

# 1 **Menopause Hormone Replacement Therapy and Lifestyle** 2 **Factors affect Metabolism and Immune System in the Serum** 3 **Proteome of Aging Individuals**

4  
5 Clemens Dierks<sup>1,2</sup>, Roza Sürme Mizrak<sup>1</sup>, Orr Shomroni<sup>1</sup>, Vadim Farztdinov<sup>1,3</sup>, Kathrin Textoris-  
6 Taube<sup>1,3</sup>, Daniela Ludwig<sup>1,3</sup>, Johannes Rainer<sup>4</sup>, Michael Mülleder<sup>3</sup>, Ilja Demuth<sup>5,6</sup>, Markus  
7 Ralser<sup>1,7,8\*</sup>

8  
9 <sup>1</sup> Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt  
10 Universität zu Berlin, Institute of Biochemistry, Charitéplatz 1, 10117 Berlin, Germany

11 <sup>2</sup> Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt  
12 Universität zu Berlin, Department of Infectious Diseases, Respiratory Medicine and Critical Care,  
13 Augustenburger Platz 1, 13353 Berlin, Germany

14 <sup>3</sup> Core Facility – High Throughput Mass Spectrometry, Charité Universitätsmedizin Berlin, 10117 Berlin,  
15 Germany

16 <sup>4</sup> Institute for Biomedicine, Eurac Research, 39100 Bolzano, Italy

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

20 <sup>6</sup> Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-  
21 Universität zu Berlin, and Berlin Institute of Health, Regenerative Immunology and Aging, BIH Center  
22 for Regenerative Therapies, 13353 Berlin, Germany

23 <sup>7</sup> Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany

24 <sup>8</sup> The Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford,  
25 Oxford OX3 7BN, UK

26

27 \* Correspondence: [markus.ralser@charite.de](mailto: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.

## 47 Introduction

48

49 Aging is a primary risk factor for numerous chronic diseases and is associated with health  
50 deterioration and functional decline of the body, making it particularly important to investigate  
51 molecular changes that are related to age <sup>1</sup>. While the underlying mechanisms are complex,  
52 the serum proteome offers huge potential for deciphering the interrelations of health and  
53 disease <sup>2,3</sup>. The serum proteome responds to changes underlying aging, such as increasing  
54 levels of inflammatory proteins and decreasing tissue repair and cell regeneration activity <sup>4</sup>.  
55 Circulating serum protein level levels can help to differentiate between healthy aging and  
56 pathological conditions <sup>1,5-7</sup>. Moreover, an increased understanding of the relationship  
57 between chronological and biological aging can contribute to delay or prevent diseases where  
58 chronological age is a major risk factor <sup>6</sup> as well as improving patient stratification <sup>1</sup>. Identifying  
59 markers of healthy aging is crucial to define appropriate lifestyle interventions and thereby  
60 enhancing lifespan <sup>8</sup>.

61

62 Broadly speaking, the serum contains two sets of proteins. Proteins that function in the blood,  
63 such as nutrient transport proteins, signaling molecules, proteins of the immune system and  
64 the coagulation machinery, and proteins, who circulate in the blood due to tissue leakage, cell  
65 necrosis, or to their presence in secreted extracellular vesicles. Several, but not all, of the  
66 biomarkers that are currently in use, are part of the former, and are often found within the  
67 highly abundant fraction of the plasma proteome, that is easily accessible and simplifies the  
68 development of routine applicable biomarker assays <sup>9-12</sup>. In recent years, the application of  
69 multiplexed proteome technologies has enhanced our ability to probe the complex molecular  
70 landscape of the proteome in plasma or serum. Recent advancements in mass spectrometry-  
71 based proteomic platforms have addressed previous challenges related to throughput and  
72 cost, enabling large-scale recording of plasma proteomes <sup>13</sup>.

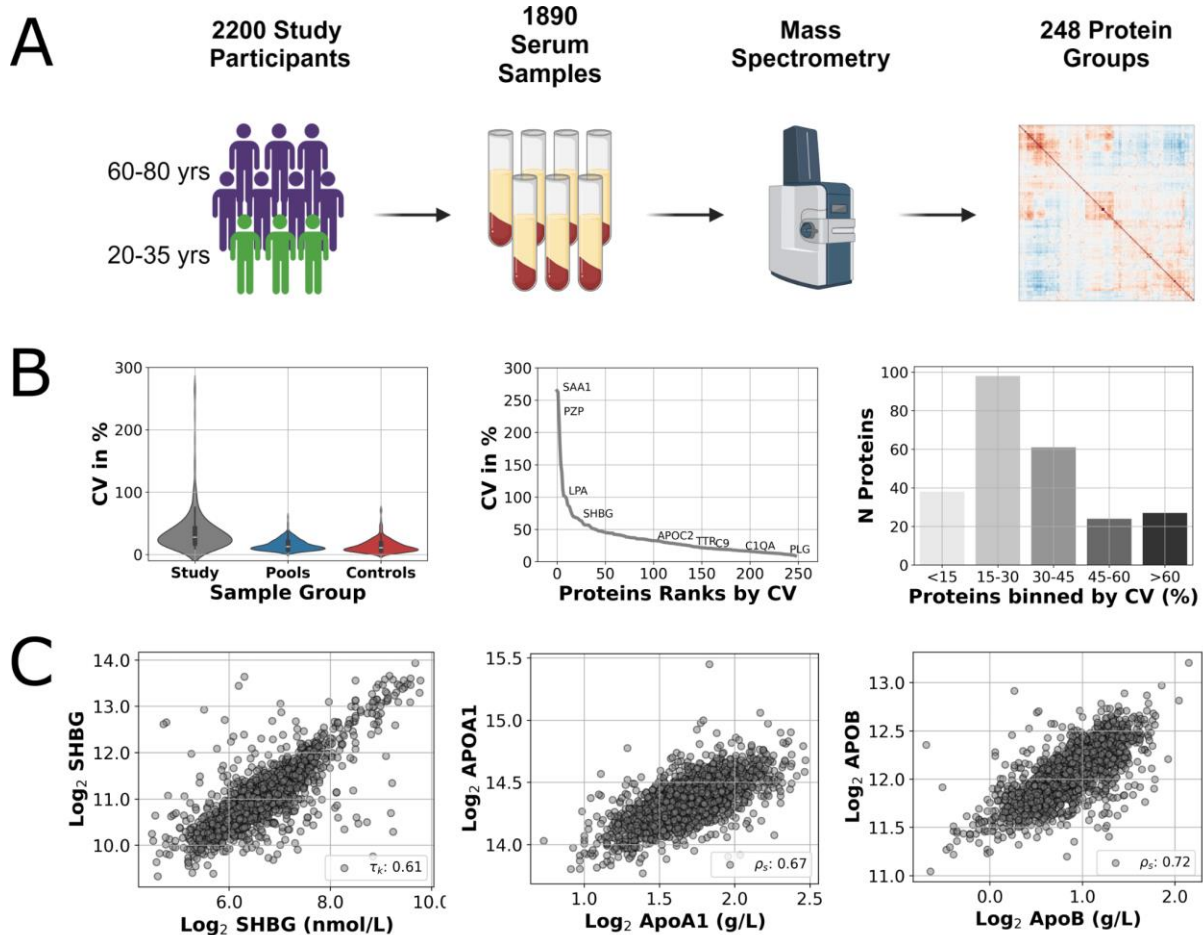
73

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

84 Results

85 Study Profile and Proteomic Assessment of the Berlin Aging Study II

86



87

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

98

99

100 We conducted MS-based proteomic measurements on a total of 2091 serum samples,  
 101 including 105 serum control and 96 pooled study samples, to generate a proteomic baseline  
 102 from the BASE-II. The samples analyzed are derived from two age groups: older adults (N =  
 103 1433) aged 60 to 85 years and younger adults (N = 426) aged 27 to 37 years. Study  
 104 demographics are given in Table 1.

105 The samples were prepared for measurement in a semi-automatic workflow that allows for  
 106 high-throughput and consistent sample quality including tryptic digestion and sample cleanup  
 107 with solid-phase extraction<sup>15</sup>. Samples were divided into 8 batches, each consisting of three  
 108 96-well plates, and were subjected to mass spectrometric measurement using a timsTOF Pro1  
 109 mass analyzer (Bruker Daltonics) operating in diaPASEF mode, using analytical flow rate  
 110 chromatography using<sup>16,17</sup>. Raw MS data was processed by DIA-NN<sup>18,19</sup>, using the  
 111 DiOGenes spectral library<sup>20</sup>. In the subsequent preprocessing the data was normalized,  
 112 filtered for outlier samples and low-presence peptides, imputed and corrected for plate effects  
 113 (N = 24). Among the data set, 31 samples were removed due to low technical quality, or  
 114 missing metadata (age and BMI) leaving the dataset with a total of 1859 proteomic samples  
 115 from BASE-II participants. After this preprocessing, the final dataset quantified 2079  
 116 proteotypic peptides derived from 248 protein groups (Table S1).  
 117

Variable	Young (N=426)	Old (N=1433)
<b>Age in years [median (IQR)]</b>	28.9 [4.5]	68.8 [5.4]
<b>Sex (N, Women in %)</b>	226 (55)	742 (52)
<b>BMI (N, %)</b>		
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)
<b>HCU (N, % women)</b>	90 (38.5)	
<b>Smoker (N, %)</b>		
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)
<b>Diabetes (N, %)</b>	1 (0.002)	177 (12.7)
<b>Metabolic Syndrome (N, %)</b>	31 (7.3)	508 (35.9)
<b>Alcohol intake (N, %)</b>		
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)
<b>MHT (N, %)</b>		
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

118  
 119  
 120 We calculated a median coefficient of variation (CV) of protein quantities of 12.74 % in study  
 121 pools and 10.89 % for serum controls samples (Zen-Bio). Reflecting the biological signal,  
 122 protein quantities in study samples deviated from the technical controls (CV of 27.79 %; Table  
 123 S1)). While most of the proteins are normally distributed, some proteins, such as lipoprotein  
 124 (a) (Lp(a)) and complement factor H-related protein 1 (CFHR1), exhibited tailed or almost  
 125 bimodal distributions (Figures S59, S158). Tailed distributions in certain proteins, particularly

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

132  
133 We analyzed functional groups, such as immunoglobulins, coagulation factors, acute-phase  
134 proteins, complement factors and lipid transport proteins for their variation. Immunoglobulins  
135 (N = 60 protein groups) had the highest median relative variation (CV 44.3 %, range: 23.6 -  
136 264.8 %), and included the proteins with the highest individual relative CVs such as IGHV3-  
137 11 (264.8 %), IGHV3-21 (183.3 %) and IGLV4-60 (150.6 %). In contrast, proteins involved in  
138 the complement system (N = 36) exhibited lower variation (median CV: 18.4 %, range: 10 -  
139 63.5 %), including some of the least variable proteins such as the two central mediators of the  
140 component system C5 and C3 (both 10.4%). Proteins related to coagulation (N = 30), lipid  
141 transport (N = 20) and acute-phase response (N = 26), showed CVs of 19.6 % (range: 11.5%  
142 - 138.2%), 23.5% (12.9 % -102%) and 22% (10.4% - 262.6%). Besides IGHV3-11, serum  
143 amyloid A1 (SAA1; 262.6%) and the pregnancy zone protein (PZP, 225.1%) showed the  
144 highest variation in the study samples, while hemopexin (HPX, 9.2%), complement factor 1R  
145 (C1R, 10%) and attractin (ATRN, 10.2%) the lowest. CVs for study samples, pooled study  
146 samples and QC samples are shown in Tables S3-S5. Furthermore, a detailed functional  
147 analysis of the proteins in the dataset is provided in Table S2.

148  
149 Next, we conducted a correlation analysis between the values obtained with routinely applied  
150 biomarker assays and the corresponding protein abundance levels. For interpreting these  
151 results, it is important to keep in mind that proteomics and the diagnostic tests are typically  
152 targeting related but not identical analytes. We obtained significant correlations, displayed as  
153 spearman's rho if not described otherwise, between proteomics and biomarker assays for  
154 apolipoprotein A1 levels (0.67) and B (0.72), low density lipoproteins (LDL) with APOB (0.72)  
155 and high-density lipoproteins (HDL) with APOA1 (0.66). Triglyceride measurements showed  
156 a moderate correlation with APOC3 levels (0.44). In contrast, diagnostic assays and protein  
157 abundance determined by proteomics correlated weakly for albumin (0.32), hemoglobin with  
158 the MS acquired hemoglobin subunit alpha 2 (HBA2; 0.31), and for fibrinogen compared to  
159 fibrinogen alpha chain levels (0.15). Notably, routine test sex hormone binding globulin  
160 (SHBG) measurements (prior to log2 transformation) displayed a robust, albeit not strictly  
161 linear relationship, with proteomic data (Kendall's tau: 0.61, see Figure S250h). The diverging  
162 results show that proteomic measurements provide complementary data to establish  
163 diagnostic assays (see Figure S250).

164

## 165 **Hormonal contraceptive usage (HCU) is a dominating covariate affecting the** 166 **serum proteome in young women**

167

168 We investigated the effect of demographic covariates, such as sex, age, BMI, and HCU due  
169 to previous findings <sup>21</sup> on the serum proteome of the 1890 BASE-II participants of both age  
170 groups by fitting multiple linear regression models to the protein abundance data. In our  
171 analysis BMI was treated as a categorical variable comparing underweight (U, BMI < 18.5),



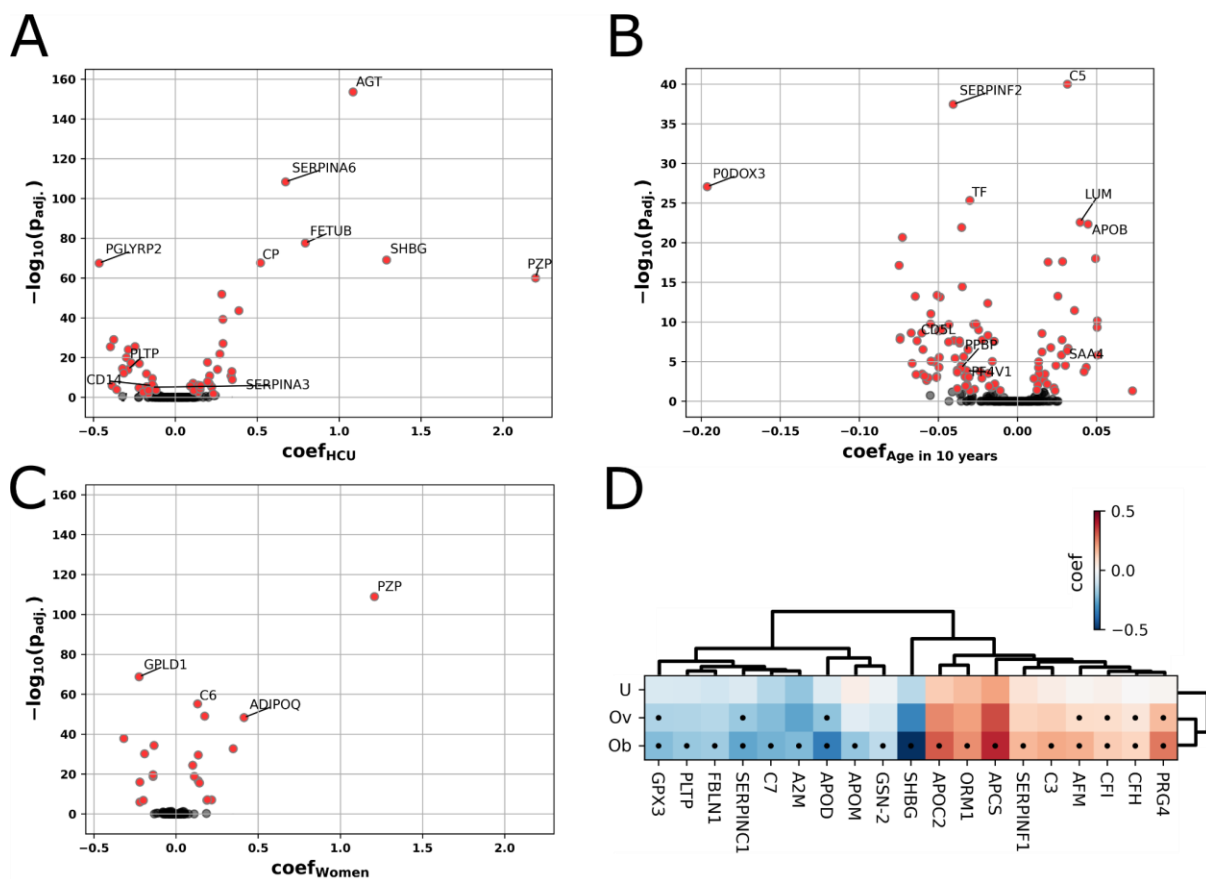
172 overweight (Or,  $25 < \text{BMI} \leq 30$ ) and obese (Ob,  $\text{BMI} > 30$ ) participants to those with a  
173 healthy BMI ( $18.5 < \text{BMI} \leq 25$ ). This categorization was chosen because these groups have  
174 higher clinical relevance than the actual, numerical BMI value and are commonly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

179  
180 We found 106 protein groups associated with age. Of these, 65 have also been identified as  
181 age-dependent in orthogonal studies<sup>21-30</sup>. Among the 41 new associations identified, 32 were  
182 immunoglobulin chains. Other newly found age-associated proteins include serum amyloid A4  
183 (SAA4), proteoglycan 4 and the two platelet-related serum proteins, platelet factor 4 variant 1  
184 (PF4V1) and pro-platelet basic protein (PPBP). Amongst the group of age-associated proteins,  
185 we report the strongest negative association, with up to 20% abundance change in 10 years,  
186 for serpin family F member 2 (SERPINF2), immunoglobulin delta heavy chain (P0DOX3) and  
187 transferrin (TF). Highest increase with progressing age was found in complement component  
188 5 (C5), apolipoprotein B (APOB) and lumican (LUM).

189  
190 The above defined age-related differences in the proteome were in part sex dependent, with  
191 23 proteins showing a different age-response in male and female study participants. We  
192 detected the highest significant differences between men and women in the pregnancy zone  
193 protein (PZP), glycosylphosphatidylinositol specific phospholipase D1 (GPLD1), complement  
194 component 6 (C6), adiponectin (ADIPOQ) and ceruloplasmin (CP). None of the proteins  
195 investigated were associated with an underweight BMI, however, 19 proteins were linked to  
196 overweight and obesity, respectively, with APOC2 and fibulin 1 (FBLN1) showing novel  
197 associations with elevated BMI<sup>21,43-47</sup>.

198  
199 We found new associations with HCU in 15 out of 55 proteins that had not been reported in  
200 the related studies before<sup>21,48-50</sup>, such as the phospholipid transfer protein (PLTP), monocyte  
201 differentiation antigen (CD14), alpha-1-antichymotrypsin (SERPINA3) and the coagulation  
202 factor V (F5) (see Figure 2). Among all covariates, highest effect sizes and significance were  
203 reported for HCU (see Figure 1D) with the top 5 candidates being angiotensinogen (AGT),  
204 peptidoglycan recognition protein 2 (PGLYRP2), serpin family A member 6 (SERPINA6),  
205 fetuin B (FETUB) and ceruloplasmin (CP). We recently investigated the plasma proteomes of  
206 3,472 participants from the CHRIS cohort<sup>21,51</sup>, an Alpine population study examining the  
207 genetic and molecular backgrounds influencing the incidence and progression of common  
208 chronic diseases. To validate our previous findings and demonstrate reproducibility, we  
209 compared the effect sizes for age, sex, and BMI categories for the overlapping proteome from  
210 the BASE-II analysis with those obtained from a similarly conducted association analysis of  
211 the CHRIS dataset, both measured on similar mass spectrometry platforms. We observed  
212 high concordance: age (Spearman's rho: 0.69), sex (0.85), BMI (obese category: 0.73), and  
213 HCU (0.9) (see Figures S256). However, the effect sizes for the underweight BMI category  
214 showed a low correlation between the studies (-0.25). Among the 125 overlapping proteins  
215 identified in both proteomic datasets, we found 53 proteins to be age-dependent and 13  
216 proteins to be sex-dependent in both studies. The age-associated proteins include various  
217 functional groups involved in disease pathology, such as components of the complement  
218 cascade (C5, C7, C9, C4A, C4BPA, C1, CFD, CFI), apolipoproteins (APOB, APOC3, APOD,  
219 APOE, APOH), and proteins involved in blood clotting, including coagulation factors X (F10)

220 and XII (F12) and antithrombin III (SERPINC1). Additionally, we identified acute-phase  
 221 proteins such as haptoglobin (HP) and transthyretin (TTR).  
 222  
 223



224  
 225 **Figure 2:** Multiple linear regression was employed to identify significant protein associations  
 226 with hormonal contraceptive use (HCU), age, sex and different BMI categories (U:  
 227 Underweight, Ov: Overweight, Ob: Obese). (A) The impact of HCU on the serum proteome.  
 228 Proteins with largest coefficients and highest significance (in negative log10 adjusted p-  
 229 values) for proteins with use of hormonal contraceptives. (B) The impact of 10-years of aging  
 230 on the serum proteome. (C) The impact of sex (men vs. women) on the plasma proteome in  
 231 Base-II. Proteins of significance are highlighted in red, and gene names are annotated for  
 232 those discussed in the text. (D) Proteins associated with the different BMI categories  
 233 (heatmap). Black dots denote significant associations after multiple testing corrections, while  
 234 color indicates effect size by its coefficient.  
 235

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

238  
 239 Building on existing knowledge of lifestyle factors impacting the serum proteome<sup>52</sup>, we  
 240 investigated the effects of smoking habits, alcohol consumption, diabetes mellitus type 2  
 241 (T2DM), or metabolic syndrome (metS) on protein abundance as covariates in our multiple  
 242 regression models in addition to age, sex, BMI and HCU. We only included HCU as a  
 243 covariate in the younger age group. Adjusted p-values, coefficients and effect sizes obtained  
 244 from the regression analyses of significantly associated proteins with lifestyle factors and



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

248

249 First, we addressed proteomic changes in current and former smokers. No significantly  
250 regulated proteins were found in former smokers that quit smoking more than a year ago. On  
251 the other hand, older participants who recently quit smoking (less than a year) had a  
252 significantly elevated level of polymeric immunoglobulin receptor (PIGR), which was also true  
253 for current smokers compared to participants who had never smoked (n = 896, see Figure  
254 3B). Additionally, immunoglobulin heavy constant gamma 2 (IGHG2) levels were significantly  
255 decreased in current smokers from the older age group. IGHG2 levels were also found  
256 lowered in the younger group, albeit not significant. Alcohol consumption of the participants  
257 was grouped into occasional (< 1.3 g/day), low volume and excessive drinkers (>30 g/day)  
258 based on their self-reported habits. Among the 8 significantly associated proteins, the most  
259 changes were obtained for apolipoprotein C3 (APOC3) and thyroxine-binding globulin  
260 (SERPINA7) (see Figure 3C). Both proteins were significantly changed in low volume and  
261 excessive drinkers compared to participants that only occasionally drink alcohol, though  
262 significance increased with age.

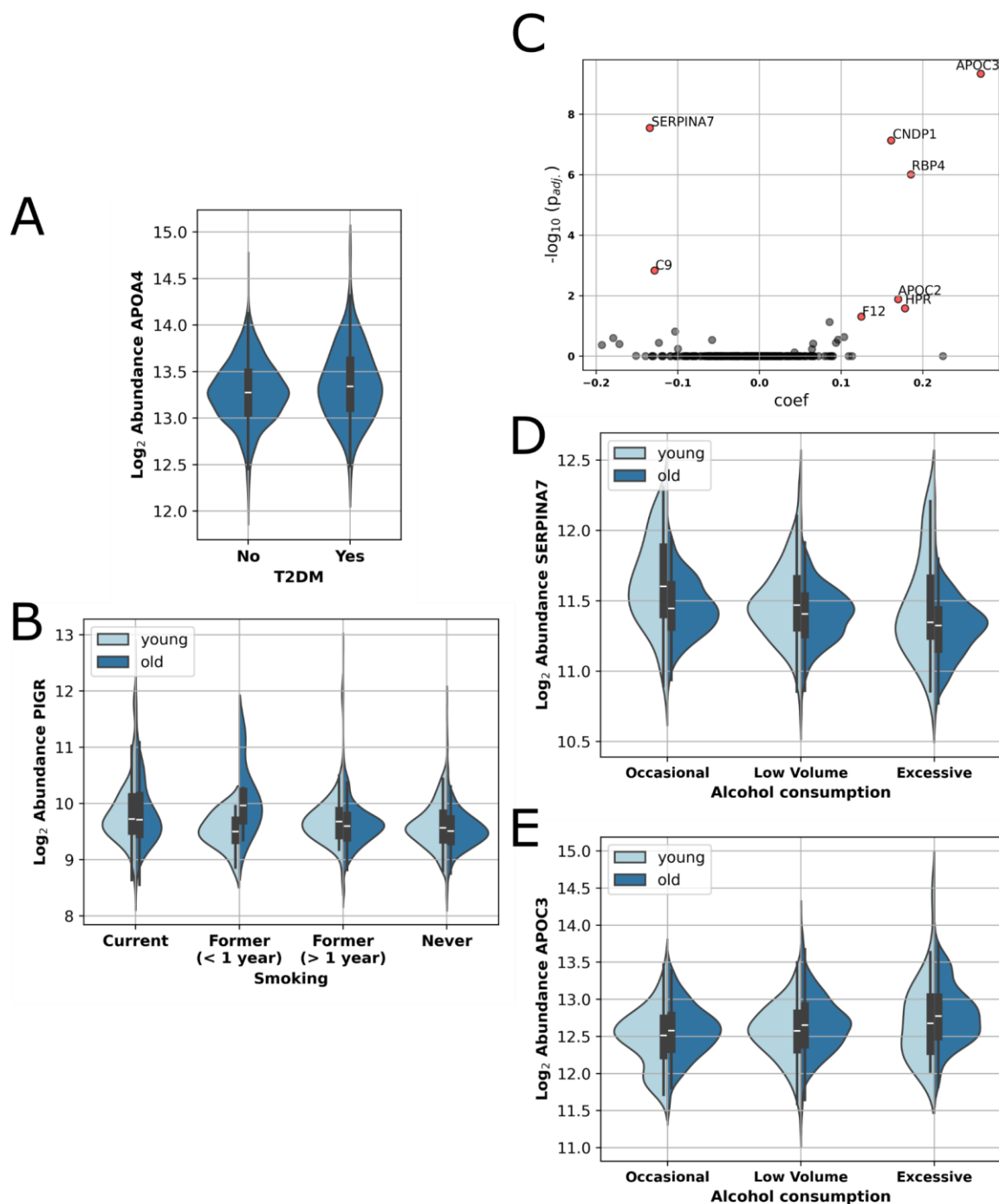
263

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

269

270 Visceral fat accumulation is a key feature of metabolic syndrome and T2DM<sup>53,54</sup>. We analyzed  
271 BMI-associated proteins in "overweight" and "obese" categories, considering T2DM and/or  
272 metabolic syndrome (metS) as covariates. Without lifestyle factors, 19 proteins were linked to  
273 "overweight" and/or "obese" BMI categories (see Table S3). Including T2DM as a covariate,  
274 dopamine-beta hydroxylase (DBH) was significantly associated with "obese". In regression  
275 analyses with T2DM or metS, PGLYRP2<sup>56</sup>, SERPINA10<sup>57</sup>, and LYZ<sup>58</sup> were significant in  
276 both elevated BMI categories, which aligns with current literature. Fibulin 1 (FBLN1) remained  
277 significantly associated ( $p < 0.05$ ) with both BMI categories, irrespective of covariates. The  
278 link between BMI and angiogenesis inhibitor FBLN1 downregulation has not been previously  
279 observed. Among the older participants N = 177, or 9.58% of all participants, were diagnosed  
280 with T2DM. This prevalence aligns closely with the previously reported rate of 9.9% for T2DM  
281 in Germany<sup>62</sup>, based on statutory insurance data<sup>63</sup>. Multiple linear regression analysis  
282 revealed a positive, significant association of APOA4 with T2DM. In an analysis of 539  
283 participants diagnosed with metS, including some T2DM diagnosed participants, APOA4 (see  
284 Figure 3A) again emerged as the only significant protein whose variance exceeded technical  
285 variability when assessing biomarkers for T2DM.

286



287  
 288 **Figure 3:** (A) Increased APOA4 abundance with age and with diabetes mellitus type II (T2DM)  
 289 (B) Effect of smoking on the PIGR abundance levels. Coefficients and adjusted log<sub>10</sub> p-values  
 290 for the proteomic profiles across different participant groups (Volcano Plot) (C). Proteins  
 291 significantly associated with alcohol consumption are marked in red and their gene names are  
 292 annotated. Accompanying split violin plots depict decreasing SERPINA7 (D) and increasing  
 293 APOC3 (E) levels in BASE-II participants with increasing alcohol consumption. Boxes in violin  
 294 plots denote the 25% and 75% quantiles, and the median is displayed by a white line.  
 295  
 296

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

299 <sup>64,65</sup>. The only protein that could exceed the technical threshold when both diseases are  
300 considered as covariates was APOB, which had a significant, positive association with metS  
301 but did not show a significant one for diabetes.

302

### 303 **Impact of menopausal hormone therapy on the high abundant serum proteome**

304

305 We recently reported that HCU has a major impact on the high abundant fraction of the plasma  
306 proteome, in women under 40 years enrolled in the CHRIS study<sup>21</sup>, and as aforementioned,  
307 we detected a similar signature in the young women within the BASE-II cohort. With the focus  
308 on aging, we herein explored the effect of menopausal hormonal replacement therapy (MHT)  
309 in older women participating in the BASE-II. For this, we categorized the BASE-II women from  
310 the older group (N = 689, aged 61-84) into three groups: those who never underwent MHT  
311 (*never*, N = 295), those who completed MHT (*completed*, N = 359) and those who underwent  
312 treatment at sampling time (*ongoing*, N = 35). Only study participants with documented start  
313 and/or termination of MHT were considered. The median treatment duration for the MHT  
314 *completed* group was 11 years (IQR 8 years), and for the *ongoing* group, it was 18 years (IQR  
315 9 years), reflecting a later onset of menopause <sup>70</sup>. Subsequently, we employed multiple  
316 regression models incorporating age, BMI, and a categorical variable denoting MHT status to  
317 evaluate associations between protein abundances and the treatment.

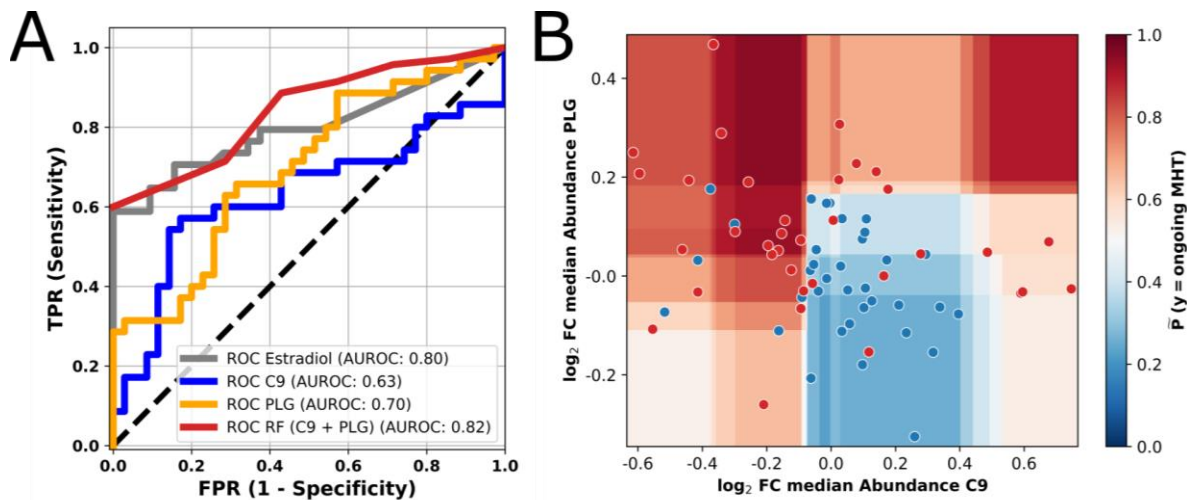
318

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

326

327 Next, we investigated the predictive power of the serum proteome for classifying MHT status  
328 of the female BASE-II participants based on the serum proteome to enhance understanding  
329 of the underlying biological processes affected by hormonal treatments and add to already  
330 available literature concerning treatment effects on the serum proteome. We first matched  
331 female study participants under ongoing treatment (N = 35) to an equally sized control group  
332 by age and BMI based on propensity scores acquired from a generalized linear model <sup>76</sup> (Age  
333 and BMI demographics are listed in Table S7). Next, we selected plasminogen (PLG) and  
334 complement component 9 (C9) as the most informative features by a cross-validated,  
335 recursive feature elimination process<sup>77</sup>. We then trained a random forest <sup>78</sup> in a 5-Fold cross  
336 validation stratified by MHT status. The two-protein classifier achieved a mean test data area  
337 under the curve (AUC) of 0.82, comparable to the AUC of 0.80 for individual estradiol levels,  
338 the active ingredient in most MHT. Moreover, since a training AUC of 0.84 was obtained, the  
339 classifier should be robust against overfitting. ROC curves for different splits in the CV  
340 procedure can be found in Figure S258. Other hormones such as SHBG or testosterone were  
341 outperformed by the random forest predictor (Figure S259). To get a better understanding of  
342 the decision boundaries of the two proteins in the random forest classifier we predicted  
343 decision probabilities for the range of observed PLG and C9 fold changes (FC) to the log2

344 median abundance values (Figure 5B). The highest predictive probability for ongoing MHT  
345 was at PLG at FC above 0.2 and C9 log<sub>2</sub> FCs below -0.1 or above 0.4, although the latter has  
346 to be interpreted with caution since it is not well displayed in our data. Conversely, we expect  
347 the highest predictive power for no MHT at PLG levels below 0.2 and C9 levels between -0.1  
348 and 0.4. Ensemble models that rely on multiple weak learners rarely produce predicted  
349 probabilities close to 0 or 1 because it requires all the individual models to make the same  
350 decision. This is particularly unlikely in random forests due to the bootstrapping (random  
351 sampling with replacement) used to create each individual tree <sup>79</sup>.  
352  
353



354  
355  
356 **Figure 4:** Machine learning classifier based on Complement Component 9 (C9) and  
357 Plasminogen (PLG) abundance levels slightly outperforms serum estradiol levels in predicting  
358 MHT status (A) Receiver operating characteristic (ROC) curve shows predictive performance  
359 of single parameter models with estradiol (coloured in gray), C9 (blue), PLG (orange) and the  
360 two protein random forest model (C9 + PLG) (red) predicting current hormonal treatment. (B)  
361 Contour plot depicting the predictive probability generated by the random forest classifier  
362 based on C9 and PLG levels. The areas highlighted in red represent the abundance ranges  
363 of PLG and C9, which support the classification of participants undergoing MHT. The added  
364 dots display the protein abundance measurements with participants currently under hormonal  
365 treatment (red) and controls (blue). Vertical and horizontal color gradations show cut-offs of  
366 the individual decision trees in the random forest model. Axis show the log<sub>2</sub> FC to the median  
367 abundance of PLG (y) and C9 (x).  
368

## 369 Discussion

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

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

381  
382 In this study, mass spectrometry-based serum proteomics of 1859 BASE-II participants  
383 consistently quantified 2917 peptides from 248 protein groups at high precision. While the  
384 number of proteins in this upper fraction of the plasma proteome may seem limited, this group  
385 of proteins is secreted in many isoforms and is particularly interesting for biomedicine.  
386 Altogether, these proteins account for more than 99% of the plasma's proteomic mass. This  
387 high abundant fraction covers many of the currently clinically used protein biomarkers, spans  
388 over many targets of FDA approved biomarkers, and closely reflects human physiology.  
389 Thereby they cover many functions in the complement system, coagulation, immunoglobulin,  
390 and acute-phase response <sup>10,15,87-91</sup>.

391  
392 Epidemiological covariates such as age, sex and BMI have a considerable impact on the  
393 general health of the population, but the mechanisms are multifactorial and complex. It has  
394 been shown that progressing age and very low or high BMI can pose a risk factor for multiple  
395 disease mechanisms and their progression <sup>92-94</sup>. With the advances of personalized medicine,  
396 these factors are important for treatments tailored to individual patients <sup>95</sup>. Specifically, the  
397 high abundant serum or plasma proteome that plays a vital role in disease pathology and  
398 metabolism, has to be investigated for these effects to reveal true relationships of treatment  
399 and disease <sup>9-12</sup>. Furthermore, differences in epidemiological factors also pose a technical  
400 challenge when analyzing proteomic data. Common techniques to address this include  
401 balancing or using epidemiological information as covariates in linear (factorial) models <sup>96</sup>.

402  
403 Previous proteomic studies into aging cohorts successfully predicted age, mortality, and made  
404 attempts to distinguish biological from chronological age ('proteomic clock') <sup>5,22-24,97</sup>. The  
405 BASE-II design is slightly different to these investigations, as sampling was not continuous  
406 over the age range, but rather categorical, and sampled from an aged group and a young  
407 control. The BASE-II is well annotated with demographic and metadata, which facilitated to  
408 probe for association of age with several demographic and physiological factors. Our  
409 association analysis did both, confirm proteins that were previously reported as age-  
410 dependent in the literature, as well as identify new proteins. Indeed, the abundance levels of  
411 106 of 248 consistently quantified proteins were changed with age. Of these, 39% (N = 41)  
412 were not reported as age-associated in previous datasets <sup>21-30</sup>. Thirty-two of these are  
413 immunoglobulin chains, which is consistent that the adaptive immune system changes as we  
414 age <sup>98</sup>. We hope our results stimulate future investigations using antigen-capture technologies,  
415 to identify the biological targets of these immunoglobulins.

416  
417 Furthermore, we report associations to some pathologically highly relevant proteins including  
418 the pro-inflammatory serum amyloid A4 (SAA4), which has been linked to various infectious  
419 diseases such as malaria or COVID19 <sup>99,100</sup> as well as rheumatic inflammation <sup>101</sup>. Another  
420 protein identified as age-dependent in our study is proteoglycan 4 (PRG4), which has been  
421 associated with tissue regeneration <sup>31</sup> and inflammation regulation <sup>32,33</sup>. It also functions as a  
422 boundary lubricant at the cartilage surface of joints, which function is known to decrease with  
423 progressing age. However, a direct connection to osteoarthritis, which can be caused due to  
424 decreased cartilage lubricants <sup>102</sup> are not yet described. In contrast to our findings, a negative  
425 regulation of PF4V1 and PPBP with age platelet activity has been described while platelet



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

431  
432 The sex dependency of the BASE-II proteome aligns well with the results obtained in other  
433 cohorts <sup>21,23,34-42</sup>. The list of n = 23 sex dependent proteins includes the pregnancy zone  
434 protein (PZP), the sex hormone binding globulin (SHBG) and angiotensinogen (AGT).  
435 Moreover, as we have reported recently, HCU had the strongest impact on the serum  
436 proteome of BASE-II participants, with PGLYRP2 being the main suppressed protein, and  
437 AGT, SERPINA6, FETUB the most induced. We found new HCU associations for 15 proteins,  
438 such as the acute-phase protein SERPINA3 <sup>106</sup>, the monocyte and macrophage cell surface  
439 protein CD14 <sup>107</sup> and the coagulation Factor 5 (F5).

440

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

452

453 We also addressed the role of body weight which is linked to various comorbidities <sup>114</sup>. We did  
454 not find significant associations with underweight, which was most likely caused by the low  
455 sample size and hence lower statistical power. Similar to previous studies<sup>115</sup>, we detected an  
456 association of apolipoproteins D and M levels with elevated BMI. We also report novel protein  
457 associations with BMI including Apolipoprotein C2 (APOC2) and fibulin 1 (FBLN1), when  
458 corrected for metabolic disorders such as T2DM and metS. When both metabolic conditions  
459 (metS and T2DM) are considered as covariates, C5 emerges as a significantly upregulated  
460 protein in obese patients. This could be due to its role in regulating inflammation in adipose  
461 tissue <sup>116</sup>, but also aligns well with our report, that the complement system is BMI dependent  
462 <sup>117</sup>. Interestingly, this association is sensitive to metS and T2DM, which suggests that the  
463 onset of metabolic disease influences the complement system. Similarly, FN1 <sup>118</sup>, VWF <sup>119</sup>,  
464 ECM1 <sup>120</sup>, and ADAMTS13 <sup>121</sup> are only significantly downregulated in obese and overweight  
465 patients when metS and T2DM are included as covariates. Furthermore, we found DBH  
466 significantly associated with "obese" BMI, when adjusted for T2DM in the older group of  
467 participants, aligning with prior findings in DBH-deficient mice <sup>55</sup>.

468

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

472 the serum proteome to varying extents across different age groups. Notably, apolipoprotein  
473 A4 (APOA4) was significantly upregulated in T2DM cases, particularly when metS was  
474 considered. APOA4, primarily synthesized in the small intestine, is crucial for lipid  
475 metabolism, insulin sensitivity, and food intake regulation. Elevated APOA4 levels are also  
476 markers for kidney disease progression and diabetic nephropathy,<sup>124</sup> highlighting its potential  
477 as a marker for T2DM monitoring and prognosis. Significant associations were also found  
478 between metS and various apolipoproteins, including C2, C4, B, and D. Apolipoprotein D  
479 (APOD) was consistently downregulated, especially in the older cohort. APOD, involved in  
480 lipid trafficking and neuroprotection, enhances insulin sensitivity and metabolic health<sup>125</sup> but is  
481 often reduced in metS patients, correlating with hyperchylomicronemia<sup>126</sup>, characterized by  
482 increased triglyceride and chylomicron levels in blood and dysfunctional APOD in metS<sup>127</sup>.  
483 Notably, this downregulation was more pronounced among older metS patients, potentially  
484 reflecting an age-related increase in APOD expression<sup>128</sup>. APOC2<sup>52</sup> and APOC4, part of a  
485 gene cluster including APOE<sup>129</sup>, are implicated in metabolic regulation. While APOE, a known  
486 marker of metabolic syndrome, was not significantly associated with metS in our study,  
487 APOC4 was upregulated in both age groups. APOC2 and APOC4 play pivotal roles in lipid  
488 metabolism and triglyceride processing associated with very low-density lipoprotein (VLDL)  
489<sup>129,130</sup>. Although the literature does not establish a direct link between APOC4 and metS, its  
490 role in lipid metabolism supports a potential relationship with metS and other cardiometabolic  
491 diseases. Prior research has also associated an increase in serum APOC2 levels with a  
492 diagnosis of metS<sup>52</sup>. APOB serves as a clinical marker of atherosclerosis due to its  
493 association with low-density lipoprotein (LDL) and is an independent predictor of metS  
494 development<sup>127,131</sup>. Previous research has shown that the clinical value of APOB in predicting  
495 metS is maintained even after accounting for the comorbidities linked with it<sup>127,132</sup>. Our results  
496 further corroborate these findings, with significant upregulation of APOB in both age groups.  
497 Recent studies have highlighted the critical role of the apolipoprotein protein family in  
498 regulating triglycerides (TGs) and cholesterol metabolism, transport, and cellular management  
499<sup>133</sup>. These findings are consistent with our observations of deregulated apolipoprotein family  
500 members in metabolic diseases, underscoring their significance as potential therapeutic and  
501 diagnostic targets. APOB was the sole protein for which the abundance change associated  
502 with disease surpassed the technical threshold in the regression analysis that considered both  
503 T2DM and metS. Serum APOB levels are directly linked to insulin secretion<sup>134</sup>, where  
504 decreased insulin sensitivity is associated with increased secretion and decreased clearance  
505 of APOB. Prediabetes, characterized by visceral fat accumulation and impaired insulin  
506 sensitivity, is considered an underlying condition of metS<sup>135</sup>. The strong link between metS  
507 and T2DM has been established in relation to both incidence and the molecular mechanisms  
508 underlying their prognosis<sup>64,65</sup>. Notably, administering oral antidiabetic drugs such as  
509 dipeptidyl peptidase-4 inhibitors (DPP4i) in T2DM patients was associated with reduced  
510 plasma APOB levels<sup>66</sup>. Among the participants diagnosed with T2DM, 85 were receiving orally  
511 administered blood glucose-lowering drugs excluding insulin, which indicates the potential  
512 utility of APOB in monitoring treatment outcomes. Our study revealed a novel link between  
513 serum fibulin 1 (FBLN1) levels and higher BMI categories. FBLN1, a regulatory component of  
514 the extracellular matrix, is downregulated in higher BMI categories, indicating potential  
515 disruptions in cellular homeostasis, mobility, and adhesion<sup>138,139</sup>. Markers of chronic  
516 inflammation and cellular degeneration such as lysozyme (LYZ) and alpha-2-macroglobulin  
517 (A2M) were also associated with individuals belonging to the "overweight" and "obese"  
518 categories, but these associations disappeared when adjusting for T2DM and metS. Only

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

521

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

537

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

541 Increased APOC3 levels with increased alcohol consumption have been previously observed  
542<sup>145,146</sup> and decreased serum thyroxine as well as SERPINA7 levels have been linked to  
543 alcoholism<sup>147</sup>. APOC3, a major HDL component released from hepatocytes, influences  
544 triglyceride catabolism and is associated with ethanol consumption, decreasing after alcohol  
545 withdrawal<sup>148,149</sup>. Alcohol-induced increases in HDL are attributed to the increased expression  
546 of cholesteryl ester transfer protein in response to ethanol intake<sup>150</sup>. On the other hand,  
547 SERPINA7 was downregulated in excessive drinkers, a phenomenon observed in the context  
548 of the gut-thyroid axis<sup>151,152</sup>. This observation could be attributed to liver dysfunction due to  
549 overworking, which impairs the liver's ability to effectively synthesize SERPINA7<sup>153</sup>.

550

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

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

572

## 573 **Conclusion**

574

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

587

## 588 **Limitations**

589

590 The study design includes two distinct age groups (27-37 and 60-84 years) and therefore we  
591 cannot claim to ultimately depict effects between these age ranges. However the high overlap  
592 of age-associated proteins between our analysis and comparable studies<sup>21-24</sup> implies that our  
593 analysis is reliable. The analysis of MHT effects is based on a relatively small sample size of  
594 positive cases (N = 35). To minimize adverse effects as much as possible, we therefore  
595 matched the treatment group to a control group with comparable age and BMI. The small  
596 sample size also limits an analysis of admission routes and active drug ingredients (Gestagen  
597 and Estrogen) of MHT. But we report this information, for future reference in the supplement  
598 (see Figure S260). Another potential limitation of our study is the categorization of alcohol  
599 consumption based on self-reported daily intake. Although we used the same thresholds for  
600 both men and women as cited in previous studies<sup>160-162</sup> (see Methods), the German Center  
601 for Addiction Issues (DHS) defines low-risk alcohol consumption differently for women (<12  
602 g/day) and men (<24 g/day)<sup>163</sup>. These differing limits complicated the definition of a moderate  
603 alcohol consumption group and led to a larger portion of the cohort being classified as  
604 excessive drinkers. Given the distribution of age and sex among participants as well as known  
605 clinical details about the group, which were also covariates in our analysis, we chose to define  
606 alcohol consumption limits based on established studies.

607

608

609

610



## 611 Methods

### 612 Berlin Aging Study II

613

614 Berlin Aging Study II (BASE-II) participants were recruited from the greater metropolitan area  
615 of Berlin, Germany. The multidisciplinary BASE-II aims to identify mechanisms and factors  
616 contributing to healthy vs unhealthy aging. In total 1,671 older participants (aged  $\geq 60$  years)  
617 and 500 younger (20-37 years) were assessed in the medical part of the study at baseline  
618 between 2009 and 2014<sup>14</sup>. The medical assessments were conducted in accordance with the  
619 Declaration of Helsinki and approved by the Ethics Committee of the Charité –  
620 Universitätsmedizin Berlin (approval number EA2/029/09) and were registered in the German  
621 Clinical Trials Registry as DRKS00009277. Blood was drawn from all participants after an >8h  
622 fasting period and kept at 4–8 °C until analysis on the same day or was stored at -80 °C  
623 (serum, plasma). Estradiol, total testosterone, and sex hormone-binding globulin (SHBG) were  
624 measured in serum probes by fluoro-, radio- and electrochemiluminescence immunoassays  
625 by an accredited laboratory. Standard laboratory parameters (including estradiol, testosterone  
626 and SHBG) and proteome analyses were performed on blood from the same draw. Smoking  
627 behavior (current smoker, former smoker, never smoker), hormonal contraceptive usage and  
628 menopause hormone replacement therapy were assessed during the 1:1 interview by trained  
629 study personnel. Alcohol consumption was assessed in g/day via food frequency  
630 questionnaire<sup>164</sup>. Body Mass Index (BMI, kg/m<sup>2</sup>) was calculated from height and weight  
631 measured with an electronic measuring station (seca 763 measuring station, SECA,  
632 Germany). Diabetes mellitus type 2 was diagnosed based on American Diabetes Association  
633 (ADA) guidelines<sup>165</sup>, and the metabolic syndrome was diagnosed according to Alberti et al.  
634<sup>166</sup>.

### 635 Sample Preparation

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

641

642 Semi-automated in-solution digestion was performed as previously described for high  
643 throughput clinical proteomics<sup>15</sup>. All stocks and stock plates were prepared in advance to  
644 reduce variability and were stored at -80°C until use. Briefly, 5  $\mu$ l of thawed samples were  
645 transferred to the denaturation and reduction solution (55  $\mu$ l 8 M Urea, 100 mM ammonium  
646 bicarbonate (ABC), 5 mM dithiothreitol per well) mixed and incubated at 30°C for 60 minutes.  
647 Five microliters were then transferred from the iodoacetamide stock solution plate (100 mM)  
648 to the sample plate and incubated in the dark at RT for 30 minutes before dilution with 100  
649 mM ABC buffer (340  $\mu$ l). 220  $\mu$ l of this solution was transferred to the pre-made trypsin stock  
650 solution plate (12.5  $\mu$ l, 0.1  $\mu$ g/ $\mu$ l) and incubated at 37°C for 17 h (Benchmark Scientific Incu-  
651 Mixer MP4). The digestion was quenched by addition of formic acid (10% v/v, 25  $\mu$ l) and  
652 cleaned using C18 solid phase extraction in 96-well plates (BioPureSPE Macro 96-Well, 100  
653 mg PROTO C18, The Nest Group). The eluent was dried under vacuum and reconstituted in  
654 60  $\mu$ l 0.1% formic acid. Insoluble particles were removed by centrifugation and the samples  
655 transferred to a new plate.



## 656 Liquid Chromatography and Mass Spectrometry

657 Digested Peptides were analyzed on a Bruker timsTOF Pro mass spectrometer coupled to a  
658 Bruker VIP-HESI electrospray source. Chromatographic separation was performed on the  
659 same column type but with only 2  $\mu\text{g}$  and a flow rate of 500  $\mu\text{L}/\text{min}$ . For diaPASEF acquisition  
660 <sup>167</sup>, the electrospray source was operated at 3000 V of capillary voltage, a drying gas flow rate  
661 of 10 L/min and 240 °C. The diaPASEF windows scheme was set up as followed: we sampled  
662 an The ion mobility range was sampled from  $1/K0 = 1.30$  to 0.7 Vs/cm<sup>2</sup> using ion accumulation  
663 times of 100ms and ramp times of 133ms in the dual TIMS analyzer, with each cycle times of  
664 1.25 s. The collision energy was lowered as a function of increasing ion mobility from 59 eV  
665 at  $1/K0 = 1.6$  Vs/cm<sup>2</sup> to 20 eV at  $1/K0 = 0.6$  Vs/cm<sup>2</sup>. For all experiments, TIMS elution  
666 voltages were calibrated linearly to obtain the reduced ion mobility coefficients (1/K0) using  
667 three Agilent ESI-L Tuning Mix ions (m/z, 1/K0: 622.0289, 0.9848 Vs/cm<sup>2</sup>; 922.0097, 1.1895  
668 Vs/cm<sup>2</sup>; and 1221.9906, 1.3820 Vs/cm<sup>2</sup>).  
669

## 670 Data processing

671 Raw MS data was processed using DIA-NN v1.8.1 <sup>19</sup>. DIA-NN mass accuracies and scan  
672 window size were fixed to ensure reproducibility (scan window size: 7; MS1/MS2 15 ppm). An  
673 external, publicly available spectral library was used for all measurements<sup>20</sup>. The spectral  
674 library was annotated using the Human UniProt<sup>168</sup> isoform sequence database (Proteome ID:  
675 3AUP000005640)<sup>20</sup>. It is accessible via the PRIDE database under ID PXD013231.

676  
677 All preprocessing steps (including normalization, batch correction and imputation) were  
678 performed in the R programming language (v4.3.1). The raw Dia-NN output was imported into  
679 R using the *msdap* package<sup>169</sup> and initially filtered for serum set QC samples and non-  
680 proteotypic peptides (peptides mapping to multiple proteins).

681  
682 Subsequently, the dataset underwent cyclic loess normalization<sup>170</sup> to normalize peptide  
683 intensities between each sample and each other sample, followed by the  
684 'modebetween\_protein' method that performs the same normalization on a protein level. The  
685 normalized data was filtered iteratively to remove samples with too many missing peptide  
686 intensities (3 median absolute deviations), followed by the filtration of low-presence peptides  
687 measured in less than 30% of all samples. The filtered data was then imputed using Bayesian  
688 PCA (bPCA) from R's *pcaMethods* package<sup>171</sup>, where the missing peptide intensities were  
689 imputed within the same batch, sex (male/female) and age-group combinations (young/old),  
690 but only for peptides in those combinations which were measured in at least 66% of the  
691 samples. Subsequently, batch effects were corrected between all 24 plates using the  
692 *removeBatchEffect* function in the *limma* Bioconductor package<sup>172</sup>.

693  
694 Finally, any remaining missing values in precursors with less than or equal to 34% missing  
695 values were imputed using the *impute.knn* k-nearest neighbor (knn) algorithm from package  
696 impute version 1.76 <sup>173</sup>, followed by a protein-level summarisation using the function  
697 *rcModelPLM* probe-level model (plm) from package preprocessCore version 1.64<sup>174</sup>. The final  
698 dataset after preprocessing consists of 248 proteins and 2917 Peptides for 2079 samples.  
699 However, the number of proteins includes multiple isoforms for Kinogen 1 (2 isoforms, KNG1),  
700 Inter-alpha-trypsin inhibitor heavy chain H4 (2, ITIH4), Gelsolin (2, GSN), alpha-1-B

701 glycoprotein (2, A1BG) and Fibulin 1 (3, FBLN1). Peptide Sequences for all proteins are  
702 reported in Tables S3-S5.  
703

#### 704 Statistical Analysis and Machine Learning

705 Functional protein analysis was performed with the PANTHER classification system<sup>175</sup>.  
706 Subsequent statistical analysis and machine learning of protein data was conducted in the  
707 Python programming language (v3.11), unless specified otherwise. Principal Component  
708 Analysis was conducted with standardized protein abundance data by employing the scikit-  
709 learn package<sup>176</sup>. The arrows added to the plot indicate both the direction and magnitude of  
710 the loadings. A significance level of  $\alpha \leq 0.05$  was employed and p-values were adjusted  
711 by bonferroni. Additionally, we employed a second significance criteria to account for technical  
712 variance, by comparing the coefficient of categorical covariates as a percentage to the CV of  
713 the study pools. We excluded 30 samples due to missing age or BMI information and one  
714 sample due to a duplicated measurement. Linear models, as implemented in statsmodels<sup>177</sup>  
715 ordinary least squares (OLS) function, were fitted to the log<sub>2</sub> protein abundance of all proteins  
716 to investigate associations. Age was treated as a continuous variable and divided by 10, thus  
717 the age-related coefficients show the change in 10 years difference. All other variables used  
718 in the association analysis were treated as categorical variables and were assigned a fixed  
719 reference category. For linear modeling, defined the BMI category “healthy” ( $18.5 < \text{BMI} \leq 25$ ;  
720 reference) as the baseline level and obtained coefficients for the contrast with “underweight”  
721 ( $\text{BMI} < 18.5$ ), “overweight” ( $25 < \text{BMI} \leq 30$ ) and “obese” ( $\text{BMI} > 30$ ). Smoking was also  
722 categorized in 4 groups, such as never smoked (never), stopped more than a year ago (former  
723  $> 1$  year), less than year ago (former  $< 1$  year) and current smokers (current). The alcohol  
724 consumption of participants was categorized according to defined ranges in the related  
725 literature<sup>160–162</sup>. In these studies, occasional drinking is defined as consuming less than 1.3 g  
726 of ethanol per day, low volume drinking is defined as consuming between 1.3 g and 30 g of  
727 ethanol per day, and excessive alcohol consumption is defined as consuming more than 30 g  
728 of ethanol per day. Furthermore, reference categories for sex, MHT, metS and diabetes were  
729 men, never treated with MHT and not diagnosed for metS and diabetes, respectively.  
730 Standardized effect sizes (ES) were obtained by subtracting the mean of every protein and  
731 dividing through its standard deviation prior to linear modeling. Spearman’s rho was used to  
732 calculate correlation coefficients as implemented in Scipy<sup>178</sup> if not indicated otherwise.  
733 Hormonal contraceptive usage (HCU) was defined as in intake within the last 3 months and  
734 considered only oral and not topical administration. Protein CVs for study, pooled and  
735 standardized serum set samples can be found in Table S1. CVs were calculated by dividing  
736 the standard deviation of the abundance data by its mean and expressed as a percentage.

737  
738 Prior to machine learning, we constructed an age and BMI matched control group (N=35) of  
739 women that were never treated with any kind of hormonal replacement therapy to ensure a  
740 balanced dataset. Matching was conducted by employing R’s matchit package<sup>76</sup> with a  
741 “nearest” matching method and by estimating propensity scores by a “mahalanobis” based  
742 distance matrix. Demographics of both, control and treatment group, are shown in Table S7.  
743 We employed random forest models to classify women based on their MHT status as  
744 implemented in the scikit-learn<sup>176</sup> package. Next, to the best predictive performance compared  
745 to other machine learning architectures such as Support Vector machines or boosting  
746 algorithms, random forests also do not need additional feature scaling and are robust to noise

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

## 762 References

- 763 1. Moaddel, R. *et al.* Proteomics in aging research: A roadmap to clinical, translational  
764 research. *Aging Cell* **20**, e13325 (2021).
- 765 2. Ping, P. *et al.* A functional annotation of subproteomes in human plasma. *Proteomics* **5**,  
766 3506–3519 (2005).
- 767 3. Bader, J. M., Albrecht, V. & Mann, M. MS-Based Proteomics of Body Fluids: The End of  
768 the Beginning. *Mol. Cell. Proteomics* **22**, 100577 (2023).
- 769 4. Rando, T. A. & Jones, D. L. Regeneration, Rejuvenation, and Replacement: Turning  
770 Back the Clock on Tissue Aging. *Cold Spring Harb. Perspect. Biol.* **13**, (2021).
- 771 5. Sathyan, S. *et al.* Plasma proteomic profile of age, health span, and all-cause mortality  
772 in older adults. *Aging Cell* **19**, e13250 (2020).
- 773 6. Kuo, C.-L. *et al.* Proteomic aging clock (PAC) predicts age-related outcomes in middle-  
774 aged and older adults. *Aging Cell* e14195 (2024).
- 775 7. Tanaka, T. *et al.* Plasma proteomic signature of age in healthy humans. *Aging Cell* **17**,  
776 e12799 (2018).
- 777 8. Santos-Lozano, A. *et al.* Successful aging: insights from proteome analyses of healthy  
778 centenarians. *Aging* **12**, 3502–3515 (2020).
- 779 9. Anderson, N. L. & Anderson, N. G. The human plasma proteome: history, character,  
780 and diagnostic prospects. *Mol. Cell. Proteomics* **1**, 845–867 (2002).
- 781 10. Vernardis, S. I. *et al.* The Impact of Acute Nutritional Interventions on the Plasma  
782 Proteome. *J. Clin. Endocrinol. Metab.* **108**, 2087–2098 (2023).
- 783 11. Macklin, A., Khan, S. & Kislinger, T. Recent advances in mass spectrometry based  
784 clinical proteomics: applications to cancer research. *Clin. Proteomics* **17**, 17 (2020).
- 785 12. Hartl, J. *et al.* Quantitative protein biomarker panels: a path to improved clinical practice  
786 through proteomics. *EMBO Mol. Med.* **15**, e16061 (2023).
- 787 13. Messner, C. B. *et al.* Mass spectrometry-based high-throughput proteomics and its role  
788 in biomedical studies and systems biology. *Proteomics* **23**, e2200013 (2023).

- 789 14. Bertram, L. *et al.* Cohort profile: The Berlin Aging Study II (BASE-II). *Int. J. Epidemiol.*  
790 **43**, 703–712 (2014).
- 791 15. Messner, C. B. *et al.* Ultra-High-Throughput Clinical Proteomics Reveals Classifiers of  
792 COVID-19 Infection. *Cell Syst* **11**, 11–24.e4 (2020).
- 793 16. Szyrwił, L., Gille, C., Mülleder, M., Demichev, V. & Ralser, M. Fast proteomics with dia-  
794 PASEF and analytical flow-rate chromatography. *Proteomics* **24**, e2300100 (2024).
- 795 17. Skowronek, P. & Meier, F. High-Throughput Mass Spectrometry-Based Proteomics with  
796 dia-PASEF. *Methods Mol. Biol.* **2456**, 15–27 (2022).
- 797 18. Kistner, F., Grossmann, J. L., Sinn, L. R. & Demichev, V. QuantUMS: uncertainty  
798 minimisation enables confident quantification in proteomics. *bioRxiv* 2023.06.20.545604  
799 (2023) doi:10.1101/2023.06.20.545604.
- 800 19. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN: neural  
801 networks and interference correction enable deep proteome coverage in high  
802 throughput. *Nat. Methods* **17**, 41–44 (2019).
- 803 20. Bruderer, R. *et al.* Analysis of 1508 Plasma Samples by Capillary-Flow Data-  
804 Independent Acquisition Profiles Proteomics of Weight Loss and Maintenance. *Mol.*  
805 *Cell. Proteomics* **18**, 1242–1254 (2019).
- 806 21. Dordevic, N. *et al.* Pervasive influence of hormonal contraceptives on the human  
807 plasma proteome in a broad population study. *medRxiv* (2023)  
808 doi:10.1101/2023.10.11.23296871.
- 809 22. Tanaka, T. *et al.* Plasma proteomic biomarker signature of age predicts health and life  
810 span. *Elife* **9**, (2020).
- 811 23. Lehallier, B. *et al.* Undulating changes in human plasma proteome profiles across the  
812 lifespan. *Nat. Med.* **25**, 1843–1850 (2019).
- 813 24. Coenen, L., Lehallier, B., de Vries, H. E. & Middeldorp, J. Markers of aging:  
814 Unsupervised integrated analyses of the human plasma proteome. *Front Aging* **4**,  
815 1112109 (2023).
- 816 25. Jenner, J. L. *et al.* Effects of age, sex, and menopausal status on plasma lipoprotein(a)



- 817 levels. The Framingham Offspring Study. *Circulation* **87**, 1135–1141 (1993).
- 818 26. Alavi, P. *et al.* Aging Is Associated With Organ-Specific Alterations in the Level and  
819 Expression Pattern of von Willebrand Factor. *Arterioscler. Thromb. Vasc. Biol.* **43**,  
820 2183–2196 (2023).
- 821 27. Huang, Y.-M. *et al.* Plasma  $\beta$ 2-microglobulin and cerebrospinal fluid biomarkers of  
822 Alzheimer's disease pathology in cognitively intact older adults: the CABLE study.  
823 *Alzheimers. Res. Ther.* **15**, 69 (2023).
- 824 28. Swanson, C. R. *et al.* Plasma apolipoprotein A1 associates with age at onset and motor  
825 severity in early Parkinson's disease patients. *Mov. Disord.* **30**, 1648–1656 (2015).
- 826 29. Richens, J. L. *et al.* Rationalising the role of Keratin 9 as a biomarker for Alzheimer's  
827 disease. *Sci. Rep.* **6**, 22962 (2016).
- 828 30. Thangaraju, K. *et al.* The Impact of Age and BMI on the VWF/ADAMTS13 Axis and  
829 Simultaneous Thrombin and Plasmin Generation in Hospitalized COVID-19 Patients.  
830 *Front. Med.* **8**, 817305 (2021).
- 831 31. Krawetz, R. J. *et al.* Proteoglycan 4 (PRG4) treatment enhances wound closure and  
832 tissue regeneration. *NPJ Regen Med* **7**, 32 (2022).
- 833 32. Iqbal, S. M. *et al.* Lubricin/Proteoglycan 4 binds to and regulates the activity of Toll-Like  
834 Receptors In Vitro. *Sci. Rep.* **6**, 18910 (2016).
- 835 33. Alquraini, A. *et al.* The interaction of lubricin/proteoglycan 4 (PRG4) with toll-like  
836 receptors 2 and 4: an anti-inflammatory role of PRG4 in synovial fluid. *Arthritis Res.*  
837 *Ther.* **17**, 353 (2015).
- 838 34. Miike, K. *et al.* Proteome profiling reveals gender differences in the composition of  
839 human serum. *Proteomics* **10**, 2678–2691 (2010).
- 840 35. Joaquín, C. *et al.* Pregnancy-associated plasma protein-A is related to gender and to  
841 adipocytokine levels: results of the Health Survey of Catalonia. *Clin. Endocrinol.* **78**,  
842 718–723 (2013).
- 843 36. Gaya da Costa, M. *et al.* Age and Sex-Associated Changes of Complement Activity and  
844 Complement Levels in a Healthy Caucasian Population. *Front. Immunol.* **9**, 2664 (2018).

- 845 37. Saunders, T. S., Jenkins, N., Blennow, K., Ritchie, C. & Muniz-Terrera, G. Interactions  
846 between apolipoprotein E, sex, and amyloid-beta on cerebrospinal fluid p-tau levels in  
847 the European prevention of Alzheimer's dementia longitudinal cohort study (EPAD  
848 LCS). *EBioMedicine* **83**, 104241 (2022).
- 849 38. Pek, S. L. T. *et al.* Association of circulating proinflammatory marker, leucine-rich- $\alpha$ 2-  
850 glycoprotein (LRG1), following metabolic/bariatric surgery. *Diabetes. Metab. Res. Rev.*  
851 **34**, e3029 (2018).
- 852 39. Riestra, P. *et al.* Gender-specific associations between ADIPOQ gene polymorphisms  
853 and adiponectin levels and obesity in the Jackson Heart Study cohort. *BMC Med.*  
854 *Genet.* **16**, 65 (2015).
- 855 40. Pek, S. L. T. *et al.* Elevation of a novel angiogenic factor, leucine-rich- $\alpha$ 2-glycoprotein  
856 (LRG1), is associated with arterial stiffness, endothelial dysfunction, and peripheral  
857 arterial disease in patients with type 2 diabetes. *J. Clin. Endocrinol. Metab.* **100**, 1586–  
858 1593 (2015).
- 859 41. Jaffa, M. A. *et al.* Longitudinal Plasma Kallikrein Levels and Their Association With the  
860 Risk of Cardiovascular Disease Outcomes in Type 1 Diabetes in DCCT/EDIC. *Diabetes*  
861 **69**, 2440–2445 (2020).
- 862 42. Trainor, P. J. *et al.* Blood Levels of Angiotensinogen and Hypertension in the Multi-  
863 Ethnic Study of Atherosclerosis (MESA). *J. Am. Coll. Cardiol.* **81**, 1248–1259 (2023).
- 864 43. Zaghlool, S. B. *et al.* Revealing the role of the human blood plasma proteome in obesity  
865 using genetic drivers. *Nat. Commun.* **12**, 1279 (2021).
- 866 44. Park, J. *et al.* Identification of tetranectin as adipogenic serum protein. *Biochem.*  
867 *Biophys. Res. Commun.* **460**, 583–588 (2015).
- 868 45. Tschoner, A. *et al.* Effect of weight loss on serum pigment epithelium-derived factor  
869 levels. *Eur. J. Clin. Invest.* **41**, 937–942 (2011).
- 870 46. Gonzalez-Quintela, A. *et al.* Determinants of serum concentrations of  
871 lipopolysaccharide-binding protein (LBP) in the adult population: the role of obesity.  
872 *PLoS One* **8**, e54600 (2013).

- 873 47. Baez-Duarte, B. G. *et al.* Serum levels of glutathione peroxidase 3 in overweight and  
874 obese subjects from central Mexico. *Arch. Med. Res.* **43**, 541–547 (2012).
- 875 48. Soria-Jasso, L. E. *et al.* Beneficial and Deleterious Effects of Female Sex Hormones,  
876 Oral Contraceptives, and Phytoestrogens by Immunomodulation on the Liver. *Int. J.*  
877 *Mol. Sci.* **20**, (2019).
- 878 49. Al-Shanqeeti, A., van Hylckama Vlieg, A., Berntorp, E., Rosendaal, F. R. & Broze, G. J.,  
879 Jr. Protein Z and protein Z-dependent protease inhibitor. Determinants of levels and risk  
880 of venous thrombosis. *Thromb. Haemost.* **93**, 411–413 (2005).
- 881 50. Kangasniemi, M. H. *et al.* Ethinylestradiol in combined hormonal contraceptive has a  
882 broader effect on serum proteome compared with estradiol valerate: a randomized  
883 controlled trial. *Hum. Reprod.* **38**, 89–102 (2023).
- 884 51. Pattaro, C. *et al.* The Cooperative Health Research in South Tyrol (CHRIS) study:  
885 rationale, objectives, and preliminary results. *J. Transl. Med.* **13**, 348 (2015).
- 886 52. Boiko, A. S. *et al.* Apolipoprotein serum levels related to metabolic syndrome in patients  
887 with schizophrenia. *Heliyon* **5**, e02033 (2019).
- 888 53. Rahmawati, N. D. *et al.* Body mass index as a dominant risk factor for metabolic  
889 syndrome among Indonesian adults: a 6-year prospective cohort study of non-  
890 communicable diseases. *BMC Nutr* **10**, 43 (2024).
- 891 54. Gray, N., Picone, G., Sloan, F. & Yashkin, A. Relation between BMI and diabetes  
892 mellitus and its complications among US older adults. *South. Med. J.* **108**, 29–36  
893 (2015).
- 894 55. Ste Marie, L., Luquet, S., Curtis, W. & Palmiter, R. D. Norepinephrine- and epinephrine-  
895 deficient mice gain weight normally on a high-fat diet. *Obes. Res.* **13**, 1518–1522  
896 (2005).
- 897 56. Sachs, S. *et al.* Plasma proteome profiles treatment efficacy of incretin dual agonism in  
898 diet-induced obese female and male mice. *Diabetes Obes. Metab.* **23**, 195–207 (2021).
- 899 57. Lee, W.-H. *et al.* Genetically predicted plasma cortisol and common chronic diseases: A  
900 Mendelian randomization study. *Clin. Endocrinol.* **100**, 238–244 (2024).

- 901 58. Moreno-Navarrete, J. M. *et al.* Lysozyme is a component of the innate immune system  
902 linked to obesity associated-chronic low-grade inflammation and altered glucose  
903 tolerance. *Clin. Nutr.* **40**, 1420–1429 (2021).
- 904 59. Moreno-Navarrete, J. M., Ortega, F. J., Bassols, J., Ricart, W. & Fernández-Real, J. M.  
905 Decreased circulating lactoferrin in insulin resistance and altered glucose tolerance as a  
906 possible marker of neutrophil dysfunction in type 2 diabetes. *J. Clin. Endocrinol. Metab.*  
907 **94**, 4036–4044 (2009).
- 908 60. Manco, M. *et al.* The decrease of serum levels of human neutrophil alpha-defensins  
909 parallels with the surgery-induced amelioration of NASH in obesity. *Obes. Surg.* **20**,  
910 1682–1689 (2010).
- 911 61. Fernández-Real, J. M. *et al.* CD14 modulates inflammation-driven insulin resistance.  
912 *Diabetes* **60**, 2179–2186 (2011).
- 913 62. Spieker, J. *et al.* Diabetes type 2 in the Berlin Aging Study II: Cross-sectional and  
914 longitudinal data on prevalence, incidence and severity over on average seven years of  
915 follow-up. *Diabet. Med.* **40**, e15104 (2023).
- 916 63. Heidemann, C. & Scheidt-Nave, C. Prevalence, incidence and mortality of diabetes  
917 mellitus in adults in Germany - A review in the framework of the Diabetes Surveillance. *J*  
918 *Health Monit* **2**, 98–121 (2017).
- 919 64. Ford, E. S., Li, C. & Sattar, N. Metabolic syndrome and incident diabetes: current state  
920 of the evidence. *Diabetes Care* **31**, 1898–1904 (2008).
- 921 65. Regufe, V. M. G., Pinto, C. M. C. B. & Perez, P. M. V. H. C. Metabolic syndrome in type  
922 2 diabetic patients: a review of current evidence. *Porto Biomed J* **5**, e101 (2020).
- 923 66. Piccirillo, F. *et al.* Novel Antidiabetic Agents and Their Effects on Lipid Profile: A Single  
924 Shot for Several Cardiovascular Targets. *Int. J. Mol. Sci.* **24**, (2023).
- 925 67. von Toerne, C. *et al.* MASP1, THBS1, GPLD1 and ApoA-IV are novel biomarkers  
926 associated with prediabetes: the KORA F4 study. *Diabetologia* **59**, 1882–1892 (2016).
- 927 68. Begum, M. *et al.* Adiponectin: A Promising Target for the Treatment of Diabetes and Its  
928 Complications. *Life* **13**, (2023).

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



- 957 81. Sun, B. B. *et al.* Plasma proteomic associations with genetics and health in the UK  
958 Biobank. *Nature* **622**, 329–338 (2023).
- 959 82. Elhadad, M. A. *et al.* Deciphering the Plasma Proteome of Type 2 Diabetes. *Diabetes*  
960 **69**, 2766–2778 (2020).
- 961 83. Suhre, K., McCarthy, M. I. & Schwenk, J. M. Genetics meets proteomics: perspectives  
962 for large population-based studies. *Nat. Rev. Genet.* **22**, 19–37 (2021).
- 963 84. Li, X. *et al.* Inflammation and aging: signaling pathways and intervention therapies.  
964 *Signal Transduction and Targeted Therapy* **8**, 1–29 (2023).
- 965 85. Di Micco, R., Krizhanovsky, V., Baker, D. & d’Adda di Fagagna, F. Cellular senescence  
966 in ageing: from mechanisms to therapeutic opportunities. *Nat. Rev. Mol. Cell Biol.* **22**,  
967 75–95 (2021).
- 968 86. Moqri, M. *et al.* Biomarkers of aging for the identification and evaluation of longevity  
969 interventions. *Cell* **186**, 3758–3775 (2023).
- 970 87. Demichev, V. *et al.* A time-resolved proteomic and prognostic map of COVID-19. *Cell*  
971 *Syst* **12**, 780–794.e7 (2021).
- 972 88. Demichev, V. *et al.* A proteomic survival predictor for COVID-19 patients in intensive  
973 care. *PLOS Digit Health* **1**, e0000007 (2022).
- 974 89. Wang, Z. *et al.* A multiplex protein panel assay for severity prediction and outcome  
975 prognosis in patients with COVID-19: An observational multi-cohort study.  
976 *EClinicalMedicine* **49**, 101495 (2022).
- 977 90. Kumar, V. *et al.* Multiplexed quantitative proteomics provides mechanistic cues for  
978 malaria severity and complexity. *Commun Biol* **3**, 683 (2020).
- 979 91. Messner, C. B. *et al.* Ultra-fast proteomics with Scanning SWATH. *Nat. Biotechnol.* **39**,  
980 846–854 (2021).
- 981 92. Laxy, M., Teuner, C., Holle, R. & Kurz, C. The association between BMI and health-  
982 related quality of life in the US population: sex, age and ethnicity matters. *Int. J. Obes.*  
983 **42**, 318–326 (2018).
- 984 93. GBD 2019 Risk Factors Collaborators. Global burden of 87 risk factors in 204 countries

- 985 and territories, 1990-2019: a systematic analysis for the Global Burden of Disease  
986 Study 2019. *Lancet* **396**, 1223–1249 (2020).
- 987 94. Cuntz, U., Quadflieg, N. & Voderholzer, U. Health Risk and Underweight. *Nutrients* **15**,  
988 (2023).
- 989 95. Stachenfeld, N. S. & Mazure, C. M. Precision medicine requires understanding how  
990 both sex and gender influence health. *Cell* **185**, 1619–1622 (2022).
- 991 96. Kahlert, J., Gribsholt, S. B., Gammelager, H., Dekkers, O. M. & Luta, G. Control of  
992 confounding in the analysis phase - an overview for clinicians. *Clin. Epidemiol.* **9**, 195–  
993 204 (2017).
- 994 97. Sathyan, S. *et al.* Plasma proteomic profile of frailty. *Aging Cell* **19**, e13193 (2020).
- 995 98. Weyand, C. M. & Goronzy, J. J. Aging of the Immune System. Mechanisms and  
996 Therapeutic Targets. *Ann. Am. Thorac. Soc.* **13 Suppl 5**, S422–S428 (2016).
- 997 99. Kassa, F. A. *et al.* New inflammation-related biomarkers during malaria infection. *PLoS*  
998 *One* **6**, e26495 (2011).
- 999 100. Zinellu, A., Paliogiannis, P., Carru, C. & Mangoni, A. A. Serum amyloid A  
1000 concentrations, COVID-19 severity and mortality: An updated systematic review and  
1001 meta-analysis. *Int. J. Infect. Dis.* **105**, 668–674 (2021).
- 1002 101. Sorić Hosman, I., Kos, I. & Lamot, L. Serum Amyloid A in Inflammatory Rheumatic  
1003 Diseases: A Compendious Review of a Renowned Biomarker. *Front. Immunol.* **11**,  
1004 631299 (2020).
- 1005 102. Li, Y. *et al.* Recent Advances in Understanding the Role of Cartilage Lubrication in  
1006 Osteoarthritis. *Molecules* **26**, (2021).
- 1007 103. Le Blanc, J. & Lordkipanidzé, M. Platelet Function in Aging. *Front Cardiovasc Med* **6**,  
1008 109 (2019).
- 1009 104. La Mendola, D., Trincavelli, M. L. & Martini, C. Angiogenesis in Disease. *Int. J. Mol. Sci.*  
1010 **23**, (2022).
- 1011 105. El-Gedaily, A., Schoedon, G., Schneemann, M. & Schaffner, A. Constitutive and  
1012 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-

- 1069 Sectional Study. *Parkinsons Dis.* **2018**, 3751516 (2018).
- 1070 129.Pirim, D. *et al.* Apolipoprotein E-C1-C4-C2 gene cluster region and inter-individual  
1071 variation in plasma lipoprotein levels: a comprehensive genetic association study in two  
1072 ethnic groups. *PLoS One* **14**, e0214060 (2019).
- 1073 130.Kim, S. Y., Park, S. M. & Lee, S.-T. Apolipoprotein C-II is a novel substrate for matrix  
1074 metalloproteinases. *Biochem. Biophys. Res. Commun.* **339**, 47–54 (2006).
- 1075 131.Li, J. *et al.* Association of apolipoprotein Cs with new-onset type 2 diabetes mellitus:  
1076 findings from the Chinese multi-provincial cohort study. *BMJ Open* **11**, e052388 (2021).
- 1077 132.Ryoo, J.-H. & Park, S. K. Association of apolipoprotein B and incidence of metabolic  
1078 syndrome in Korean men: a 5-years' follow-up study. *Atherosclerosis* **226**, 496–501  
1079 (2013).
- 1080 133.Basavaraju, P. *et al.* Genetic Regulatory Networks of Apolipoproteins and Associated  
1081 Medical Risks. *Front Cardiovasc Med* **8**, 788852 (2021).
- 1082 134.Haas, M. E., Attie, A. D. & Biddinger, S. B. The regulation of ApoB metabolism by  
1083 insulin. *Trends Endocrinol. Metab.* **24**, 391–397 (2013).
- 1084 135.Babu, A. & Fogelfeld, L. Metabolic syndrome and prediabetes. *Dis. Mon.* **52**, 55–144  
1085 (2006).
- 1086 136.Nesic, J. *et al.* Adiponectin and Interleukin-33: Possible Early Markers of Metabolic  
1087 Syndrome. *J. Clin. Med. Res.* **12**, (2022).
- 1088 137.Snijder, M. B. *et al.* Associations of adiponectin levels with incident impaired glucose  
1089 metabolism and type 2 diabetes in older men and women: the hoorn study. *Diabetes*  
1090 *Care* **29**, 2498–2503 (2006).
- 1091 138.Roark, E. F. *et al.* The association of human fibulin-1 with elastic fibers: an  
1092 immunohistological, ultrastructural, and RNA study. *J. Histochem. Cytochem.* **43**, 401–  
1093 411 (1995).
- 1094 139.Jie, Y. *et al.* The clinical value of Fibulin-1 for prognosis and its prospective mechanism  
1095 in intrahepatic cholangiocarcinoma. *HPB* **21**, 499–507 (2019).
- 1096 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. Glinoeer, 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-Soehy, 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](mailto: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 repository<sup>179</sup> 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 Mülleider 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.