1   **Unmasking Hidden Systemic Effects of Neurodegenerative Diseases: A Two-Pronged**

2   **Approach to Biomarker Discovery**

3   Sandra I. Anjo[1,2,3,4,*,†], Miguel Rosado[1,2,3,5,*], Inês Baldeiras[1,2,6], Andreia Gomes[7], Diana Pires[7], Cátia

4   Santa[1,2,3], Joana Pinto[1,2], Cristina Januário[6], Isabel Santana[1,2,6], Ana Verdelho[8,9], Alexandre de

5   Mendonça[9], Miguel Castelo-Branco[7,#], Bruno Manadas[1,2,3#,†]

6   [1] CNC - Centro de Neurociências e Biologia Celular, Universidade de Coimbra, Rua Larga - Faculdade de

7   Medicina, 1ºandar - POLO I, 3004-504 Coimbra, Portugal.

8   [2] CIBB - Centre for Innovative Biomedicine and Biotechnology, University of Coimbra, Rua Larga -

9   Faculdade de Medicina, 1ºandar - POLO I, 3004-504 Coimbra, Portugal.

10   [3] Instituto de Investigação Interdisciplinar, Universidade de Coimbra (IIIUC), Casa Costa Alemão, R. Dom

11   Francisco de Lemos, 3030-789 Coimbra, Portugal.

12   [4] Faculdade de Medicina, Universidade de Coimbra, Azinhaga de Santa Comba, 3000-548 Coimbra,

13   Portugal.

14   [5] - University of Coimbra, Institute for Interdisciplinary Research, Doctoral Programme in Experimental

15   Biology and Biomedicine (PDBEB), Portugal

16   [6] Serviço de Neurologia, Centro Hospitalar e Universitário de Coimbra, Praceta Professor Mota Pinto,

17   3004-561 Coimbra, Portugal.

18   [7] Coimbra Institute for Biomedical Imaging and Translational Research (CIBIT), Azinhaga de. Santa

19   Comba, 3000-548 Coimbra, Portugal.

20   [8] Department of Neurosciences and Mental Health, Santa Maria Hospital - CHLN, ISAMB, University of

21   Lisbon, Av. Prof. Egas Moniz MB, 1649-028 Lisboa, Portugal.

22   [9] Instituto de Medicina Molecular e Faculdade de Medicina, Universidade de Lisboa, Av. Prof. Egas

23   Moniz, 1649-028 Lisboa, Portugal.

24   * - Equal contribution

25   # - Senior equal contribution

26   † - Corresponding authors: Bruno Manadas (bmanadas@cnc.uc.pt, +351231249170) and Sandra Anjo

27   (sandra.anjo@cnc.uc.pt, +351231249170), Center for Neuroscience and Cell Biology – University of

28   Coimbra, UC Biotech - Parque Tecnológico de Cantanhede, Núcleo 04, Lote 8, 3060-197 Cantanhede –

29   Portugal

30   E-mail addresses: Miguel Rosado - mmva.rosado@gmail.com; Ines Baldeiras - ibaldeiras@cnc.uc.pt;

31   Andreia Gomes - andmaggomes@gmail.com; Diana Pires - diana.santos.pires@gmail.com; Cátia Santa -

32   catiajmsanta@gmail.com; Joana Pinto - joana.amaralpinto@gmail.com; Cristina Januário -

33   cristinajanuario@gmail.com; Isabel Santana - isabeljsantana@gmail.com; Ana Verdelho -

34   averdelho@medicina.ulisboa.pt; Alexandre de Mendonça - mendonca@medicina.ulisboa.pt; Miguel

35   Castelo-Branco   -   mcbranco@fmed.uc.pt.

36   **Abstract**

37   **Background:** The identification of reliable blood biomarkers for neurodegenerative Diseases (NDs) has

38   been of pivotal importance in translational/clinical research. However, conventional omics struggle with

39   the complexity of blood samples, which makes it difficult to achieve the desired goal. To address this, in

40   this work the potential of High Molecular Weight (HMW) fractionation under non-denaturing conditions

41   as a complementary approach to the conventional proteomics for identifying serum biomarkers in NDs

42   was explored.

43   **Methods:** A cohort of 58 serum samples of Alzheimer's disease (AD), Parkinson's disease (PD) patients

44   and control (CT) individuals was used to compare the two proteomics strategies: i) direct analysis of

45   whole serum and ii) non-denaturing fractionation using 300 kDa cut-off filters (HMW serum). Univariate

46   analysis was applied to the proteins quantified by each approach to identify the subset of proteins

47   altered among the different groups (AD, PD and CT). Subsequently, linear discriminant analysis was

48   performed using each subset of differently altered proteins, either individually or in combination, to

49   construct the predictive models for the diseases under study and to identify a panel of potential

50   biomarkers that could aid in the diagnosis of AD and PD.

51   **Results:** Although both approaches quantified a similar set of proteins, it was observed that each

52   approach capture a distinct subset of differentially altered proteins, suggesting that HMW fractionation

53   is capable of identifying additional types of alterations beyond conventional protein level changes. The

54   discriminant model created by combining altered proteins from both datasets demonstrated an

55   impressive efficacy in distinguishing between the three groups (AUC = 0.999 and, median sensitivity and

56   specificity of 97.4% and 91.7%, respectively). Importantly, this performance surpassed that of any model

57   created using each method individually.

58   **Conclusions:** Altogether, this work demonstrated that HMW fractionation can be a valuable

59   complementary method to direct serum analysis and could enhance biomarker discovery. The 10

60   proteins included in the model (5 from each strategy), comprise clear evidence for the contribution of

61   apolipoproteins for the diagnosis of NDs, revealing potential changes within lipid metabolism and the

62 organization of macromolecules and their complexes, thereby uncovering effects that remain hidden

63 from a conventional serum proteome analysis.

64   **Keywords**

65   Serum Proteomics

66   High Molecular Weight Fractionation

67   Neurodegenerative Diseases

68   SWATH-MS/DIA

69   Protein complexes/Proteoforms

70   Apolipoproteins

71   Lipoproteins

72   **Background**

73   Proteins can be directly or indirectly related to a myriad of diseases, thereby being important targets of

74   biomarker research, which remains a pivotal aspect of clinical research. However, biomarker discovery

75   based on conventional proteomics strategies has not yet yielded substantial practical applications. Blood

76   and its fractions are the most studied samples for biomarker identification, not only due to easy

77   accessibility but also because of its interactions with most tissues in the body, potentially reflecting

78   disease-related alterations (1, 2). However, both plasma and serum proteomes have a large dynamic

79   range of protein concentrations (1), with the detection of those least abundant being masked by the

80   most abundant. Despite all the technical developments in proteomics quantification by mass

81   spectrometry (MS) (3), the high dynamic range still precludes the complete characterization of these

82   samples, which may be one of the reasons why the identification of relevant biomarkers in

83   plasma/serum by MS remains challenging (4). Therefore, it is essential to have alternative approaches to

84   the direct analysis of these highly demanding biofluids. Several approaches have been developed to

85   reduce sample complexity (5), such as size exclusion chromatography and electrophoretic separation

86   methods. When dealing with numerous samples, a more straightforward approach that could serve as

87   an alternative to direct sample analysis is centrifugal ultrafiltration (5). This approach relies on the

88   separation of sample components based on size exclusion, which is correlated to molecular weight

89   (MW). This method has typically been used to study the low molecular weight proteome of serum and

90   plasma samples, using filters between 20 kDa and 40 kDa (6, 7), but it may also be used to study other

91   proteome fractions. Alzheimer's Disease (AD) and Parkinson's Disease (PD) are inherently linked to

92   protein aggregation (8) and, therefore, to the formation of high molecular weight (HMW) protein

93   complexes. Fractionation of peripheral fluid samples from patients with these diseases, focused on the

94   HMW proteome, could be useful to eliminate high abundance proteins and to study protein aggregates

95   that may be present in circulation. In this study, we propose using centrifugal ultrafiltration focused on

96   the HMW proteome, not as an alternative but as a complementary analysis to the investigation of

97   unfractionated samples, to potentially assess the restructuring of molecular complexes as a distinctive

98   feature of disease. To evaluate our hypothesis, the same set of samples was subjected to the

99   proteomics analysis of the whole serum and fractionation using centrifugal ultrafiltration with 300 kDa

100   molecular weight cut-off (MWCO) filters in a non-denaturing environment (henceforth referred to as

101   HMW fractionation). The data obtained from both approaches were directly compared and merged to

102   identify potential biomarkers. This study used a cohort consisting of patients with neurodegenerative

103   disorders, specifically Alzheimer's and Parkinson's disease patients, and healthy controls (CT) to test the

104   significance of using HMW fractionation as a complementary tool for biomarker discovery. These

105   disorders were chosen as they are considered as potential proteinopathies, making them promising

106   targets for this research.

107   **Methods**

108   **Participants**

109   A total of 58 serum samples were used in this study, comprising 3 groups of individuals: AD (n = 22), PD

110   (n = 24), and CT (n = 12). The PD patients were recruited at the Movement Disorders Units of the

111   Neurological Department of the CHUC, where they were assessed by a movement disorders specialist

112   and were diagnosed according to the criteria defined by the UK Parkinsons's Disease Society Brain Bank

113   (9). The exclusion criteria for these patients consisted of severe dementia (as indicated by a Mini-Mental

114   State Exam score below 15), any psychiatric disorder, or other forms of Parkinsonism. The clinical group

115   of individuals with AD diagnosis was recruited and prospectively evaluated by two experienced

116   neurologists at Memoclínica and the Neurology Department of the CHUC. The standard criteria for the

117   diagnosis of AD were the Diagnostic and Statistical Manual of Mental Disorders—fourth edition (DSM-

118   IV-TR) (10) and the National Institute on Aging and the Alzheimer's Association Workgroup (11). To

119   ensure the homogeneity of the sample, only patients who met the following criteria were included: they

120   were in a stable condition, did not sustain recent changes in medication, and did not have

121   ophthalmological or neurological/psychiatric conditions other than AD. The CT group was composed of

122   age- and gender-matched individuals from the community with no history of cognitive deterioration,

123   neurological or acquired central nervous system (CNS) disorders, traumatic brain injury, or psychiatric

124   disorders. The CT group was also submitted to a brief cognitive assessment to exclude the presence of

125   cognitive impairment.

126   **Serum processing for proteomics analysis**

127   Two different strategies were used to obtain a more comprehensive proteomic characterization of

128   serum samples, namely: i) direct analysis of whole serum and ii) HMW fractionation through

129   ultrafiltration using 300 kDa cut-off filters (HMW serum).

130   For each sample, 5 μL were used for direct serum analysis and 82.5 μL for HMW serum fractionation.

131   Additionally, three sample pools were prepared by combining aliquots of all the samples, for the AD

132   pool, PD pool, and CT pool, respectively. Pooled samples were used for Data-dependent acquisition

133 (DDA) experiments to build a specific protein library to be used in data-independent acquisition (DIA)

134 analysis and were subjected to the same sample processing as the individual samples. Before

135 processing, all samples were spiked with the same amount of an internal standard (IS) to account for

136 sample loss (12). Different internal standards were used depending on the type of analysis: MBP-GFP

137 (12) in the case of the whole serum approach, while equine ferritin, commonly available as one of the

138 standards in the Gel Filtration Calibrants Kit for High Molecular Weight proteins (GE28-4038-42), for the

139 HMW fractionation approach.

140 For the direct analysis of whole serum, the samples were diluted in Laemmli buffer, followed by

141 denaturation for 5 min at 95°C and cysteine alkylation with acrylamide, and the total volume in all

142 samples was subjected to in-gel digestion using the Short-GeLC for subsequent quantitative analysis by

143 LC-MS/MS-DIA (13).

144 Samples subjected to HMW fractionation were ultrafiltrated using 300 kDa cut-off filters (Vivaspin® 500

145 Polyethersulfone, 300 kDa (Sartorius)) pre-conditioned to PBS. Serum samples were diluted into 200 µL

146 of PBS and subjected to 20 min centrifugation at 14,500× g at 4 °C followed by an additional washing

147 step with another 200 µL of PBS. In some cases, the washing step was repeated until the retentate

148 volume did not exceed 50 µL. The resulting retentates, the HMW fraction, were collected into a new

149 LoBind® microcentrifuge tube and precipitated with ice-cold acetone (14). The precipitated pellets were

150 resuspended into 30 µL of a solution containing 2% SDS (v/v) and 1 M of Urea, always aided by

151 sonication (VibraCell 750 watt-Sonics ®) with ice in the cup horn (2 min. pulse duration, at 1 second

152 intervals, and with 40% amplitude). Afterward, concentrated Laemmli Buffer was added to the samples,

153 followed by a 30 min incubation to reduce the samples and a 20 min incubation with iodoacetamide for

154 cysteine alkylation (15). The total volume in all samples was subjected to in-gel digestion as previously

155 specified (13).

156 **Mass spectrometry data acquisition**

157 Samples were analyzed on a NanoLC™ 425 System (Eksigent®) couple to a TripleTOF™ 6600 System

158 (Sciex®) using DDA for each fraction of the pooled samples for protein identification and SWATH-MS

159 acquisition of each sample for protein quantification. Peptides were resolved by micro-flow liquid

9

160   chromatography on a MicroLC column ChromXP™ C18CL (300 µm ID × 15 cm length, 3 µm particles, 120

161   Å pore size, Eksigent®) at 5 µL/min. The liquid chromatography program was performed as follows with

162   a multistep gradient: 2 % to 5 % mobile phase B (0-2 min), 5 % to 28 % B (2-50 min), 28% to 35% B (50-

163   51 min), 35 to 98% of B (50–52 min), 98% of B (52-61 min), 98 to 2% of B (61–62 min), 2% of B (68 min).

164   Mobile phase A was composed of 0.1 % formic acid (FA) with 5% dimethyl sulfoxide (DMSO), and mobile

165   phase B was composed of 0.1 % FA and 5% DMSO in acetonitrile. Peptides were eluted into the mass

166   spectrometer using an electrospray ionization source (DuoSpray™ Source, ABSciex®) with a 25 µm

167   internal diameter hybrid PEEKsil/stainless steel emitter (ABSciex®). The ionization source was operated

168   in the positive mode set to an ion spray voltage of 5 500 V, 25 psi for nebulizer gas 1 (GS1) and 25 psi for

169   the curtain gas (CUR).

170   For DDA experiments, the mass spectrometer was set to scan full spectra (m/z 350-1250) for 250 ms,

171   followed by up to 100 MS/MS scans (m/z 100–1500) per cycle to maintain a cycle time of 3.309 s. The

172   accumulation time of each MS/MS scan was adjusted in accordance with the precursor intensity

173   (minimum of 30 ms for precursor above the intensity threshold of 1000). Candidate ions with a charge

174   state between +2 and +5 and counts above a minimum threshold of 10 counts per second were isolated

175   for fragmentation and one MS/MS spectrum was collected before adding those ions to the exclusion list

176   for 25 seconds (mass spectrometer was operated by Analyst® TF 1.7, ABSciex®). The rolling collision

177   energy (CE) was used with a collision energy spread (CES) of 5.

178   For SWATH-MS-based experiments, the mass spectrometer was operated in a looped product ion mode

179   (16) and the same chromatographic conditions were used as in the DDA experiments described above. A

180   set of 60 windows of variable width (containing 1 m/z for the window overlap) was constructed,

181   covering the precursor mass range of m/z 350-1250. A 250 ms survey scan (m/z 350-1500 m/z) was

182   acquired at the beginning of each cycle for instrument calibration and SWATH-MS/MS spectra were

183   collected from the precursors ranging from m/z 350 to 1250 for m/z 100–1500 for 20 ms resulting in a

184   cycle time of 3.304 s. The CE for each window was determined according to the calculation for a charge

185   +2 ion centered upon the window with variable CES according to the window.

186   **Mass spectrometry data processing**

187 A specific library of precursor masses and fragment ions was created by combining all files from the DDA

188 experiments and used for subsequent SWATH processing. Libraries were obtained using ProteinPilot™

189 software (v5.1, ABSciex®), using the following parameters: i) search against a database from SwissProt

190 composed by Homo Sapiens (released in March 2019), and MBP-GFP (15) and horse ferritin light and

191 heavy chains sequences ii) acrylamide or iodoacetamide alkylated cysteines, for whole serum or HMW

192 respectively, as fixed modification; iii) trypsin as digestion enzyme and iv) urea denaturation as a special

193 factor in the case of the HMW samples. An independent False Discovery Rate (FDR) analysis using the

194 target-decoy approach provided with Protein Pilot software was used to assess the quality of the

195 identifications, and positive identifications were considered when identified proteins and peptides

196 reached a 5% local FDR (17, 18).

197 Data processing was performed using the SWATH™ processing plug-in for PeakView™ (v2.0.01, AB

198 Sciex®) (19). After adjustment of retention time using a combination of IS and endogenous peptides, up

199 to 15 peptides, with up to 5 fragments each, were chosen per protein, and quantitation was attempted

200 for all proteins in the library file identified from ProteinPilot™ searches.

201 Protein levels were estimated based on peptides that met the 1% FDR threshold with at least 3

202 transitions in at least six samples in a group, and the peak areas of the target fragment ions of those

203 peptides were extracted across the experiment using an extracted-ion chromatogram (XIC) window of 5

204 minutes with 100 ppm XIC width. Protein levels were estimated by summing all the transitions from all

205 the peptides for a given protein (an adaptation of (20) and further normalized to the levels of the IS

206 (15)).

207 **Statistical analysis and biological interpretation**

208 Pearson's Chi-squared Test for Count Data was performed in R version 4.2.1, using the chisq.test

209 function available in the native stats package in R to determine if there were significant differences in

210 the gender proportion of the groups within the studied cohort.

211 To assess the variation of the serum proteins (either Whole serum or the HMW fraction of the serum)

212 among the three groups, a Kruskal–Wallis H test was followed by the Dunn's Test for pairwise

11

213   comparison. Dunn's p-values were corrected using the Benjamini-Hochberg FDR adjustment, and

214   statistical significance was considered for p-values below 0.05.

215   Stepwise Linear Discriminant Analysis (LDA) was performed to select the proteins responsible for the

216   best separation of the groups being studied. LDA was performed using IBM® SPSS® Statistics Version 22

217   (Trial). LDA was attempted considering the proteins only altered at the Whole Serum or HMW Serum,

218   and for the combination of both results. The evaluation of the models obtained from each analysis was

219   performed by comparison of the Receiver operating characteristic (ROC) curves obtained using each

220   model. A comparison of the ROC curves was performed using MedCalc Statistical Software version

221   20.106 (MedCalc Software Ltd;(21); Trial). The Delong et al. (1988) (22) method was used for the

222   calculation of the Standard Error (SE) of the Area Under the Curve (AUC) and of the difference between

223   two AUCs, and the Confidence Interval (CI) for the AUCs were calculated using the exact Binomial

224   Confidence Intervals which are calculated as the following AUC ± 1.96 SE.

225   Violin plots were used to present the distribution of the individual protein levels among each condition,

226   and Pearson's correlation analysis was performed to evaluate the similarity between the profiles of the

227   proteins highlighted in the study. Violin plots were generated using GraphPad Prism 8.0.1 (Trial) and the

228   Pearson's correlation was performed using Morpheus software (23). Heatmap and hierarchical

229   clustering analyses were computed using PermutMatrix version 1.9.3 (24, 25) using the Euclidean

230   distance and McQuitty's criteria.

231   Physical protein-protein interactions between the highlighted analytes were predicted by GeneMANIA

232   webserver (Gene Function Prediction using a Multiple Association Network Integration Algorithm; (26,

233   27) together with a gene ontology (GO) analysis of the formed network. In addition to the proteins

234   imported from this study, 28 additional related genes were allowed to create the interaction network

235   using equal weighting by network. An additional GO enrichment analysis considering the term

236   "biological process" was also performed. On the other hand, functional protein association networks

237   were evaluated using the Search Tool for Retrieval of Interacting Genes/Proteins (STRING) version 11.5

238   (28) with a medium confidence of 0.4 (29).

239   Pathway enrichment analyses were performed using the FunRich software (version 3.1.3) (30),

240   considering two different databases: the FunRich or the Reactome database. In both instances, a

241   statistical analysis employing a hypergeometric test was conducted, using the FunRich human genome

242   database as the background. Enriched pathways were considered for a non-corrected or a Bonferroni-

243   corrected p-value below 0.05 for Funrich or Reactome database, respectively.

244   **Results**

245   **High molecular weight fractionation as a reproducible method for biomarker discovery**

246   To explore the potential of HMW fractionation in peripheral biomarker research, serum samples

247   underwent either ultrafiltration using a 300 kDa MWCO filter followed by protein precipitation of the

248   retentate (referred to as HMW serum), before protein digestion and MS analysis, or were directly

249   analyzed (whole serum) (see Figure 1). In the proposed pipeline, the HMW fractionation was performed

250   under non-denaturing conditions, such that the filter would retain native HMW protein complexes and

251   large molecular structures.

252

253   **Figure 1:** Pipeline of sample preparation, data acquisition and data analysis.

254   As previously reported, employing an appropriate IS is crucial for delineating effective proteome

255   changes among different groups (12). Given the nature of the fractionation proposed in this study, an

256   ideal IS for this procedure would be a HMW protein that remains retained by the filter and bears no

257   similarity to other human proteins. In this context, equine globular protein Ferritin (~440 kDa size)

258   emerged as a suitable IS for several reasons: it lacks similarity with any other protein from the human

259   proteome (Additional File 1: Supplementary Figure 1a), the peptides monitored in the SWATH-MS

260   analysis are easily distinguishable from the matrix (the human serum proteome; Additional File 1:

261   Supplementary Figure 1b), and it exhibits a coefficient of variation similar to the one obtained by the

262   MBP-GFP used in the unfractionated analysis and previously characterized (12) (Additional File 1:

263   Supplementary Figure 1c). Additionally, the overall reproducibility of the fractionation was also

264   inspected. Similarly, to what was observed for the IS (Additional File 1: Supplementary Figure 1c), the

265   overall coefficient of variation of the proteins quantified using technical replicates (Additional File 1:

266   Supplementary Figure 1d) revealed that this procedure did not induce an appreciable increase in the

267   variability of the quantification when compared with the conventional protocol (unfractionated

268   samples). Moreover, the variation caused by the sample processing steps may be reverted by the

269   normalization of the values to the IS since the coefficient of variation of the IS is similar to the one

270   observed for the majority of the proteins, indicating that the selected IS is a good predictor of the

271   alterations induced by the method.

272   **High molecular weight fractionation increases the detection of proteomic alterations in NDs**

273   As proof of concept, this procedure was applied to serum samples of a cohort comprised of 3 different

274   groups: AD (n = 22), PD (n = 24), and CT (n = 12). No statistically significant differences were found in the

275   gender distribution between the three groups; however, differences were observed concerning age

276   distribution, with PD individuals being slightly younger, on average, than both other groups (Table 1).

277   Patients with neurodegenerative diseases, both AD and PD, were selected for this study. These

278   conditions are frequently associated with the development of abnormal protein complexes and protein

279   aggregation. Hence, they offer a fitting context to evaluate the fractionation approach's effectiveness in

280   exploring potentially altered protein interactions—an aspect frequently neglected in conventional

281   proteomics analyses.

282 **Table 1.** Study population distribution across age and sex.

	AD (n = 22)	PD (n = 24)	CT (n = 12)	p-value				
				CT vs. Diseases	CT vs. AD vs. PD	AD vs. PD	CT vs. PD	CT vs. AD
Male (n)	9	11	3	0.197[*]	-	0.643[*]	0.148[*]	0.262[*]
Female (n)	13	13	9					
Age (mean ± SD)	68.4 ± 8.2	60.3 ± 10.4	67.5 ± 7.2	-	0.014[$]	0.009[#]	0.033[#]	0.396[#]

283　[*] Determined by χ^2 test according to each group's male/female proportions. [$] Determined by Kruskal-Wallis rank

284　sum test to compare age between all groups. [#] Determined by Dunn's test (Benjamini-Hochberg correction) to

285　compare age between group pairs. Diseases refer to AD and PD grouped together.

286　A total of 203 and 186 proteins were quantified in whole serum and HMW serum, respectively (Figure

287　2a, solid lines; Additional File 2: Supplementary Tables 1 and 2 for detailed information). A large overlap

288　was observed between both sample preparation procedures (168 proteins were shared, which

289　corresponds to more than 70% of all the quantified proteins). Additionally, quantification in the whole

290　serum of the proteins identified in the HMW serum library did not lead to a discernible increase in

291　proteome coverage (Figure 2a, dashed line; Additional File 2: Supplementary Table 3). Altogether, these

292　results indicate that there are no major differences in terms of the proteins being quantified in the two

293　approaches, revealing that the main aim of the HMW fractionation presented in this work –

294　fractionation under non-denaturing condition – is not the overall improvement of the proteome

295　coverage but the possibility of interrogating the samples considering protein

296　interactions/macromolecular organization.

297

298　**Figure 2 – Comparative differential proteomic analysis of whole and HMW serum of AD and PD**

299　**patients. (a)** Venn diagram comparing the number of quantified proteins in each sample type using

300　sample-specific libraries (solid lines; Additional File 2: Supplementary Tables 1-2). In addition,

301   quantification in whole serum was also performed for the proteins identified in the HMW-specific library

302   (dashed line; Serum&HMWLib condition; Additional File 2: Supplementary Table 3) to evaluate the

303   possible use of HMW fractionation as a tool to improve the proteome coverage of serum samples. A

304   total of 224 proteins were quantified in the serum samples using the three strategies referred above;

305   from those, 162 proteins were commonly quantified independently of the strategy, corresponding to

306   nearly three-quarters of all the quantified proteins (72.3%). Only three new proteins were quantified in

307   whole serum using the HMW-specific library. **(b)** Venn diagram comparing the total number of proteins

308   considered altered among the three experimental groups using the two different serum-processing

309   strategies used in this work (Additional File 1: Supplementary Figure 2 and Additional File 2:

310   Supplementary Tables 1-2). A total of 69 proteins were considered altered among the three

311   experimental groups; of those, only 11 proteins (the respective gene names are indicated) were

312   consistently considered altered independently of the strategy used. **(c)** Comparison of the levels of the

313   11 proteins commonly considered altered in both whole serum and HMW-fractionated serum. The

314   proteins were arranged considering the group comparison where the statistical differences were

315   observed (Additional File 1: Supplementary Figure 3), and the alterations were presented as the median

316   fold change observed in each sample type. Proteins considered altered in more than one group are

317   indicated in italic and with a grey shadow. Only one protein, the beta-2-glycoprotein 1 (indicated in

318   bold), presented a divergent tendency when considering its values in the whole serum versus after the

319   HMW fractionation of the samples. # - non-statistically significant difference.

320   This was further confirmed by the fact that only 14.7% of all the proteins considered altered in this study

321   (11 out of 69 proteins) were consistently altered in the fractionated and unfractionated samples (Figure

322   2b; Additional File 2: Supplementary Tables 1-2). Moreover, taking into consideration the changes of

323   those 11 commonly altered proteins, it is possible to observe that all proteins, except for the protein

324   beta-2-glycoprotein 1 (Figure 2c, bold), presented the same tendency in both fractionated and

325   unfractionated samples (Figure 2c). These results, in combination with the previous observations,

326   demonstrate that, although the two approaches quantified the same set of proteins, each method

327   interrogated the samples in a particular context, resulting in the identification of a different set of

328   altered proteins and the possibility to identify different regulatory mechanisms: while the direct analysis

17

329   of the serum mainly represents the alteration at the protein level, the HMW fractionation under non-

330   denaturing conditions may allow the evaluation of the physical interaction of the proteins.

331   These results are further supported by the analysis of the MW distribution of the proteins being studied,

332   which reveals a similar profile between both approaches (Additional File 1: Supplementary Figure 4a, 4b,

333   and 4c), with most of the proteins detected in HMW serum presenting a MW below 150 kDa. Moreover,

334   a similar distribution is observed for the proteins altered exclusively in the HMW approach (Additional

335   File 1: Supplementary Figure 4d), indicating that this strategy is not biased towards only the HMW

336   proteins and supporting the idea that those proteins altered in this fraction may correspond to proteins

337   being organized into different complexes.

338   **High molecular weight fractionation revealed differences in the macromolecular organization of the**

339   **serum proteome**

340   To test the hypothesis that the HMW approach is capable of evaluating the re-organization of protein

341   complexes, the nature of the 28 proteins altered exclusively in HMW serum was evaluated. This analysis

342   revealed that most of those proteins had been described to interact physically either with each other or

343   with other proteins (Figure 3a). Upon immediate observation of the GeneMania network, which

344   primarily evaluates the interactions between the proteins under study (Additional File 2: Supplementary

345   Table 4), it becomes apparent that the majority of the proteins participate in established physical

346   interactions. Moreover, it can be observed that these proteins form a large and interconnected network

347   comprising 20 of the 28 altered proteins centered around the interactions between the apolipoproteins

348   and lecithin-cholesterol acyltransferase (encoded by the *LCAT* gene). Although not participating in any

349   known interactions with another protein from the 28 altered proteins, the proteins Glutathione

350   peroxidase 3 (*GPX3*), Pigment epithelium-derived factor (*SERPINF1*), Thyroxine-binding globulin

351   (*SERPINA7*), and Alpha-1B-glycoprotein (*A1BG*) are also known to establish physical interactions

352   including with the proteins identified in this analysis. Only four of the 28 altered proteins, namely SRR1-

353   like protein (*SRRD*), carnosine dipeptidase 1 (*CNDP1*), peptidoglycan recognition protein 2 (*PGLYRP2*),

354   and serum amyloid A-4 protein (*SAA4*), were found to have no disclosed interactions in this particular

355   analysis. However, interactors for those proteins have already been pointed out in some screening

356   assays, as confirmed in BioGRID (Biological General Repository for Interaction Datasets, Additional File 2:

357   Supplementary Table 5). Furthermore, as revealed by the functional analysis, several of these 28

358   proteins are involved in the formation of complexes with lipids and platelet components (Additional File

359   2: Supplementary Table 6), indicating that those proteins can form complexes not only via the

360   interaction with other proteins but also with other molecules, and thus be organized in large complexes.

361   This involvement in the potential formation of macromolecular complexes is even more evident when

362   the functional pathways enriched in each of the two lists of proteins (30 and 28 altered proteins

363   exclusively in the whole serum or HMW serum, respectively) are directly compared (Figure 3b). This

364   comparison highlights the fact that all the pathways that are either only, or at least more, enriched in

365   the HMW dataset in comparison to the whole serum dataset (pathways indicated in bold) are related to

366   the formation/regulation of large complexes/macrostructures, namely amyloids, fibrin clot formation

367   and dissolution pathways and lipoprotein-related metabolism. On the other hand, the pathways that are

368   particularly enriched or unique in the whole serum dataset are mainly related to transcription factor

369   networks (Additional File 2: Supplementary Table 7).

370

371   **Figure 3 – Characterization of the proteins exclusively altered in whole serum or HMW serum. (a)**

372   GeneMania Network of the 28 proteins altered only at the HMW serum sample (listed in larger circles).

373   The analysis was performed with network weighting equal by network, allowing a maximum of 28 extra

374   resultant genes (non-listed small circles). The seven most enriched GeneMania Functions were

375   highlighted in the network (color code; complete results in Additional File 2: Supplementary Table 4).

376   Only protein-protein physical interactions (red edges) were considered in this analysis, demonstrating

377   that most of these 28 altered proteins have known interactors and can be involved in the formation of

378   large protein-protein complexes. **(b)** FunRich Biological Pathways enriched in the whole serum (30

379   proteins; Additional File 2: Supplementary Table 5) and HMW serum (28 proteins; Additional File 2:

380   Supplementary Table 6) proteomes. All GO analyses considered a p <0.05. Pathways uniquely enriched

381   at the HMW serum or particularly enriched in this type of sample when compared to the whole serum

382   are indicated in bold.

383   **Evaluation of the potential of this combined strategy for biomarker discovery**

384   The previous set of results demonstrates that both approaches can provide complementary information.

385   In line with this evidence, the potential to use this combined strategy for biomarker discovery was also

386   evaluated. In general, both approaches result in nearly 40 proteins altered in at least one pair of

387   comparisons (Figure 4a-b; Additional File 1: Supplementary Figure 3 for details regarding each pair of

388   comparisons), with a tendency to have more proteins altered in the comparisons involving the AD group

389   (at least 20 proteins were altered compared to CT against a maximum of 14 altered proteins in PD vs.

390   CT, Additional File 1: Supplementary Figure 2a-b) and only a small subset of proteins altered in only one

391   comparison (15/41 in the whole serum and 11/39 in HMW serum, Figure 4a and 4b, respectively).

392   Besides those similarities, different profiles of altered proteins were observed depending on the

393   approach used. Hence, it can also be highlighted that while the whole serum strategy (Figure 4a) mainly

394   found proteins altered between the two disease groups (39 proteins in AD vs. PD compared to 20 and 12

395   proteins in AD vs. CT and PD vs. CT, respectively), the HMW approach (Figure 4b) captured more

396   differences between AD and CT samples (36 out of the 39 altered proteins). Due to this

397   complementarity, the combination of results from both approaches resulted in a more comprehensive

398   profile (Additional File 1: Supplementary Figure 2a-c), with a general increase in the number of altered

399   proteins per group (total of 69 altered proteins, Figure 4c). This improvement is particularly evident in

400   the case of proteins altered between PD and CT samples, for which only one protein was considered

401   commonly altered in the two approaches (Figure 2c and Additional File 1: Supplementary Figure 2b),

402   thus resulting in the duplication of the list of proteins with the potential to serve as biomarkers for PD

403   versus CT individuals. Common to all mapped profiles was the low number of proteins altered between

404   all three groups (4, 1, and 6 for Whole serum, HMW serum and the combination of both, Figure 4a-c,

405   respectively) and the absence of proteins altered exclusively between PD and CT samples.

406

407   **Figure 4 – Identification of potential circulating biomarkers of AD and PD. (a-c)** Venn diagrams

408   representing the distribution of the altered proteins among the different comparisons (AD vs. CT; PD vs.

409   CT; AD vs. PD) considering the whole serum, HMW serum and the combination of the two (Whole +

20

410   HMW serum) types of samples, respectively. **(d-f)** LDA using all altered proteins per sample type or the

411   combination of the two. The number of proteins used in each model is indicated above the graphic, and

412   for each model it is indicated their specificity (percentage of healthy individuals correctly identified;

413   indicated in black) and the sensitivity per disease condition (percentage of AD or PD patients correctly

414   identified; indicated in red and blue, respectively). Specificity and sensitivity values were also

415   summarized in Additional File 2: Supplementary Table 7, and the generated LDA discriminant functions

416   and their respective statistical confidence were summarized in Table 2. **(g-h)** Comparative ROC curves of

417   the discriminant functions generated using all altered proteins per sample type or a combination of the

418   two. An independent evaluation was performed for each disease group being studied. The AUC of the

419   ROC curves, their 95% CI and the pairwise comparisons are summarized in **(i)**. CI was calculated as

420   follows AUC ± 1.96 SE. n.s., non-significant alterations. * and # indicate a $p < 0.05$ for statistically

421   significant differences in comparison to the Whole Serum or HMW+Whole Serum in comparison to

422   HMW Serum, respectively, using the method of Delong et al. (1988) (22).

423   To further confirm the biomarker potential of the altered proteins from the three strategies presented

424   above, they were used as input to build discriminant models that could differentiate between the

425   studied groups (Figure 4d-f). From each dataset, candidates whose combination resulted in the best

426   possible discriminant model were automatically selected (detailed information regarding the methods in

427   Table 2), resulting in three distinct and statistically valid models (all with $p < 0.0032$) capable of

428   discriminating the three groups being studied. The whole serum dataset resulted in a reasonable model

429   composed of 4 proteins (Figure 4d), with a median sensitivity (the capacity to classify the individuals

430   from each disease group correctly) of 86.95% (sensitivity and specificity are summarized in Additional

431   File 1: Supplementary Table 8). On the other hand, the model created with six proteins from the HMW

432   approach (Figure 4e) had lower performance, with a median sensitivity of only 80.3% (corresponding to

433   83.3% predicting capacity for PD samples and 77.3% for AD samples). Moreover, neither of the models

434   was particularly good in the classification of CT samples, resulting in a specificity of only 66.7% and 75%

435   in the whole serum and HMW serum models, respectively. Remarkably, the combined model (created

436   from the dataset containing the altered proteins from both approaches - Figure 4f) clearly outperformed

437   the two models based only on proteins from a single approach. For this combined model, 10 proteins

21

438 were selected and integrated, creating a discriminant model capable of correctly classifying more than

439 96% of all tested samples (93% in a cross-validation test; Table 2). This included the correct classification

440 of 100% of samples from the PD group (Figure 4f), which encompassed samples from younger patients

441 compared to the other studied groups (Table 1). However, it was confirmed that this age difference did

442 not interfere with the classification of samples when employing the discriminant model since: i) age by

443 itself did not yield good classification performance (Additional File 1: Supplementary Figure 5) and ii)

444 none of the 10 proteins included in the combined model were strongly correlated with age (Additional

445 File 1: Supplementary Figure 6). Altogether, this combined method presented a median sensitivity of

446 97.75% and a specificity of 91.7%.

447 The diagnostic capacity of these models was further assessed by ROC curves, which measured their

448 ability to accurately classify AD and PD patients from among all other samples (Figure 4g-i). This analysis

449 confirmed that the best model is the one created with the combination of proteins from both

450 approaches and that, in general, the model using only proteins from the whole serum approach is better

451 than the model from the HMW approach. The respective statistics (Figure 4i) further support that the

452 whole serum model performed better than the HMW serum model but without statistically significant

453 differences between the two ROC curves. Additionally, the statistical analysis also confirmed that the

454 combined model (AUC = 0.999 for the classification of AD and PD patients) is the best model and that it

455 performed significantly better (p < 0.05) than both other models for PD classification (HMW serum, AUC

456 = 0.888; whole serum, AUC = 0.960) and better than the HMW serum model (AUC = 0.919) in the case of

457 AD classification. The robustness of the combined model is further evidenced by the confidence interval

458 (CI) of the AUC, which has a lower limit above 0.93 for both diseases, in contrast with the values

459 achieved for the other two methods, whose lower limits are all below 0.9.

Table 2. Linear Discriminant models, respective statistical analysis and classification results.

	Gene Name	Protein name	MW (KDa)	Whole Serum		HMW serum		Whole + HMW serum	
				Dis 1	Dis 2	Dis 1	Dis 2	Dis 1	Dis 2
#	**FOXM1**	**Forkhead box protein M1**	**84.283**	---	---	---	---	**245.36**	**-10.93**
#	PROC	Vitamin K-dependent protein C	52.071	-1133.77	1212.34	---	---	-1557.39	623.28
#	HBB	Hemoglobin subunit beta (Beta-globin)	15.998	1.49	0.58	-0.06	-0.13	0.42	-1.50
#	APOA1	Apolipoprotein A1 (Apo-AI)	30.778	3.63	0.53	---	---	-0.16	-3.08
#	**IGLV3-19**	**Immunoglobulin lambda variable 3-19**	**12.042**	---	---	---	---	**62.89**	**52.70**
§	SRRD	SRR1-like protein	38.573	---	---	20.37	7.55	16.69	11.52
§	APOC1	Apolipoprotein C1 (Apo-CI)	9.332	---	---	-9.74	19.21	-15.77	1.33
§	APOE	Apolipoprotein E (Apo-E)	36.154	---	---	1.46	-0.42	0.84	0.53
§	**SERPINF1**	**Pigment epithelium-derived factor (PEDF)**	**46.312**	---	---	---	---	**6.85**	**2.63**
§	KRT9	Keratin, type I cytoskeletal 9	62.064	---	---	-8.80	5.05	-12.73	-4.91
	PPBP	**Platelet basic protein (PBP)**	**13.894**	---	---	**8.89**	**-7.88**	---	---
	TF	**Serotransferrin (Transferrin)**	**77.050**	**-0.99**	**-1.66**	---	---	---	---
		(Constant)		2.11	-0.81	-1.11	-2.82	3.34	-1.10

Statistics (Wilks' Lambda, Chi-square and p-value)	λ=	0.253	0.773	0.242	0.629	0.058	0.292
	χ2=	73.473	13.473	74.458	24.361	143.675	62.244
	p<	9.97×10^{-13}	3.23×10^{-03}	4.65×10^{-11}	1.85×10^{-04}	1.01×10^{-20}	4.94×10^{-10}

Overall classification results (% of cases correctly classified)	Original	82.8	79.3	96.6
	Cross-validation	81.0	72.4	93.1

selected from the data from the Whole serum approach for the Whole + HMW serum model; § selected from the data from the HMW serum approach for the Whole + HMW serum model;

Dis – discriminant function; Note that proteins are sorted by order of inclusion into the Whole + HMW serum discriminant model. Proteins exclusive to one of the models appear in bold.

463   By looking at the proteins selected to build the different methods (Table 2), it was observed that the

464   combined method is not the simple combination of the proteins previously selected from each of the

465   individual methods. The combined model is built by the combination of 10 proteins, five from each

466   dataset, including three [forkhead box protein M1 (*FOXM1*), immunoglobulin lambda variable 3-19

467   (*IGLV3-19*) and pigment epithelium-derived factor (*SERPINF1*)] that were not selected on the database-

468   specific models. On the other hand, some previously selected proteins [serotransferrin (*TF*) and platelet

469   basic protein (*PPBP*)] were not included in the combined model. Finally, the Hemoglobin subunit beta

470   (*HBB*) was selected in both approach-specific models, although only data from the whole serum dataset

471   was used in the combined model. These results demonstrate that the increase in the initial amount of

472   data provided for the discriminant analysis has an important impact on the generated model by making

473   it possible to test different combinations of proteins and, thus, allowing for the identification of better

474   combinations than those highlighted in the analysis of individual datasets. Interestingly, all ten proteins

475   selected in the combined model have a MW below 90 kDa (Table 2). This finding confirms that all the

476   proteins selected from the HMW approach have a MW below the theoretical cut-off of the filters used

477   for fractionation. This observation supports the hypothesis that this approach might effectively assess

478   the remodeling of molecular complexes. By plotting the individual values of each of the ten proteins

479   selected in the combined model (Figure 5a), it is possible to observe that, as expected, those values

480   present some variation characteristic of the individuality of each patient. Nevertheless, considering that

481   the model was able to correctly classify more than 90% of all the patients (Figure 4f), it is possible to

482   infer that the combinations performed in the model could diminish the impact of the biological

483   variability, proving that the combination of different markers can overcome their individual weaknesses.

484   The analysis of these results immediately reveals that i) only three proteins from the model (the

485   proteins encoded by the genes *FOXM1, HBB,* and *SRRD*) are significantly altered between all three

486   groups and ii) only one protein, the apolipoprotein C1 (Apo-CI, encoded by the *APOC1* gene), is altered

487   between a single comparison, in this case between AD and PD which may indicate that this protein may

488   have a particularly important role in this model for distinguishing AD from PD patients. Among the

489   remaining 6 proteins: i) three, apolipoprotein E, pigment epithelium-derived factor and KRT9, are

490   altered between both disease groups and control sample (all three found in the HMW fraction); ii) two,

491   apolipoprotein A1 and immunoglobulin lambda variable 3-19, are altered in AD in comparison to both

24

492   PD and CT, and one, the Vitamin K-dependent protein C (encoded by the PROC gene), is altered in PD

493   patients in comparison to the other two groups. Moreover, it is noteworthy that while there was a

494   tendency to incorporate proteins that were increased in AD compared to CT from the whole serum

495   approach (three out of the five proteins from the whole serum model, encoded by the genes *FOXM1*,

496   *HBB*, and *APOA1*), the opposite trend was observed in the case of proteins from the HMW approach.

497   Specifically, three out of the five proteins (the proteins encoded by the genes *SRRD*, *APOE*, and

498   *SERPINF1*) were found to be decreased in the AD vs. CT comparison. On the contrary, for the PD vs. CT

499   comparison, there were no major differences in terms of tendencies when considering the proteins

500   captured in the whole serum or the HMW serum.

501   Finally, all proteins from the whole serum dataset, in addition to SRR1-like protein (encoded by the

502   *SRRD* gene) and Apo-CI from the HMW dataset, were significantly altered between both disease groups.

503   From these, three proteins [the protein Forkhead box protein M1 (encoded by the *FOXM1* gene),

504   hemoglobin subunit beta (encoded by *HBB* gene), and apolipoprotein A1 (Apo-AI, encoded by the

505   *APOA1* gene)] are less abundant in PD samples than in AD samples, while the remaining four are

506   increased.

507

508   **Figure 5 – Characterization of the protein panel selected by the combined LDA model. (a)** Violin plots

509   representing the protein level distribution of the ten proteins from the model. The dashed lines indicate

510   the first, second (median) and third quartiles. *, **, and *** indicate a p<0.05, p<0.01, and p<0.001

511   for statistically significant differences. * denotes the comparison to control and # the differences

512   between disease groups. Statistical analysis was performed using the Kruskal–Wallis H test followed by

513   the Dunn's Test for pairwise comparison. **(b)** Pearson's correlation analysis between the overall

514   regulation profile of the proteins included in the model. **(c)** Heatmap and hierarchical clustering analysis

515   of the proteins from the model. Clustering was performed for both proteins and individuals. Three

516   different clusters (Cluster PD, CT and AD) containing the large majority of the individuals from a given

517   group can be highlighted from the analysis. The average profile of each cluster is indicated on the right

518   and can be considered as the profile of expression of those proteins within the groups considered in this

25

519 study. **(d)** Interaction network of these proteins carried out with STRING with a medium confidence (0.4)

520 score. The color of the edges indicates the type of evidence that supports the interaction, while the

521 color of the nodes represents the categorization of the proteins considering UniProt Keywords. The

522 calculated PPI enrichment p-value is 2.07e-05. Three clusters (Cluster 1 to 3) can be identified within the

523 network, with the dashed edges indicating the separation between clusters. Cluster 1 corresponds to

524 proteins whose interactions are experimentally confirmed, cluster 2 is composed of theoretically related

525 proteins, and cluster 3 corresponds to non-related proteins. **(e)** Reactome pathways enrichment analysis

526 using the proteins from the model. The red line indicates Bonferroni's corrected p-value with the

527 corrected $p < 0.05$, meaning a significant enrichment. The grey dashed line indicates the reference line

528 ($p = 0.05$).

529 The correlation analysis of the protein abundances among groups confirmed that, in general, there is no

530 particularly evident correlation between the profiles and the magnitude of regulation of these proteins

531 (Figure 5b). There were, however, some observed exceptions, including a strong positive correlation

532 between the proteins encoded by the genes *APOA1* and *FOXM1* (r = 0.8) and, to a lesser extent, the

533 proteins encoded by *APOC1, APOE and SERPINF1* (r = 0.5-0.6). Interestingly, the proteins showing a

534 positive correlation originated from the same approach, the gene products of *APOA1* and *FOXM1* were

535 both highlighted in the whole serum approach, suggesting that the overall levels of these two proteins

536 were similarly altered. Meanwhile, the products of the *APOC1, APOE,* and *SERPINF1* genes were

537 identified as altered in the HMW fractionation strategy, which may indicate that these proteins might be

538 part of the same complex and thus regulated in a similar manner. No remarkable negative correlations

539 were found, with the strongest correlation observed between the proteins encoded by the *APOA1* and

540 *SRRD* genes, which indicates that none of the proteins in the model present a completely opposite

541 regulation profile. Additionally, an unsupervised clustering analysis using these ten proteins (Figure 5c)

542 confirmed their capacity to partially distinguish the three groups being studied, revealing that, besides

543 the existence of individual variability, it was possible to identify three independent clusters composed

544 exclusively or mainly of samples from one of the three groups. This analysis also demonstrates that this

545 set of proteins is particularly efficient in isolating the AD patients from the remaining individuals from

546 the study: the AD cluster was composed exclusively of AD patients, and only 6 out of the 22 AD patients

547   were not included in this cluster. On the other hand, a slightly lower separation capacity was observed

548   for both PD and CT samples. These two clusters contained few samples that did not belong to their

549   respective groups, resulting in a higher percentage of individuals not properly grouped (10 out of 24 and

550   4 out of 12 samples for PD and CT, respectively). The discrepancies observed between the clustering

551   analysis and the discriminant model results, where the latter correctly classified over 90% of the

552   samples, can be attributed to the fact that the clustering analysis relied solely on the individual protein

553   distribution profiles across the samples. In contrast, the discriminant analysis employed equations with

554   different weightings for each protein, resulting in a single model that effectively reduces the intragroup

555   variability while promoting a better separation between the analyzed groups. Despite that, the

556   clustering analysis remains an important approach for understanding how the proteins are modulated

557   within the samples. Thus, from the three different clusters highlighted in the analysis, it was possible to

558   infer the median protein abundance profile of these proteins among the three groups. For instance, the

559   gene products of *SERPINF1, APOE* and *APOC1* tend to be less abundant in both disease groups compared

560   to CT samples. Furthermore, some proteins are more abundant in each disease group, namely the gene

561   products of *PROC* and *KRT9* in PD samples and the gene products of *HBB, APOA1* and *FOXM1* in AD

562   samples. Another disease-specific observation was the smaller amount of immunoglobulin lambda

563   variable 3-19 (*IGLV3-19*) in AD samples compared to both other groups. Overall, these tendencies

564   characterize the unique profiles determined for each disease group, which may be a precursor to a

565   potential future biomarker panel that could be more informative than the analysis based on any single

566   protein.

567   Finally, STRING analysis (Figure 5d and Additional File 2: Supplementary Table 9) revealed that these ten

568   proteins have more interactions among themselves than what would be expected for a random set of

569   proteins of the same size and degree of distribution, indicating that this set of proteins is, at least

570   partially, biologically connected (PPI enrichment $p = 2.07 \times 10^{-5}$). This result may be mainly due to the

571   strong network involving apolipoproteins and Hemoglobin subunit beta (cluster 1). Again, two out of the

572   ten proteins selected for the discriminant method revealed to be associated with high-density

573   lipoproteins (HDL) and chylomicron (ultra-low-density lipoproteins particles) remodeling and assembly

574   (Figure 5e and Additional File 2: Supplementary Table 10), highlighting the potential importance of these

575   mechanisms in the neurodegenerative process, and confirming that the proteins related with these

576   mechanisms could be good biomarker candidates for their diagnosis.

577   Given the central role that apolipoproteins appear to play in this model, a discriminant analysis was

578   performed using data from ten altered proteins involved in apolipoprotein-related mechanisms to

579   investigate if the diagnostic model could be limited to this set of functionally related proteins (Additional

580   File 1: Supplementary Figure 7). However, the generated model exhibited lower diagnostic capacity

581   compared to the combined approach, with only 82.75% of the samples being correctly classified. The

582   model demonstrated a specificity of 66.7%, and a sensitivity of 86.4% and 87.5% for AD and PD,

583   respectively, yielding ROC curves with AUCs equal to or below 0.955. Thus, besides the importance of

584   apolipoproteins, the results from these proteins alone are not enough to distinguish the three groups,

585   emphasizing the importance of having diagnostic models based on several complementary candidates

586   instead of a single or just a few candidates. Nonetheless, the identification of this robust core of

587   functionally related proteins underscores the significance of the combined approach for identifying new

588   potential biomarkers. For instance, while the dysregulation of Apo-C1 and Apolipoprotein E (Apo-E,

589   encoded by the *APOE* gene) was discovered using the HMW fractionation approach, the dysregulation of

590   Apo-AI was identified using the whole serum approach.

591   **Discussion**

592   The present study presents a proof of concept of a novel two-pronged approach to biomarker discovery

593   in complex peripheral biological fluids. More specifically, it was demonstrated that through the

594   combination of two complementary proteomics strategies, the direct analysis of whole serum and the

595   analysis of serum HMW fraction (above 300 kDa) in non-denaturing conditions, another level of

596   proteome characterization of the samples could be achieved. This resulted in more robust diagnostic

597   models and insights into disease mechanisms. In this sense, when applied to serum samples from a

598   cohort of control individuals and individuals afflicted by neurodegenerative diseases, this strategy

599   allowed for a strong discriminant model to be built, able to distinguish all studied groups more

600   effectively than the models generated from a single proteomics analysis. The most noticeable findings

601   from this model showed that several, otherwise overlooked, proteins may yet serve as potential

602   biomarkers of disease, in this case, AD and PD, particularly when analyzed together in a model created

603   using the two different approaches. Thus, these results confirm the importance of having a panel of

604   potential candidates rather than a single protein biomarker. Furthermore, they also demonstrate that

605   the biomarker discovery field will benefit from combining data from the sample obtained through

606   different sample processing strategies. While not individually sufficient to be considered as biomarkers,

607   the significant influence of apolipoproteins, particularly Apo-AI, Apo-CI, and Apo-E, in the

608   aforementioned discriminant model, suggests a potential disease-specific dysregulation of lipoprotein

609   metabolism in AD and PD patients.

610   **High molecular weight fractionation may reveal a potentially altered macromolecular and**

611   **macromolecular-complex organization**

612   In this work, it was demonstrated that interrogating serum samples with the HMW fractionation

613   method adds an extra layer of information capable of bringing new insight into the behavior of the

614   serum proteins, particularly regarding their potential macromolecular organization. Because the

615   fractionation procedure took place under non-denaturing conditions and since aberrant protein

616   aggregation (8) is a common hallmark of both AD and PD, it was hypothesized that macromolecular

617   complexes, potentially altered between the studied groups, could be captured through the HMW

618   fractionation approach.

619   The present results confirmed this premise, as evidenced by the fact that although no variation was

620   observed in the overall serum protein captured by both strategies, proteins exclusively altered in the

621   HMW fraction accounted for 40% of the total list of altered proteins. Furthermore, with the exception of

622   one protein (Centrosome-associated protein CEP250), all proteins had a MW below 90 kDa, which is

623   considerably lower than the 300 kDa cut-off filter used. It is worth noting that 72% of proteins altered in

624   the HMW fraction did not exhibit alterations in their total levels. This supports the possibility that

625   different regulatory mechanisms of these proteins, apart from expression and degradation, are being

626   revealed and studied using this approach.

627   Moreover, the results show that most of these proteins have several reported interactors and thus may

628   be involved in the formation of large complexes. An illustration of this phenomenon is found in the

629   protein clusterin (CLU gene), also known as apolipoprotein J, which has been implicated in the

630   metabolism of aggregation-prone proteins, including those associated with NDs (31-33). Notably,

631   clusterin's interaction with Aβ42 has been demonstrated to enhance its clearance from the brain

632   through the blood-brain barrier (BBB) (31). Furthermore, clusterin has been implicated in various stages

633   of PD, potentially exerting a neuroprotective effect through its interaction with α-synuclein aggregates

634   (32). Additionally, the interaction between clusterin and α-synuclein has been detected in plasma

635   samples (33). Beyond this specific example, the generic functional analysis of altered proteins,

636   particularly those from the HMW serum strategy, suggests their close association with amyloids and clot

637   formation, which may involve large structures.

638   On the other hand, some of these proteins may instead, or additionally, be present in large biological

639   structures not composed exclusively of proteins, like exosomes or lipoproteins, which would not only

640   likewise justify their presence in this HMW serum fraction but also give further understanding of the

641   potentially altered mechanisms related to the diseases being studied. Such may be the case of the

642   proteins clusterin and serum amyloid A-4 (SAA4 gene), which were found to be altered in serum neuron-

643   derived exosomes of AD patients (34). Additionally, exosomal clusterin was found to be altered in

644   patients at different stages of PD when compared to controls (35). Thus, although the presence of

645   exosomes in the HMW fraction was not confirmed, given the MWCO of the filters used in this work, it is

646   feasible that some of the proteins being analyzed in the HMW fraction may correspond to proteins

647   linked to the extracellular vesicles.

648   Altogether, these findings support the notion that the HMW fractionation approach can provide a new

649   level of information that may offer new insights into how proteins are organized within a given sample.

650   **Altered lipoprotein metabolism can be a peripherical marker of AD and PD**

651   The combination of the two approaches in this study led to a robust and promising potential biomarker

652   panel composed of ten proteins quantified in whole serum or HMW serum. A major finding revealed by

653   this model was the involvement of several lipoproteins in discriminating between the studied groups.

654   Among the ten proteins used in the best discriminant model, three are apolipoproteins: Apo-AI from

655   whole serum, and Apo-CI and Apo-E from HMW serum. Besides those three proteins, other altered

656   apolipoproteins were observed in this study but not included in the model, namely: i) the Apo-AII, Apo-

657   LI and clusterin highlighted in the HMW serum strategy; ii) the Apo-AIV from the whole serum; and, iii)

658   beta-2-glycoprotein 1 in both approaches. Furthermore, a lipoprotein-related enzyme, lecithin-

659   cholesterol acyltransferase, was also altered in HMW serum. This is further supported by the functional

660   enrichment analysis of the altered proteins discovered in both strategies that highlight the involvement

661   of those proteins in lipoprotein metabolism and HDL-mediated lipid transport pathways. Cumulatively,

662   all these findings point to the relevance of lipoproteins in the context of NDs and, although not

663   absolutely clear, the link between these diseases, in particularly AD, and apolipoproteins has been the

664   focus of many studies (31, 33, 36-46).

665   Apo-AI and Apo-E have an established relation to toxic species clearance from the brain in AD and PD

666   (31, 33, 37, 47-50). Additionally, regarding AD, our findings for both proteins contradict what can be

667   found in the literature (38, 43). For Apo-AI, we found an increase in abundance in AD patients as

668   opposed to the decrease reported for most studies (38). However, in another study where no significant

669   alterations in total serum Apo-AI content of AD patients were reported, further investigation revealed

670   that some proteoforms of this protein were significantly increased compared to the levels observed in

671   the controls (39). This has been suggested as a possible explanation for the different observations

672   regarding this protein in the context of AD, which might be related to the use of different detection

673   methods within different studies (38). Regarding this protein's connection to clearance mechanisms,

674   evidence suggests that for HDLs containing Apo-AI (APOA1-HDL), the structure seems to influence not

675   only the disaggregation of Aβ fibrils but also its ability to cross the BBB, with lipid-poor discoidal APOA1-

676   HDL having the best performance when compared to APOA1-HDL in other lipidation states (37).

677   Moreover, in the present study, phosphatidylcholine-sterol acyltransferase, an enzyme known to affect

678   HDL structure through lipidation of Apo-AI in plasma (45), was also found to be altered in AD patients.

679   As this enzyme participates in HDL maturation in plasma (51), this result could reflect the dysregulation

680   of lipoproteins in AD. For Apo-E, we found a decrease in abundance in AD patients as opposed to the

681   increase reported for MCI patients in a previous study (43). Isoform and lipidation status of Apo-E is also

682   crucial for the Aβ clearance, with the Apo-E4 isoform, the genetic risk factor most associated with the

683   onset of AD (31, 48), and higher lipidation having detrimental effects on the process (31, 48). Further

684   investigation using the combined approach presented in this study, particularly in the context of AD,

685   should also involve a lipid profile analysis and APOE genotyping of the participants to enable a more

686   comprehensive characterization of serum HDLs and Apo-E content, respectively.

687   Although systemic lipid abnormalities have also been implicated in PD, there are fewer findings

688   connecting it to HDL-related proteins, as compared to AD (32). Nonetheless, a previous report observed

689   significantly decreased values of Apo-AI in mild PD patients compared to healthy controls, but much like

690   what we observed for this protein, a less impactful and non-significant decrease was observed in

691   moderate/severe PD patients (52). In fact, most research indicates that Apo-AI may have a protective

692   role in PD (32) and it has been hypothesized that APOA1-HDL could take part in the efflux of α-synuclein

693   from the brain (33). Additionally, both Apo-AI and Apo-E have been reported to interact with α-

694   synuclein (33).

695   Finally, HDL size and plasma levels have been shown to be dependent on the levels of Apo-CI (53),

696   another apolipoprotein included in the best diagnostic model generated in this study. Besides that, a

697   previous study also shows that the dysregulation of this protein can lead to impaired memory processes

698   in mice (54). This suggests that the regulation of Apo-CI can be pivotal in the brain and that a systemic

699   disruption of this process could have effects detectable beyond the CNS, particularly in lipoprotein

700   metabolism, that could be observable in peripheral biofluids. Interestingly, this study found that Apo-CI

701   significantly decreased in the HMW serum fraction of AD compared to PD patients, but only a slight and

702   non-significant decrease was observed when compared to CT patients. Similarly, only a small non-

703   significant decrease in HMW serum Apo-CI was observed for controls in comparison to PD patients,

704   which is in accordance to what was already reported for whole plasma (40). Given that memory

705   impairment is a hallmark of AD and not a predominant feature among PD patients (55), these results

706   may be influenced, at least in part, by the age disparities observed between the PD group and the other

707   two groups. In fact, PD patients are, on average, younger than those in the other two groups, with

708   similar age distributions. Additionally, the results regarding Apo-CI may also be affected by individual

709   Apo-CI and Apo-E isoforms. Notably, the *APOE* and *APOC1* genes are in linkage disequilibrium (56), and

710   carriers of the *APOE(ε4)* and *APOC1(H2)* alleles have been demonstrated to have an increased risk of

711   developing AD (46). This was further confirmed in a study using human *APOE*-carrying mice, which

712   demonstrated that those animals carrying the *APOE(ε4)* allele were found to have decreased serum

713   Apo-CI content when compared to those carrying the *APOE(ε3)* allele (57). However, Apo-CI has also

714   been suggested to potentially play a modulatory role in the development of AD, with reported effects

715   on mice cognitive function independent of Apo-E expression (58). Again, these observations strengthen

716   the importance of combining these results with further characterization of the individuals, including

717   genotyping of the apolipoproteins' isoforms.

718   Although these three apolipoproteins have been linked to AD and PD, as evidenced by previously

719   mentioned studies, and their relevance for the discriminant model, their use alone or in combination

720   with other proteins associated with apolipoprotein-related mechanisms did not result in a robust

721   diagnostic model capable of effectively distinguishing between the studied groups (as shown in

722   Additional File 1: Supplementary Figure 7). This indicates that these three apolipoproteins had to be

723   combined with other seemingly unrelated proteins to be used as potential biomarkers. Further studies

724   should be directed towards elucidating this potential relationship to understand: i) the importance of

725   the identified proteins/mechanisms for the pathophysiology of the studied NDs, and ii) to which extent

726   these mechanisms are differently altered between the two diseases.

727   **Conclusion**

728   In summary, this work demonstrated that combining two complementary sample processing

729   approaches is a more effective strategy to reach potential biomarkers than a single approach. Besides

730   that, the strategies used here, which combine the analysis of the whole serum and the HMW

731   fractionation of non- denaturing serum, can also identify proteins being differentially modulated besides

732   the conventional alteration in their total levels. In this work, this new strategy was applied to a cohort of

733   NDs patients and respective CT individuals, being able to build a good predictive model capable of

734   distinguishing all the three groups studied (AD, PD and CT). This predictive model highlighted the linkage

735   of the apolipoprotein family and NDs, with three out of the ten proteins included in this model being

736   apolipoproteins. Nevertheless, further validation in a larger and independent cohort is needed to

737   confirm the robustness and reliability of the model, as well as more studies to link the alterations

738   observed and these pathologies. Controlling the lipid profile in future studies is also advised, as altered

739   lipid metabolism was a major finding of the present work. Another aspect to be further explored would

740   be the identification of protein complexes in the HMW fraction to better understand the origin of the

741   protein alterations observed.

742   Overall, the results of this study demonstrate that HMW fractionation under non-denaturing conditions

743   could be a valuable addition to routine biofluid analysis, particularly regarding the detection of changes

744   in macromolecular organization. Furthermore, this study highlights the importance of exploring the

745   involvement of apolipoproteins and lipid metabolism in greater detail within the context of NDs.

746 **List of abbreviations**

747 AD, Alzheimer's Disease; APOA1-HDL, High-Density Lipoproteins containing Apolipoprotein A1; AUC,

748 Area Under the Curve; Aβ, Amyloid-β; BBB, Blood-Brain Barrier; BioGRID, Biological General Repository

749 for Interaction Datasets; CE, Collision Energy; CES, Collision Energy Spread; CHUC, Centro Hospitalar e

750 Universitário de Coimbra; CI, Confidence Interval; CNS, Central Nervous System; CT, Healthy Controls;

751 DDA, Data-dependent Acquisition; DMSO, Dimethyl Sulfoxide; FA Formic Acid; FDR, False Discovery

752 Rate; GO, Gene Ontology; HDL, High-Density Lipoproteins; HMW, High Molecular Weight; IS, Internal

753 Standard; LDA, Linear Discriminant Analysis; MCI, Mild Cognitive Impairment; MS, Mass Spectrometry;

754 MW, Molecular Weight; MWCO, Molecular Weight Cut-Off; NDs, Neurodegenerative Diseases; PD,

755 Parkinson's Disease; ROC, Receiver Operating Curve; SE, Standard Error; STRING, Search Tool for

756 Retrieval of Interacting Genes/Proteins; XIC, Extracted Ion Chromatogram.

757   **Declarations**

758   **Ethics approval and consent to participate**

759   The study was approved by the Ethics Committee of the Faculty of Medicine of the University of

760   Coimbra (reference CE_010.2017) and the Ethics Committee of the Centro Hospitalar e Universitário de

761   Coimbra (CHUC) (reference 34 /CES-CHUC-024-18) and was conducted according to the principles stated

762   in the Declaration of Helsinki (59). Written informed consent was obtained from all participants.

763   **Consent for publication**

764   Not applicable

765   **Availability of data and materials**

766   The MS proteomics data have been deposited to the ProteomeXchange Consortium (60) via the PRIDE

767   (61) partner repository with the dataset identifier PXD034077.

768   **Competing interests**

769   The authors declare that they have no competing interests

770   **Funding**

781 **Authors' contributions**

782 SIA and BM conceived and designed the study. SIA and MR performed all the experiments, data analysis,

783 and wrote the manuscript. IB, AG, DP, CS, JP, CJ, IS, AV, AM, and MCB were involved in the selection and

784 collection of the plasma samples used in this study. MCB and BM were responsible for funding. The

785 author(s) read and approved the final manuscript.

786 **Acknowledgements**

787 Not applicable

788 **Authors' information (optional)**

789 Not applicable

790 **Additional files**

791 **Additional File 1 (.docx)** – Supplementary Figure 1: Internal standard proteins and ultrafiltration

792 reproducibility; Supplementary Figure 2: Altered proteins quantified using the whole serum dataset and

793 HMW serum dataset; Supplementary Figure 3: Volcano plots comparing all proteins quantified across

794 the three different experimental groups using the two different serum-processing strategies;

795 Supplementary Figure 4: Mass distribution of proteins identified or quantified using different

796 approaches; Supplementary Figure 5: Linear discriminant analysis using patient's age; Supplementary

797 Figure 6: Correlation between quantification of the 10 proteins included in the combined model and

798 patient age; Supplementary Figure 7: Linear discriminant analysis using 10 altered proteins associated

799 with apolipoprotein related mechanisms; Supplementary Table 8 – Summary of the specificity and

800 sensitivity of the LDA methods obtained using all the altered proteins per sample type (whole serum or

801 HMW serum) or the combination of the two (Whole + HMW serum).

802 **Additional File 2 (.xlsx)** – Supplementary Table 1 - Evaluation of the total levels of proteins at the

803 unfractionated serum (Whole Serum) of Control and Alzheimer's and Parkinson's Disease patients;

804 Supplementary Table 2 - Evaluation of the total levels of proteins at the high molecular fraction of the

805 serum (HMW Serum) of Control and Alzheimer's and Parkinson's Disease patients; Supplementary Table

806 3 - Evaluation of the total levels of proteins at the unfractionated serum (Whole Serum) of Control and

807 Alzheimer's and Parkinson's Disease patients; Supplementary Table 4 - Descriptive table of the

808 GeneMania functional enrichment analysis of the 28 proteins exclusively quantified in the HMW serum

809 samples; Supplementary Table 5 - Descriptive table of protein physical interactions reported in BioGRID

810 for the four proteins (marked in yellow) with no reported interactions according to the GeneMania

811 network (Figure 3a); Supplementary Table 6 - Descriptive table of the Funrich Biological Pathway

812 Analysis of the 30 proteins exclusively quantified in Whole serum samples; Supplementary Table 7 -

813 Descriptive table of the Funrich Biological Pathway Analysis of the 30 proteins exclusively quantified in

814 HMW serum samples; Supplementary Table 9 - Descriptive table of the String Enrichment Analysis of the

815 10 proteins used in the LDA combined model; Supplementary Table 10 - Descriptive table of the

816 Reactome Pathway enrichment analysis of the 10 proteins used in the LDA combined model.

References

1.	Anderson NL, Anderson NG. The Human Plasma Proteome: History, Character, and Diagnostic Prospects*. Molecular & Cellular Proteomics. 2002;1(11):845-67.

2.	Antoranz A, Sakellaropoulos T, Saez-Rodriguez J, Alexopoulos LG. Mechanism-based biomarker discovery. Drug Discovery Today. 2017;22(8):1209-15.

3.	Anjo SI, Santa C, Manadas B. SWATH-MS as a tool for biomarker discovery: From basic research to clinical applications. PROTEOMICS. 2017;17(3-4):1600278.

4.	Anjo SI, Manadas B. A translational view of cells' secretome analysis - from untargeted proteomics to potential circulating biomarkers. Biochimie. 2018;155:37-49.

5.	Rosado M, Silva R, G. Bexiga M, G. Jones J, Manadas B, Anjo SI. Chapter Four - Advances in biomarker detection: Alternative approaches for blood-based biomarker detection. In: Makowski GS, editor. Advances in Clinical Chemistry. 92: Elsevier; 2019. p. 141-99.

6.	Greening DW, Simpson RJ. A centrifugal ultrafiltration strategy for isolating the low-molecular weight (≤25K) component of human plasma proteome. Journal of Proteomics. 2010;73(3):637-48.

7.	Greening DW, Simpson RJ. Low-Molecular Weight Plasma Proteome Analysis Using Centrifugal Ultrafiltration. In: Simpson RJ, Greening DW, editors. Serum/Plasma Proteomics: Methods and Protocols. Totowa, NJ: Humana Press; 2011. p. 109-24.

8.	Espay AJ, Vizcarra JA, Marsili L, Lang AE, Simon DK, Merola A, et al. Revisiting protein aggregation as pathogenic in sporadic Parkinson and Alzheimer diseases. Neurology. 2019;92(7):329-37.

9.	Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. Journal of Neurology, Neurosurgery & Psychiatry. 1992;55(3):181-4.

10.	American Psychiatric Association A, Association AP. Diagnostic and statistical manual of mental disorders: DSM-IV: American psychiatric association Washington, DC; 1994.

11.	McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack Jr CR, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's & dementia. 2011;7(3):263-9.

845 12.   Anjo SI, Simoes I, Castanheira P, Graos M, Manadas B. Use of recombinant proteins as a simple

846 and robust normalization method for untargeted proteomics screening: exhaustive performance

847 assessment. Talanta. 2019;205:120163.

848 13.   Anjo SI, Santa C, Manadas B. Short GeLC-SWATH: a fast and reliable quantitative approach for

849 proteomic screenings. Proteomics. 2015;15(4):757-62.

850 14.   Santa C, Anjo SI, Manadas B. Protein precipitation of diluted samples in SDS-containing buffer

851 with acetone leads to higher protein recovery and reproducibility in comparison with TCA/acetone

852 approach. Proteomics. 2016;16(13):1847-51.

853 15.   Anjo SI, Melo MN, Loureiro LR, Sabala L, Castanheira P, Grãos M, et al. oxSWATH: An integrative

854 method for a comprehensive redox-centered analysis combined with a generic differential proteomics

855 screening. Redox biology. 2019;22:101130.

856 16.   Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, et al. Targeted data extraction of the

857 MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate

858 proteome analysis. Molecular & Cellular Proteomics. 2012;11(6).

859 17.   Tang WH, Shilov IV, Seymour SL. Nonlinear fitting method for determining local false discovery

860 rates from decoy database searches. Journal of proteome research. 2008;7(9):3661-7.

861 18.   Sennels L, Bukowski-Wills JC, Rappsilber J. Improved results in proteomics by use of local and

862 peptide-class specific false discovery rates. BMC Bioinformatics. 2009;10:179.

863 19.   Lambert J-P, Ivosev G, Couzens AL, Larsen B, Taipale M, Lin Z-Y, et al. Mapping differential

864 interactomes by affinity purification coupled with data-independent mass spectrometry acquisition.

865 Nature methods. 2013.

866 20.   Collins BC, Gillet LC, Rosenberger G, Röst HL, Vichalkovski A, Gstaiger M, et al. Quantifying

867 protein interaction dynamics by SWATH mass spectrometry: application to the 14-3-3 system. Nature

868 methods. 2013.

869 21.   MedCalc. https://www.medcalc.org. Accessed 19 April 2022

870 22.   DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated

871 receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988:837-45.

872 23.   MORPHEUS. https://software.broadinstitute.org/morpheus/. Accessed 24 May 2022

873   24.   Caraux G, Pinloche S. PermutMatrix: a graphical environment to arrange gene expression
874   profiles in optimal linear order. Bioinformatics. 2005;21(7):1280-1.
875   25.   PermutMatrix. http://www.atgc-montpellier.fr/permutmatrix/. Accessed 24 May 2022
876   26.   Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA
877   prediction server: biological network integration for gene prioritization and predicting gene function.
878   Nucleic Acids Research. 2010;38(suppl_2):W214-W20.
879   27.   GeneMANIA. https://genemania.org/. Accessed 18 May 2022
880   28.   STRING. http://string-db.org. Accessed 22 June 2022
881   29.   Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in
882   2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids
883   Res. 2017;45(D1):D362-d8.
884   30.   Pathan M, Keerthikumar S, Ang CS, Gangoda L, Quek CY, Williamson NA, et al. FunRich: An open
885   access standalone functional enrichment and interaction network analysis tool. Proteomics.
886   2015;15(15):2597-601.
887   31.   Bell RD, Sagare AP, Friedman AE, Bedi GS, Holtzman DM, Deane R, et al. Transport pathways for
888   clearance of human Alzheimer's amyloid beta-peptide and apolipoproteins E and J in the mouse central
889   nervous system. Journal of cerebral blood flow and metabolism : official journal of the International
890   Society of Cerebral Blood Flow and Metabolism. 2007;27(5):909-18.
891   32.   Berdowska I, Matusiewicz M, Krzystek-Korpacka M. HDL Accessory Proteins in Parkinson's
892   Disease—Focusing on Clusterin (Apolipoprotein J) in Regard to Its Involvement in Pathology and
893   Diagnostics—A Review. Antioxidants. 2022;11(3).
894   33.   Emamzadeh FN, Allsop D. α-Synuclein Interacts with Lipoproteins in Plasma. Journal of
895   Molecular Neuroscience. 2017;63(2):165-72.
896   34.   Arioz BI, Tufekci KU, Olcum M, Durur DY, Akarlar BA, Ozlu N, et al. Proteome profiling of
897   neuron-derived exosomes in Alzheimer's disease reveals hemoglobin as a potential biomarker.
898   Neuroscience Letters. 2021;755:135914.
899   35.   Kitamura Y, Kojima M, Kurosawa T, Sasaki R, Ichihara S, Hiraku Y, et al. Proteomic Profiling of
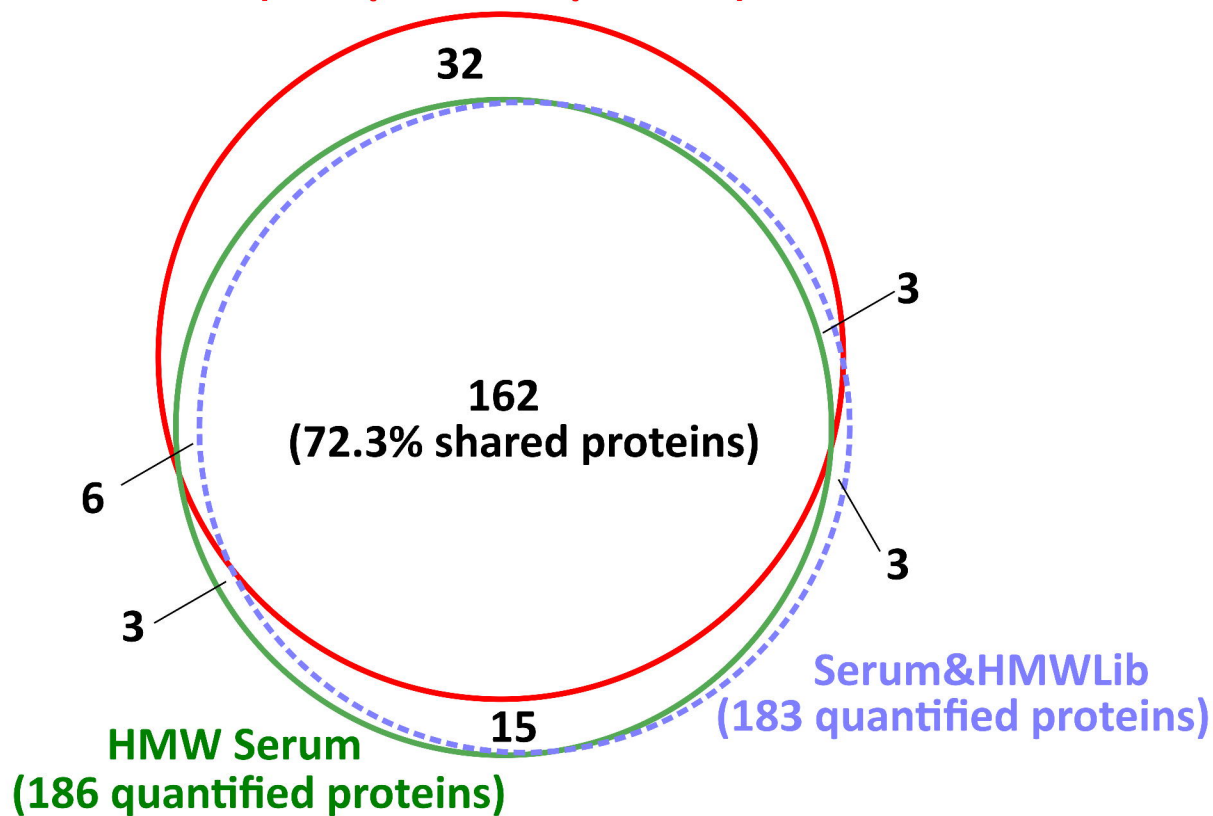900   Exosomal Proteins for Blood-based Biomarkers in Parkinson's Disease. Neuroscience. 2018;392:121-8.

901   36.   Berdowska I, Matusiewicz M, Krzystek-Korpacka M. HDL Accessory Proteins in Parkinson's

902   Disease–Focusing on Clusterin (Apolipoprotein J) in Regard to Its Involvement in Pathology and

903   Diagnostics–A Review. Antioxidants. 2022;11(3).

904   37.   Dal Magro R, Simonelli S, Cox A, Formicola B, Corti R, Cassina V, et al. The Extent of Human

905   Apolipoprotein A-I Lipidation Strongly Affects the β-Amyloid Efflux Across the Blood-Brain Barrier in

906   vitro. Frontiers in neuroscience. 2019;13:419.

907   38.   Endres K. Apolipoprotein A1, the neglected relative of Apolipoprotein E and its potential role in

908   Alzheimer's disease. Neural Regeneration Research. 2021;16(11).

909   39.   Fania C, Arosio B, Capitanio D, Torretta E, Gussago C, Ferri E, et al. Protein signature in

910   cerebrospinal fluid and serum of Alzheimer's disease patients: The case of apolipoprotein A-1

911   proteoforms. PLOS ONE. 2017;12(6):e0179280.

912   40.   Paslawski W, Zareba-Paslawska J, Zhang X, Hölzl K, Wadensten H, Shariatgorji M, et al. α-

913   synuclein–lipoprotein interactions and elevated ApoE level in cerebrospinal fluid from Parkinson's

914   disease patients. Proceedings of the National Academy of Sciences. 2019;116(30):15226-35.

915   41.   Paula-Lima AC, Tricerri MA, Brito-Moreira J, Bomfim TR, Oliveira FF, Magdesian MH, et al.

916   Human apolipoprotein A–I binds amyloid-β and prevents Aβ-induced neurotoxicity. The International

917   Journal of Biochemistry & Cell Biology. 2009;41(6):1361-70.

918   42.   Qiang JK, Wong YC, Siderowf A, Hurtig HI, Xie SX, Lee VM-Y, et al. Plasma apolipoprotein A1 as a

919   biomarker for Parkinson disease. Annals of Neurology. 2013;74(1):119-27.

920   43.   Song F, Poljak A, Crawford J, Kochan NA, Wen W, Cameron B, et al. Plasma Apolipoprotein

921   Levels Are Associated with Cognitive Status and Decline in a Community Cohort of Older Individuals.

922   PLOS ONE. 2012;7(6):e34078.

923   44.   Swanson CR, Berlyand Y, Xie SX, Alcalay RN, Chahine LM, Chen-Plotkin AS. Plasma

924   apolipoprotein A1 associates with age at onset and motor severity in early Parkinson's disease patients.

925   Movement Disorders. 2015;30(12):1648-56.

926   45.   van der Vorst EPC. High-Density Lipoproteins and Apolipoprotein A1. In: Hoeger U, Harris JR,

927   editors. Vertebrate and Invertebrate Respiratory Proteins, Lipoproteins and other Body Fluid Proteins.

928   Cham: Springer International Publishing; 2020. p. 399-420.

929   46.   Zhou Q, Zhao F, Lv Z-p, Zheng C-g, Zheng W-d, Sun L, et al. Association between APOC1

930   Polymorphism and Alzheimer's Disease: A Case-Control Study and Meta-Analysis. PLOS ONE.

931   2014;9(1):e87017.

932   47.   Castellano JM, Kim J, Stewart FR, Jiang H, DeMattos RB, Patterson BW, et al. Human apoE

933   Isoforms Differentially Regulate Brain Amyloid-β Peptide Clearance. Science Translational Medicine.

934   2011;3(89):89ra57-89ra57.

935   48.   Deane R, Sagare A, Hamm K, Parisi M, Lane S, Finn MB, et al. apoE isoform-specific disruption of

936   amyloid beta peptide clearance from mouse brain. The Journal of clinical investigation.

937   2008;118(12):4002-13.

938   49.   Merino-Zamorano C, Fernández-de Retana S, Montañola A, Batlle A, Saint-Pol J, Mysiorek C, et

939   al. Modulation of Amyloid-β 1–40 Transport by ApoA1 and ApoJ Across an in vitro Model of the Blood-

940   Brain Barrier. Journal of Alzheimer's Disease. 2016;53:677-91.

941   50.   Verghese PB, Castellano JM, Garai K, Wang Y, Jiang H, Shah A, et al. ApoE influences amyloid-β

942   (Aβ) clearance despite minimal apoE/Aβ association in physiological conditions. Proceedings of the

943   National Academy of Sciences. 2013;110(19):E1807-E16.

944   51.   Rye K-A, Clay MA, Barter PJ. Remodelling of high density lipoproteins by plasma factors.

945   Atherosclerosis. 1999;145(2):227-38.

946   52.   Zhang X, Yin X, Yu H, Liu X, Yang F, Yao J, et al. Quantitative proteomic analysis of serum

947   proteins in patients with Parkinson's disease using an isobaric tag for relative and absolute

948   quantification labeling, two-dimensional liquid chromatography, and tandem mass spectrometry.

949   Analyst. 2012;137(2):490-5.

950   53.   de Haan W, Out R, Berbée JFP, van der Hoogt CC, Dijk KWv, van Berkel TJC, et al. Apolipoprotein

951   CI inhibits scavenger receptor BI and increases plasma HDL levels in vivo. Biochemical and Biophysical

952   Research Communications. 2008;377(4):1294-8.

953   54.   Berbée JFP, Vanmierlo T, Abildayeva K, Blokland A, Jansen PJ, Lütjohann D, et al. Apolipoprotein

954   CI Knock-Out Mice Display Impaired Memory Functions. Journal of Alzheimer's Disease. 2011;23:737-47.

955   55.   Albert AD, Brad R. Parkinson disease and cognitive impairment. Neurology: Clinical Practice.

956   2016;6(5):452.

957   56.   Fuior EV, Gafencu AV. Apolipoprotein C1: Its Pleiotropic Effects in Lipid Metabolism and

958   Beyond. International Journal of Molecular Sciences. 2019;20(23):5939.

959   57.   Cudaback E, Li X, Yang Y, Yoo T, Montine KS, Craft S, et al. Apolipoprotein C-I is an APOE

960   genotype-dependent suppressor of glial activation. Journal of Neuroinflammation. 2012;9(1):192.

961   58.   Abildayeva K, Berbée JFP, Blokland A, Jansen PJ, Hoek FJ, Meijer O, et al. Human apolipoprotein

962   C-I expression in mice impairs learning and memory functions. Journal of Lipid Research.

963   2008;49(4):856-69.

964   59.   Association GAotWM. World Medical Association Declaration of Helsinki: ethical principles for

965   medical research involving human subjects. The Journal of the American College of Dentists.

966   2014;81(3):14.

967   60.   Deutsch EW, Bandeira N, Perez-Riverol Y, Sharma V, Carver Jeremy J, Mendoza L, et al. The

968   ProteomeXchange consortium at 10 years: 2023 update. Nucleic Acids Research. 2022;51(D1):D1539-

969   D48.

970   61.   Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, Hewapathirana S, Kamatchinathan S, et al.

971   The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences.
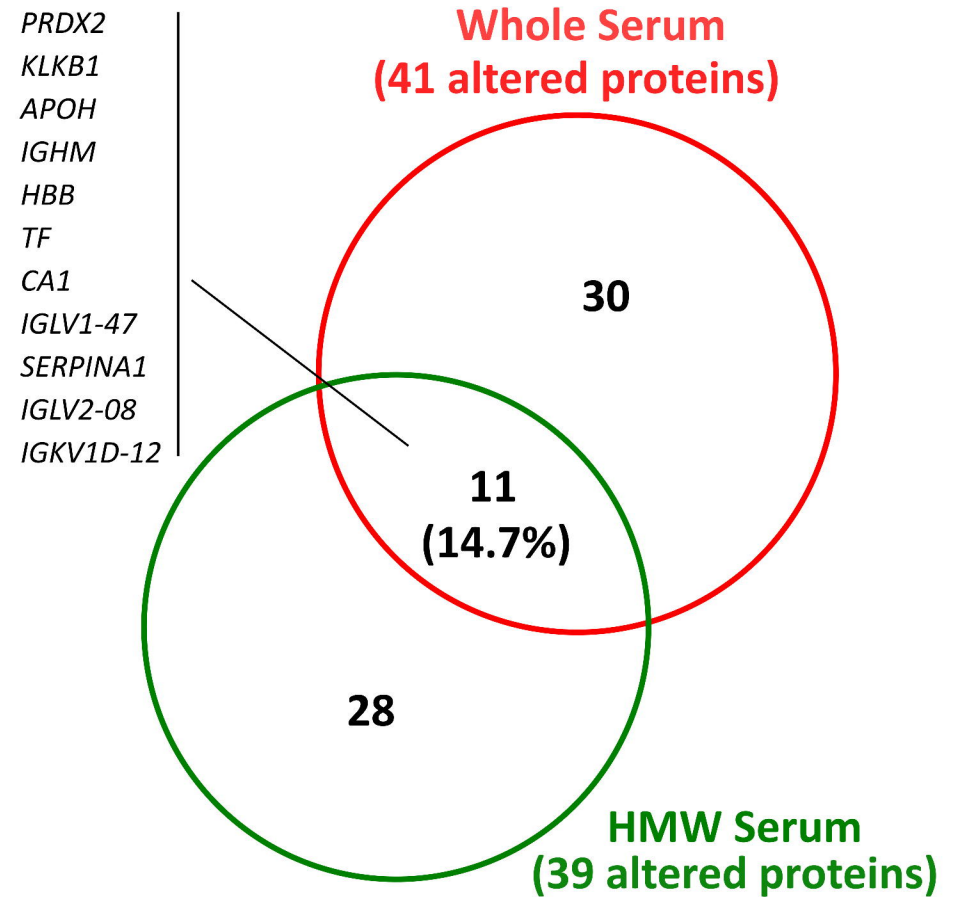
972   Nucleic Acids Research. 2021;50(D1):D543-D52.

(a)

Whole Serum
(203 quantified proteins)

32

3

162
(72.3% shared proteins)

6

3

3

15

Serum&HMWLib
(183 quantified proteins)

HMW Serum
(186 quantified proteins)

Total: 224 quantified proteins

(b)

PRDX2
KLKB1
APOH
IGHM
HBB
TF
CA1
IGLV1-47
SERPINA1
IGLV2-08
IGKV1D-12

Whole Serum
(41 altered proteins)

30

11
(14.7%)

28

HMW Serum
(39 altered proteins)

Total: 69 altered proteins

(c)

Group comparison	Gene Name	Protein Name	Fold Change Serum	HMW
AD vs CT	APOH	Beta-2-glycoprotein 1 (APC inhibitor)	1.008#	0.676
	CA1	Carbonic anhydrase 1	2.467	2.397
	IGKV1D-12	Immunoglobulin kappa variable 1D-12	0.799	0.762
	IGLV1-47	Immunoglobulin lambda variable 1-47	0.767	0.786
	IGHM	Immunoglobulin heavy constant mu	0.372	0.692
	TF	Serotransferrin	0.856	0.792
	KLKB1	Plasma kallikrein	0.807	0.707
	PRDX2	Peroxiredoxin-2	2.691	2.979
	HBB	Hemoglobin subunit beta	3.958	3.129
PD vs CT	SERPINA1	Alpha-1-antitrypsin	0.877	0.752
AD vs PD	APOH	Beta-2-glycoprotein 1 (APC inhibitor)	1.238	0.803#
	CA1	Carbonic anhydrase 1	1.977	1.943
	IGLV1-47	Immunoglobulin lambda variable 1-47	0.894	0.785
	IGLV2-8	Immunoglobulin lambda variable 2-8	0.777	0.788
	PRDX2	Peroxiredoxin-2	1.749	2.568
	HBB	Hemoglobin subunit beta	2.622	2.094

Max~4.00

Min~0.300

(a)
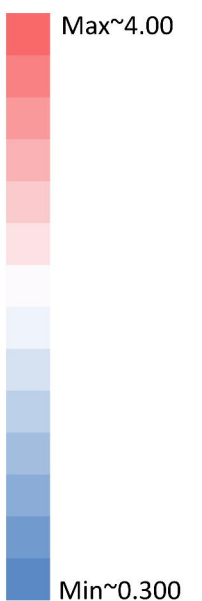
Proteins (nodes)

⬤ (hatched) Proteins from the study

⬤ Related proteins

Networks (edges)

— Physical Interactions

Functions

- 🟥 Lipoprotein particle
- 🟦 Plasma lipoprotein particle
- 🟧 Protein-lipid complex
- 🟪 Platelet alpha granule
- 🟩 Triglyceride-rich plasma lipoprotein particle
- 🟪 Protein-lipid complex remodeling
- 🟩 Plasma lipoprotein particle remodeling

(b)

FunRich Biological Pathway Analysis

-Log₁₀(p-value) scale: $-\text{Log}_{10}(\text{p-value})$ from 2 to 7

% of genes scale: 1, 10, 20, 30, 40, 50

Abbreviations:
- **ab** - antibody
- **PTM** - Post-translational modification
- **IGF** - Insulin-like Growth Factor
- **IGFBPs** - Insulin-like Growth Factor Binding Proteins
- **ER** - endoplasmic reticulum

Pathways (top to bottom):
- *Amyloids*
- Intrinsic Pathway
- Chylomicron-mediated lipid transport
- *Dissolution of Fibrin Clot*
- Classical ab-mediated complement activation
- Cell surface interactions at the vascular wall
- Angiopoietin receptor Tie2-mediated signaling
- Lipid digestion, mobilization, and transport
- *Formation of Fibrin Clot (Clotting Cascade)*
- *Lipoprotein metabolism*
- Common Pathway
- HDL-mediated lipid transport
- PTM: γ-carboxylation and hypusine formation
- Regulation of IGF Activity by IGFBPs
- γ-carboxylation, transport, and amino-terminal cleavage
- Hemostasis
- Removal of aminoterminal propeptides from γ-carboxylated proteins
- γ-carboxylation of protein precursors
- Transport of γ-carboxylated protein precursors from the ER to the Golgi
- FOXA2 and FOXA3 transcription factor networks
- FOXA transcription factor networks

Columns: **Whole**, **HMW**

(a) PD vs CT (12 altered proteins)

AD vs CT (20 altered proteins)

AD vs PD (39 altered proteins)

1 · 1 · 4 · 7 · 14 · 14

Whole Serum
Total: 41 altered proteins

(b) PD vs CT (14 altered proteins)

AD vs CT (36 altered proteins)

AD vs PD (19 altered proteins)

12 · 1 · 1 · 8 · 15 · 2

HMW Serum
Total: 39 altered proteins

(c) PD vs CT (25 altered proteins)

AD vs CT (48 altered proteins)

AD vs PD (53 altered proteins)

12 · 7 · 4 · 6 · 26 · 14

Whole + HMW Serum
Total: 69 altered proteins

(d) Whole Serum — 4 proteins

87.5% · 86.4% · 66.7%

★ Group Centroid
CT
AD
PD

(e) HMW Serum — 6 proteins

83.3% · 77.3% · 75.0%

★ Group Centroid
CT
AD
PD

(f) Serum + HMW Serum — 10 proteins

100% · 91.7% · 95.5%

★ Group Centroid
CT
AD
PD

(g) AD

(h) PD

(i) ROC curve - statistics

		Whole Serum	HMW serum	Whole + HMW serum
AD	AUC	0.957	0.919	0.999
	95% CI	[0.895-0.998]	[0.817-0.974][n.s]	[0.936-1.000][n.s;#]
PD	AUC	0.960	0.888	0.999
	95% CI	[0.872-0.994]	[0.778-0.956][n.s]	[0.936-1.000][*;#]

(a) *Whole Serum* — Violin plots of relative protein levels for FOXM1, PROC, HBB, APOA1, and IGLV3-19 across CT, AD, and PD groups. *HMW Serum* — Violin plots of relative protein levels for SRRD, APOC1, APOE, SERPINF1, and KRT9 across CT, AD, and PD groups.

(b) Pearson's correlation matrix among FOXM1, IGLV3-19, APOA1, PROC, HBB, APOE, APOC1, KRT9, SERPINF1, and SRRD.

(c) Hierarchical clustering heatmap with Cluster PD, Cluster CT, and Cluster AD. Protein names: Hemoglobin subunit beta (Beta-globin); Vitamin K-dependent protein C; Immunoglobulin lambda variable 3-19; Apolipoprotein A-I (Apo-AI); Forkhead box protein M1; Keratin, type I cytoskeletal 9; SRR1-like protein; Pigment epithelium-derived factor (PEDF); Apolipoprotein E (Apo-E); Apolipoprotein C-I (Apo-CI). Average profile of Cluster: PD, CT, AD. Min = -4.76, 0.00, Max = 4.76.

(d) UniProt Keywords: Glycation, Lipid transport, VLDL, HDL, Oxidation, Amyloidosis, Cholesterol metabolism, Secreted. Edges — Known Interactions: Experimentally determined, From curated databases; Other: Textmining, Co-expression. Cluster 1: APOE, APOC1, HBB, APOA1. Cluster 2: PROC, FOXM1. Cluster 3: SERPINF1, SRRD, KRT9.

(e) Percentage of genes, p-value, p = 0.05 reference. Pathways: HDL remodeling; Chylomicron remodeling; Chylomicron assembly; Scavenging of heme from plasma; Post-translational protein phosphorylation; Regulation of IGF transport and uptake by IGFBPs; Scavenging by Class A Receptors; Retinoid metabolism and transport. X-axis: -log10(Bonferroni corrected p-values).