

1 Identifying Metabolomic and Proteomic Biomarkers for Age-Related Morbidity in a Population-
2 Based Cohort - the Cooperative Health Research in South Tyrol (CHRIS) study
3

4 Essi Hantikainen^{1*}, Christian X. Weichenberger¹, Nikola Dordevic¹, Vinicius Verri Hernandez², Luisa
5 Foco¹, Martin Gögele¹, Roberto Melotti¹, Cristian Pattaro¹, Markus Ralser³, Fatma Amari³, Vadim
6 Farztdinov⁴, Michael Mülleder⁴, Peter P. Pramstaller¹, Johannes Rainer¹, Francisco S. Domingues¹

7

8 ¹ Institute for Biomedicine, Eurac Research, Bolzano, Italy

9 ² Department of Food Chemistry and Toxicology, University of Vienna, Vienna, Austria

10 ³ Department of Biochemistry, Charité – Universitätsmedizin Berlin, Corporate Member of Freie
11 Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

12 ⁴ Core Facility, High-Throughput Mass Spectrometry, Charité, Universitätsmedizin Berlin, Corporate
13 Member of Freie Universität Berlin and Humboldt-Universität Zu Berlin, Berlin, Germany

14

15 *Corresponding author

16 E-mail: essimarjatta.hantikainen@eurac.edu

17

18

19 **Abstract**

20 Identifying biomarkers able to discriminate individuals on different health trajectories is crucial to
21 understand the molecular basis of age-related morbidity. We investigated multi-omics signatures of
22 general health and organ-specific morbidity, as well as their interconnectivity. We examined cross-
23 sectional metabolome and proteome data from 3,142 adults of the Cooperative Health Research in
24 South Tyrol (CHRIS) study, an Alpine population study designed to investigate how human biology,
25 environment, and lifestyle factors contribute to people's health over time. We had 174 metabolites and
26 148 proteins quantified from fasting serum and plasma samples. We used the Cumulative Illness
27 Rating Scale (CIRS) Comorbidity Index (CMI), which considers morbidity in 14 organ systems, to
28 assess health status (any morbidity vs. healthy). Omics-signatures for health status were identified
29 using random forest (RF) classifiers. Linear regression models were fitted to assess directionality of
30 omics markers and health status associations, as well as to identify omics markers related to organ-
31 specific morbidity.

32 Next to age, we identified 21 metabolites and 10 proteins as relevant predictors of health status and
33 results confirmed associations for serotonin and glutamate to be age-independent. Considering organ-
34 specific morbidity, several metabolites and proteins were jointly related to endocrine, cardiovascular,
35 and renal morbidity. To conclude, circulating serotonin was identified as a potential novel predictor
36 for overall morbidity.

37

38

39 Key words: Health Status, Metabolomics, Proteomics, Comorbidity, Aging, CHRIS study

40

41

42 1. Introduction

43 Non-communicable diseases (NCDs) are the leading cause of morbidity and premature mortality
44 globally¹. Age itself is the leading predictor for most NCDs: NCD prevalence increases with age and
45 multiple diseases tend to cluster among older individuals². Common biological processes triggered by
46 molecular damage and modified by cellular and systemic responses drive biological aging and modify
47 risks for multiple diseases in a tissue-, organ- and system-specific manner³. In turn, these health
48 outcomes feed back into the underlying biological processes impacting the rate of aging and
49 enhancing the risk for further disease⁴. It is therefore important to enhance our understanding of the
50 molecular basis of age-related diseases to improve measures for disease prevention and general
51 health⁵. Omics-based biomarkers provide insights into the molecular processes driving functional
52 decline, they also help monitoring health trajectories, age-related physiological decline and disease
53 onset⁶. Such biomarkers can also support the development of prevention strategies targeting those
54 processes and provide surrogate endpoints in intervention studies⁵. The goal is early identification of
55 individuals at higher risk of diseases, who will benefit most from such preventive interventions⁷.
56 Protein biomarkers have the advantage of being direct biological effectors of the underlying genomic
57 background⁷. Serum metabolomics on the other hand provides a snapshot of general physiological
58 state of an organism, and is influenced by genetic, epigenetic, and environmental factors^{8,9}. For
59 instance, Tanaka et al. (2020)⁷ identified a proteomic signature of aging involving 76 proteins and
60 predicting accumulation of chronic diseases and all-cause mortality. You et al. (2023)¹⁰ developed a
61 disease specific proteomic risk score, which stratified the risk for 45 common disease conditions,
62 resulting into an equivalent predictive performance over established clinical indicators for almost all
63 endpoints. Similarly, Gadd et al. (2024)¹¹ demonstrated the utility of proteomic scores in predicting
64 several 10-year incident outcomes beyond factors, such as age, sex, lifestyle and clinically relevant
65 biomarkers, showing the relevance of early proteomic contributions to major age-related diseases.
66 Pietzner et al. (2021)¹² used untargeted metabolomics to investigate signatures of multimorbidity and
67 found that 420 metabolites are shared between at least two chronic diseases. A recent work
68 demonstrated the potential of metabolomic profiles as a multi-disease assay to inform on the risk of
69 many common diseases simultaneously. For 10-year outcome prediction of 15 selected endpoints a
70 combination of age, sex and the metabolomic state was equal or outperformed established
71 predictors¹³.

72

73 Advances in different omics technologies and computing capabilities have also enabled the
74 integration of multi-omics data to capture the complex molecular interplay of health and disease¹⁴. In
75 TwinsUK, using data from 510 women, Zierer et al. (2016)¹⁵ integrated four high-throughput omics
76 datasets and demonstrated the interconnectivity of age-related diseases by highlighting molecular
77 markers of the aging process, which might drive disease comorbidities.

78

79 Here, we aimed to identify multi-omics signatures of general health among adult individuals using
80 cross sectional data from the population-based Cooperative Health Research in South Tyrol (CHRIS)
81 study, an Alpine population study designed to investigate how human biology, environment, and
82 lifestyle factors contribute to people's health over time. We specifically included targeted serum
83 metabolomics¹⁶ and plasma proteomics data¹⁷. The Cumulative Illness Rating Scale (CIRS) based
84 Comorbidity Index (CMI)¹⁸, reflecting disease status and severity in 14 relevant organ systems, was
85 used to assess health status by classifying individuals into having any morbidity (CMI \geq 1) or being
86 healthy (CMI=0). We then applied predictive models using a random forest classifier to determine
87 health status. The analyses were complemented by multiple linear regression models investigating
88 associations and inter-dependencies of individual proteins or metabolites with morbidity assessed by
89 each organ-specific CIRS domains, such as the endocrine-metabolic or the renal domains.

90 **2. Results**

91 The main analytic sample consisted of n=3,142 adult individuals from the CHRIS study with
92 available metabolomics and proteomics data. The AbsoluteIDQ® p180 kit from Biocrates (Biocrates
93 Life Sciences AG, Innsbruck, Austria) was used for metabolite quantification in fasting serum
94 samples¹⁶. The high abundance plasma proteome was determined using the Scanning SWATH mass
95 spectrometry-based approach¹⁷. We first present the study sample's main characteristics by health
96 status (any morbidity vs. healthy) and the relationships between the co-occurring comorbidities. Next,
97 we provide findings from the random forest (RF) analysis. The analyses were complemented by the
98 integration of multiple linear regression models to better characterize actual associations of each
99 significant feature from the RF analysis with health status. Finally, we investigated associations of
100 omics signatures with organ specific morbidity using linear regression models.

101

102 **2.1 Health status and characteristics of the study sample**

103 The characteristics for the main analytic sample are presented in **Table 1**. The CIRS organ domains
104 most completely described by the available data (completeness \geq 50%) were the hypertension, cardiac,
105 respiratory, neurological, renal, vascular, endocrine-metabolic, hepatic and psychiatric/behavioral
106 domains (**Table 1**). The remaining domains had <50% completeness and in general a lower
107 proportion of unhealthy individuals was observed for these domains (**Table 1**). Among all
108 individuals, 56% (n=1,751) were affected by at least one morbidity condition (CMI \geq 1). As expected,
109 these were on average older than healthy individuals (CMI=0; **Table 1, Figure 1**). The top five organ
110 domains affected by health problems were the hepatic, vascular, hypertension, endocrine-metabolic
111 and the respiratory domain, with 27.4%, 14.9%, 12.8%, 10.3%, and 8.8% of morbidity prevalence
112 estimates, respectively. We additionally provide the characteristics for the CHRIS cohort, regardless

113 of available omics data, to confirm the robustness of main characteristics and estimated disease
 114 prevalences in our analytic sample. These are reported in **Supplementary Table S1 and Figure S1**.

115

116 **Table 1.** Characteristics of the main analytic sample with information on health status (any morbidity
 117 vs. healthy)^a and CIRS domain specific morbidity^b.

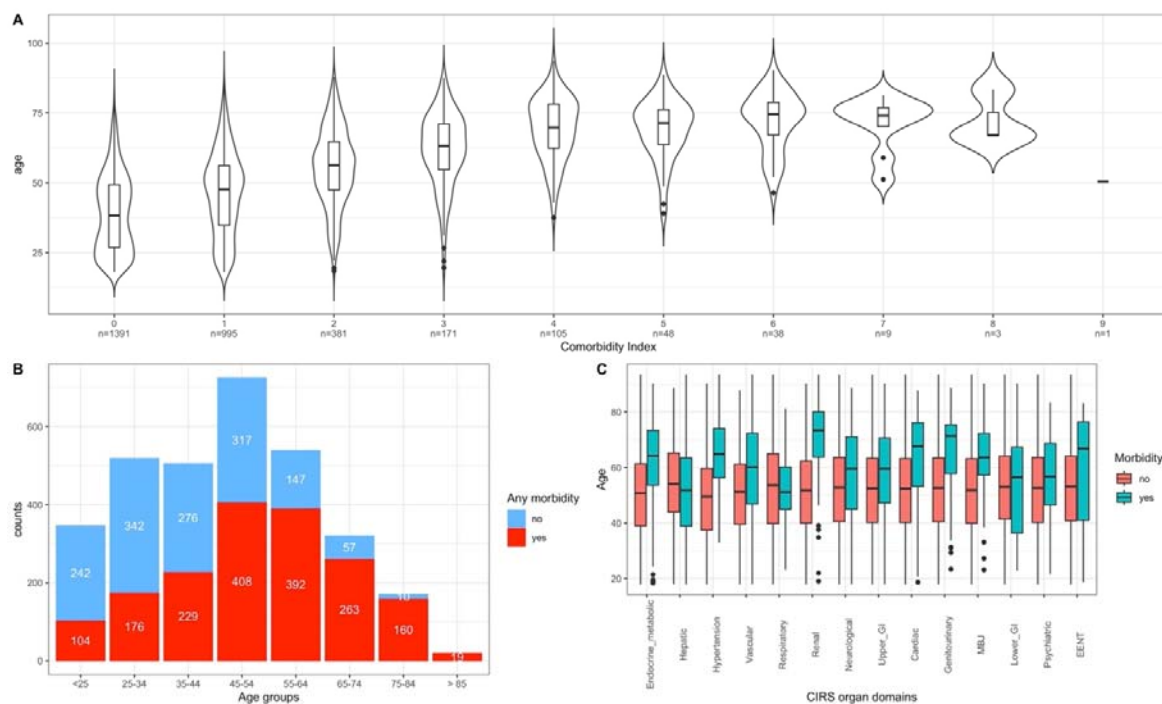
	Overall	Healthy	Any morbidity	Completeness^c
Sample size, n (%)	3,142	1,391 (44%)	1,751 (56%)	
Age, mean (SD)	46.5 (16.7)	39.0 (13.9)	52.5 (16.3)	
Comorbidity Index, mean (SD)	1.04 (1.34)	0	1.86 (1.30)	
Sex, n (%)				
Females	1748 (55.6%)	868 (62.4%)	880 (50.3%)	
Males	1394 (44.4%)	523 (37.6%)	871 (49.7%)	
Morbidity in CIRS domains, yes (%)				
Hepatic	861 (27.4%)		861 (49.2%)	59
Vascular	467 (14.9%)		467 (26.7%)	62
Hypertension	403 (12.8%)		403 (23.0%)	100
Endocrine-Metabolic	325 (10.3%)		325 (18.6%)	74
Respiratory	278 (8.8%)		278 (15.9%)	71
Upper Gastrointestinal	168 (5.3%)		168 (9.6%)	44
Psychiatric and behavioral	151 (4.8%)		151 (8.6%)	53
MBJ ^d	141 (4.5%)		141 (8.1%)	48
Renal	132 (4.2%)		132 (7.5%)	59
Cardiac	110 (3.5%)		110 (6.3%)	71
Neurological	82 (2.6%)		82 (4.7%)	66
Lower Gastrointestinal	67 (2.1%)		67 (3.8%)	44
Genitourinary	58 (1.8%)		58 (3.3%)	33
EENT ^e	11 (0.4%)		8 (0.6%)	32

^aHealth status was assessed through the Cumulative Illness Rating Scale (CIRS) Comorbidity Index (CMI) by classifying individuals as having any morbidity (CMI \geq 1) or being healthy (CMI=0). ^bCIRS domain specific morbidity is defined as having a score \geq 2 in the given domain. ^cCompleteness presents the coverage of the necessary information available in the CHRIS Study with regard to the CIRS guidelines expressed in percentages. ^dMBJ=Musculoskeletal, bones and joints. ^eEENT= Ears, eyes, nose and throat.

118

119

120



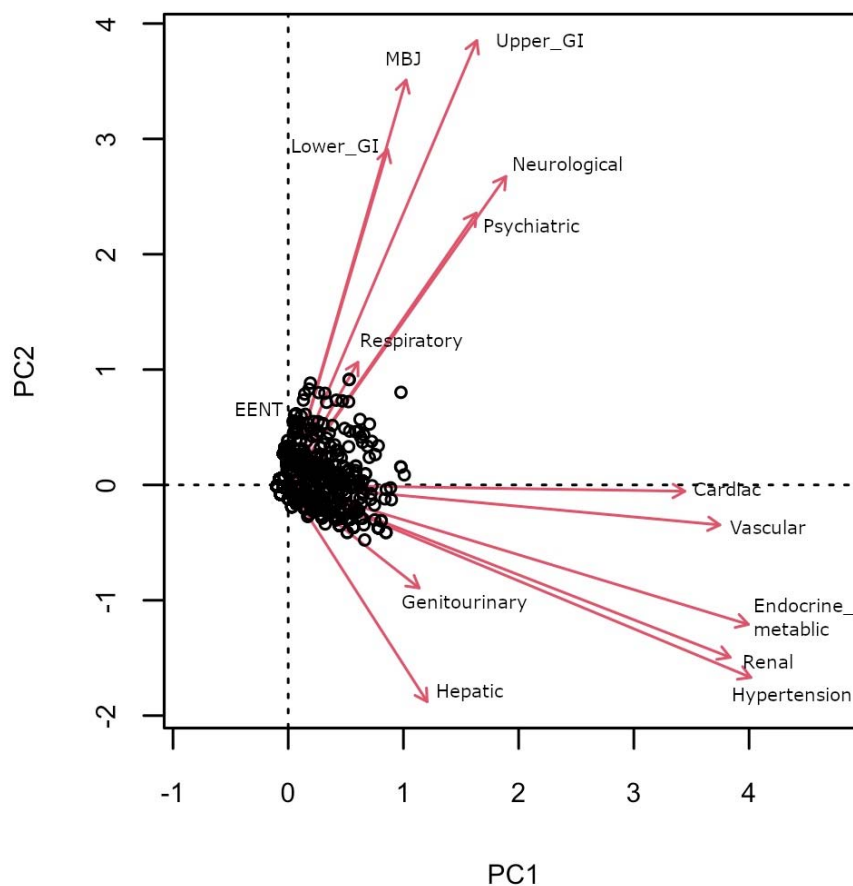
121

122 **Figure 1.** Distribution of health status and age in the CIRSHS sample. **A)** Age distribution by Comorbidity
 123 Index, which is expressed as the total number of CIRS organ domains scoring ≥ 2 . **B)** Absolute distribution of
 124 “Any morbidity” (yes,no) by age-group. The relative proportion of red-bars within each stacked column-bar
 125 represents the age-group specific prevalence of any morbidity. **C)** Age distribution by morbidity conditions in
 126 the specific CIRS domains.

127

128 Next, we explored relationships between the 14 CIRS domains through ordinary correspondence
 129 analysis (OCA; **Figure 2**). We observed proximity between the hypertension, renal and endocrine
 130 domains, the cardiac and the vascular domains, and between the neurological and the psychiatric
 131 domains. To compare the robustness of the comorbidity relationships we present the OCA analysis
 132 results for the CHRIS cohort in **Supplementary Figure S2**.

133



134

135 **Figure 2.** Biplot of ordinary correspondence analysis considering all subjects with available metabolomics and
136 proteomics data, presenting relations between the 14 CIRS domains. CIRS domains with a stronger relation
137 have longer (size consistency) and closer (direction consistency) loadings.

138

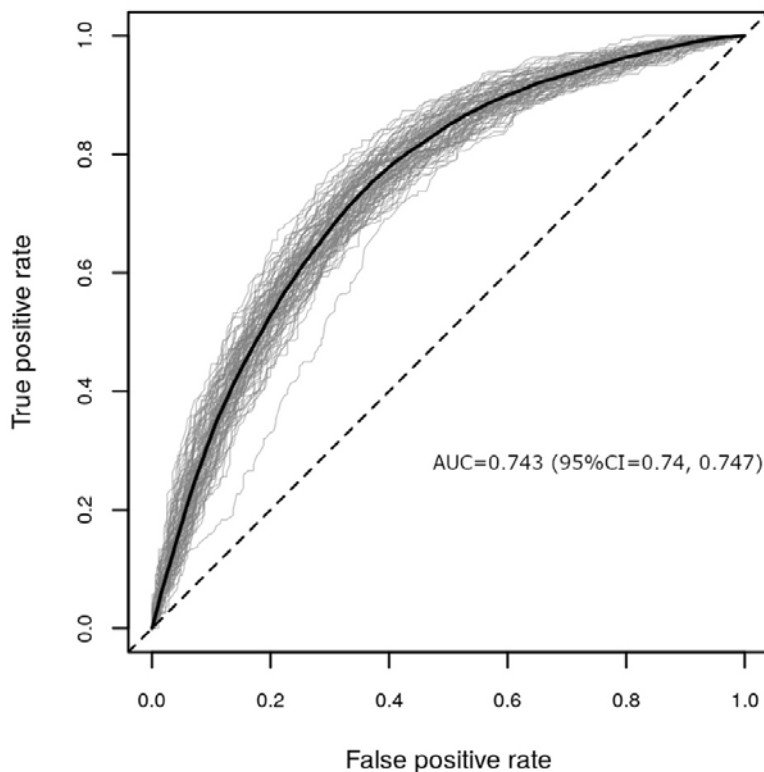
139 2.2 Multi omics signatures of health status

140 To avoid confounding of results due to the impact of medication¹⁷, we performed the analysis on
141 metabolites and protein abundances adjusted for use of medications that were not considered in the
142 CIRS definition (**Supplementary Table S2**).

143

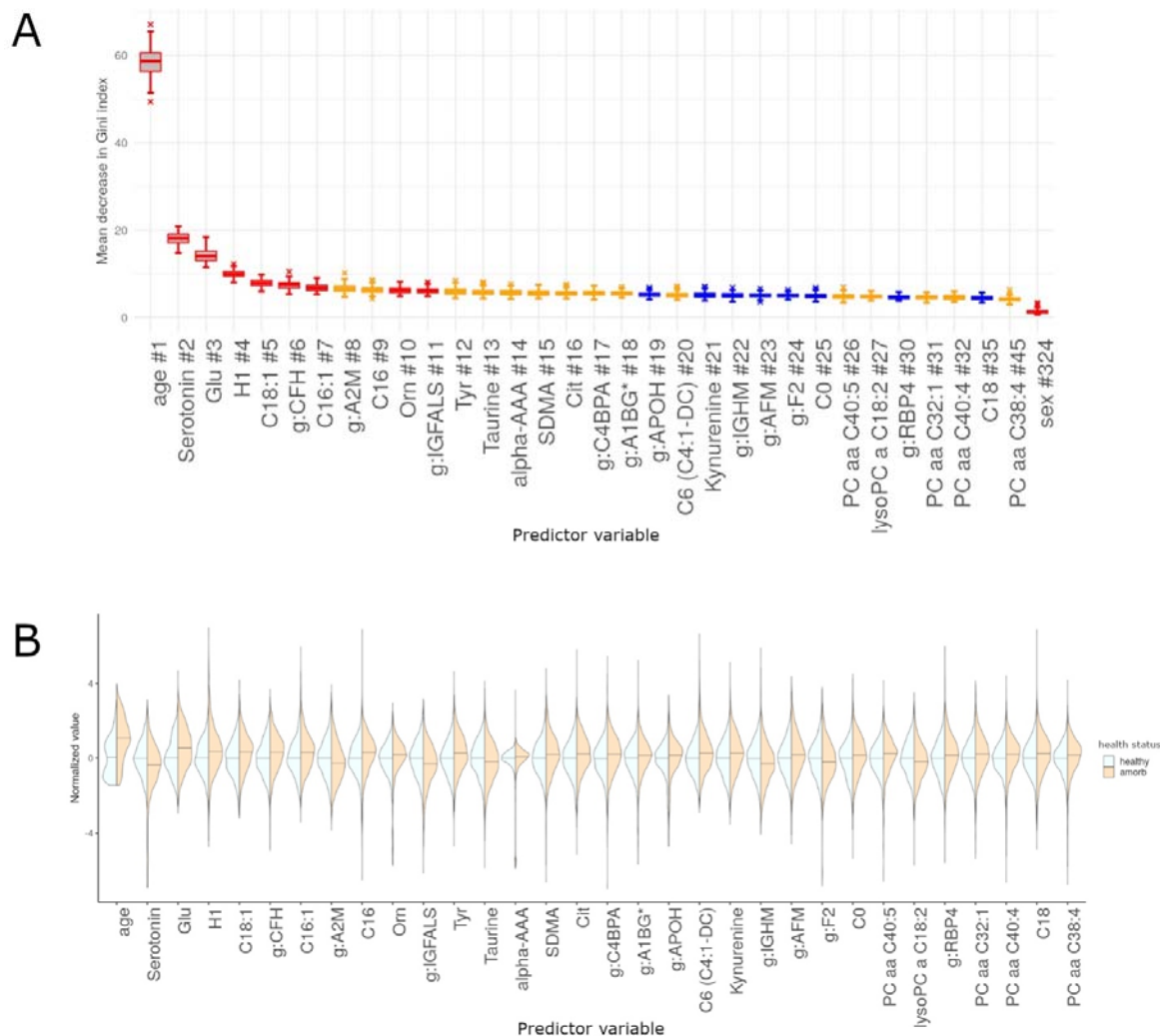
144 For the RF analysis we built a model including age, sex, 174 metabolites and 148 proteins as
145 predictors. Overall, 100 RF models were generated, each containing 500 trees per model, using
146 repeated random subsampling with 80% training and 20% validation set sizes, respectively. We
147 compared the performance measures using the area under the receiver operating curve (ROC AUC,
148 as well as the Matthew's correlation coefficient (MCC) and MCC-F1. The RF model showed
149 moderate performance (AUC=0.743, 95% Confidence Interval (CI)=0.740, 0.747; **Figure 3**; MCC and
150 MCC-F1 are presented in **Supplementary Figure S3**). In addition, we built models that included
151 varying sets of the predictors, which were a) age and sex, b) age, sex and metabolites, c) age, sex and

152 proteins, and d) age, sex, metabolites, proteins. The performance comparison of these to the full
153 model are presented in **Supplementary Text S1**. The metabolomics/proteomics based models (b,c,d)
154 show greater performance but the differences are not statistically significant.
155



156
157 **Figure 3.** Performance evaluation of 100 random forest models including as predictors age, sex, 174 metabolites
158 and 148 proteins to classify health status (any morbidity vs. healthy) represented as ROC AUCs. In the plots
159 each validation run is shown as a gray line with the average curve shown in black.

160
161 We next estimated the significance of the selected individual features using permutation testing. We
162 obtained 33 significant features, including 21 metabolites and 10 proteins, with age being the most
163 important variable and sex the least relevant (**Figure 4A**). Among the top ten omics features were the
164 metabolites serotonin, glutamate, hexose, three acylcarnitines (C18:1, C16:1, C16), ornithine, and the
165 proteins CFH, A2M and IGFALS. The medication-adjusted abundance distributions of these
166 metabolites and proteins stratified by health status are presented in **Figure 4B**. Individuals with any
167 morbidity had lower mean abundance of serotonin, taurine and lysoPC C18:2, and higher mean
168 abundance of all other metabolites. Regarding proteins, individuals with any morbidity had lower
169 mean abundances of A2M, IGFALS, IGHM, and F2, and higher abundance of CFH, C4BPA, A1BG,
170 APOH, AFM, and RBP4.
171



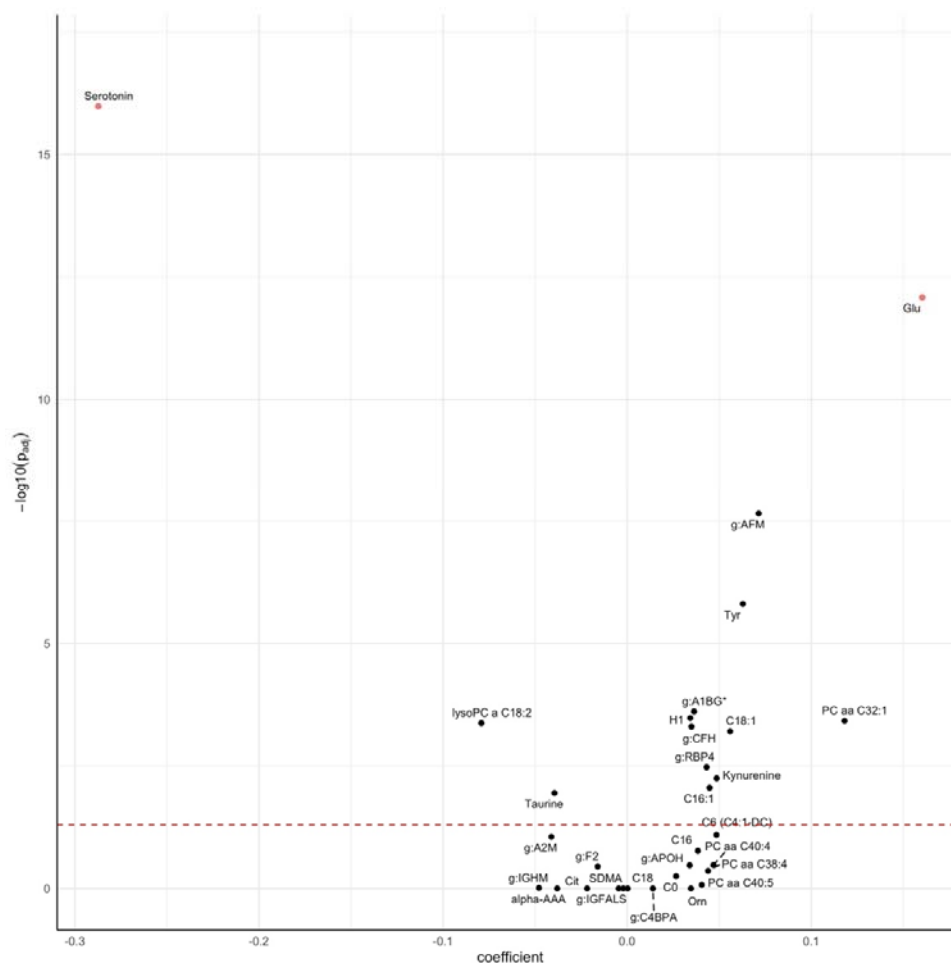
172

173 **Figure 4.** Evaluation of importance for significant features identified through the random forest models
 174 including as predictors age, sex, 174 metabolites and 148 proteins, for health status (any morbidity vs healthy).
 175 #: Rank of the specific feature. Predictor significance was estimated by permutation testing, which generates
 176 100 background distributions, yielding 100 p -values for each predictor variable. Box plots are color-coded by
 177 the number of times a p -value was significant for such a variable ($p < 0.05$): red, all 100 runs returned a
 178 significant p -value; orange, between 80 and 99 runs returned a significant p -value; blue, between 50 and 79 runs
 179 returned a significant p -value. Panels are restricted to features that are significant in at least 50% of all
 180 permutation runs. Density plots are scaled to have the same width, a median of zero (0) (horizontal bar) and
 181 standard deviation of one (1) for the healthy group. For comparability, the any morbidity group data have been
 182 scaled to the standard deviation of the healthy group. **A)** Mean decrease in Gini Index for significant features
 183 **B)** Violin plots presenting scaled abundance distributions stratified by health status for significant features
 184 identified through RF, ordered by importance from left to right.

185

186 To better characterize the actual association of each significant feature from the RF analysis with
 187 health status, we performed separate regression analyses with the abundances of each identified
 188 molecular feature as the response variable and health status (any morbidity vs. healthy), age and sex
 189 as the explanatory variables, allowing us to evaluate the influence of the feature on health status

190 independently of age or sex. Coefficients for health status were extracted from these models and are
191 presented in **Figure 5** and **Supplementary Table S3**. Individuals with any morbidity had, on average,
192 22% lower mean abundance of serotonin and 12% higher abundance of serum glutamate. Twelve
193 other features (C18:1, C16:1, lyso PC a C18:2, PC aa C32:1, tyrosine, taurine, hexose, kynurenine,
194 AFM, CFH, RBP4, and A1BG) passed the multiple-testing correction, however, the observed
195 differences in abundances did fall within the range of the technical variability and were thus not
196 considered significant. Given that age was the strongest predictor for health status in the RF model,
197 we further investigated and compared the coefficients for age and health association from the
198 regression models. For some features, such as citrulline, abundances were almost entirely explained
199 by age and the coefficient for the association with health status from the regression model was only
200 very small, and not significant. For others, such as serotonin and glutamate, associations with health
201 status were strong, even in these age-adjusted models, suggesting an age-independent association of
202 these features with health.
203



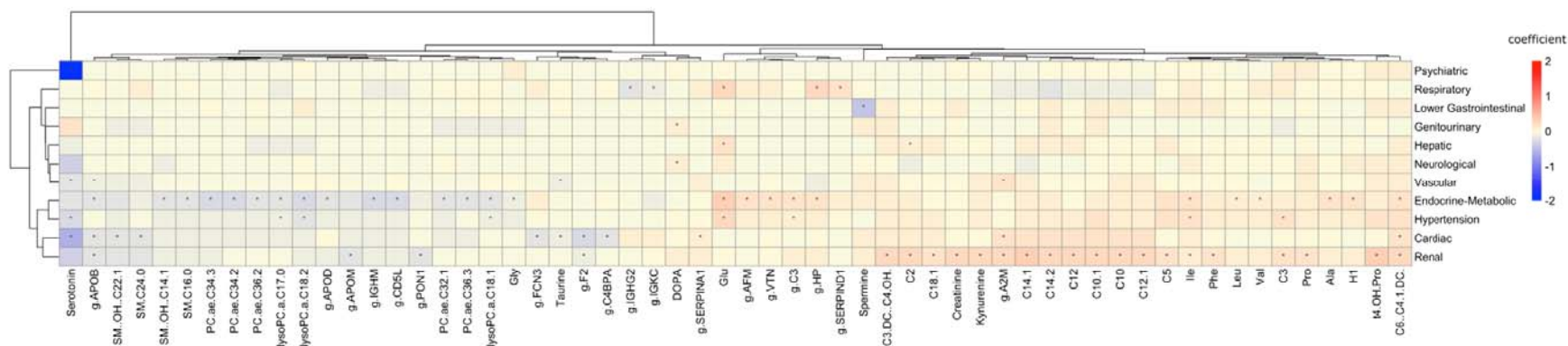
204
205 **Figure 5.** Volcano plot for the differential abundance of metabolomic and proteomic features of health status
206 (any morbidity vs. healthy). Coefficients represent the \log_2 difference in average concentrations between the
207 groups.

208

209 **2.3 Metabolomic and proteomic signatures related to CIRS domain specific morbidity**

210 We further evaluated morbidity markers, implementing separate regression models for medication
211 adjusted abundances of all 174 metabolites and 148 proteins with each CIRS domain as well as age
212 and sex as explanatory variables, and evaluated whether markers were shared across - or specific for
213 any domain (**Figure 6; Supplementary Table S4**). In total, 83 significant omics-disease associations
214 were identified, with 40 metabolites and 17 proteins being significant for ≥ 1 CIRS domain.
215 Associations were observed with the cardiac, vascular, hypertension, endocrine-metabolic, renal,
216 hepatic, psychiatric, neurological, respiratory and lower gastrointestinal and genitourinary domains
217 (**Figure 6**). Eleven metabolites (serotonin, glutamate, isoleucine, taurine, dihydroxyphenylalanine,
218 several glycerophospholipids and acylcarnitines) and three proteins (F2, C3, A2M) were shared across
219 multiple domains. For example, serotonin was related to the cardiac, vascular, hypertension and
220 psychiatric systems, and glutamate to the hypertension, endocrine-metabolic, respiratory and hepatic
221 domains. The proteins F2 and A2M were related to the cardiac, vascular, and the renal domains, and
222 C3 to the hypertension and endocrine-metabolic domains. Overall, phosphatidylcholines and
223 sphingolipids were all negatively associated to morbidity conditions in the related CIRS domain,
224 whereas for acylcarnitines positive associations were observed. The directionality of biogenic amines
225 and amino acids was not consistent across classes but remained consistent across CIRS domains. For
226 proteins, negative associations were observed with APOB, APOD, APOM, IGHM, CD5L, PON1,
227 FCN3, F2, C4BPA, IGHG2 and IGKC, whereas positive associations were found with SERPINA1,
228 AFM, VTN, C3, HP, SERPIND1 and A2M. In general, all metabolites and proteins that were
229 significantly associated with multiple domains showed consistent effect directions, being either
230 negative or positive.

231



232

233

234

235

236

237

Figure 6. Heatmap presenting associations between metabolites, proteins and CIRS organ domains. Associations were obtained from multiple linear regression models adjusted for age and sex. Hierarchical clustering using the Euclidean distance as the similarity measure was applied. Only metabolites and proteins that were significantly associated with at least one specific CIRS domain are presented. *: highlights significant associations. β -coefficients are color-coded: red hues represent positive relations and blue hues represent negative relations.

238

239 **3. Discussion**

240 In this cross-sectional analysis of CHRIS study data we identified age and 31 molecular features
241 predicting overall health status (any morbidity vs healthy) in adults using a random forest classifier.
242 These features included 21 metabolites (5 biogenic amines; 4 amino acids; 6 acylcarnitines; 5
243 glycerophospholipids; and hexose monosaccharides), and 10 plasma proteins. Subsequent regression
244 analyses confirmed a sizeable association of health status with both serotonin and glutamate, which
245 was independent of age and sex. Analyses on single CIRS domains further identified multiple
246 metabolite- and protein-disease associations, most being related to cardiovascular, hypertension,
247 endocrine-metabolic and renal morbidity, revealing strong molecular interconnectivity across these
248 related domains.

249 Several studies have investigated omics markers of aging, longevity^{6,19}, aging-related chronic
250 diseases^{7,11-13,20,21} but only few have integrated multi-omics data with regards to general health
251 assessment^{15,22}. Independent of the approach used, great heterogeneity exists in the included omics
252 technologies, protein and/or metabolite coverage, analytic tools as well as the outcome of interest,
253 which makes comparison of studies challenging. In our study, multi-omics RF models identified
254 several molecular features relevant for predicting health status (any morbidity vs. healthy).

255 Among the age-independent predictive features was serotonin, which is involved in the regulation of
256 energy, glucose and lipid metabolism. Changes in the serotonin system are known risk factors for
257 many age-related diseases, such as diabetes and cardiovascular disease^{23,24}, which was also observed
258 in our study. Up to 95% of serotonin is produced in the gut, and only 5% of serotonin is synthesized
259 by neurons, mainly in the central nervous system²⁴. Although serotonin does not cross the blood-brain
260 barrier, intestinal serotonin release causes neuronal activation in the brain stem, thus indirectly
261 affecting the brain²³. No other study has linked serotonin as a healthy aging marker per se. However,
262 serotonin is a tryptophan derivative, and as inflammation and stress activate the tryptophan
263 metabolism through the kynurenine pathway²⁵, this consequently causes decreased production of
264 serotonin²⁶. The age associated upregulation of kynurenine and downregulation of serotonin therefore
265 indicate a relevant role of tryptophan metabolism in inflammaging and aging²⁷. Although tryptophan
266 was included in the metabolomics panel in this study, it was not identified among the selected features
267 in the RF model. These results indicate a robust relation between health and circulating serotonin
268 levels, but the role of tryptophan metabolism and related pathways, and the putative causal
269 relationships deserves further investigation.

270 Glutamate was also identified as an age-independent predictor for health status. This amino acid has
271 been previously associated with physical frailty in elderly individuals²⁸, supporting our findings. It has
272 further been linked to cardiovascular disease and been suggested to be a potential biomarker of
273 abdominal obesity and metabolic risk²⁹. Glutamate is an important excitatory neurotransmitter in the
274 brain, and although concentrations in the brain are much higher than in plasma as the blood brain
275 barrier is not very permeable to glutamate³⁰, it is also one of the most abundant amino acids in the
276 liver, kidney and skeletal muscle, showing great metabolic versatility³¹. Glutamate plays a key role in
277 protein synthesis and degradation³² and is a by-product of the catabolism of branched chain amino
278 acids²⁹. By linking amino acid and carbohydrate metabolism glutamate supports energy production,
279 which has further implications for insulin secretion³². More specifically, glutamate is also a source of
280 alpha-ketoglutarate, which plays a key role in energy metabolism and aging processes and that has
281 been implicated in improved life and health span³³⁻³⁶. In this study the morbidity group is associated
282 with higher levels of circulating glutamate, which might reflect a depletion of alpha-ketoglutarate
283 relative to healthy individuals.

284 Other relevant molecular features (hexose, C18:1, C16:1, lyso PC a C18:2, PC aa C32:1, tyrosine,
285 taurine, kynurenine, AFM, CFH, RBP4, A1BG) were determined with the RF model, and the age
286 independent associations confirmed by linear regression, but the difference in abundances was lower
287 than the cut-off set relative to the coefficient of variation. We provide further discussion of those
288 features in **Supplementary Text S2**.

289 When looking into organ-specific morbidity, most health conditions in our study were related to the
290 hepatic domain, followed by the vascular, hypertension, endocrine-metabolic and the renal domains.
291 Using OCA, we observed closer relatedness between the renal, hypertension and endocrine-metabolic,
292 as well as the vascular and cardiac domains. Our analyses linking omics markers to specific CIRS
293 domains supported such connections. For example, we observed the metabolites serotonin, glutamate,
294 taurine, isoleucine, three glycerophospholipids (lysoPC a C17.0/18.1/18.2), and two acylcarnitines
295 (C3, C6 (C4)-1-DC) and the proteins C3, APOB, F2, A2M and HP as common markers among the
296 cardiac, vascular, hypertension, endocrine-metabolic, renal but also the respiratory domains. The co-
297 occurrence of type-2 diabetes and cardiovascular diseases is very common, and its high degree of
298 connectivity has been found with other diseases as well³⁷. In addition, cardiovascular and kidney
299 disease are closely interrelated and disease of one organ is known to cause dysfunction of the other³⁸.
300 Remaining molecular signatures were domain specific showing no connections among each other,
301 such as the proteins VTN, APOD and a set of other amino acids (valine, leucine, alanine) being only
302 related to the endocrine-metabolic domain. Such associations have also been reported previously in
303 the literature³⁹⁻⁴¹. Untargeted metabolomics analysis in the prospective, population-based EPIC-
304 Norfolk study identified 420 metabolites that were shared among two or more chronic diseases¹²,

305 further observing high connectivity among cardiometabolic and respiratory diseases across different
306 biochemical classes of metabolites. Those findings highlight potential biological pathways related to
307 the onset of multiple chronic diseases, such as liver and kidney function, lipid and glucose
308 metabolism and low-grade inflammation, among others¹².

309

310 Strengths of our investigation are the wealth of the available data resource, foremost with the
311 availability of both plasma metabolomics and proteomics data among the same participants.
312 Additionally available phenotypic parameters, both quantitative (such as blood parameters), self-
313 reported or collected by trained study-nurses allowed a detailed characterization of the participants
314 and enabled the assessment of the health status through CIRS, which has been shown to be a useful
315 tool to measure morbidity in clinical research⁴².

316 A first limitations is given by the study design, due to which participants with severe morbidity might
317 have been underrepresented. In addition, metabolite and protein concentrations might be influenced
318 by lifestyle, hormonal changes such as introduced by menopausal status and its treatment, as well as
319 medication. Available information on medication use, irrelevant for the CIRS assessment, allowed
320 adjustments to exclude spurious influence of this factor on the results. However, as the CIRS itself
321 considers medication status to characterize certain diseases, it was not possible to distinguish whether
322 the observed associations were driven by the disease itself or by the treatment for the given or related
323 diseases. The present analyses are also limited to the set of metabolites and proteins that are possible
324 to quantify by the analytical approaches used. Finally, given the cross-sectional design of the study
325 we are only able to assess associations, hence no conclusions on temporal antecedence and causality
326 can be drawn. Finally, we did not validate our models on an independent testing set. Despite these
327 limitations, we were able to replicate several findings from previous studies, which supports reliable
328 data and procedural quality within this study. Overall, using multi-omics data for profiling health
329 status has great potential to identify changes in health trajectories at an earlier stage in life. This could
330 help to develop new, effective target therapies for treating related as well as seemingly unrelated
331 diseases occurring at the same time by uncovering common biological pathways connecting different
332 underlying pathogenic mechanisms⁴³.

333

334 To conclude, we identified several molecular signatures of overall health status. Specifically,
335 circulating serotonin is suggested as a promising novel predictor for health and morbidity independent
336 of age, implicating a potential key role of tryptophan metabolism and serotonin related pathways in
337 sustaining health. The results also point to glutamate as another predictor for health and morbidity in
338 adults, in agreement with previous studies relating this amino acid to frailty, metabolic and
339 cardiovascular health. Future studies are needed to investigate the mediating role of these signatures
340 in relation to lifestyle and the environment to promote healthy aging. In this regard the application of

341 mendelian randomization approaches should be considered to further investigate causal links between
342 circulating serotonin, serotonin metabolism and chronic disease.

343

344 **4. Methods**

345 **4.1 Study cohort**

346 The CHRIS study is a population-based cohort of 13,393 adults aged 18 and over recruited from 13
347 towns in the alpine Val Venosta/Vinschgau district in the Bolzano-South Tyrol province of northern
348 Italy. The study was designed to investigate the genetic and molecular basis of age-related common
349 chronic conditions and their interaction with lifestyle and environment in the general population⁴⁴.
350 The study was approved by the Ethics Committee of the Healthcare System of the Autonomous
351 Province of Bolzano. The study conforms to the Declaration of Helsinki, and with national and
352 institutional legal and ethical requirements.

353

354 Metabolomics data were available for n=6,415 individuals and proteomics data for n=3,541. After
355 excluding participants who were not fasting (n=475) or had missing information on fasting status
356 (n=2), as well as women who were pregnant (n=25) or unsure about pregnancy (n=9), n=3,142
357 individuals with overlapping omics data were available for analysis.

358

359 **4.2 Data collection**

360 Data, including collection of anthropometric measurements, blood and urine standard laboratory tests,
361 blood pressure measurement and lifestyle information were collected at the study center following
362 overnight fasting⁴⁴. Laboratory test data included all main cardiovascular and metabolic risk factors,
363 and markers of iron metabolism, coagulation, renal damage, thyroid, and liver function. Blood (serum
364 and plasma) and urine samples were collected and stored in a biobank.

365

366 To obtain information on disease history, participants were interviewed by trained study assistants.
367 Specific clinical domains covered by the questionnaires were the circulatory and nervous system, as
368 well as psychiatric disorders, cognition, autonomic and genitourinary function, endocrine, nutritional,
369 and metabolic diseases. A detailed overview of the mode of assessment can be found elsewhere⁴⁴. In
370 addition, each questionnaire contained a section for “other diseases”, where participants could report
371 any other condition not explicitly included in the domains, as free text. Detailed medication
372 information was collected by scanning the barcodes of the boxes of the medication used within the
373 seven days prior to study center visit and brought by the study participants to the interview. The
374 Anatomical Therapeutic Chemical (ATC) medicinal product classification coding system⁴⁵ and the
375 mode, frequency, and duration of drug administration was recorded for each scanned medication.

376

377 **4.3 Assessment of the health status**

378 To assess health in the CHRIS study sample, we used the Cumulative Illness Rating Scale (CIRS),
379 which is a clinically relevant tool to measure the chronic medical illness burden by taking the number
380 and the severity of chronic diseases into account⁴⁶. For the purpose of this study we used the revised
381 CIRS¹⁸, which assesses 14 organ related domains rating each of them according to the degree of
382 severity ranging from: grade 0 (no impairment) to grade 4 (extremely severe impairment). Based on
383 this guideline, we determined for all CHRIS participants the CIRS score for each domain by screening
384 our questionnaire and interview data, as well as laboratory and clinical parameters for any given
385 medical condition. We next derived the CIRS Comorbidity Index (CMI), which calculates the total
386 number of CIRS organ domains with a score ≥ 2 (domains with moderate or severe morbidity). To
387 define health status we classified individuals as being healthy if the CMI=0 and as having any
388 morbidity if CMI ≥ 1 , which means that at least one CIRS organ domain was identified with moderate
389 or severe morbidity. We additionally considered organ specific health by classifying individuals as
390 being healthy if the CIRS organ domain scored < 2 , or having morbidity if the respective domain
391 scored ≥ 2 .

392 We further assessed the completeness of the CIRS score with regard to the guidelines¹⁸ based on
393 available CHRIS data for each domain. Detailed information on the completeness assessment is
394 presented in **Supplementary Text S3**.

395

396 **Metabolite and protein quantification**

397 The AbsoluteIDQ® p180 kit from Biocrates (Biocrates Life Sciences AG, Innsbruck, Austria) was
398 used for metabolite quantification in fasting serum samples. Details on data generation, quality
399 assessment and normalization are provided in Verri Hernandez et al. (2022)¹⁶. In total, concentrations
400 of 175 metabolites and lipids were quantified. Due to the large number of missing values, sarcosine
401 was excluded from the present analysis. An overview of the included metabolites is presented in the
402 **Supplementary Table S5**. Prior to the main analysis all metabolite concentrations were log₂
403 transformed.

404 The high abundance plasma proteome was determined using the Scanning SWATH mass
405 spectrometry-based approach. Details on sample processing, data acquisition and normalization are
406 provided in Dordevic et al. (2024)¹⁷. In total 148 highly abundant proteins were quantified and
407 included in this analysis. An overview of the included proteins is presented in **Supplementary Table**
408 **S6**.

409

410 Metabolite and protein abundances were adjusted for frequent medication use with a linear model
411 based approach: models were fitted separately for each metabolite or protein with their abundance as

412 response, and medication (as individual binary variables) as explanatory variables. Only medications
413 taken at least twice per week and not considered for CIRS scoring were used. The residuals from
414 these models were used to construct medication-independent abundances for the RF and subsequent
415 regression analyses. **Supplementary Table S2** lists the medications for which abundances were
416 adjusted.

417

418 **4.4 Statistical analysis**

419 Differences in characteristics were presented as mean (standard deviation, SD) for continuous
420 variables and as percentages for categorical variables.

421

422 To investigate the relations between CIRS-domain specific morbidity (morbidity vs healthy), ordinary
423 correspondence analysis (OCA) was performed using the rda function in the vegan package for R⁴⁷.

424

425 *Random forest*

426 We used a random forest (RF) regression model to predict health status (any morbidity vs healthy),
427 including as predictors age, sex, 174 metabolites and 148 proteins, respectively. Overall, 100 RF
428 models were generated, each containing 500 trees per model, considering 80% training set sizes.
429 Predictions were validated by 100 times stratified repeated random subsampling with 20% test set
430 sizes, such that the ratio between the healthy and the any morbidity group of the original set sizes
431 remained the same for the subsampled sets. We used the randomForest R package⁴⁸ to perform the RF
432 analysis.

433

434 Model performance was measured based on the receiver operating characteristic (ROC) area under
435 the curve (AUC), the Matthew's correlation coefficient (MCC), and the unit-normalized Matthews
436 correlation coefficient (MCC-F1).

437

438 The Gini Index was used to measure feature importance. We further estimated the significance of
439 importance metrics for the random forest models by permuting the response variable using the
440 rfpermute R package⁴⁹. This generates a background distribution and recalculates the 100 random
441 forests, each with 500 trees. Usually, a background distribution is calculated only once; however, to
442 increase the robustness of our results, we created 100 background distributions, yielding 100 *p*-values
443 for each predictor variable. We considered associations with a predictor to be significant if ≥ 50 out of
444 100 test runs yielded a *p*-value < 0.05 .

445

446 To further investigate the influence of the significant features obtained from the RF analysis we
447 additionally fitted multiple linear regression models to estimate the association between metabolites

448 and proteins from the selected significant features, each implemented as the response variable, with
449 health status as the explanatory variable and by further adjusting the models for age (years) and sex.

450

451 *Investigation of omics-markers related to CIRS domains*

452 To aid interpretation from the RF analysis, we further investigated the relation between metabolites,
453 proteins and the specific CIRS domains using again multiple linear regression models with
454 implementing all available metabolites and proteins (response variable) one by one with each CIRS
455 domain (explanatory variable), adjusted for age and sex.

456 We adjusted all *p*-values for multiple hypothesis testing using the Bonferroni correction method
457 (metabolites: 0.05/(14 CIRS domains*174 metabolites); proteins: 0.05/(14 CIRS domains*148
458 proteins)), thus an adjusted *p*-value <0.05 was considered statistically significant. For all linear
459 regression analyses we additionally required the difference in abundance between categories to be
460 ≥ 2 * coefficient of variation (CV) for the given metabolite and >1 *CV for the given protein to
461 consider the markers significant. All analyses were conducted using the R statistical software, version
462 4.1.0 (www.R-project.org).

463

464

465 **References**

- 466 1. NCD Countdown 2030 collaborators. NCD Countdown 2030: worldwide trends in non-
467 communicable disease mortality and progress towards Sustainable Development Goal target 3.4.
468 *Lancet Lond. Engl.* **392**, 1072–1088 (2018).
- 469 2. Tchkonina, T. & Kirkland, J. L. Aging, Cell Senescence, and Chronic Disease: Emerging
470 Therapeutic Strategies. *JAMA* **320**, 1319 (2018).
- 471 3. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. Hallmarks of aging: An
472 expanding universe. *Cell* **186**, 243–278 (2023).
- 473 4. Martin-Ruiz, C. & von Zglinicki, T. Biomarkers of healthy ageing: expectations and validation.
474 *Proc. Nutr. Soc.* **73**, 422–429 (2014).
- 475 5. Deelen, J. Targeting multimorbidity: Using healthspan and lifespan to identify biomarkers of
476 ageing that pinpoint shared disease mechanisms. *EBioMedicine* **67**, 103364 (2021).
- 477 6. Menni, C. *et al.* Metabolomic markers reveal novel pathways of ageing and early development in
478 human populations. *Int. J. Epidemiol.* **42**, 1111–1119 (2013).
- 479 7. Tanaka, T. *et al.* Plasma proteomic biomarker signature of age predicts health and life span. *eLife*
480 **9**, e61073 (2020).
- 481 8. Chaleckis, R., Murakami, I., Takada, J., Kondoh, H. & Yanagida, M. Individual variability in
482 human blood metabolites identifies age-related differences. *Proc. Natl. Acad. Sci. U. S. A.* **113**,
483 4252–4259 (2016).

- 484 9. Johnson, C. H. & Gonzalez, F. J. Challenges and opportunities of metabolomics. *J. Cell. Physiol.*
485 **227**, 2975–2981 (2012).
- 486 10. You, J. *et al.* Plasma proteomic profiles predict individual future health risk. *Nat. Commun.* **14**,
487 7817 (2023).
- 488 11. Gadd, D. A. *et al.* Blood protein assessment of leading incident diseases and mortality in the UK
489 Biobank. *Nat. Aging* (2024) doi:10.1038/s43587-024-00655-7.
- 490 12. Pietzner, M. *et al.* Plasma metabolites to profile pathways in noncommunicable disease
491 multimorbidity. *Nat. Med.* **27**, 471–479 (2021).
- 492 13. Buerge, T. *et al.* Metabolomic profiles predict individual multidisease outcomes. *Nat. Med.* **28**,
493 2309–2320 (2022).
- 494 14. Babu, M. & Snyder, M. Multi-Omics Profiling for Health. *Mol. Cell. Proteomics* **22**, 100561
495 (2023).
- 496 15. Zierer, J. *et al.* Exploring the molecular basis of age-related disease comorbidities using a multi-
497 omics graphical model. *Sci. Rep.* **6**, 37646 (2016).
- 498 16. Verri Hernandez, V. *et al.* Age, Sex, Body Mass Index, Diet and Menopause Related Metabolites
499 in a Large Homogeneous Alpine Cohort. *Metabolites* **12**, 205 (2022).
- 500 17. Dordevic, N. *et al.* *Hormonal Contraceptives Are Shaping the Human Plasma Proteome in a*
501 *Large Population Cohort.* <http://medrxiv.org/lookup/doi/10.1101/2023.10.11.23296871> (2023)
502 doi:10.1101/2023.10.11.23296871.
- 503 18. Salvi, F. *et al.* A manual of guidelines to score the modified cumulative illness rating scale and its
504 validation in acute hospitalized elderly patients. *J. Am. Geriatr. Soc.* **56**, 1926–1931 (2008).
- 505 19. Menni, C. *et al.* Circulating Proteomic Signatures of Chronological Age. *J. Gerontol. Ser. A* **70**,
506 809–816 (2015).
- 507 20. Orwoll, E. S. *et al.* Proteomic assessment of serum biomarkers of longevity in older men. *Aging*
508 *Cell* **19**, (2020).
- 509 21. Santos-Lozano, A. *et al.* Successful aging: insights from proteome analyses of healthy
510 centenarians. *Aging* **12**, 3502–3515 (2020).
- 511 22. Zhang, W. *et al.* A population-based study of precision health assessments using multi-omics
512 network-derived biological functional modules. *Cell Rep. Med.* **3**, 100847 (2022).
- 513 23. Fidalgo, S., Ivanov, D. K. & Wood, S. H. Serotonin: from top to bottom. *Biogerontology* **14**, 21–
514 45 (2013).
- 515 24. Martin, A. M. *et al.* The Diverse Metabolic Roles of Peripheral Serotonin. *Endocrinology* **158**,
516 1049–1063 (2017).
- 517 25. Calder, P. C. *et al.* Health relevance of the modification of low grade inflammation in ageing
518 (inflammageing) and the role of nutrition. *Ageing Res. Rev.* **40**, 95–119 (2017).
- 519 26. Kanova, M. & Kohout, P. Tryptophan: A Unique Role in the Critically Ill. *Int. J. Mol. Sci.* **22**,
520 11714 (2021).

- 521 27. Lassen, J. K. *et al.* LARGE-SCALE metabolomics: Predicting biological age using 10,133 routine
522 untargeted LC-MS measurements. *Aging Cell* (2023) doi:10.1111/ace1.13813.
- 523 28. Calvani, R. *et al.* A Distinct Pattern of Circulating Amino Acids Characterizes Older Persons
524 with Physical Frailty and Sarcopenia: Results from the BIOSPHERE Study. *Nutrients* **10**, 1691
525 (2018).
- 526 29. Maltais-Payette, I., Allam-Ndoul, B., Pérusse, L., Vohl, M.-C. & Tchernof, A. Circulating
527 glutamate level as a potential biomarker for abdominal obesity and metabolic risk. *Nutr. Metab.*
528 *Cardiovasc. Dis.* **29**, 1353–1360 (2019).
- 529 30. Hawkins, R. A. The blood-brain barrier and glutamate. *Am. J. Clin. Nutr.* **90**, 867S-874S (2009).
- 530 31. Brosnan, J. T. Glutamate, at the Interface between Amino Acid and Carbohydrate Metabolism. *J.*
531 *Nutr.* **130**, 988S-990S (2000).
- 532 32. Kelly, A. & Stanley, C. A. Disorders of glutamate metabolism. *Ment. Retard. Dev. Disabil. Res.*
533 *Rev.* **7**, 287–295 (2001).
- 534 33. Treberg, J. R., Banh, S., Pandey, U. & Weihrauch, D. Intertissue differences for the role of
535 glutamate dehydrogenase in metabolism. *Neurochem. Res.* **39**, 516–526 (2014).
- 536 34. Canfield, C.-A. & Bradshaw, P. C. Amino acids in the regulation of aging and aging-related
537 diseases. *Transl. Med. Aging* **3**, 70–89 (2019).
- 538 35. Nassar, K. *et al.* The significance of caloric restriction mimetics as anti-aging drugs. *Biochem.*
539 *Biophys. Res. Commun.* **692**, 149354 (2024).
- 540 36. Kurhaluk, N. Tricarboxylic Acid Cycle Intermediates and Individual Ageing. *Biomolecules* **14**,
541 260 (2024).
- 542 37. Guasch-Ferré, M. *et al.* Metabolomics in Prediabetes and Diabetes: A Systematic Review and
543 Meta-analysis. *Diabetes Care* **39**, 833–846 (2016).
- 544 38. Liu, M. *et al.* Cardiovascular disease and its relationship with chronic kidney disease. *Eur. Rev.*
545 *Med. Pharmacol. Sci.* **18**, 2918–2926 (2014).
- 546 39. Cai, X. *et al.* Population serum proteomics uncovers a prognostic protein classifier for metabolic
547 syndrome. *Cell Rep. Med.* **4**, 101172 (2023).
- 548 40. Coan, P. M. *et al.* Complement Factor B Is a Determinant of Both Metabolic and Cardiovascular
549 Features of Metabolic Syndrome. *Hypertension* **70**, 624–633 (2017).
- 550 41. Kollerits, B. *et al.* Plasma Concentrations of Afamin Are Associated With Prevalent and Incident
551 Type 2 Diabetes: A Pooled Analysis in More Than 20,000 Individuals. *Diabetes Care* **40**, 1386–
552 1393 (2017).
- 553 42. de Groot, V., Beckerman, H., Lankhorst, G. & Bouter, L. How to measure comorbidity: a critical
554 review of available methods. *J. Clin. Epidemiol.* **56**, 221–229 (2003).
- 555 43. Dash, P., Mohapatra, S. R. & Pati, S. Metabolomics of Multimorbidity: Could It Be the Quo
556 Vadis? *Front. Mol. Biosci.* **9**, 848971 (2022).

- 557 44. Pattaro, C. *et al.* The Cooperative Health Research in South Tyrol (CHRIS) study: rationale,
558 objectives, and preliminary results. *J. Transl. Med.* **13**, 348 (2015).
- 559 45. WHO Collaborating Centre for Drug Statistics Methodology, ATC classification index with
560 DDDs, 2022. Oslo, Norway 2021.
- 561 46. Linn, B. S., Linn, M. W. & Gurel, L. Cumulative illness rating scale. *J. Am. Geriatr. Soc.* **16**,
562 622–626 (1968).
- 563 47. Oksanen J, Simpson G, Blanchet F, Kindt R, Legendre P, Minchin P, O’Hara R, Solymos P,
564 Stevens M, Szoecs E, Wagner H, Barbour M, Bedward M, Bolker B, Borcard D, Carvalho G,
565 Chirico M, De Caceres M, Durand S, Evangelista H, FitzJohn R, Friendly M, Furneaux B,
566 Hannigan G, Hill M, Lahti L, McGlenn D, Ouellette M, Ribeiro Cunha E, Smith T, Stier A, Ter,
567 & Braak C, Weedon J. *_vegan: Community Ecology Package_*. R package version 2.6-4. (2022).
- 568 48. Liaw A, Wiener M. Classification and Regression by randomForest. *R News*, 2(3), 18-22.
569 <https://CRAN.R-project.org/doc/Rnews/>. (2002).
- 570 49. Archer, E. rfPermute: Estimate Permutation p-Values for Random Forest Importance Metrics.
571 2.5.2 <https://doi.org/10.32614/CRAN.package.rfPermute> (2011).

572

573 **Acknowledgments**

574 CHRIS study investigators thank all study participants, the Healthcare System of the Autonomous
575 Province of Bolzano-South Tyrol, and all Eurac Research staff involved in the study. Bioresource
576 Impact Factor Code: BRIF6107.

577

578 **Author contributions** All authors contributed to the study conception and design. Material
579 preparation and data collection were performed by J.R, V.H, F.D, P.P, M.G, L.F, C.P, F.A, V.F, M.R,
580 M.M and E.H. Formal analysis and investigation was performed by C.W, E.H, N.D, J.R and F.D. All
581 authors contributed to the interpretation of the findings. The first draft of the manuscript was written
582 by E.H, C.W, N.D, J.R and F.D and all authors commented on previous versions of the manuscript.
583 All authors read and approved the final manuscript.

584 **Data availability** The data are not openly available due to reasons of sensitivity. CHRIS study data
585 can be requested for research purposes by submitting a dedicated request to the CHRIS Access
586 Committee (access.request.biomedicine@eurac.edu).

587 **Conflict of interest** The authors have no conflict of interest to declare.

588

589 **Informed Consent Statement** Informed consent was obtained from all subjects involved in the study.

590

591 **Financial Disclosure Statement** The CHRIS study was funded by the Department of Innovation,
592 Research and University of the Autonomous Province of Bolzano-South Tyrol and supported by the
593 European Regional Development Fund (FESR1157). The funders had no role in study design, data
594 collection and analysis, decision to publish, or preparation of the manuscript.

595

596

597

598