dot plots highlighting the Tier 1 proteins (**j**) and metabolites (**k**)

1103 Fig. 7

1104





1105



1107 a Measurements of classical metabolic risk indicators, including HDL, BMI, SBP, TG and Gluc, 1108 compared between high risk and low risk groups. b Measurements of ALAT, GGT, urate, TNT and 1109 WBC for the high risk and low risk groups. c Area under the curve (AUC) of prediction models based 1110 on six different combinations of plasma proteins and metabolites: i) all proteins and metabolites; ii) tier

1111 1 proteins and metabolites; iii) all proteins; iv) tier 1 proteins; v) all metabolites; vi) tier 1 metabolites. 1112 The AUC values were calculated by performing random sampling 100 times to account for variability. 1113 d Average receiver operating characteristic (ROC) curves for the predictive models for the six 1114 combination of plasma proteins and metabolites for predicting metabolic risk levels. e, f Linear model 1115 coefficients obtained using Tier 1 proteins (e) and metabolites (f) as independent variables and a dummy 1116 variable to indicate if the sample belonged to the high risk or ow risk group. g Tier 1 proteins' profiles 1117 of healthy individuals, compared with obesity, T2D, gout or hyperthyroidism in the UK Biobank. h 1118 Average ROC curves obtained using Tier 1 proteins to predict individuals with either obesity, type 2 1119 diabetes (T2D), gout, hyperthyroidism in the UK Biobank database. i Alluvial plot showing the risk 1120 level of individual samples across visits. The y-axis represents the individuals involved in the study, 1121 with colors indicating different risk levels. *j* BMI levels of the individual W0010 during the study visits. 1122 k Changes in Tier 1 proteins over time in individual W0010 who lost a significant amount of weight 1123 (15.4kg) between visit 3 and visit 4. I Average ROC curve of predictive models using Tier 1 proteins to 1124 predict future obesity in the UK Biobank. The 95% confidence intervals (CI) of ROC curves were 1125 plotted in **d**, **h**, **k** Red; High risk group, blue; Low risk group. High density lipoprotein, HDL; body 1126 mass index, BMI; systolic blood pressure, SBP; triglycerides, TG; fasting glucose, Gluc; alanine 1127 aminotransferase, ALAT; gamma glutamyltransferase, GGT, troponin T, TNT; white blood cells count, 1128 WBC.

1129

1130 Additional file 1

- Supplementary Methods and Figures. Additional methods on experimental validation and figures S1-S6.
- 1133 Fig. S1 Co-expression clustering of 527 analyzed metabolites.
- 1134 Fig. S2 Seasonal variation analysis of plasma metabolites and proteins.
- 1135 Fig. S3 Mendelian randomization protein-metabolite network.
- 1136 Fig. S4 Associations between Tier 1 proteins and metabolic diseases.
- 1137 Fig. S5 Association between Tier 1 proteins and obesity.

Fig. S6 Prediction of future obesity in the UK Biobank.

1139 Additional file 2

- 1140 Table S1-S15.
- **Table S1.** Complete list of analyzed metabolites, proteins, and clinical parameters.
- **Table S2.** Classification and co-expression patterns of 527 analyzed metabolites.
- **Table S3.** Functional enrichment analysis of metabolite co-expression clusters.
- **Table S4.** Inter- and intra-individual variations of 527 analyzed metabolites.
- **Table S5.** Plasma metabolites with significant seasonal variation.
- 1146 Table S6. Temporal co-expression patterns of seasonal associated metabolites
- **Table S7.** Pathway enrichment analysis of metabolite co-expression clusters.
- **Table S8.** Temporal co-expression patterns of seasonally associated proteins.
- **Table S9.** Sex-associated plasma metabolites.
- **Table S10.** BMI-associated plasma metabolites.
- 1151 Table S11. Summary of identified independent mQTLs
- **Table S12.** Meta-analysis of identified independent mQTLs.
- 1153 Table S13. Contributions of genetics, lifestyle, and clinical factors to the variability of plasma
- 1154 metabolites.
- **Table S14.** Comprehensive mixed-effect modeling analysis of protein-metabolite associations.
- **Table S15.** Summary of one-sample mendelian randomization analysis.