

Menopause Hormone Replacement Therapy and Lifestyle Factors affect Metabolism and Immune System in the Serum Proteome of Aging Individuals

Clemens Dierks^{1,2}, Roza Sürme Mizrak¹, Orr Shomroni¹, Vadim Farztdinov^{1,3}, Kathrin Textoris-Taube^{1,3}, Daniela Ludwig^{1,3}, Johannes Rainer⁴, Michael Mülleder³, Ilja Demuth^{5,6}, Markus Ralser^{1,7,8*}

¹ Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt Universität zu Berlin, Institute of Biochemistry, Charitéplatz 1, 10117 Berlin, Germany

² Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt Universität zu Berlin, Department of Infectious Diseases, Respiratory Medicine and Critical Care, Augustenburger Platz 1, 13353 Berlin, Germany

³ Core Facility – High Throughput Mass Spectrometry, Charité Universitätsmedizin Berlin, 10117 Berlin, Germany

⁴ Institute for Biomedicine, Eurac Research, 39100 Bolzano, Italy

⁵ Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Department of Endocrinology and Metabolic Diseases (including Division of Lipid Metabolism), Biology of Aging working group, Augustenburger Platz 1, 13353 Berlin, Germany

⁶ Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Regenerative Immunology and Aging, BIH Center for Regenerative Therapies, 13353 Berlin, Germany

⁷ Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany

⁸ The Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7BN, UK

* Correspondence: markus.ralser@charite.de

28 Abstract

29 Aging is a fundamental risk factor for a wide array of diseases. The Berlin Aging Study II
 30 (BASE-II) is a cohort study designed to investigate the physical, mental, and social
 31 determinants of successful aging. We utilized high-throughput mass spectrometry to measure
 32 the proteomes of 1890 BASE-II participants, divided into two age groups: 27-37 years and 60-
 33 85 years. We employed multiple linear regression analyses to explore the effects of
 34 demographic factors such as age, sex, and BMI, along with hormonal treatments and lifestyle
 35 factors, on the serum proteome. We identify new associations and confirm previously
 36 described proteins linked to age, sex, BMI and hormonal contraceptive use (HCU). Notably,
 37 we observed that the abundance of nutrient transport proteins, particularly apolipoproteins, is
 38 linked to metabolic diseases in aged individuals, including metabolic syndrome and type 2
 39 diabetes. Additionally, we identified specific alterations explained by lifestyle factors, such as
 40 smoking and alcohol consumption. We further report a significant proteome signature in
 41 female study participants corresponding to menopause hormone replacement therapy (MHT).
 42 We successfully classified these participants based on MHT status with an AUROC of 0.82
 43 using two proteins, Complement Component 9 and Plasminogen, slightly outperforming
 44 estradiol (AUROC: 0.80), the active ingredient in most MHT preparations. Overall, our study
 45 underscores the impact of lifestyle and hormonal therapies on the serum proteome during
 46 aging, primarily affecting components of the immune system and metabolism.

Introduction

Aging is a primary risk factor for numerous chronic diseases and is associated with health deterioration and functional decline of the body, making it particularly important to investigate molecular changes that are related to age¹. While the underlying mechanisms are complex, the serum proteome offers huge potential for deciphering the interrelations of health and disease^{2,3}. The serum proteome responds to changes underlying aging, such as increasing levels of inflammatory proteins and decreasing tissue repair and cell regeneration activity⁴. Circulating serum protein level levels can help to differentiate between healthy aging and pathological conditions^{1,5-7}. Moreover, an increased understanding of the relationship between chronological and biological aging can contribute to delay or prevent diseases where chronological age is a major risk factor⁶ as well as improving patient stratification¹. Identifying markers of healthy aging is crucial to define appropriate lifestyle interventions and thereby enhancing lifespan⁸.

Broadly speaking, the serum contains two sets of proteins. Proteins that function in the blood, such as nutrient transport proteins, signaling molecules, proteins of the immune system and the coagulation machinery, and proteins, who circulate in the blood due to tissue leakage, cell necrosis, or to their presence in secreted extracellular vesicles. Several, but not all, of the biomarkers that are currently in use, are part of the former, and are often found within the highly abundant fraction of the plasma proteome, that is easily accessible and simplifies the development of routine applicable biomarker assays⁹⁻¹². In recent years, the application of multiplexed proteome technologies has enhanced our ability to probe the complex molecular landscape of the proteome in plasma or serum. Recent advancements in mass spectrometry-based proteomic platforms have addressed previous challenges related to throughput and cost, enabling large-scale recording of plasma proteomes¹³.

In this study, we recorded the serum proteomes of 1890 participants from the Berlin Aging Study II (BASE-II)¹⁴. The study aims to uncover factors contributing to healthy aging, involving in total 2200, with about two thirds belonging to the older group aged 60 to 85 years and about one third representing the younger group of participants aged between 27 to 37 years from the metropolitan area of Berlin, Germany. We obtained a quantitative dataset focussing on the neat serum proteome and the high abundant fraction of the serum proteome, and quantifies acute phase proteins, lipid transport proteins, proteins involved in the innate or adaptive immune system and coagulation factors. We report their dependency on age, common disease, selected lifestyle factors, and discover strong signals that are attributable to hormonal replacement therapy.

Results

Study Profile and Proteomic Assessment of the Berlin Aging Study II

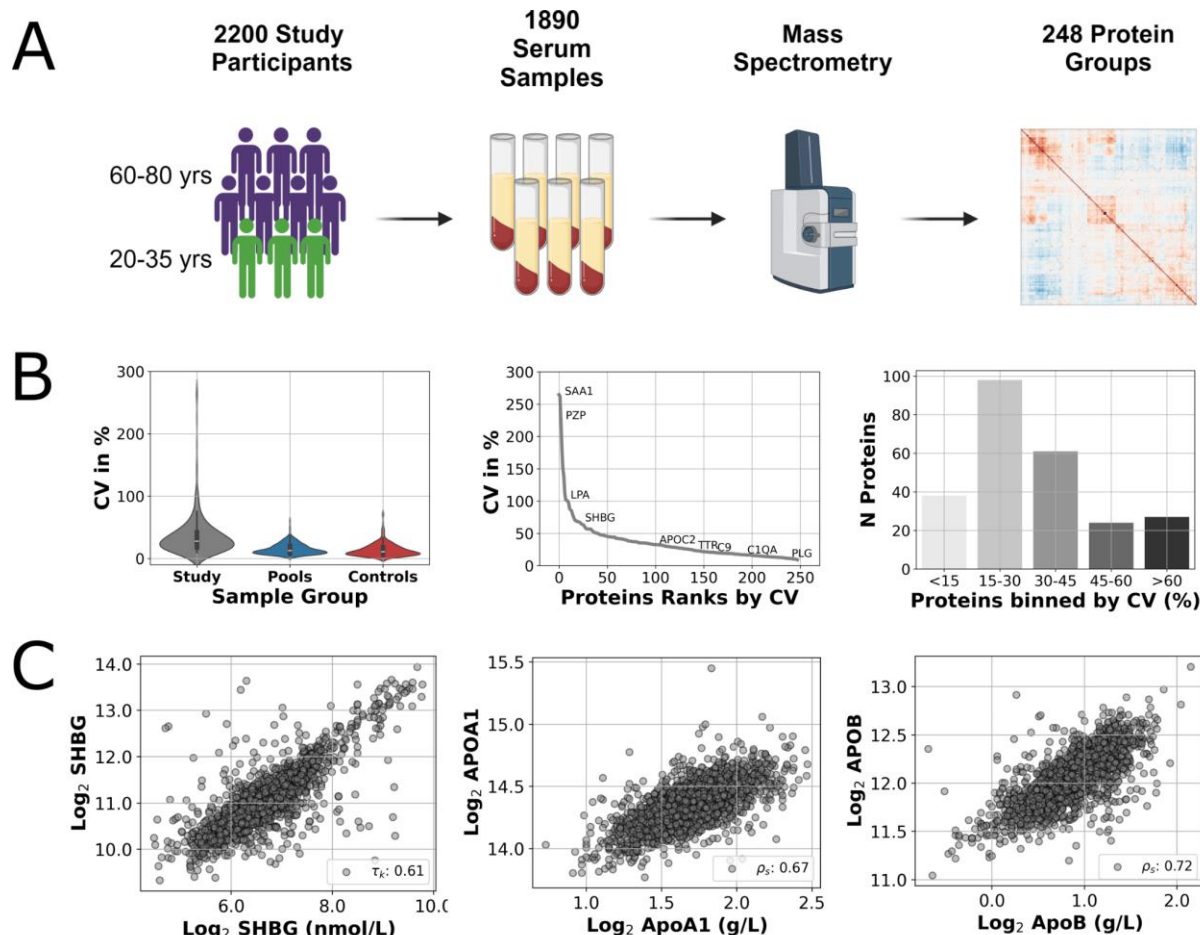


Figure 1: (A) Study Design. Schematic representation of the 2200 BASE-II participants, divided into an older and a younger control group. Serum samples of 1890 participants were analyzed using a high-throughput proteomics platform combining semi-automated sample preparation and latest trapped ion mobility spectrometry. (B) Distribution of coefficient of variation (CV) per protein for study data (gray), pooled study samples (blue) and standardized control serum samples (red) (Violin plots). The line plot depicts proteins ranked by their CV, while barplots show the CV distribution across all proteins. (C) Correlations of sex hormone binding globulin (SHBG), apolipoprotein A1 (APOA1) and B (APOB) concentrations on the y-axis versus the values obtained with related or corresponding routine diagnostic tests (left to right). Diagnostic test values were transformed to log2 for visualization purposes.

We conducted MS-based proteomic measurements on a total of 2091 serum samples, including 105 serum control and 96 pooled study samples, to generate a proteomic baseline from the BASE-II. The samples analyzed are derived from two age groups: older adults (N = 1433) aged 60 to 85 years and younger adults (N = 426) aged 27 to 37 years. Study demographics are given in Table 1.

The samples were prepared for measurement in a semi-automatic workflow that allows for high-throughput and consistent sample quality including tryptic digestion and sample cleanup with solid-phase extraction¹⁵. Samples were divided into 8 batches, each consisting of three 96-well plates, and were subjected to mass spectrometric measurement using a timsTOF Pro1 mass analyzer (Bruker Daltonics) operating in diaPASEF mode, using analytical flow rate chromatography using^{16,17}. Raw MS data was processed by DIA-NN^{18,19}, using the DiOGenes spectral library²⁰. In the subsequent preprocessing the data was normalized, filtered for outlier samples and low-presence peptides, imputed and corrected for plate effects (N = 24). Among the data set, 31 samples were removed due to low technical quality, or missing metadata (age and BMI) leaving the dataset with a total of 1859 proteomic samples from BASE-II participants. After this preprocessing, the final dataset quantified 2079 proteotypic peptides derived from 248 protein groups (Table S1).

| Variable | Young (N=426) | Old (N=1433) |
|------------------------------------|---------------|--------------|
| Age in years [median (IQR)] | 28.9 [4.5] | 68.8 [5.4] |
| Sex (N, Women in %) | 226 (55) | 742 (52) |
| BMI (N, %) | | |
| Underweight (BMI < 18.5) | 29 (7) | 6 (0.4) |
| Healthy (18.5 ≤ BMI < 25) | 278 (67.3) | 508 (35.6) |
| Overweight (25 ≤ BMI < 30) | 81 (19.6) | 644 (45.1) |
| Obese (BMI ≥ 30) | 25 (6.1) | 269 (18.9) |
| HCU (N, % women) | 90 (38.5) | |
| Smoker (N, %) | | |
| Never | 214 (50.2) | 682 (47.7) |
| Former (> 1 year) | 64 (15.0) | 597 (41.8) |
| Former (< 1 year) | 15 (3.5) | 12 (0.8) |
| Current | 133 (31.2) | 138 (9.7) |
| Diabetes (N, %) | 1 (0.002) | 177 (12.7) |
| Metabolic Syndrome (N, %) | 31 (7.3) | 508 (35.9) |
| Alcohol intake (N, %) | | |
| Occasional (≤ 1.3 g/day) | 60 (14.2) | 154 (11.2) |
| Low Volume (1.3 < X ≤ 30 g/day) | 334 (79.2) | 1020 (74.1) |
| Excessive (> 30 g/day) | 28 (6.6) | 202 (14.7) |
| MHT (N, %) | | |
| Never | | 295 (42.8) |
| Completed | | 359 (52.1) |
| Ongoing | | 35 (5.1) |

Table 1: Study Demographics of study participants subjected to the analysis stratified by age groups. HCU: Hormonal Contraceptive Use, MHT: Menopause Hormone Replacement Therapy

We calculated a median coefficient of variation (CV) of protein quantities of 12.74 % in study pools and 10.89 % for serum controls samples (Zen-Bio). Reflecting the biological signal, protein quantities in study samples deviated from the technical controls (CV of 27.79 %; Table S1)). While most of the proteins are normally distributed, some proteins, such as lipoprotein (a) (Lp(a)) and complement factor H-related protein 1 (CFHR1), exhibited tailed or almost bimodal distributions (Figures S59, S158). Tailed distributions in certain proteins, particularly

pronounced among the younger age group, were found for proteins that we recently associated to the intake of hormonal contraceptives in younger women (including angiotensinogen (AGT), fetuin B (FETUB), sex hormone binding globulin (SHBG), plasminogen (PLG) or peptidoglycan recognition protein 2 (PGLYRP2))²¹. (Figures S2 - S254 for Log2 transformed protein abundance distributions of all proteins across the entire cohort, as well as stratified by age groups).

We analyzed functional groups, such as immunoglobulins, coagulation factors, acute-phase proteins, complement factors and lipid transport proteins for their variation. Immunoglobulins (N = 60 protein groups) had the highest median relative variation (CV 44.3 %, range: 23.6 - 264.8 %), and included the proteins with the highest individual relative CVs such as IGHV3-11 (264.8 %), IGHV3-21 (183.3 %) and IGLV4-60 (150.6 %). In contrast, proteins involved in the complement system (N = 36) exhibited lower variation (median CV: 18.4 %, range: 10 - 63.5 %), including some of the least variable proteins such as the two central mediators of the component system C5 and C3 (both 10.4%). Proteins related to coagulation (N = 30), lipid transport (N = 20) and acute-phase response (N = 26), showed CVs of 19.6 % (range: 11.5% - 138.2%), 23.5% (12.9 % -102%) and 22% (10.4% - 262.6%). Besides IGHV3-11, serum amyloid A1 (SAA1; 262.6%) and the pregnancy zone protein (PZP, 225.1%) showed the highest variation in the study samples, while hemopexin (HPX, 9.2%), complement factor 1R (C1R, 10%) and attractin (ATRN, 10.2%) the lowest. CVs for study samples, pooled study samples and QC samples are shown in Tables S3-S5. Furthermore, a detailed functional analysis of the proteins in the dataset is provided in Table S2.

Next, we conducted a correlation analysis between the values obtained with routinely applied biomarker assays and the corresponding protein abundance levels. For interpreting these results, it is important to keep in mind that proteomics and the diagnostic tests are typically targeting related but not identical analytes. We obtained significant correlations, displayed as spearman's rho if not described otherwise, between proteomics and biomarker assays for apolipoprotein A1 levels (0.67) and B (0.72), low density lipoproteins (LDL) with APOB (0.72) and high-density lipoproteins (HDL) with APOA1 (0.66). Triglyceride measurements showed a moderate correlation with APOC3 levels (0.44). In contrast, diagnostic assays and protein abundance determined by proteomics correlated weakly for albumin (0.32), hemoglobin with the MS acquired hemoglobin subunit alpha 2 (HBA2; 0.31), and for fibrinogen compared to fibrinogen alpha chain levels (0.15). Notably, routine test sex hormone binding globulin (SHBG) measurements (prior to log2 transformation) displayed a robust, albeit not strictly linear relationship, with proteomic data (Kendall's tau: 0.61, see Figure S250h). The diverging results show that proteomic measurements provide complementary data to establish diagnostic assays (see Figure S250).

Hormonal contraceptive usage (HCU) is a dominating covariate affecting the serum proteome in young women

We investigated the effect of demographic covariates, such as sex, age, BMI, and HCU due to previous findings²¹ on the serum proteome of the 1890 BASE-II participants of both age groups by fitting multiple linear regression models to the protein abundance data. In our analysis BMI was treated as a categorical variable comparing underweight (U, BMI < 18.5),

overweight (Or, $25 < \text{BMI} \leq 30$) and obese (Ob, $\text{BMI} > 30$) participants to those with a healthy BMI ($18.5 < \text{BMI} \leq 25$). This categorization was chosen because these groups have higher clinical relevance than the actual, numerical BMI value and are commonly used in clinical studies. Besides the significance level of 0.05, we also required the abundance difference of the protein to be higher than the CV of the pooled study samples. Statistical summaries of the association analysis for the entire cohort and separated by age groups are reported in Tables S3-S5.

We found 106 protein groups associated with age. Of these, 65 have also been identified as age-dependent in orthogonal studies^{21–30}. Among the 41 new associations identified, 32 were immunoglobulin chains. Other newly found age-associated proteins include serum amyloid A4 (SAA4), proteoglycan 4 and the two platelet-related serum proteins, platelet factor 4 variant 1 (PF4V1) and pro-platelet basic protein (PPBP). Amongst the group of age-associated proteins, we report the strongest negative association, with up to 20% abundance change in 10 years, for serpin family F member 2 (SERPINF2), immunoglobulin delta heavy chain (P0DOX3) and transferrin (TF). Highest increase with progressing age was found in complement component 5 (C5), apolipoprotein B (APOB) and lumican (LUM).

The above defined age-related differences in the proteome were in part sex dependent, with 23 proteins showing a different age-response in male and female study participants. We detected the highest significant differences between men and women in the pregnancy zone protein (PZP), glycosylphosphatidylinositol specific phospholipase D1 (GPLD1), complement component 6 (C6), adiponectin (ADIPOQ) and ceruloplasmin (CP). None of the proteins investigated were associated with an underweight BMI, however, 19 proteins were linked to overweight and obesity, respectively, with APOC2 and fibulin 1 (FBLN1) showing novel associations with elevated BMI^{21,43–47}.

We found new associations with HCU in 15 out of 55 proteins that had not been reported in the related studies before^{21,48–50}, such as the phospholipid transfer protein (PLTP), monocyte differentiation antigen (CD14), alpha-1-antichymotrypsin (SERPINA3) and the coagulation factor V (F5) (see Figure 2). Among all covariates, highest effect sizes and significance were reported for HCU (see Figure 1D) with the top 5 candidates being angiotensinogen (AGT), peptidoglycan recognition protein 2 (PGLYRP2), serpin family A member 6 (SERPINA6), fetuin B (FETUB) and ceruloplasmin (CP). We recently investigated the plasma proteomes of 3,472 participants from the CHRIS cohort^{21,51}, an Alpine population study examining the genetic and molecular backgrounds influencing the incidence and progression of common chronic diseases. To validate our previous findings and demonstrate reproducibility, we compared the effect sizes for age, sex, and BMI categories for the overlapping proteome from the BASE-II analysis with those obtained from a similarly conducted association analysis of the CHRIS dataset, both measured on similar mass spectrometry platforms. We observed high concordance: age (Spearman's rho: 0.69), sex (0.85), BMI (obese category: 0.73), and HCU (0.9) (see Figures S256). However, the effect sizes for the underweight BMI category showed a low correlation between the studies (-0.25). Among the 125 overlapping proteins identified in both proteomic datasets, we found 53 proteins to be age-dependent and 13 proteins to be sex-dependent in both studies. The age-associated proteins include various functional groups involved in disease pathology, such as components of the complement cascade (C5, C7, C9, C4A, C4BPA, C1, CFD, CFI), apolipoproteins (APOB, APOC3, APOD, APOE, APOH), and proteins involved in blood clotting, including coagulation factors X (F10)

and XII (F12) and antithrombin III (SERPINC1). Additionally, we identified acute-phase proteins such as haptoglobin (HP) and transthyretin (TTR).

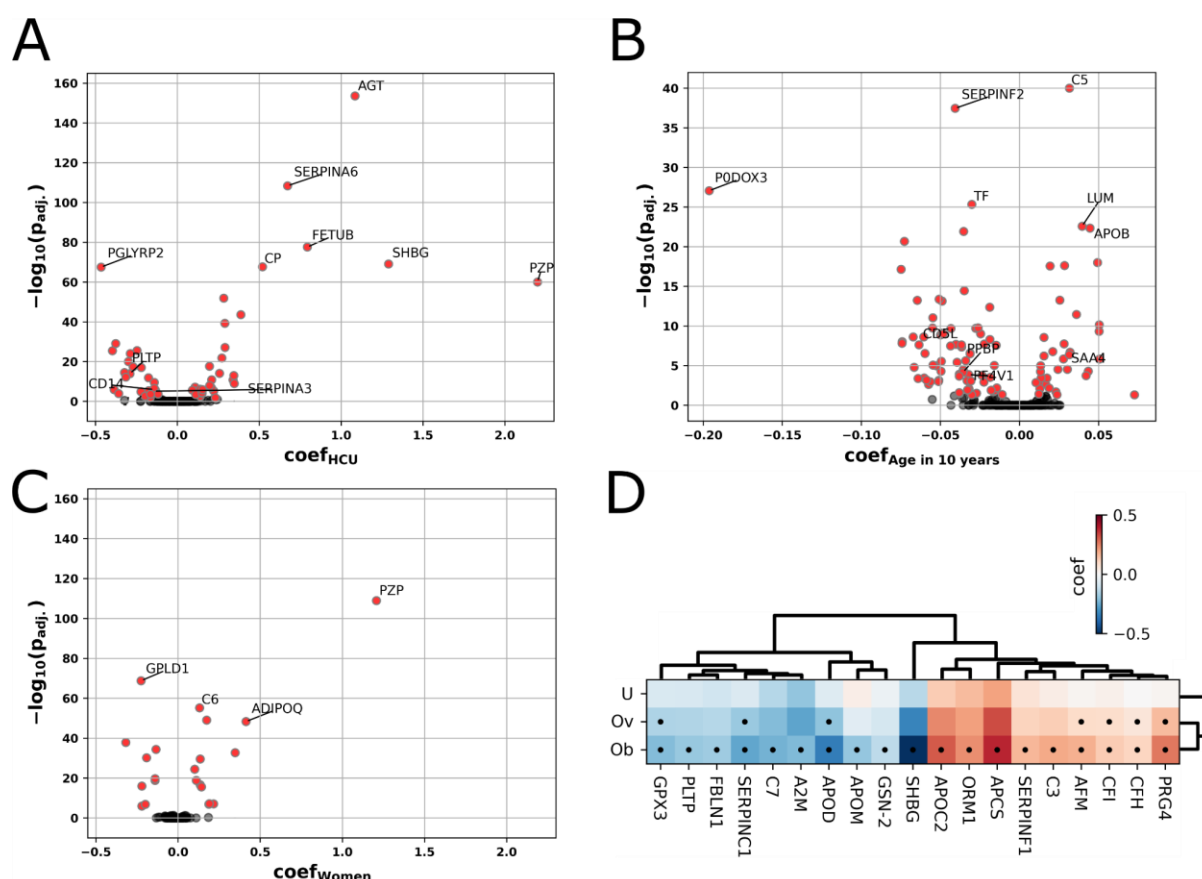


Figure 2: Multiple linear regression was employed to identify significant protein associations with hormonal contraceptive use (HCU), age, sex and different BMI categories (U: Underweight, Ov: Overweight, Ob: Obese). (A) The impact of HCU on the serum proteome. Proteins with largest coefficients and highest significance (in negative log10 adjusted p-values) for proteins with use of hormonal contraceptives. (B) The impact of 10-years of aging on the serum proteome. (C) The impact of sex (men vs. women) on the plasma proteome in Base-II. Proteins of significance are highlighted in red, and gene names are annotated for those discussed in the text. (D) Proteins associated with the different BMI categories (heatmap). Black dots denote significant associations after multiple testing corrections, while color indicates effect size by its coefficient.

Lifestyle factors show strong associations with apolipoproteins and inflammation markers in the serum proteome

Building on existing knowledge of lifestyle factors impacting the serum proteome⁵², we investigated the effects of smoking habits, alcohol consumption, diabetes mellitus type 2 (T2DM), or metabolic syndrome (metS) on protein abundance as covariates in our multiple regression models in addition to age, sex, BMI and HCU. We only included HCU as a covariate in the younger age group. Adjusted p-values, coefficients and effect sizes obtained from the regression analyses of significantly associated proteins with lifestyle factors and

comorbidities are reported in Tables S8 - S12. On this basis we addressed the association of selected individual lifestyle factors, i.e. smoking or alcohol consumption, with the serum proteome.

First, we addressed proteomic changes in current and former smokers. No significantly regulated proteins were found in former smokers that quit smoking more than a year ago. On the other hand, older participants who recently quit smoking (less than a year) had a significantly elevated level of polymeric immunoglobulin receptor (PIGR), which was also true for current smokers compared to participants who had never smoked ($n = 896$, see Figure 3B). Additionally, immunoglobulin heavy constant gamma 2 (IGHG2) levels were significantly decreased in current smokers from the older age group. IGHG2 levels were also found lowered in the younger group, albeit not significant. Alcohol consumption of the participants was grouped into occasional (< 1.3 g/day), low volume and excessive drinkers (>30 g/day) based on their self-reported habits. Among the 8 significantly associated proteins, the most changes were obtained for apolipoprotein C3 (APOC3) and thyroxine-binding globulin (SERPINA7) (see Figure 3C). Both proteins were significantly changed in low volume and excessive drinkers compared to participants that only occasionally drink alcohol, though significance increased with age.

Next, we continued to address protein associations with metS in the entire cohort and type 2 diabetes in the older group. The abundance levels of four different apolipoproteins were found to be significantly associated with metS. Specifically, apolipoprotein D (APOD) and apolipoprotein A1 (APOA1) were decreased in participants diagnosed with metS, whereas apolipoprotein C2 (APOC2), and apolipoprotein C4 (APOC4) were increased (see Table S8).

Visceral fat accumulation is a key feature of metabolic syndrome and T2DM^{53,54}. We analyzed BMI-associated proteins in "overweight" and "obese" categories, considering T2DM and/or metabolic syndrome (metS) as covariates. Without lifestyle factors, 19 proteins were linked to "overweight" and/or "obese" BMI categories (see Table S3). Including T2DM as a covariate, dopamine-beta hydroxylase (DBH) was significantly associated with "obese". In regression analyses with T2DM or metS, PGLYRP2⁵⁶, SERPINA10⁵⁷, and LYZ⁵⁸ were significant in both elevated BMI categories, which aligns with current literature. Fibulin 1 (FBLN1) remained significantly associated ($p < 0.05$) with both BMI categories, irrespective of covariates. The link between BMI and angiogenesis inhibitor FBLN1 downregulation has not been previously observed. Among the older participants $N = 177$, or 9.58% of all participants, were diagnosed with T2DM. This prevalence aligns closely with the previously reported rate of 9.9% for T2DM in Germany⁶², based on statutory insurance data⁶³. Multiple linear regression analysis revealed a positive, significant association of APOA4 with T2DM. In an analysis of 539 participants diagnosed with metS, including some T2DM diagnosed participants, APOA4 (see Figure 3A) again emerged as the only significant protein whose variance exceeded technical variability when assessing biomarkers for T2DM.

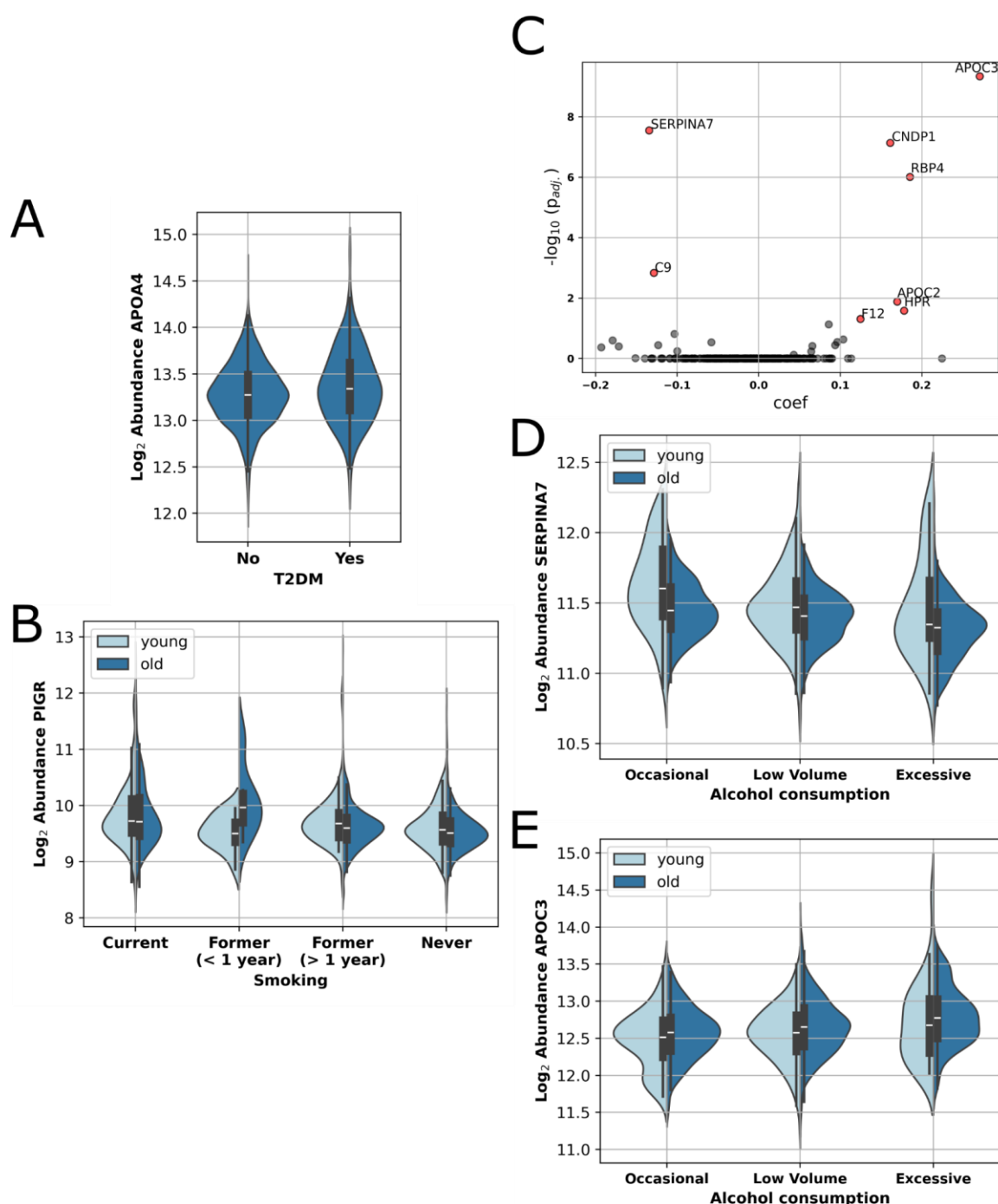


Figure 3: (A) Increased APOA4 abundance with age and with diabetes mellitus type II (T2DM) (B) Effect of smoking on the PIGR abundance levels. Coefficients and adjusted log₁₀ p-values for the proteomic profiles across different participant groups (Volcano Plot) (C). Proteins significantly associated with alcohol consumption are marked in red and their gene names are annotated. Accompanying split violin plots depict decreasing SERPINA7 (D) and increasing APOC3 (E) levels in BASE-II participants with increasing alcohol consumption. Boxes in violin plots denote the 25% and 75% quantiles, and the median is displayed by a white line.

Considering the established association between metS and T2DM, we performed multiple regression analysis on the complete cohort by including for metS and T2DM as covariates

^{64,65}. The only protein that could exceed the technical threshold when both diseases are considered as covariates was APOB, which had a significant, positive association with metS but did not show a significant one for diabetes.

Impact of menopausal hormone therapy on the high abundant serum proteome

We recently reported that HCU has a major impact on the high abundant fraction of the plasma proteome, in women under 40 years enrolled in the CHRIS study²¹, and as aforementioned, we detected a similar signature in the young women within the BASE-II cohort. With the focus on aging, we herein explored the effect of menopausal hormonal replacement therapy (MHT) in older women participating in the BASE-II. For this, we categorized the BASE-II women from the older group (N = 689, aged 61-84) into three groups: those who never underwent MHT (*never*, N = 295), those who completed MHT (*completed*, N = 359) and those who underwent treatment at sampling time (*ongoing*, N = 35). Only study participants with documented start and/or termination of MHT were considered. The median treatment duration for the MHT *completed* group was 11 years (IQR 8 years), and for the *ongoing* group, it was 18 years (IQR 9 years), reflecting a later onset of menopause⁷⁰. Subsequently, we employed multiple regression models incorporating age, BMI, and a categorical variable denoting MHT status to evaluate associations between protein abundances and the treatment.

We identified three proteins - angiotensinogen (AGT)⁷¹⁻⁷⁴, peptidoglycan recognition protein 2 (PGLYRP2)⁷⁵ and plasminogen (PLG) - to be significantly changed between study participants that are currently undergoing MHT treatment and those who have never received treatment. Additionally, we observed non-significant upregulation of ceruloplasmin (CP) and non-significant downregulation of vitamin K-dependent protein S (PROS1), C4b-binding protein alpha chain (C4BPA), lumican (LUM), and complement component 8A (C8A with MHT in our analysis).

Next, we investigated the predictive power of the serum proteome for classifying MHT status of the female BASE-II participants based on the serum proteome to enhance understanding of the underlying biological processes affected by hormonal treatments and add to already available literature concerning treatment effects on the serum proteome. We first matched female study participants under ongoing treatment (N = 35) to an equally sized control group by age and BMI based on propensity scores acquired from a generalized linear model⁷⁶ (Age and BMI demographics are listed in Table S7). Next, we selected plasminogen (PLG) and complement component 9 (C9) as the most informative features by a cross-validated, recursive feature elimination process⁷⁷. We then trained a random forest⁷⁸ in a 5-Fold cross validation stratified by MHT status. The two-protein classifier achieved a mean test data area under the curve (AUC) of 0.82, comparable to the AUC of 0.80 for individual estradiol levels, the active ingredient in most MHT. Moreover, since a training AUC of 0.84 was obtained, the classifier should be robust against overfitting. ROC curves for different splits in the CV procedure can be found in Figure S258. Other hormones such as SHBG or testosterone were outperformed by the random forest predictor (Figure S259). To get a better understanding of the decision boundaries of the two proteins in the random forest classifier we predicted decision probabilities for the range of observed PLG and C9 fold changes (FC) to the log2

median abundance values (Figure 5B). The highest predictive probability for ongoing MHT was at PLG at FC above 0.2 and C9 log2 FCs below -0.1 or above 0.4, although the latter has to be interpreted with caution since it is not well displayed in our data. Conversely, we expect the highest predictive power for no MHT at PLG levels below 0.2 and C9 levels between -0.1 and 0.4. Ensemble models that rely on multiple weak learners rarely produce predicted probabilities close to 0 or 1 because it requires all the individual models to make the same decision. This is particularly unlikely in random forests due to the bootstrapping (random sampling with replacement) used to create each individual tree ⁷⁹.

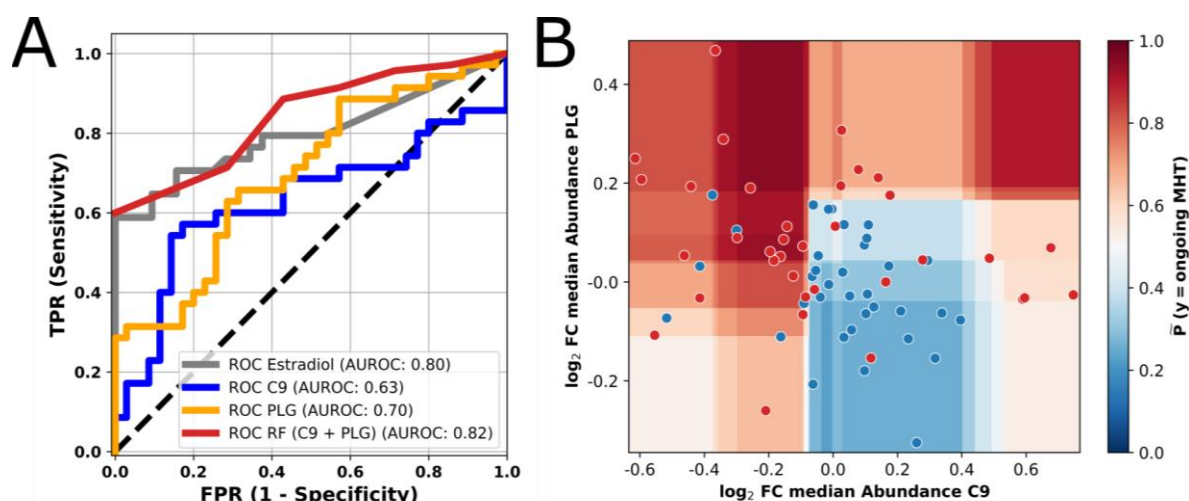


Figure 4: Machine learning classifier based on Complement Component 9 (C9) and Plasminogen (PLG) abundance levels slightly outperforms serum estradiol levels in predicting MHT status (A) Receiver operating characteristic (ROC) curve shows predictive performance of single parameter models with estradiol (coloured in gray), C9 (blue), PLG (orange) and the two protein random forest model (C9 + PLG) (red) predicting current hormonal treatment. (B) Contour plot depicting the predictive probability generated by the random forest classifier based on C9 and PLG levels. The areas highlighted in red represent the abundance ranges of PLG and C9, which support the classification of participants undergoing MHT. The added dots display the protein abundance measurements with participants currently under hormonal treatment (red) and controls (blue). Vertical and horizontal color gradations show cut-offs of the individual decision trees in the random forest model. Axis show the log₂ FC to the median abundance of PLG (y) and C9 (x).

Discussion

Large-scale serum and plasma proteome studies ^{21,43,80,81} have revealed disease prevalence, risk factors, and healthcare needs across diverse populations. These studies have identified putative biomarkers of aging ^{1,21-24,80}, common lifestyle factors like Type II Diabetes, and administered drugs ^{43,82,83}. Advancing age significantly heightens the risk for various chronic ailments alongside physical and cognitive decline, ultimately leading to mortality. Consequently, age directly impacts critical physiological processes, including inflammation ⁸⁴, often referred to as 'inflammaging', and cellular senescence ⁸⁵. However, biomarkers associated with aging have yet to gain widespread acceptance within the biomedical field,

partly due to the non-trivial nature of validating them ⁸⁶. The BASE-II ¹⁴ specifically aims to explore factors influencing a healthy aging process, presenting a rich resource for further investigations of age-related proteins.

In this study, mass spectrometry-based serum proteomics of 1859 BASE-II participants consistently quantified 2917 peptides from 248 protein groups at high precision. While the number of proteins in this upper fraction of the plasma proteome may seem limited, this group of proteins is secreted in many isoforms and is particularly interesting for biomedicine. Altogether, these proteins account for more than 99% of the plasma's proteomic mass. This high abundant fraction covers many of the currently clinically used protein biomarkers, spans over many targets of FDA approved biomarkers, and closely reflects human physiology. Thereby they cover many functions in the complement system, coagulation, immunoglobulin, and acute-phase response ^{10,15,87–91}.

Epidemiological covariates such as age, sex and BMI have a considerable impact on the general health of the population, but the mechanisms are multifactorial and complex. It has been shown that progressing age and very low or high BMI can pose a risk factor for multiple disease mechanisms and their progression ^{92–94}. With the advances of personalized medicine, these factors are important for treatments tailored to individual patients ⁹⁵. Specifically, the high abundant serum or plasma proteome that plays a vital role in disease pathology and metabolism, has to be investigated for these effects to reveal true relationships of treatment and disease ^{9–12}. Furthermore, differences in epidemiological factors also pose a technical challenge when analyzing proteomic data. Common techniques to address this include balancing or using epidemiological information as covariates in linear (factorial) models ⁹⁶.

Previous proteomic studies into aging cohorts successfully predicted age, mortality, and made attempts to distinguish biological from chronological age ('proteomic clock') ^{5,22–24,97}. The BASE-II design is slightly different to these investigations, as sampling was not continuous over the age range, but rather categorical, and sampled from an aged group and a young control. The BASE-II is well annotated with demographic and metadata, which facilitated to probe for association of age with several demographic and physiological factors. Our association analysis did both, confirm proteins that were previously reported as age-dependent in the literature, as well as identify new proteins. Indeed, the abundance levels of 106 of 248 consistently quantified proteins were changed with age. Of these, 39% (N = 41) were not reported as age-associated in previous datasets ^{21–30}. Thirty-two of these are immunoglobulin chains, which is consistent that the adaptive immune system changes as we age ⁹⁸. We hope our results stimulate future investigations using antigen-capture technologies, to identify the biological targets of these immunoglobulins.

Furthermore, we report associations to some pathologically highly relevant proteins including the pro-inflammatory serum amyloid A4 (SAA4), which has been linked to various infectious diseases such as malaria or COVID19 ^{99,100} as well as rheumatic inflammation ¹⁰¹. Another protein identified as age-dependent in our study is proteoglycan 4 (PRG4), which has been associated with tissue regeneration ³¹ and inflammation regulation ^{32,33}. It also functions as a boundary lubricant at the cartilage surface of joints, which function is known to decrease with progressing age. However, a direct connection to osteoarthritis, which can be caused due to decreased cartilage lubricants ¹⁰² are not yet described. In contrast to our findings, a negative regulation of PF4V1 and PPBP with age platelet activity has been described while platelet

counts were shown to increase with age ¹⁰³. Platelet factor 4 variant 1 (PF4V1)¹⁰⁴ has been associated with angiogenesis, which, if pathogenic, can serve as a marker for autoimmune diseases, cancer, and cardiovascular diseases. Additionally, pro-platelet basic protein (PPBP) has been reported to have a wide range of functions, including neutrophil activation and antibacterial activity ¹⁰⁵.

The sex dependency of the BASE-II proteome aligns well with the results obtained in other cohorts ^{21,23,34–42}. The list of $n = 23$ sex dependent proteins includes the pregnancy zone protein (PZP), the sex hormone binding globulin (SHBG) and angiotensinogen (AGT). Moreover, as we have reported recently, HCU had the strongest impact on the serum proteome of BASE-II participants, with PGLYRP2 being the main suppressed protein, and AGT, SERPINA6, FETUB the most induced. We found new HCU associations for 15 proteins, such as the acute-phase protein SERPINA3 ¹⁰⁶, the monocyte and macrophage cell surface protein CD14 ¹⁰⁷ and the coagulation Factor 5 (F5).

Furthermore, we compared effect sizes for age, sex, BMI, and hormonal contraceptive usage with those from our recently presented analysis of the CHRIS dataset ²¹. We found correlations of up to 0.9 for HCU (Spearman's rho), indicating a consistent effect on the plasma and serum proteome across both studies. Found age-associated proteins ($N=53$) in both studies included various components of the complement system ($N=8$) and apolipoproteins (5). Hence, the previously described age association of these mechanisms aligns with our results ^{36,109,111–113}. The most prominent sex-dependent proteins in both the CHRIS and BASE-II proteomic datasets include SHBG, which is recognized as a key proteomic marker for sex differences, complement component 6 (C6), TTR, and ceruloplasmin (CP). Effect sizes for underweight BMI showed a negative correlation (-0.25), likely due to the small number of participants in this category ($N=35$), resulting in less consistent results.

We also addressed the role of body weight which is linked to various comorbidities ¹¹⁴. We did not find significant associations with underweight, which was most likely caused by the low sample size and hence lower statistical power. Similar to previous studies¹¹⁵, we detected an association of apolipoproteins D and M levels with elevated BMI. We also report novel protein associations with BMI including Apolipoprotein C2 (APOC2) and fibulin 1 (FBLN1), when corrected for metabolic disorders such as T2DM and metS. When both metabolic conditions (metS and T2DM) are considered as covariates, C5 emerges as a significantly upregulated protein in obese patients. This could be due to its role in regulating inflammation in adipose tissue ¹¹⁶, but also aligns well with our report, that the complement system is BMI dependent ¹¹⁷. Interestingly, this association is sensitive to metS and T2DM, which suggests that the onset of metabolic disease influences the complement system. Similarly, FN1 ¹¹⁸, VWF ¹¹⁹, ECM1 ¹²⁰, and ADAMTS13 ¹²¹ are only significantly downregulated in obese and overweight patients when metS and T2DM are included as covariates. Furthermore, we found DBH significantly associated with “obese” BMI, when adjusted for T2DM in the older group of participants, aligning with prior findings in DBH-deficient mice ⁵⁵.

BASE-II offers a valuable insight into proteomic signatures of metabolic disease like metS and T2DM as the incidence within our cohort aligns with previously reported national numbers ^{63,122,123}. Our findings indicate that lifestyle factors like T2DM and metS diagnosis, influence

the serum proteome to varying extents across different age groups. Notably, apolipoprotein A4 (APOA4) was significantly upregulated in T2DM cases, particularly when metS was considered. APOA4, primarily synthesized in the small intestine, is crucial for lipid metabolism, insulin sensitivity, and food intake regulation. Elevated APOA4 levels are also markers for kidney disease progression and diabetic nephropathy,¹²⁴ highlighting its potential as a marker for T2DM monitoring and prognosis. Significant associations were also found between metS and various apolipoproteins, including C2, C4, B, and D. Apolipoprotein D (APOD) was consistently downregulated, especially in the older cohort. APOD, involved in lipid trafficking and neuroprotection, enhances insulin sensitivity and metabolic health¹²⁵ but is often reduced in metS patients, correlating with hyperchylomicronemia¹²⁶, characterized by increased triglyceride and chylomicron levels in blood and dysfunctional APOD in metS¹²⁷. Notably, this downregulation was more pronounced among older metS patients, potentially reflecting an age-related increase in APOD expression¹²⁸. APOC2⁵² and APOC4, part of a gene cluster including APOE¹²⁹, are implicated in metabolic regulation. While APOE, a known marker of metabolic syndrome, was not significantly associated with metS in our study, APOC4 was upregulated in both age groups. APOC2 and APOC4 play pivotal roles in lipid metabolism and triglyceride processing associated with very low-density lipoprotein (VLDL)^{129,130}. Although the literature does not establish a direct link between APOC4 and metS, its role in lipid metabolism supports a potential relationship with metS and other cardiometabolic diseases. Prior research has also associated an increase in serum APOC2 levels with a diagnosis of metS⁵². APOB serves as a clinical marker of atherosclerosis due to its association with low-density lipoprotein (LDL) and is an independent predictor of metS development^{127,131}. Previous research has shown that the clinical value of APOB in predicting metS is maintained even after accounting for the comorbidities linked with it^{127,132}. Our results further corroborate these findings, with significant upregulation of APOB in both age groups. Recent studies have highlighted the critical role of the apolipoprotein protein family in regulating triglycerides (TGs) and cholesterol metabolism, transport, and cellular management¹³³. These findings are consistent with our observations of deregulated apolipoprotein family members in metabolic diseases, underscoring their significance as potential therapeutic and diagnostic targets. APOB was the sole protein for which the abundance change associated with disease surpassed the technical threshold in the regression analysis that considered both T2DM and metS. Serum APOB levels are directly linked to insulin secretion¹³⁴, where decreased insulin sensitivity is associated with increased secretion and decreased clearance of APOB. Prediabetes, characterized by visceral fat accumulation and impaired insulin sensitivity, is considered an underlying condition of metS¹³⁵. The strong link between metS and T2DM has been established in relation to both incidence and the molecular mechanisms underlying their prognosis^{64,65}. Notably, administering oral antidiabetic drugs such as dipeptidyl peptidase-4 inhibitors (DPP4i) in T2DM patients was associated with reduced plasma APOB levels⁶⁶. Among the participants diagnosed with T2DM, 85 were receiving orally administered blood glucose-lowering drugs excluding insulin, which indicates the potential utility of APOB in monitoring treatment outcomes. Our study revealed a novel link between serum fibulin 1 (FBLN1) levels and higher BMI categories. FBLN1, a regulatory component of the extracellular matrix, is downregulated in higher BMI categories, indicating potential disruptions in cellular homeostasis, mobility, and adhesion^{138,139}. Markers of chronic inflammation and cellular degeneration such as lysozyme (LYZ) and alpha-2-macroglobulin (A2M) were also associated with individuals belonging to the "overweight" and "obese" categories, but these associations disappeared when adjusting for T2DM and metS. Only

FBLN1 remained significantly linked to BMI, indicating its potential as a robust biomarker, regardless of T2DM or metS.

Lifestyle factors are modifiers of the blood protein profile. Plasma/serum proteomics might thus be able to complement patient questionnaires in clinical studies, as these have the reputation to be unreliable, specifically, when they address questions related to stigmatized lifestyle components, such as alcohol consumption or smoking¹⁴⁰. One notable marker for current and former smoking habits identified in our analysis was an increased abundance of PIGR in smokers. PIGR transports IgA through mucosal epithelia,¹⁴⁰ and has been linked to smoking previously^{141,142}. In contrast to previous studies, we also report higher PIGR levels in some former smokers, specifically older participants who quit recently (less than a year)¹⁴³. Although imprecise responses about smoking habits in the questionnaires might partly explain these results (e.g., some individuals who identified as ex-smokers might still be smoking), the findings could also suggest that some individuals experience a longer recovery time from smoking-related chronic inflammation. Similarly, downregulation of IGGs, specifically IGHG2, relates to smoke-related inflammation^{143,144}. Our observations reinforce the idea of pulmonary inflammation as a consequence of smoking, which we show, might be traceable even in former smokers to a certain degree.

Excessive alcohol consumption (defined as more than 30 g/day) led to deregulated serum levels of several key metabolic proteins. The two with the highest significance, APOC3 and SERPINA7, were respectively upregulated and downregulated.

Increased APOC3 levels with increased alcohol consumption have been previously observed^{145,146} and decreased serum thyroxine as well as SERPINA7 levels have been linked to alcoholism¹⁴⁷. APOC3, a major HDL component released from hepatocytes, influences triglyceride catabolism and is associated with ethanol consumption, decreasing after alcohol withdrawal^{148,149}. Alcohol-induced increases in HDL are attributed to the increased expression of cholesteryl ester transfer protein in response to ethanol intake¹⁵⁰. On the other hand, SERPINA7 was downregulated in excessive drinkers, a phenomenon observed in the context of the gut-thyroid axis^{151,152}. This observation could be attributed to liver dysfunction due to overworking, which impairs the liver's ability to effectively synthesize SERPINA7¹⁵³.

Eventually, we investigated the influence of MHT on the data subset of women from the older age group in the BASE-II. Statistical significant associations with MHT were found for peptidoglycan recognition protein 2 (PGLYRP2), angiotensinogen (AGT) and plasminogen (PLG), which have been described before to be influenced by estradiol-based hormone therapy or medication, such as hormonal contraceptives^{21,71,154,155}. Eight other proteins^{74,75,156,157} showed trends similar to those described in the literature but did not reach significance levels. However, while other studies reported the dysregulation of Lp(A), APOA and APOB^{158,159}, these proteins were not identified in our analysis (Figure S257 and Table S6). Furthermore, no protein was found significant between women with previous MHT treatment and those that never received treatment. Thus, in our dataset, we did not find any significant long lasting changes in the plasma proteome by MHT. The impact of MHT on the proteome (with N = 3 significant protein associations) compared to HCU (with N = 55) aligns well with the lower dosage of estradiol in MHT and the predominant use of bioidentical estradiol as an active ingredient in MHT compared to classical HCU, which predominantly contains the more powerful synthetic ethinylestradiol and much higher dosages. Employing

machine learning, we successfully differentiated between participants who had undergone MHT during sampling from those who had not, achieving an AUROC of 0.82%. Interestingly, we obtained best separation with a random forest model trained solely on abundance levels of PLG and C9, which were identified beforehand in a recursive feature elimination process, even slightly outperforming the predictive performance of estradiol (AUROC: 0.8), a major component of most used MHTs.

Conclusion

In summary, our study reports a significant serum proteome remodeling during aging. Our data confirms previous, and reports new associations, of serum proteins that change during aging dependent on age, sex, and BMI, and highlights the influence of lifestyle factors such as smoking and alcohol consumption. In particular, our findings underscore the impact of hormonal therapies, HCU in young female study participants, and MHT in older female participants, which both have a quantitative impact on the serum proteome. Especially HCU had a stronger impact than any other covariates on the serum proteome in the younger female participants. Our study did not assess whether the proteomic changes resulting from HCU or MHT are related to health outcomes. We hope however our study will stimulate studies to clarify this question. Overall, our findings provide valuable insights into the factors influencing serum proteome remodeling during aging and pave the way for future research to explore the health implications and potential therapeutic interventions.

Limitations

The study design includes two distinct age groups (27-37 and 60-84 years) and therefore we cannot claim to ultimately depict effects between these age ranges. However the high overlap of age-associated proteins between our analysis and comparable studies²¹⁻²⁴ implies that our analysis is reliable. The analysis of MHT effects is based on a relatively small sample size of positive cases (N = 35). To minimize adverse effects as much as possible, we therefore matched the treatment group to a control group with comparable age and BMI. The small sample size also limits an analysis of admission routes and active drug ingredients (Gestagen and Estrogen) of MHT. But we report this information, for future reference in the supplement (see Figure S260). Another potential limitation of our study is the categorization of alcohol consumption based on self-reported daily intake. Although we used the same thresholds for both men and women as cited in previous studies¹⁶⁰⁻¹⁶² (see Methods), the German Center for Addiction Issues (DHS) defines low-risk alcohol consumption differently for women (<12 g/day) and men (<24 g/day)¹⁶³. These differing limits complicated the definition of a moderate alcohol consumption group and led to a larger portion of the cohort being classified as excessive drinkers. Given the distribution of age and sex among participants as well as known clinical details about the group, which were also covariates in our analysis, we chose to define alcohol consumption limits based on established studies.

Methods

Berlin Aging Study II

Berlin Aging Study II (BASE-II) participants were recruited from the greater metropolitan area of Berlin, Germany. The multidisciplinary BASE-II aims to identify mechanisms and factors contributing to healthy vs unhealthy aging. In total 1,671 older participants (aged ≥ 60 years) and 500 younger (20-37 years) were assessed in the medical part of the study at baseline between 2009 and 2014¹⁴. The medical assessments were conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the Charité – Universitätsmedizin Berlin (approval number EA2/029/09) and were registered in the German Clinical Trials Registry as DRKS00009277. Blood was drawn from all participants after an >8 h fasting period and kept at $4-8^{\circ}\text{C}$ until analysis on the same day or was stored at -80°C (serum, plasma). Estradiol, total testosterone, and sex hormone-binding globulin (SHBG) were measured in serum probes by fluoro-, radio- and electrochemiluminescence immunoassays by an accredited laboratory. Standard laboratory parameters (including estradiol, testosterone and SHBG) and proteome analyses were performed on blood from the same draw. Smoking behavior (current smoker, former smoker, never smoker), hormonal contraceptive usage and menopause hormone replacement therapy were assessed during the 1:1 interview by trained study personnel. Alcohol consumption was assessed in g/day via food frequency questionnaire¹⁶⁴. Body Mass Index (BMI, kg/m^2) was calculated from height and weight measured with an electronic measuring station (seca 763 measuring station, SECA, Germany). Diabetes mellitus type 2 was diagnosed based on American Diabetes Association (ADA) guidelines¹⁶⁵, and the metabolic syndrome was diagnosed according to Alberti et al.¹⁶⁶.

Sample Preparation

In this study, a total of 2,191 samples were subjected to measurement. Among these, 105 standardized, commercially available serum set samples (SER-SPL, Zen-Bio, Durham, NC, USA) and 96 were pooled study samples that were utilized to monitor measurement quality and control technical variation. Measurements were performed in 8 batches, each consisting of 3 96-well plates. Serum samples were randomly distributed across all plates.

Semi-automated in-solution digestion was performed as previously described for high throughput clinical proteomics¹⁵. All stocks and stock plates were prepared in advance to reduce variability and were stored at -80°C until use. Briefly, $5\ \mu\text{l}$ of thawed samples were transferred to the denaturation and reduction solution ($55\ \mu\text{l}$ 8 M Urea, 100 mM ammonium bicarbonate (ABC), 5 mM dithiothreitol per well) mixed and incubated at 30°C for 60 minutes. Five microliters were then transferred from the iodoacetamide stock solution plate (100 mM) to the sample plate and incubated in the dark at RT for 30 minutes before dilution with 100 mM ABC buffer ($340\ \mu\text{l}$). $220\ \mu\text{l}$ of this solution was transferred to the pre-made trypsin stock solution plate ($12.5\ \mu\text{l}$, $0.1\ \mu\text{g}/\mu\text{l}$) and incubated at 37°C for 17 h (Benchmark Scientific Incu-Mixer MP4). The digestion was quenched by addition of formic acid (10% v/v, $25\ \mu\text{l}$) and cleaned using C18 solid phase extraction in 96-well plates (BioPureSPE Macro 96-Well, 100 mg PROTO C18, The Nest Group). The eluent was dried under vacuum and reconstituted in $60\ \mu\text{l}$ 0.1% formic acid. Insoluble particles were removed by centrifugation and the samples transferred to a new plate.

Liquid Chromatography and Mass Spectrometry

Digested Peptides were analyzed on a Bruker timsTOF Pro mass spectrometer coupled to a Bruker VIP-HESI electrospray source. Chromatographic separation was performed on the same column type but with only 2 μ g and a flow rate of 500 μ L/min. For diaPASEF acquisition¹⁶⁷, the electrospray source was operated at 3000 V of capillary voltage, a drying gas flow rate of 10 L/min and 240 °C. The diaPASEF windows scheme was set up as followed: we sampled an The ion mobility range was sampled from $1/K_0 = 1.30$ to 0.7 Vs/cm² using ion accumulation times of 100ms and ramp times of 133ms in the dual TIMS analyzer, with each cycle times of 1.25 s. The collision energy was lowered as a function of increasing ion mobility from 59 eV at $1/K_0 = 1.6$ Vs/cm² to 20 eV at $1/K_0 = 0.6$ Vs/cm². For all experiments, TIMS elution voltages were calibrated linearly to obtain the reduced ion mobility coefficients ($1/K_0$) using three Agilent ESI-L Tuning Mix ions (m/z , $1/K_0$: 622.0289, 0.9848 Vs/cm²; 922.0097, 1.1895 Vs/cm²; and 1221.9906, 1.3820 Vs/cm²).

Data processing

Raw MS data was processed using DIA-NN v1.8.1¹⁹. DIA-NN mass accuracies and scan window size were fixed to ensure reproducibility (scan window size: 7; MS1/MS2 15 ppm). An external, publicly available spectral library was used for all measurements²⁰. The spectral library was annotated using the Human UniProt¹⁶⁸ isoform sequence database (Proteome ID: 3AUP000005640)²⁰. It is accessible via the PRIDE database under ID PXD013231.

All preprocessing steps (including normalization, batch correction and imputation) were performed in the R programming language (v4.3.1). The raw Dia-NN output was imported into R using the *msdap* package¹⁶⁹ and initially filtered for serum set QC samples and non-proteotypic peptides (peptides mapping to multiple proteins).

Subsequently, the dataset underwent cyclic loess normalization¹⁷⁰ to normalize peptide intensities between each sample and each other sample, followed by the 'modebetween_protein' method that performs the same normalization on a protein level. The normalized data was filtered iteratively to remove samples with too many missing peptide intensities (3 median absolute deviations), followed by the filtration of low-presence peptides measured in less than 30% of all samples. The filtered data was then imputed using Bayesian PCA (bpca) from R's *pcaMethods* package¹⁷¹, where the missing peptide intensities were imputed within the same batch, sex (male/female) and age-group combinations (young/old), but only for peptides in those combinations which were measured in at least 66% of the samples. Subsequently, batch effects were corrected between all 24 plates using the *removeBatchEffect* function in the *limma* Bioconductor package¹⁷².

Finally, any remaining missing values in precursors with less than or equal to 34% missing values were imputed using the *impute.knn* k-nearest neighbor (knn) algorithm from package *impute* version 1.76¹⁷³, followed by a protein-level summarisation using the function *rcModelPLM* probe-level model (plm) from package *preprocessCore* version 1.64¹⁷⁴. The final dataset after preprocessing consists of 248 proteins and 2917 Peptides for 2079 samples. However, the number of proteins includes multiple isoforms for Kinogen 1 (2 isoforms, KNG1), Inter-alpha-trypsin inhibitor heavy chain H4 (2, ITIH4), Gelsolin (2, GSN), alpha-1-B

glycoprotein (2, A1BG) and Fibulin 1 (3, FBLN1). Peptide Sequences for all proteins are reported in Tables S3-S5.

Statistical Analysis and Machine Learning

Functional protein analysis was performed with the PANTHER classification system¹⁷⁵. Subsequent statistical analysis and machine learning of protein data was conducted in the Python programming language (v3.11), unless specified otherwise. Principal Component Analysis was conducted with standardized protein abundance data by employing the scikit-learn package¹⁷⁶. The arrows added to the plot indicate both the direction and magnitude of the loadings. A significance level of $\alpha \leq 0.05$ was employed and p-values were adjusted by bonferroni. Additionally, we employed a second significance criteria to account for technical variance, by comparing the coefficient of categorical covariates as a percentage to the CV of the study pools. We excluded 30 samples due to missing age or BMI information and one sample due to a duplicated measurement. Linear models, as implemented in statsmodels¹⁷⁷ ordinary least squares (OLS) function, were fitted to the log2 protein abundance of all proteins to investigate associations. Age was treated as a continuous variable and divided by 10, thus the age-related coefficients show the change in 10 years difference. All other variables used in the association analysis were treated as categorical variables and were assigned a fixed reference category. For linear modeling, defined the BMI category “healthy” ($18.5 < \text{BMI} \leq 25$; reference) as the baseline level and obtained coefficients for the contrast with “underweight” ($\text{BMI} < 18.5$), “overweight” ($25 < \text{BMI} \leq 30$) and “obese” ($\text{BMI} > 30$). Smoking was also categorized in 4 groups, such as never smoked (never), stopped more than a year ago (former > 1 year), less than year ago (former < 1 year) and current smokers (current). The alcohol consumption of participants was categorized according to defined ranges in the related literature^{160–162}. In these studies, occasional drinking is defined as consuming less than 1.3 g of ethanol per day, low volume drinking is defined as consuming between 1.3 g and 30 g of ethanol per day, and excessive alcohol consumption is defined as consuming more than 30 g of ethanol per day. Furthermore, reference categories for sex, MHT, metS and diabetes were men, never treated with MHT and not diagnosed for metS and diabetes, respectively. Standardized effect sizes (ES) were obtained by subtracting the mean of every protein and dividing through its standard deviation prior to linear modeling. Spearman’s rho was used to calculate correlation coefficients as implemented in Scipy¹⁷⁸ if not indicated otherwise. Hormonal contraceptive usage (HCU) was defined as in intake within the last 3 months and considered only oral and not topical administration. Protein CVs for study, pooled and standardized serum set samples can be found in Table S1. CVs were calculated by dividing the standard deviation of the abundance data by its mean and expressed as a percentage.

Prior to machine learning, we constructed an age and BMI matched control group (N=35) of women that were never treated with any kind of hormonal replacement therapy to ensure a balanced dataset. Matching was conducted by employing R’s matchit package⁷⁶ with a “nearest” matching method and by estimating propensity scores by a “mahalanobis” based distance matrix. Demographics of both, control and treatment group, are shown in Table S7. We employed random forest models to classify women based on their MHT status as implemented in the scikit-learn¹⁷⁶ package. Next, to the best predictive performance compared to other machine learning architectures such as Support Vector machines or boosting algorithms, random forests also do not need additional feature scaling and are robust to noise

and overfitting due to bootstrapping. To reduce model complexity and only select the most informative features, we employed a cross-validated recursive feature elimination approach (as implemented in scikit-learn) with 10 features removed at every iteration. We finetuned the model using grid search, which constrained the model to have 10 trees and a maximum node depth of 2. Prediction scores were obtained in a 5-Fold cross validation procedure, where training and test sets were balanced by MHT status. To obtain predicted probabilities for Figure 5B, we trained a random forest with identical hyperparameters on all 70 individuals that have been used in the cross validation before. The predicted probabilities of random forest models are calculated by dividing the number of votes for a particular class by the total number of trees. Receiver Operating Curves (ROC), the area under the receiver operating curve (AUROC) were calculated with scikit-learn. To evaluate model performance across multiple cross validation sets, we aggregated the false positive rates (FPR) from each test set, creating a comprehensive set of unique and sorted FPR values. We then interpolated the true positive rates (TPR) for these FPR values from each test set, averaged the interpolated TPRs, and calculated the mean ROC curve.

References

1. Moaddel, R. *et al.* Proteomics in aging research: A roadmap to clinical, translational research. *Aging Cell* **20**, e13325 (2021).
2. Ping, P. *et al.* A functional annotation of subproteomes in human plasma. *Proteomics* **5**, 3506–3519 (2005).
3. Bader, J. M., Albrecht, V. & Mann, M. MS-Based Proteomics of Body Fluids: The End of the Beginning. *Mol. Cell. Proteomics* **22**, 100577 (2023).
4. Rando, T. A. & Jones, D. L. Regeneration, Rejuvenation, and Replacement: Turning Back the Clock on Tissue Aging. *Cold Spring Harb. Perspect. Biol.* **13**, (2021).
5. Sathyan, S. *et al.* Plasma proteomic profile of age, health span, and all-cause mortality in older adults. *Aging Cell* **19**, e13250 (2020).
6. Kuo, C.-L. *et al.* Proteomic aging clock (PAC) predicts age-related outcomes in middle-aged and older adults. *Aging Cell* e14195 (2024).
7. Tanaka, T. *et al.* Plasma proteomic signature of age in healthy humans. *Aging Cell* **17**, e12799 (2018).
8. Santos-Lozano, A. *et al.* Successful aging: insights from proteome analyses of healthy centenarians. *Aging* **12**, 3502–3515 (2020).
9. Anderson, N. L. & Anderson, N. G. The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell. Proteomics* **1**, 845–867 (2002).
10. Vernardis, S. I. *et al.* The Impact of Acute Nutritional Interventions on the Plasma Proteome. *J. Clin. Endocrinol. Metab.* **108**, 2087–2098 (2023).
11. Macklin, A., Khan, S. & Kislinger, T. Recent advances in mass spectrometry based clinical proteomics: applications to cancer research. *Clin. Proteomics* **17**, 17 (2020).
12. Hartl, J. *et al.* Quantitative protein biomarker panels: a path to improved clinical practice through proteomics. *EMBO Mol. Med.* **15**, e16061 (2023).
13. Messner, C. B. *et al.* Mass spectrometry-based high-throughput proteomics and its role in biomedical studies and systems biology. *Proteomics* **23**, e2200013 (2023).

14. Bertram, L. *et al.* Cohort profile: The Berlin Aging Study II (BASE-II). *Int. J. Epidemiol.* **43**, 703–712 (2014).
15. Messner, C. B. *et al.* Ultra-High-Throughput Clinical Proteomics Reveals Classifiers of COVID-19 Infection. *Cell Syst* **11**, 11–24.e4 (2020).
16. Szyrwił, L., Gille, C., Mülleder, M., Demichev, V. & Ralser, M. Fast proteomics with dia-PASEF and analytical flow-rate chromatography. *Proteomics* **24**, e2300100 (2024).
17. Skowronek, P. & Meier, F. High-Throughput Mass Spectrometry-Based Proteomics with dia-PASEF. *Methods Mol. Biol.* **2456**, 15–27 (2022).
18. Kistner, F., Grossmann, J. L., Sinn, L. R. & Demichev, V. QuantUMS: uncertainty minimisation enables confident quantification in proteomics. *bioRxiv* 2023.06.20.545604 (2023) doi:10.1101/2023.06.20.545604.
19. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods* **17**, 41–44 (2019).
20. Bruderer, R. *et al.* Analysis of 1508 Plasma Samples by Capillary-Flow Data-Independent Acquisition Profiles Proteomics of Weight Loss and Maintenance. *Mol. Cell. Proteomics* **18**, 1242–1254 (2019).
21. Dordevic, N. *et al.* Pervasive influence of hormonal contraceptives on the human plasma proteome in a broad population study. *medRxiv* (2023) doi:10.1101/2023.10.11.23296871.
22. Tanaka, T. *et al.* Plasma proteomic biomarker signature of age predicts health and life span. *Elife* **9**, (2020).
23. Lehallier, B. *et al.* Undulating changes in human plasma proteome profiles across the lifespan. *Nat. Med.* **25**, 1843–1850 (2019).
24. Coenen, L., Lehallier, B., de Vries, H. E. & Middeldorp, J. Markers of aging: Unsupervised integrated analyses of the human plasma proteome. *Front Aging* **4**, 1112109 (2023).
25. Jenner, J. L. *et al.* Effects of age, sex, and menopausal status on plasma lipoprotein(a)

- levels. The Framingham Offspring Study. *Circulation* **87**, 1135–1141 (1993).
26. Alavi, P. *et al.* Aging Is Associated With Organ-Specific Alterations in the Level and Expression Pattern of von Willebrand Factor. *Arterioscler. Thromb. Vasc. Biol.* **43**, 2183–2196 (2023).
27. Huang, Y.-M. *et al.* Plasma β 2-microglobulin and cerebrospinal fluid biomarkers of Alzheimer's disease pathology in cognitively intact older adults: the CABLE study. *Alzheimers. Res. Ther.* **15**, 69 (2023).
28. Swanson, C. R. *et al.* Plasma apolipoprotein A1 associates with age at onset and motor severity in early Parkinson's disease patients. *Mov. Disord.* **30**, 1648–1656 (2015).
29. Richens, J. L. *et al.* Rationalising the role of Keratin 9 as a biomarker for Alzheimer's disease. *Sci. Rep.* **6**, 22962 (2016).
30. Thangaraju, K. *et al.* The Impact of Age and BMI on the VWF/ADAMTS13 Axis and Simultaneous Thrombin and Plasmin Generation in Hospitalized COVID-19 Patients. *Front. Med.* **8**, 817305 (2021).
31. Krawetz, R. J. *et al.* Proteoglycan 4 (PRG4) treatment enhances wound closure and tissue regeneration. *NPJ Regen Med* **7**, 32 (2022).
32. Iqbal, S. M. *et al.* Lubricin/Proteoglycan 4 binds to and regulates the activity of Toll-Like Receptors In Vitro. *Sci. Rep.* **6**, 18910 (2016).
33. Alquraini, A. *et al.* The interaction of lubricin/proteoglycan 4 (PRG4) with toll-like receptors 2 and 4: an anti-inflammatory role of PRG4 in synovial fluid. *Arthritis Res. Ther.* **17**, 353 (2015).
34. Miike, K. *et al.* Proteome profiling reveals gender differences in the composition of human serum. *Proteomics* **10**, 2678–2691 (2010).
35. Joaquín, C. *et al.* Pregnancy-associated plasma protein-A is related to gender and to adipocytokine levels: results of the Health Survey of Catalonia. *Clin. Endocrinol.* **78**, 718–723 (2013).
36. Gaya da Costa, M. *et al.* Age and Sex-Associated Changes of Complement Activity and Complement Levels in a Healthy Caucasian Population. *Front. Immunol.* **9**, 2664 (2018).

37. Saunders, T. S., Jenkins, N., Blennow, K., Ritchie, C. & Muniz-Terrera, G. Interactions between apolipoprotein E, sex, and amyloid-beta on cerebrospinal fluid p-tau levels in the European prevention of Alzheimer's dementia longitudinal cohort study (EPAD LCS). *EBioMedicine* **83**, 104241 (2022).
38. Pek, S. L. T. *et al.* Association of circulating proinflammatory marker, leucine-rich- α 2-glycoprotein (LRG1), following metabolic/bariatric surgery. *Diabetes. Metab. Res. Rev.* **34**, e3029 (2018).
39. Riestra, P. *et al.* Gender-specific associations between ADIPOQ gene polymorphisms and adiponectin levels and obesity in the Jackson Heart Study cohort. *BMC Med. Genet.* **16**, 65 (2015).
40. Pek, S. L. T. *et al.* Elevation of a novel angiogenic factor, leucine-rich- α 2-glycoprotein (LRG1), is associated with arterial stiffness, endothelial dysfunction, and peripheral arterial disease in patients with type 2 diabetes. *J. Clin. Endocrinol. Metab.* **100**, 1586–1593 (2015).
41. Jaffa, M. A. *et al.* Longitudinal Plasma Kallikrein Levels and Their Association With the Risk of Cardiovascular Disease Outcomes in Type 1 Diabetes in DCCT/EDIC. *Diabetes* **69**, 2440–2445 (2020).
42. Trainor, P. J. *et al.* Blood Levels of Angiotensinogen and Hypertension in the Multi-Ethnic Study of Atherosclerosis (MESA). *J. Am. Coll. Cardiol.* **81**, 1248–1259 (2023).
43. Zaghlool, S. B. *et al.* Revealing the role of the human blood plasma proteome in obesity using genetic drivers. *Nat. Commun.* **12**, 1279 (2021).
44. Park, J. *et al.* Identification of tetranectin as adipogenic serum protein. *Biochem. Biophys. Res. Commun.* **460**, 583–588 (2015).
45. Tschoner, A. *et al.* Effect of weight loss on serum pigment epithelium-derived factor levels. *Eur. J. Clin. Invest.* **41**, 937–942 (2011).
46. Gonzalez-Quintela, A. *et al.* Determinants of serum concentrations of lipopolysaccharide-binding protein (LBP) in the adult population: the role of obesity. *PLoS One* **8**, e54600 (2013).

47. Baez-Duarte, B. G. *et al.* Serum levels of glutathione peroxidase 3 in overweight and obese subjects from central Mexico. *Arch. Med. Res.* **43**, 541–547 (2012).
48. Soria-Jasso, L. E. *et al.* Beneficial and Deleterious Effects of Female Sex Hormones, Oral Contraceptives, and Phytoestrogens by Immunomodulation on the Liver. *Int. J. Mol. Sci.* **20**, (2019).
49. Al-Shanqeeti, A., van Hylckama Vlieg, A., Berntorp, E., Rosendaal, F. R. & Broze, G. J., Jr. Protein Z and protein Z-dependent protease inhibitor. Determinants of levels and risk of venous thrombosis. *Thromb. Haemost.* **93**, 411–413 (2005).
50. Kangasniemi, M. H. *et al.* Ethinylestradiol in combined hormonal contraceptive has a broader effect on serum proteome compared with estradiol valerate: a randomized controlled trial. *Hum. Reprod.* **38**, 89–102 (2023).
51. Pattaro, C. *et al.* The Cooperative Health Research in South Tyrol (CHRIS) study: rationale, objectives, and preliminary results. *J. Transl. Med.* **13**, 348 (2015).
52. Boiko, A. S. *et al.* Apolipoprotein serum levels related to metabolic syndrome in patients with schizophrenia. *Heliyon* **5**, e02033 (2019).
53. Rahmawati, N. D. *et al.* Body mass index as a dominant risk factor for metabolic syndrome among indonesian adults: a 6-year prospective cohort study of non-communicable diseases. *BMC Nutr* **10**, 43 (2024).
54. Gray, N., Picone, G., Sloan, F. & Yashkin, A. Relation between BMI and diabetes mellitus and its complications among US older adults. *South. Med. J.* **108**, 29–36 (2015).
55. Ste Marie, L., Luquet, S., Curtis, W. & Palmiter, R. D. Norepinephrine- and epinephrine-deficient mice gain weight normally on a high-fat diet. *Obes. Res.* **13**, 1518–1522 (2005).
56. Sachs, S. *et al.* Plasma proteome profiles treatment efficacy of incretin dual agonism in diet-induced obese female and male mice. *Diabetes Obes. Metab.* **23**, 195–207 (2021).
57. Lee, W.-H. *et al.* Genetically predicted plasma cortisol and common chronic diseases: A Mendelian randomization study. *Clin. Endocrinol.* **100**, 238–244 (2024).

58. Moreno-Navarrete, J. M. *et al.* Lysozyme is a component of the innate immune system linked to obesity associated-chronic low-grade inflammation and altered glucose tolerance. *Clin. Nutr.* **40**, 1420–1429 (2021).
59. Moreno-Navarrete, J. M., Ortega, F. J., Bassols, J., Ricart, W. & Fernández-Real, J. M. Decreased circulating lactoferrin in insulin resistance and altered glucose tolerance as a possible marker of neutrophil dysfunction in type 2 diabetes. *J. Clin. Endocrinol. Metab.* **94**, 4036–4044 (2009).
60. Manco, M. *et al.* The decrease of serum levels of human neutrophil alpha-defensins parallels with the surgery-induced amelioration of NASH in obesity. *Obes. Surg.* **20**, 1682–1689 (2010).
61. Fernández-Real, J. M. *et al.* CD14 modulates inflammation-driven insulin resistance. *Diabetes* **60**, 2179–2186 (2011).
62. Spieker, J. *et al.* Diabetes type 2 in the Berlin Aging Study II: Cross-sectional and longitudinal data on prevalence, incidence and severity over on average seven years of follow-up. *Diabet. Med.* **40**, e15104 (2023).
63. Heidemann, C. & Scheidt-Nave, C. Prevalence, incidence and mortality of diabetes mellitus in adults in Germany - A review in the framework of the Diabetes Surveillance. *J Health Monit* **2**, 98–121 (2017).
64. Ford, E. S., Li, C. & Sattar, N. Metabolic syndrome and incident diabetes: current state of the evidence. *Diabetes Care* **31**, 1898–1904 (2008).
65. Regufe, V. M. G., Pinto, C. M. C. B. & Perez, P. M. V. H. C. Metabolic syndrome in type 2 diabetic patients: a review of current evidence. *Porto Biomed J* **5**, e101 (2020).
66. Piccirillo, F. *et al.* Novel Antidiabetic Agents and Their Effects on Lipid Profile: A Single Shot for Several Cardiovascular Targets. *Int. J. Mol. Sci.* **24**, (2023).
67. von Toerne, C. *et al.* MASP1, THBS1, GPLD1 and ApoA-IV are novel biomarkers associated with prediabetes: the KORA F4 study. *Diabetologia* **59**, 1882–1892 (2016).
68. Begum, M. *et al.* Adiponectin: A Promising Target for the Treatment of Diabetes and Its Complications. *Life* **13**, (2023).

69. Fisman, E. Z. & Tenenbaum, A. Adiponectin: a manifold therapeutic target for metabolic syndrome, diabetes, and coronary disease? *Cardiovasc. Diabetol.* **13**, 103 (2014).
70. El Khoudary, S. R. Age at menopause onset and risk of cardiovascular disease around the world. *Maturitas* **141**, 33–38 (2020).
71. Schunkert, H. *et al.* Effects of estrogen replacement therapy on the renin-angiotensin system in postmenopausal women. *Circulation* **95**, 39–45 (1997).
72. Umeda, M., Ichikawa, S., Kanda, T., Sumino, H. & Kobayashi, I. Hormone replacement therapy increases plasma level of angiotensin II in postmenopausal hypertensive women. *Am. J. Hypertens.* **14**, 206–211 (2001).
73. Victor J. Dzau and Howard C. Herrmann. Hormonal Control of Angiotensinogen Production. *Life Sciences* (1982).
74. Katayama, H. *et al.* Application of serum proteomics to the Women’s Health Initiative conjugated equine estrogens trial reveals a multitude of effects relevant to clinical findings. *Genome Med.* **1**, 47 (2009).
75. Pitteri, S. J. *et al.* Postmenopausal estrogen and progestin effects on the serum proteome. *Genome Med.* **1**, 121 (2009).
76. Ho, D. E., Imai, K., King, G. & Stuart, E. A. MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *J. Stat. Softw.* **42**, (2011).
77. Guyon, I., Weston, J., Barnhill, S. & Vapnik, V. Gene Selection for Cancer Classification using Support Vector Machines. *Mach. Learn.* **46**, 389–422 (2002).
78. Ho, T. K. Random decision forests. in *Proceedings of 3rd International Conference on Document Analysis and Recognition* (IEEE Comput. Soc. Press, 2002). doi:10.1109/icdar.1995.598994.
79. Niculescu-Mizil, A. & Caruana, R. Predicting good probabilities with supervised learning. in *Proceedings of the 22nd international conference on Machine learning* 625–632 (Association for Computing Machinery, New York, NY, USA, 2005).
80. Eldjarn, G. H. *et al.* Large-scale plasma proteomics comparisons through genetics and disease associations. *Nature* **622**, 348–358 (2023).

81. Sun, B. B. *et al.* Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329–338 (2023).
82. Elhadad, M. A. *et al.* Deciphering the Plasma Proteome of Type 2 Diabetes. *Diabetes* **69**, 2766–2778 (2020).
83. Suhre, K., McCarthy, M. I. & Schwenk, J. M. Genetics meets proteomics: perspectives for large population-based studies. *Nat. Rev. Genet.* **22**, 19–37 (2021).
84. Li, X. *et al.* Inflammation and aging: signaling pathways and intervention therapies. *Signal Transduction and Targeted Therapy* **8**, 1–29 (2023).
85. Di Micco, R., Krizhanovsky, V., Baker, D. & d’Adda di Fagagna, F. Cellular senescence in ageing: from mechanisms to therapeutic opportunities. *Nat. Rev. Mol. Cell Biol.* **22**, 75–95 (2021).
86. Moqri, M. *et al.* Biomarkers of aging for the identification and evaluation of longevity interventions. *Cell* **186**, 3758–3775 (2023).
87. Demichev, V. *et al.* A time-resolved proteomic and prognostic map of COVID-19. *Cell Syst* **12**, 780–794.e7 (2021).
88. Demichev, V. *et al.* A proteomic survival predictor for COVID-19 patients in intensive care. *PLOS Digit Health* **1**, e0000007 (2022).
89. Wang, Z. *et al.* A multiplex protein panel assay for severity prediction and outcome prognosis in patients with COVID-19: An observational multi-cohort study. *EClinicalMedicine* **49**, 101495 (2022).
90. Kumar, V. *et al.* Multiplexed quantitative proteomics provides mechanistic cues for malaria severity and complexity. *Commun Biol* **3**, 683 (2020).
91. Messner, C. B. *et al.* Ultra-fast proteomics with Scanning SWATH. *Nat. Biotechnol.* **39**, 846–854 (2021).
92. Laxy, M., Teuner, C., Holle, R. & Kurz, C. The association between BMI and health-related quality of life in the US population: sex, age and ethnicity matters. *Int. J. Obes.* **42**, 318–326 (2018).
93. GBD 2019 Risk Factors Collaborators. Global burden of 87 risk factors in 204 countries

- and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* **396**, 1223–1249 (2020).
94. Cuntz, U., Quadflieg, N. & Voderholzer, U. Health Risk and Underweight. *Nutrients* **15**, (2023).
95. Stachenfeld, N. S. & Mazure, C. M. Precision medicine requires understanding how both sex and gender influence health. *Cell* **185**, 1619–1622 (2022).
96. Kahlert, J., Gribsholt, S. B., Gammelager, H., Dekkers, O. M. & Luta, G. Control of confounding in the analysis phase - an overview for clinicians. *Clin. Epidemiol.* **9**, 195–204 (2017).
97. Sathyan, S. *et al.* Plasma proteomic profile of frailty. *Aging Cell* **19**, e13193 (2020).
98. Weyand, C. M. & Goronzy, J. J. Aging of the Immune System. Mechanisms and Therapeutic Targets. *Ann. Am. Thorac. Soc.* **13 Suppl 5**, S422–S428 (2016).
99. Kassa, F. A. *et al.* New inflammation-related biomarkers during malaria infection. *PLoS One* **6**, e26495 (2011).
100. Zinellu, A., Paliogiannis, P., Carru, C. & Mangoni, A. A. Serum amyloid A concentrations, COVID-19 severity and mortality: An updated systematic review and meta-analysis. *Int. J. Infect. Dis.* **105**, 668–674 (2021).
101. Sorić Hosman, I., Kos, I. & Lamot, L. Serum Amyloid A in Inflammatory Rheumatic Diseases: A Compendious Review of a Renowned Biomarker. *Front. Immunol.* **11**, 631299 (2020).
102. Li, Y. *et al.* Recent Advances in Understanding the Role of Cartilage Lubrication in Osteoarthritis. *Molecules* **26**, (2021).
103. Le Blanc, J. & Lordkipanidzé, M. Platelet Function in Aging. *Front Cardiovasc Med* **6**, 109 (2019).
104. La Mendola, D., Trincavelli, M. L. & Martini, C. Angiogenesis in Disease. *Int. J. Mol. Sci.* **23**, (2022).
105. El-Gedaily, A., Schoedon, G., Schneemann, M. & Schaffner, A. Constitutive and regulated expression of platelet basic protein in human monocytes. *J. Leukoc. Biol.* **75**,

1013 495–503 (2004).

1014 106.Eidet, J. R. *et al.* The acute phase response protein SERPINA3 is increased in tear fluid
1015 from the unaffected eyes of patients with unilateral acute anterior uveitis. *J. Ophthalmic*
1016 *Inflamm. Infect.* **11**, 19 (2021).

1017 107.Ziegler-Heitbrock, H. W. & Ulevitch, R. J. CD14: cell surface receptor and differentiation
1018 marker. *Immunol. Today* **14**, 121–125 (1993).

1019 108.Hammond, G. L. Diverse roles for sex hormone-binding globulin in reproduction. *Biol.*
1020 *Reprod.* **85**, 431–441 (2011).

1021 109.Ingenbleek, Y. & Bernstein, L. H. Plasma Transthyretin as a Biomarker of Lean Body
1022 Mass and Catabolic States. *Adv. Nutr.* **6**, 572–580 (2015).

1023 110.Walshe, J. M. & Yealland, M. Not Wilson's disease: a review of misdiagnosed cases.
1024 *QJM* **88**, 55–59 (1995).

1025 111.Kurachi, S., Deyashiki, Y., Takeshita, J. & Kurachi, K. Genetic mechanisms of age
1026 regulation of human blood coagulation factor IX. *Science* **285**, 739–743 (1999).

1027 112.Perdomo, G. & Henry Dong, H. Apolipoprotein D in lipid metabolism and its functional
1028 implication in atherosclerosis and aging. *Aging* **1**, 17–27 (2009).

1029 113.Chiellini, C. *et al.* Serum haptoglobin: a novel marker of adiposity in humans. *J. Clin.*
1030 *Endocrinol. Metab.* **89**, 2678–2683 (2004).

1031 114.Kivimäki, M. *et al.* Body-mass index and risk of obesity-related complex multimorbidity:
1032 an observational multicohort study. *Lancet Diabetes Endocrinol* **10**, 253–263 (2022).

1033 115.Huang, M., Zheng, J., Chen, L., You, S. & Huang, H. Role of apolipoproteins in the
1034 pathogenesis of obesity. *Clin. Chim. Acta* **545**, 117359 (2023).

1035 116.Phieler, J. *et al.* The complement anaphylatoxin C5a receptor contributes to obese
1036 adipose tissue inflammation and insulin resistance. *J. Immunol.* **191**, 4367–4374 (2013).

1037 117.Farztdinov, V. *et al.* Role of the complement system in Long COVID.
1038 doi:10.1101/2024.03.14.24304224.

1039 118.Yu, Y. *et al.* Bioinformatics analysis of candidate genes and potential therapeutic drugs
1040 targeting adipose tissue in obesity. *Adipocyte* **11**, 1–10 (2022).

1041 119.Schaefer, B. A., Cheng, D. & Kouides, P. Impact of obesity on factor VIII and von
1042 Willebrand factor levels in patients with Type 1 von Willebrand disease and low von
1043 Willebrand factor: An analysis of the ATHNdataset. *Haemophilia* **28**, 109–116 (2022).

1044 120.Xu, K. *et al.* Elevated extracellular matrix protein 1 in circulating extracellular vesicles
1045 supports breast cancer progression under obesity conditions. *Nat. Commun.* **15**, 1685
1046 (2024).

1047 121.Zanato, V. *et al.* Weight loss reduces anti-ADAMTS13 autoantibodies and improves
1048 inflammatory and coagulative parameters in obese patients. *Endocrine* **56**, 521–527
1049 (2017).

1050 122.Moebus, S. *et al.* Assessing the prevalence of the Metabolic Syndrome according to
1051 NCEP ATP III in Germany: feasibility and quality aspects of a two step approach in 1550
1052 randomly selected primary health care practices. *Ger. Med. Sci.* **4**, Doc07 (2006).

1053 123.Moebus, S. *et al.* Regional differences in the prevalence of the metabolic syndrome in
1054 primary care practices in Germany. *Dtsch. Arztebl. Int.* **105**, 207–213 (2008).

1055 124.Cheng, C.-W., Chang, C.-C., Chen, H.-W., Lin, C.-Y. & Chen, J.-S. Serum ApoA4 levels
1056 predicted the progression of renal impairment in T2DM. *Eur. J. Clin. Invest.* **48**, e12937
1057 (2018).

1058 125.Desmarais, F. *et al.* High ApoD protein level in the round ligament fat depot of severely
1059 obese women is associated with an improved inflammatory profile. *Endocrine* **61**, 248–
1060 257 (2018).

1061 126.Curry, M. D., McConathy, W. J. & Alaupovic, P. Quantitative determination of human
1062 apolipoprotein D by electroimmunoassay and radial immunodiffusion. *Biochim. Biophys.*
1063 *Acta* **491**, 232–241 (1977).

1064 127.Lim, Y. *et al.* Apolipoprotein B Is Related to Metabolic Syndrome Independently of Low
1065 Density Lipoprotein Cholesterol in Patients with Type 2 Diabetes. *Endocrinol Metab*
1066 *(Seoul)* **30**, 208–215 (2015).

1067 128.Waldner, A., Dassati, S., Redl, B., Smania, N. & Gandolfi, M. Apolipoprotein D
1068 Concentration in Human Plasma during Aging and in Parkinson's Disease: A Cross-

- Sectional Study. *Parkinsons Dis.* **2018**, 3751516 (2018).
- 129.Pirim, D. *et al.* Apolipoprotein E-C1-C4-C2 gene cluster region and inter-individual variation in plasma lipoprotein levels: a comprehensive genetic association study in two ethnic groups. *PLoS One* **14**, e0214060 (2019).
- 130.Kim, S. Y., Park, S. M. & Lee, S.-T. Apolipoprotein C-II is a novel substrate for matrix metalloproteinases. *Biochem. Biophys. Res. Commun.* **339**, 47–54 (2006).
- 131.Li, J. *et al.* Association of apolipoprotein Cs with new-onset type 2 diabetes mellitus: findings from the Chinese multi-provincial cohort study. *BMJ Open* **11**, e052388 (2021).
- 132.Ryoo, J.-H. & Park, S. K. Association of apolipoprotein B and incidence of metabolic syndrome in Korean men: a 5-years' follow-up study. *Atherosclerosis* **226**, 496–501 (2013).
- 133.Basavaraju, P. *et al.* Genetic Regulatory Networks of Apolipoproteins and Associated Medical Risks. *Front Cardiovasc Med* **8**, 788852 (2021).
- 134.Haas, M. E., Attie, A. D. & Biddinger, S. B. The regulation of ApoB metabolism by insulin. *Trends Endocrinol. Metab.* **24**, 391–397 (2013).
- 135.Babu, A. & Fogelfeld, L. Metabolic syndrome and prediabetes. *Dis. Mon.* **52**, 55–144 (2006).
- 136.Nesic, J. *et al.* Adiponectin and Interleukin-33: Possible Early Markers of Metabolic Syndrome. *J. Clin. Med. Res.* **12**, (2022).
- 137.Snijder, M. B. *et al.* Associations of adiponectin levels with incident impaired glucose metabolism and type 2 diabetes in older men and women: the hoorn study. *Diabetes Care* **29**, 2498–2503 (2006).
- 138.Roark, E. F. *et al.* The association of human fibulin-1 with elastic fibers: an immunohistological, ultrastructural, and RNA study. *J. Histochem. Cytochem.* **43**, 401–411 (1995).
- 139.Jie, Y. *et al.* The clinical value of Fibulin-1 for prognosis and its prospective mechanism in intrahepatic cholangiocarcinoma. *HPB* **21**, 499–507 (2019).
- 140.Liangpunsakul, S. *et al.* Novel serum biomarkers for detection of excessive alcohol use.

1097 *Alcohol. Clin. Exp. Res.* **39**, 556–565 (2015).

1098 141.Ohlmeier, S. *et al.* Sputum proteomics identifies elevated PIGR levels in smokers and
1099 mild-to-moderate COPD. *J. Proteome Res.* **11**, 599–608 (2012).

1100 142.McGrath, J. J. C. *et al.* Cigarette smoke exposure attenuates the induction of antigen-
1101 specific IgA in the murine upper respiratory tract. *Mucosal Immunol.* **14**, 1067–1076
1102 (2021).

1103 143.Poussin, C. *et al.* Blood and urine multi-omics analysis of the impact of e-vaping,
1104 smoking, and cessation: from exposome to molecular responses. *Sci. Rep.* **14**, 4286
1105 (2024).

1106 144.Tangada, S. D. *et al.* The effect of smoking on serum IgG2 reactive with *Actinobacillus*
1107 *actinomycetemcomitans* in early-onset periodontitis patients. *J. Periodontol.* **68**, 842–
1108 850 (1997).

1109 145.Tilly, P. *et al.* Biological and genetic determinants of serum apoC-III concentration:
1110 reference limits from the Stanislas Cohort. *J. Lipid Res.* **44**, 430–436 (2003).

1111 146.Wilkens, T. L. *et al.* Associations between Alcohol Consumption and HDL Subspecies
1112 Defined by ApoC3, ApoE and ApoJ: the Cardiovascular Health Study. *Curr. Probl.*
1113 *Cardiol.* **48**, 101395 (2023).

1114 147.Grimaldi, F., De Marchi, S. & Cecchin, E. [Low T4 syndrome in alcoholism: role of the
1115 decrease in TBG]. *Minerva Med.* **77**, 381–385 (1986).

1116 148.Song, Y. *et al.* Associations of the APOC3 rs5128 polymorphism with plasma APOC3
1117 and lipid levels: a meta-analysis. *Lipids Health Dis.* **14**, 32 (2015).

1118 149.Lecomte, E. *et al.* Changes in serum apolipoprotein and lipoprotein profile induced by
1119 chronic alcohol consumption and withdrawal: determinant effect on heart disease? *Clin.*
1120 *Chem.* **42**, 1666–1675 (1996).

1121 150.Shaish, A. *et al.* Alcohol increases plasma levels of cholesterol diet-induced atherogenic
1122 lipoproteins and aortic atherosclerosis in rabbits. *Arterioscler. Thromb. Vasc. Biol.* **17**,
1123 1091–1097 (1997).

1124 151.Geurts, J. *et al.* Alterations in circulating thyroid hormones and thyroxine binding

1125 globulin in chronic alcoholism. *Clin. Endocrinol.* **14**, 113–118 (1981).

1126 152. Marino, L., Kim, A., Ni, B. & Celi, F. S. Thyroid hormone action and liver disease, a
1127 complex interplay. *Hepatology* (2023) doi:10.1097/HEP.0000000000000551.

1128 153. Glinoe, D., Gershengorn, M. C. & Robbins, J. Thyroxine-binding globulin biosynthesis
1129 in isolated monkey hepatocytes. *Biochim. Biophys. Acta* **418**, 232–244 (1976).

1130 154. Josse, A. R., Garcia-Bailo, B., Fischer, K. & El-Sohemy, A. Novel effects of hormonal
1131 contraceptive use on the plasma proteome. *PLoS One* **7**, e45162 (2012).

1132 155. Briggs, M. H. & Briggs, M. Oral Contraceptives and plasma protein metabolism. *Journal*
1133 *of Steroid Biochemistry* **11**, 525–428 (1979).

1134 156. Clemente, C. *et al.* Ceruloplasmin serum level in post-menopausal women treated with
1135 oral estrogens administered at different times. *Horm. Metab. Res.* **24**, 191–193 (1992).

1136 157. Melissari, E. & Kakkar, V. V. The effects of oestrogen administration on the plasma free
1137 protein S and C4b-binding protein. *Thromb. Res.* **49**, 489–495 (1988).

1138 158. Godsland, I. F. Effects of postmenopausal hormone replacement therapy on lipid,
1139 lipoprotein, and apolipoprotein (a) concentrations: analysis of studies published from
1140 1974–2000. *Fertil. Steril.* **75**, 898–915 (2001).

1141 159. Bayrak, A., Aldemir, D. A., Bayrak, T., Corakçi, A. & Dursun, P. The effect of hormone
1142 replacement therapy on the levels of serum lipids, apolipoprotein AI, apolipoprotein B
1143 and lipoprotein (a) in Turkish postmenopausal women. *Arch. Gynecol. Obstet.* **274**,
1144 289–296 (2006).

1145 160. Yang, H. K. *et al.* Different effect of alcohol consumption on hypertension according to
1146 metabolic health status. *J. Hum. Hypertens.* **30**, 591–598 (2016).

1147 161. Wakabayashi, I. Associations of alcohol drinking and cigarette smoking with serum lipid
1148 levels in healthy middle-aged men. *Alcohol Alcohol* **43**, 274–280 (2008).

1149 162. Burger, M., Brönstrup, A. & Pietrzik, K. Derivation of tolerable upper alcohol intake
1150 levels in Germany: a systematic review of risks and benefits of moderate alcohol
1151 consumption. *Prev. Med.* **39**, 111–127 (2004).

1152 163. German Federal Ministry of Health. Alcohol and alcohol dependence.

1153 <https://gesund.bund.de/en/alcohol>.

1154 164.Nöthlings, U., Hoffmann, K., Bergmann, M. M. & Boeing, H. Fitting portion sizes in a
1155 self-administered food frequency questionnaire. *J. Nutr.* **137**, 2781–2786 (2007).

1156 165.American Diabetes Association. 2. Classification and Diagnosis of Diabetes: Standards
1157 of Medical Care in Diabetes-2019. *Diabetes Care* **42**, S13–S28 (2019).

1158 166.Alberti, K. G. M. M. *et al.* Harmonizing the metabolic syndrome: a joint interim statement
1159 of the International Diabetes Federation Task Force on Epidemiology and Prevention;
1160 National Heart, Lung, and Blood Institute; American Heart Association; World Heart
1161 Federation; International Atherosclerosis Society; and International Association for the
1162 Study of Obesity. *Circulation* **120**, 1640–1645 (2009).

1163 167.Meier, F. *et al.* diaPASEF: parallel accumulation-serial fragmentation combined with
1164 data-independent acquisition. *Nat. Methods* **17**, 1229–1236 (2020).

1165 168.UniProt Consortium. The universal protein resource (UniProt). *Nucleic Acids Res.* **36**,
1166 D190–5 (2008).

1167 169.Koopmans, F., Li, K. W., Klaassen, R. V. & Smit, A. B. MS-DAP Platform for
1168 Downstream Data Analysis of Label-Free Proteomics Uncovers Optimal Workflows in
1169 Benchmark Data Sets and Increased Sensitivity in Analysis of Alzheimer’s Biomarker
1170 Data. *J. Proteome Res.* **22**, 374–386 (2023).

1171 170.Ballman, K. V., Grill, D. E., Oberg, A. L. & Therneau, T. M. Faster cyclic loess:
1172 normalizing RNA arrays via linear models. *Bioinformatics* **20**, 2778–2786 (2004).

1173 171.Stacklies, W., Redestig, H., Scholz, M., Walther, D. & Selbig, J. pcaMethods -- a
1174 Bioconductor package providing PCA methods for incomplete data. *Bioinformatics* vol.
1175 23 1164–1167 Preprint at (2007).

1176 172.Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing
1177 and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

1178 173.Troyanskaya, O. *et al.* Missing value estimation methods for DNA microarrays.
1179 *Bioinformatics* **17**, 520–525 (2001).

1180 174.Bolstad, B. preprocessCore: a collection of pre-processing functions. R package version

- 1181 1.46. 0. Preprint at (2019).
- 1182 175.Mi, H., Muruganujan, A., Casagrande, J. T. & Thomas, P. D. Large-scale gene function
1183 analysis with the PANTHER classification system. *Nat. Protoc.* **8**, 1551–1566 (2013).
- 1184 176.Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**,
1185 2825–2830 (2011).
- 1186 177.Seabold, S. & Perktold, J. Statsmodels: Econometric and statistical modeling with
1187 python. in *Proceedings of the 9th Python in Science Conference* (SciPy, 2010).
1188 doi:10.25080/majora-92bf1922-011.
- 1189 178.Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python.
1190 *Nat. Methods* **17**, 261–272 (2020).
- 1191 179.Perez-Riverol, Y. *et al.* The PRIDE database and related tools and resources in 2019:
1192 improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2019).

1193 Supplementary Material

1194 PDF document contains Supplementary Figures S1-S260 and Supplementary Tables S1 and
1195 S6-S12. Spreadsheets (xlsx) with Supplementary Table S2 contains functional annotations of
1196 protein groups present in the dataset. Tables S3-S5 contain results of association analysis for
1197 the whole cohort and the younger and older age group, respectively.

1198 Data Availability Statement

1199 BASE-II proteome data is available only upon reasonable request. Please contact Ludmila
1200 Müller, scientific coordinator, at lmuller@mpib-berlin.mpg.de, for additional information.
1201 The raw mass spectrometry data of study pools and standardized, commercially available
1202 serum set sample as well as the raw DIA-NN output and used FASTA file for spectral library
1203 annotation will be deposited to the ProteomeXchange Consortium via the PRIDE partner
1204 repository¹⁷⁹ upon publication.

1205 Acknowledgements

1206 The authors of this paper thank all study participants of the Berlin Aging Study II, as well as
1207 all people involved in planning this study, executing sample collection, conducting
1208 questionnaire survey and preparing the data. We thank the Charité Core Facility High
1209 Throughput Mass Spectrometry for conducting the measurements. We thank Vivien Bahr for
1210 clinical background information regarding glucose lowering drugs and Oliver Lemke for
1211 inspiration on how to visualize a 2-feature random forest.

1212 Funding

1213 BASE-II was supported by the German Federal Ministry of Education and Research under
 1214 grant numbers #01UW0808; #16SV5536K, #16SV5537, #16SV5538, #16SV5837,
 1215 #01GL1716A, and #01GL1716B. This work was supported by a grant of the Deutsche
 1216 Forschungsgemeinschaft (grant number 460683900 to ID). The proteomic work presented
 1217 here was supported by the Ministry of Education and Research (BMBF), as part of the National
 1218 Research Node 'Mass spectrometry in Systems Medicine (MSCoresys), under grant
 1219 agreement 031L0220 and 16LW023K, as well as the the German Cancer Consortium (DKTK)
 1220 under agreement BE01 1020000483.

1221

1222

1223 Competing Interest Statement

1224 Markus Ralser is founder and shareholder of Eliptica Ltd. Michael Mülleder is a consultant and
 1225 shareholder of Eliptica Ltd.

1226 Ethics: Institutional Review Board Statement

1227 All participants gave written informed consent. The Ethics Committee of the Charité –
 1228 Universitätsmedizin Berlin approved the study (approval number EA2/029/09). The study was
 1229 conducted in accordance with the Declaration of Helsinki and was registered in the German
 1230 Clinical Trials Registry as DRKS00009277.

1231 Author Contributions

1232 Conceptualization: M.R., I.D., C.D., R.S.M.; Investigation: K.T., D.L., O.S., V.F., M.M.; Formal
 1233 Analysis: C.D., R.S.M., O.S., V.F., M.M., J.R.; Writing – original draft preparation: C.D.,
 1234 R.S.M., O.S., K.T.; Writing – review and editing: C.D., R.S.M., O.S., J.R., I.D., M.R.; Funding
 1235 acquisition: I.D., M.R.