bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Unmasking Hidden Systemic Effects of Neurodegenerative Diseases: A Two-Pronged Approach to Biomarker Discovery

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 Santa[1,2,3], Joana Pinto[1,2], Cristina Januário[6], Isabel Santana[1,2,6], Ana Verdelho[8,9], Alexandre de Mendonça[9], Miguel Castelo-Branco[7,#], Bruno Manadas[1,2,3,#,†]

Provide the full postal address of each affiliation

[1] CNC - Centro de Neurociências e Biologia Celular, Universidade de Coimbra, Rua Larga - Faculdade de Medicina, 1°andar - POLO I, 3004-504 Coimbra, Portugal.

[2] CIBB - Centre for Innovative Biomedicine and Biotechnology, University of Coimbra, Rua Larga - Faculdade de Medicina, 1°andar - POLO I, 3004-504 Coimbra, Portugal.

[3] Instituto de Investigação Interdisciplinar, Universidade de Coimbra (IIIUC), Casa Costa Alemão, R. Dom Francisco de Lemos, 3030-789 Coimbra, Portugal.

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

[5] - University of Coimbra, Institute for Interdisciplinary Research, Doctoral Programme in Experimental Biology and Biomedicine (PDBEB), Portugal

[6] Serviço de Neurologia, Centro Hospitalar e Universitário de Coimbra, Praceta Professor Mota Pinto, 3004-561 Coimbra, Portugal.

[7] Coimbra Institute for Biomedical Imaging and Translational Research (CIBIT), Azinhaga de. Santa Comba, 3000-548 Coimbra, Portugal.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

[8] Department of Neurosciences and Mental Health, Santa Maria Hospital - CHLN, ISAMB, University of Lisbon, Av. Prof. Egas Moniz MB, 1649-028 Lisboa, Portugal.

[9] Instituto de Medicina Molecular e Faculdade de Medicina, Universidade de Lisboa, Av. Prof. Egas Moniz, 1649-028 Lisboa, Portugal.

* - Equal contribution

# - Senior equal contribution

† - Corresponding authors: Bruno Manadas (bmanadas@cnc.uc.pt, +351231249170) and Sandra Anjo (sandra.anjo@cnc.uc.pt, +351231249170), Center for Neuroscience and Cell Biology – University of Coimbra, UC Biotech - Parque Tecnológico de Cantanhede, Núcleo 04, Lote 8, 3060-197 Cantanhede – Portugal

E-mail addresses: Miguel Rosado - mmva.rosado@gmail.com; Ines Baldeiras - ibaldeiras@cnc.uc.pt; Andreia Gomes - andmaggomes@gmail.com; Diana Pires - diana.santos.pires@gmail.com; Cátia Santa - catiajmsanta@gmail.com; Joana Pinto - joana.amaralpinto@gmail.com; Cristina Januário - cristinajanuario@gmail.com; Isabel Santana - isabeljsantana@gmail.com; Ana Verdelho - averdelho@medicina.ulisboa.pt; Alexandre de Mendonça - mendonca@medicina.ulisboa.pt; Miguel Castelo-Branco – mcbranco@fmed.uc.pt.

Abbreviations: AD, Alzheimer's Disease; APOA1-HDL, High-Density Lipoproteins containing Apolipoprotein A1; AUC, Area Under the Curve; Aβ, Amyloid-β; BBB, Blood-Brain Barrier; BioGRID, Biological General Repository for Interaction Datasets; CE, Collision Energy; CES, Collision Energy Spread; CHUC, Centro Hospitalar e Universitário de Coimbra; CI, Confidence Interval; CNS, Central Nervous System; CT, Healthy Controls; DDA, Data-dependent Acquisition; DMSO, Dimethyl Sulfoxide; FA Formic Acid; FDR, False Discovery Rate; GO, Gene Ontology; HDL, High-Density Lipoproteins; HMW, High Molecular Weight; IS, Internal Standard; LDA, Linear Discriminant Analysis; MCI, Mild Cognitive Impairment; MS, Mass Spectrometry; MW, Molecular Weight; MWCO, Molecular

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Weight Cut-Off; PD, Parkinson's Disease; ROC, Receiver Operating Curve; SE, Standard Error; STRING, Search Tool for Retrieval of Interacting Genes/Proteins; XIC, Extracted Ion Chromatogram.

Keywords

Serum Proteomics

High Molecular Weight Fractionation

Neurodegenerative Diseases

SWATH-MS/DIA

Protein complexes/Proteoforms

Apolipoproteins

Lipoproteins

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Abstract

Neurodegenerative Diseases (NDs) are a major health challenge. Thus, finding reliable blood biomarkers for them has been of pivotal importance in translational/clinical research. However, conventional omics struggle with the complexity of blood samples making it difficult to achieve the desired goal. To address this, the potential of High Molecular Weight (HMW) serum fractionation under non-denaturing conditions as a complementary approach to the direct analysis of whole serum proteomics was explored in this work. To achieve this, a total of 58 serum samples of Alzheimer's disease (AD), Parkinson's disease (PD) patients and control individuals underwent both strategies: i) direct analysis of whole serum and ii) non-denaturing fractionation using 300 kDa cut-off filters (HMW serum). As expected, each approach was able to capture different sets of differentially regulated proteins since most of the altered proteins were not shared between them. More importantly, it was possible to create a discriminant model using the altered proteins from both datasets capable of successfully distinguishing the three groups (AUC = 0.999 and, median sensitivity and specificity of 97.4% and 91.7%, respectively). Among the 10 proteins included in the model (5 from each strategy), a clear evidence for the contribution of proteins from the apolipoprotein family for the diagnosis of NDs was revealed. Furthermore, HMW fractionation exposed potential changes within the organization of macromolecules and their complexes, thereby uncovering hidden effects in serum. Altogether, this work demonstrated that HMW fractionation can be a valuable complementary method to direct serum analysis and could enhance biomarker discovery.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Introduction

Proteins can be directly or indirectly related to a myriad of diseases, thereby being important targets of biomarker research, which remains a pivotal aspect of clinical research. However, biomarkers discovery based on conventional proteomics strategies have not yet yielded substantial practical applications. Blood and its fractions are the most studied samples for biomarker identification, not only due to easy accessibility but also because of its interactions with most tissues in the body, potentially reflecting disease-related alterations [1, 2]. However, both plasma and serum proteomes have a very wide dynamic range of protein concentrations [1], with the detection of those least abundant being masked by the most abundant. Despite all the technical developments in proteomics quantification by mass spectrometry (MS) [3], the high dynamic range still precludes the complete characterization of these samples may be one of the reasons why the identification of relevant biomarkers in plasma/serum by MS has been difficult [4]. Therefore, it is essential to have alternative approaches to the direct analysis of these highly demanding biofluids. Several approaches have been developed to reduce sample complexity [5] like size exclusion chromatography and electrophoretic separation methods. When working with many samples, a simpler approach that could serve as an alternative to direct sample analysis is centrifugal ultrafiltration [5], which relies on the separation of sample components based on size exclusion, correlated to molecular weight (MW). This method has typically been used to study the low molecular weight proteome of serum and plasma samples, using filters between 20 kDa and 40 kDa [6, 7], but it may also be used to study other proteome fractions. Alzheimer's Disease (AD) and Parkinson's Disease (PD) are inherently linked to protein aggregation [8] and, therefore, to the formation of high molecular weight (HMW) protein complexes. Fractionation of peripheral fluid samples from patients with these diseases, focused on the HMW proteome, could be useful to not only eliminate high abundance proteins, but also study

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

protein aggregates that may be present in circulation. In this study, we propose the use of centrifugal ultrafiltration focused on the HMW proteome, not as an alternative but as a complementary analysis to the investigation of unfractionated samples. To evaluate our hypothesis, the same set of samples was subjected to the proteomics analysis of the whole serum and to fractionation using centrifugal ultrafiltration with 300 kDa molecular weight cut-off (MWCO) filters in a non-denaturing environment (henceforth referred to as HMW fractionation). The data obtained in the two approaches were not only directly compared but also combined to identify potential biomarkers. This study used a cohort comprising patients with neurodegenerative disorders, specifically Alzheimer's and Parkinson's disease patients, and healthy controls (CT) to test the significance of using HMW fractionation as a complementary tool for biomarker discovery. These disorders were selected as they are considered proteinopathies and may offer ideal targets for this purpose.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Material and Methods

Participants

A total of 58 serum samples were used in this study, comprising 3 groups of individuals: AD (n = 22), PD (n = 24), and CT (n = 12). The study was approved by the Ethics Committee of the Faculty of Medicine of the University of Coimbra (reference CE_010.2017) and the Ethics Committee of the Centro Hospitalar e Universitário de Coimbra (CHUC) (reference 34 /CES-CHUC-024-18) and was conducted according to the principles stated in the Declaration of Helsinki [9]. Written informed consent was obtained from all participants. The PD patients were recruited at the Movement Disorders Units of the Neurological Department of the CHUC, where they were assessed by a movement disorders specialist and were diagnosed according to the criteria defined by the UK Parkinsons's Disease Society Brain Bank [10]. The exclusion criteria for these patients consisted of severe dementia (as indicated by a Mini-Mental State Exam score below 15), any psychiatric disorder, or other forms of parkinsonism. The clinical group of individuals with AD diagnosis was recruited and prospectively evaluated by two experienced neurologists at Memoclínica and the Neurology Department of the CHUC. The standard criteria for the diagnosis of AD were the Diagnostic and Statistical Manual of Mental Disorders—fourth edition (DSM-IV-TR) and the National Institute on Aging and the Alzheimer's Association Workgroup [11]. To ensure the homogeneity of the sample, only patients who met the following criteria were included: they were in a stable condition, did not sustain recent changes in medication, and did not have ophthalmological or neurological/psychiatric conditions other than AD. The CT group was composed of age- and gender-matched individuals from the community with no history of cognitive deterioration, neurological or acquired central nervous system (CNS) disorders, traumatic brain injury, or psychiatric disorders. The CT group was also submitted to a brief cognitive assessment to exclude the presence of cognitive impairment.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Serum processing for proteomics analysis

Two different strategies were used to obtain a more comprehensive proteomic characterization of serum samples, namely: i) direct analysis of whole serum, and ii) HMW fractionation through ultrafiltration using 300 kDa cut-off filters (HMW serum).

For each sample, 5 µL were used for direct serum analysis and 82.5 µL for HMW serum fractionation. Additionally, three sample pools were prepared by combining aliquots of all the samples, for the AD pool, PD pool or CT pool, respectively. Pooled samples were used for Data-dependent acquisition (DDA) experiments to build a specific protein library to be used in Data independent acquisition (DIA) analysis and were subjected to the same sample processing as the individual samples. Before processing, all samples were spiked with the same amount of an internal standard (IS) to account for sample loss [12]. Different internal standards were used depending on the type of analysis: MBP-GFP [12] in the case of the whole serum approach, while equine ferritin, commonly available as one of the standards in the Gel Filtration Calibrants Kit for High Molecular Weight proteins (GE28-4038-42), for the HMW fractionation approach.

For the direct analysis of whole serum, the samples were diluted in Laemmli buffer, followed by denaturation for 5 min at 95°C and cysteine alkylation with acrylamide, and the total volume in all samples was subjected to in-gel digestion using the Short-GeLC for subsequent quantitative analysis by LC-MS/MS-DIA [13].

Samples subjected to HMW fractionation were ultrafiltrated using 300 kDa cut-off filters (Vivaspin® 500 Polyethersulfone, 300 kDa (Sartorius)) pre-conditioned to PBS. Serum samples were diluted into 200 µL of PBS and subjected to 20 min centrifugation at 14,500× g at 4 °C followed by an additional washing step with another 200 µL of PBS. In some cases, the washing step was repeated until the retentate volume did not exceed 50 µL. The

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

resulting retentates, the HMW fraction, were collected to a new LoBind® microcentrifuge tube and precipitated with ice-cold acetone [14]. The precipitated pellets were resuspended into 30 µL of a solution containing 2% SDS (v/v) and 1 M of Urea, always aided by sonication (VibraCell 750 watt-Sonics ®) with ice in the cup horn (2 min. pulse duration, at 1 second intervals, and with 40% amplitude). Afterward, concentrated Laemmli Buffer was added to the samples, followed by a 30 min incubation to reduce the samples and a 20 min incubation with iodoacetamide for cysteine alkylation [15]. The total volume in all samples was subjected to in-gel digestion as previously specified [13].

Mass spectrometry data acquisition

Samples were analyzed on a NanoLC™ 425 System (Eksigent®) couple to a TripleTOF™ 6600 System (Sciex®) using DDA for each fraction of the pooled samples for protein identification and SWATH-MS acquisition of each individual sample for protein quantification. Peptides were resolved by micro-flow liquid chromatography on a MicroLC column ChromXP™ C18CL (300 µm ID × 15 cm length, 3 µm particles, 120 Å pore size, Eksigent®) at 5 µL/min. The liquid chromatography program was performed as follows with a multistep gradient: 2 % to 5 % mobile phase B (0-2 min), 5 % to 28 % B (2-50 min), 28% to 35% B (50-51 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). Mobile phase A, composed of 0.1 % formic acid (FA) with 5% dimethyl sulfoxide (DMSO), and mobile phase B with 0.1 % FA and 5% DMSO in acetonitrile. Peptides were eluted into the mass spectrometer using an electrospray ionization source (DuoSpray™ Source, ABSciex®) with a 25 µm internal diameter hybrid PEEKsil/stainless steel emitter (ABSciex®). The ionization source was operated 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 the curtain gas (CUR).

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

For DDA experiments, the mass spectrometer was set to scan full spectra (m/z 350-1250) for 250 ms, followed by up to 100 MS/MS scans (m/z 100–1500) per cycle to maintain a cycle time of 3.309 s. The accumulation time of each MS/MS scan was adjusted in accordance with the precursor intensity (minimum of 30 ms for precursor above the intensity threshold of 1000). Candidate ions with a charge state between +2 and +5 and counts above a minimum threshold of 10 counts per second were isolated for fragmentation and one MS/MS spectrum was collected before adding those ions to the exclusion list for 25 seconds (mass spectrometer operated by Analyst® TF 1.7, ABSciex®). The rolling collision energy (CE) was used with a collision energy spread (CES) of 5.

For SWATH-MS-based experiments, the mass spectrometer was operated in a looped product ion mode [16] and the same chromatographic conditions were used as in the DDA experiments described above. A set of 60 windows of variable width (containing 1 m/z for the window overlap) was constructed, covering the precursor mass range of m/z 350-1250. A 250 ms survey scan (m/z 350-1500 m/z) was acquired at the beginning of each cycle for instrument calibration and SWATH-MS/MS spectra were collected from the precursors ranging from *m/z* 350 to 1250 for *m/z* 100–1500 for 20 ms resulting in a cycle time of 3.304 s. The CE for each window was determined according to the calculation for a charge +2 ion centered upon the window with variable CES according to the window.

Mass spectrometry data processing

A specific library of precursor masses and fragment ions was created by combining all files from the DDA experiments and used for subsequent SWATH processing. Libraries were obtained using ProteinPilot™ software (v5.1, ABSciex®), using the following parameters: i) search against a database from SwissProt composed by Homo Sapiens (released in March 2019), and MBP-GFP [15] and horse ferritin light and heavy chains sequences ii) acrylamide or

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

iodoacetamide alkylated cysteines, for whole serum or HMW respectively, as fixed modification; iii) trypsin as digestion enzyme and iv) urea denaturation as a special factor in the case of the HMW samples. An independent False Discovery Rate (FDR) analysis using the target-decoy approach provided with Protein Pilot software was used to assess the quality of the identifications, and positive identifications were considered when identified proteins and peptides reached a 5% local FDR [17, 18].

Data processing was performed using the SWATH™ processing plug-in for PeakView™ (v2.0.01, AB Sciex®) [19]. After retention time adjustment using a combination of IS and endogenous peptides, up to 15 peptides, with up to 5 fragments each, were chosen per protein, and quantitation was attempted for all proteins in the library file that were identified from ProteinPilot™ searches.

Protein levels were estimated based on peptides that met the 1% FDR threshold with at least 3 transitions in at least six samples in a group, and the peak areas of the target fragment ions of those peptides were extracted across the experiment using an extracted-ion chromatogram (XIC) window of 5 minutes with 100 ppm XIC width. Protein levels were estimated by summing all the transitions from all the peptides for a given protein (an adaptation of [20] and further normalized to the levels of the IS [15]).

The MS proteomics data have been deposited to the ProteomeXchange Consortium (22) via the PRIDE (23) partner repository with the dataset identifier PXD034077 (for review: Username: reviewer_pxd034077@ebi.ac.uk; Password: jhF8zQaR).

Statistical analysis and biological interpretation

Pearson's Chi-squared Test for Count Data was performed in R version 4.2.1, using the chisq.test function available in the native stats package in R to

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

determine if there were significant differences in the gender proportion of the groups within the studied cohort.

To assess the variation of the serum proteins (either Whole serum or the HMW fraction of the serum) among the three groups, a Kruskal–Wallis H test was followed by the Dunn's Test for pairwise comparison. Dunn's p-values were corrected using the Benjamini-Hochberg FDR adjustment, and statistical significance was considered for p-values below 0.05.

Stepwise Linear Discriminant Analysis (LDA) was performed to select the proteins responsible for the best separation of the groups being studied. LDA was performed using the software IBM® SPSS® Statistics Version 22 (Trial). LDA was attempted considering the proteins only altered at the Whole Serum or HMW Serum, and for the combination of both results. The evaluation of the models obtained from each analysis was performed by comparison of the Receiver operating characteristic (ROC) curves obtained using each model. ROC curves comparison was performed using the MedCalc Statistical Software version 20.106 (MedCalc Software Ltd; https://www.medcalc.org; Trial). The Delong et al. (1988) [21] method was used for the calculation of the Standard Error (SE) of the Area Under the Curve (AUC) and of the difference between two AUCs, and the Confidence Interval (CI) for the AUCs were calculated using the exact Binomial Confidence Intervals which are calculated as the following AUC ± 1.96 SE.

Violin plots were used to present the distribution of the individual protein levels among each condition, and Pearson’s correlation analysis was performed to evaluate the similarity between the profiles of the proteins highlighted in the study. Violin plots were generated using GraphPad Prism 8.0.1 (Trial) and the Pearson’s correlation was performed using Morpheus software (https://software.broadinstitute.org/Morpheus). Heatmap and hierarchical clustering analyses were computed by using PermutMatrix version 1.9.3

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

(http://www.atgc-montpellier.fr/permutmatrix/) [22] using the Euclidean distance and McQuitty's criteria.

Physical protein-protein interactions between the highlighted analytes were predicted by GeneMANIA webserver (Gene Function Prediction using a Multiple Association Network Integration Algorithm; https://genemania.org/) [23] together with a gene ontology (GO) analysis of the formed network. In addition to the proteins imported from this study, 28 additional related genes were allowed to create the interaction network using equal weighting by network. An additional GO enrichment analysis considering the term "biological process" was also performed. On the other hand, functional protein association networks were evaluated using the Search Tool for Retrieval of Interacting Genes/Proteins (STRING) version 11.5 (http://string-db.org/) with a medium confidence of 0.4 [24].

Pathway enrichment analyses were performed using the FunRich software (version 3.1.3) [25], considering two different databases: the FunRich or the Reactome database. In both cases, a statistically analyzed with a hypergeometric test using the FunRich human genome database as the background was performed. Enriched pathways were considered for a non-corrected, or a Bonferroni-corrected p-value below 0.05, for Funrich or Reactome database, respectively.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Results

High molecular weight fractionation as a complementary method to the conventional whole serum proteomics

To investigate the applicability of HMW fractionation to peripheral biomarker research, serum samples were subjected to either ultrafiltration using a 300 kDa MWCO filter and subsequent protein precipitation of the retentate (hereinafter referred as HMW serum) prior to protein digestion and MS analysis, or directly analyzed (whole serum) (Figure 1). In the proposed pipeline, the HMW fractionation is performed under non-denaturing conditions, such that the filter would retain native HMW protein complexes and large molecular structures.

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

As previously reported, the use of a proper IS is of utmost importance to uncover the effective proteome changes between different groups [12]; thus, considering the nature of the fractionation proposed in this work, an adequate IS for this procedure would be a HMW protein that is retained by the filter and has no similarity with the remaining human proteins. In this sense, two proteins with a MW higher than the MWCO of the filter (i.e., above 300 kDa) commonly used in a commercially available Gel Filtration Calibrants Kit for size exclusion chromatography were tested, and the equine globular protein Ferritin (~440 kDa size) proved to be a good IS for this procedure as it has no similarity with any other protein from the human proteome (Supplementary Figure 1a), the peptides monitored in the SWATH-MS analysis are easily distinguished from the matrix (the human serum proteome; Supplementary Figure 1b), and present a coefficient of variation similar to the one obtained by the MBP-GFP used in the unfractionated analysis and previously characterized [12] (Supplementary Figure 1c). Additionally, the overall reproducibility of the fractionation was also inspected, and

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

similarly to what was observed for the IS (Supplementary Figure 1c), the overall coefficient of variation of the proteins quantified using technical replicates (Supplementary Figure 1d) revealed that this procedure did not induce an appreciable increase in the variability of the quantification when compared with the conventional protocol (unfractionated samples). Moreover, the variation caused by the sample processing steps may be reverted by the normalization of the values to the IS, since the coefficient of variation of the IS is similar to the one observed for the majority of the proteins, indicating that the selected IS is a good predictor of the alterations induced by the method.

As proof of concept, this procedure was applied to serum samples of a cohort comprised of 3 different groups: AD (n = 22), PD (n = 24), and CT (n = 12). No statistically significant differences were found in the gender distribution between the three groups; however, differences were observed concerning age distribution, with PD individuals being slightly younger, on average, than both other groups (Table 1). Patients with neurodegenerative diseases, both AD and PD, were selected for this study, since these pathologies are commonly linked with the formation of abnormal protein complexes and protein aggregation, making these disorders suitable to test the fractionation approach as a means to investigate potentially altered protein interactions, which are lost in the conventional proteomics analysis.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

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#

* Determined by χ^2 test according to each group's male/female proportions. $ Determined by Kruskal-Wallis rank sum test to compare age between all groups. # Determined by Dunn's test (Benjamini-Hochberg correction) to compare age between group pairs. Diseases refer to AD and PD grouped together.

A total of 203 and 186 proteins were quantified in whole serum and HMW serum, respectively (Figure 2a, solid lines; Supplementary Tables 1 and 2 for detailed information). A large overlap was observed between both sample preparation procedures (168 proteins were shared, which corresponds to more than 70% of all the quantified proteins). Additionally, quantification in the whole serum of the proteins identified in the HMW serum library did not lead to a discernible increase in proteome coverage (Figure 2a, dashed line; Supplementary Table 3). Altogether, these results indicate that there are no major differences in terms of the proteins being quantified in the two approaches, revealing that the main aim of the HMW fractionation presented in this work – fractionation under non-denaturing condition – is not the overall improvement of the proteome coverage but the possibility of interrogating the samples considering protein interactions/macromolecular organization.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 2 – Comparative differential proteomic analysis of whole and HMW serum of AD and PD patients. (a) Venn diagram comparing the number of quantified proteins in each sample type using sample-specific libraries (solid lines; Supplementary Tables 1-2). In addition, quantification in whole serum was also performed for the proteins identified in the HMW specific library (dashed line; Serum&HMWLib condition; Supplementary Table 3) to evaluate the possible use of HMW fractionation as a tool to improve the proteome coverage of serum samples. A total of 224 proteins were quantified in the serum samples using the three strategies referred above, from those 162 of the proteins were commonly quantified independently of the strategy, corresponding to near three quarters of all the quantified proteins (72.3%). Only three new proteins were quantified in whole serum using the HMW-specific library. **(b)** Venn diagram comparing the total number of proteins considered as altered among the three experimental groups using the two different serum-processing strategies used in this work (Supplementary Figure 4 and Supplementary Tables 1-2). A total of 69 proteins were considered altered among the three experimental groups, of those only 11 proteins (the respective gene names are indicated) were consistently considered as being altered independently of the strategy used. **(c)** Comparison of the levels of the 11 proteins commonly considered as altered in both whole serum and HMW-fractionated serum. The proteins were arranged considering the group comparison where the statistical differences were observed (Supplementary Figure 5), and the alterations were presented as the median fold change observed in each sample-type. Proteins considered altered in more than one group are indicated in italic and with a grey shadow. Only one protein, the beta-2-glycoprotein 1 (indicated in bold), presented a divergent tendency when considering its values in the whole serum versus after the HMW-fractionation of the samples. # - non-statistically significant difference.

This was further confirmed by the fact that only 14.7% of all the proteins considered altered in this study (11 out of 69 proteins) were consistently

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

altered in the fractionated and unfractionated samples (Figure 2b; Supplementary Tables 1-2). Moreover, taking into consideration the changes of those 11 commonly altered proteins, it is possible to observe that all proteins, except for the protein beta-2-glycoprotein 1 (Figure 2c, bold), presented the same tendency in both fractionated and unfractionated samples (Figure 2c). These results, in combination with the previous observations, demonstrate that, although the two approaches quantified the same set of proteins, each method interrogated the samples in a particular context, resulting in the identification of a different set of altered proteins and the possibility to identify different regulatory mechanisms: while the direct analysis of the serum mainly represents the alteration at the protein level, the HMW fractionation under non-denaturing conditions may allow the evaluation of the physical interaction of the proteins.

These results are further supported by the analysis of the MW distribution of the proteins being studied, which reveals a similar profile between both approaches (Supplementary Figure 2a, 2b, and 2c), with most of the proteins detected in HMW serum presenting a MW below 150 kDa. Moreover, a similar distribution is observed for the proteins altered exclusively in the HMW approach (Supplementary Figure 2d), indicating that this strategy is not biased towards only the HMW proteins and supporting the idea that those proteins altered in this fraction may correspond to proteins being organized into different complexes.

To test the hypothesis that the HMW approach is capable of evaluating the re-organization of protein complexes, the nature of the 28 proteins altered exclusively in HMW serum was evaluated. This analysis revealed that most of those proteins have been described to interact physically either with each other or with other proteins (Figure 3a). Upon immediate observation of the GeneMania network, which primarily evaluates the interactions between the proteins under study (Supplementary Table 4), it becomes apparent that the majority of the proteins participate in established physical interactions.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Moreover, it can be observed that these proteins form a large, interconnected network comprising 20 of the 28 altered proteins, centered around the interactions between the apolipoproteins and lecithin-cholesterol acyltransferase (encoded by the *LCAT* gene). Although not participating in any known interactions with another protein from the 28 altered proteins, the proteins Glutathione peroxidase 3 (*GPX3*), Pigment epithelium-derived factor (*SERPINF1*), Thyroxine-binding globulin (*SERPINA7*), and Alpha-1B-glycoprotein (*A1BG*) are also known to establish physical interactions including with the proteins identified in this analysis. Only four of the 28 altered proteins, namely SRR1-like protein (*SRRD*), carnosine dipeptidase 1 (*CNDP1*), peptidoglycan recognition protein 2 (*PGLYRP2*), and serum amyloid A-4 protein (*SAA4*), were found to have no disclosed interactions in this particular analysis. However, interactors for those proteins were already pointed out in some screening assays, as confirmed in BioGRID (Biological General Repository for Interaction Datasets, Supplementary Table 5). Furthermore, as revealed by the functional analysis, several of these 28 proteins are involved in the formation of complexes with lipids and platelet components (Supplementary Table 6), indicating that those proteins can form complexes not only via the interaction with other proteins but also with other molecules, and thus be organized in large complexes. This involvement in the potential formation of macromolecular complexes is even more evident when the functional pathways enriched in each of the two lists of proteins (30 and 28 altered proteins exclusively in the whole serum or HMW serum, respectively) are directly compared (Figure 3b). This comparison highlights the fact that all the pathways that are either only, or at least more, enriched in the HMW dataset in comparison to the whole serum dataset (pathways indicated in bold) are related to the formation/regulation of large complexes/macrostructures, namely amyloids, fibrin clot formation and dissolution pathways and lipoprotein-related metabolism. On the other hand, the pathways that are particularly enriched or unique in the whole serum

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

dataset are mainly related to transcription factor networks (Supplementary Table 7).

Figure 3 – Characterization of the proteins exclusively altered in whole serum or HMW serum. (a) GeneMania Network of the 28 proteins altered only at the HMW serum sample (listed in larger circles). The analysis was performed with network weighting equal by the network, allowing a maximum of 28 extra resultant genes (non-listed small circles). The seven most enriched GeneMania Functions were highlighted in the network (color code; complete results in Supplementary Table 4). Only protein-protein physical interactions (red edges) were considered in this analysis, demonstrating that most of these 28 altered proteins have known interactors and can be involved in the formation of large protein-protein complexes. **(b)** FunRich Biological Pathways enriched in the whole serum (30 proteins; Supplementary Table 5) and HMW serum (28 proteins; Supplementary Table 6) proteomes. All GO analyses considered a $p < 0.05$. Pathways uniquely enriched at the HMW serum or particularly enriched in this type of sample when compared to the whole serum are indicated in bold.

Evaluation of the potential of this combined strategy for biomarker discovery

The previous set of results demonstrates that both approaches can provide complementary information. In line with this evidence, the potential to use this combined strategy for biomarker discovery was also evaluated. In general, both approaches result in nearly 40 proteins being altered in at least one pair of comparisons (Figure 4a-b; Supplementary Figure 3 for details regarding each pair of comparisons), with a tendency to have more proteins being altered in the comparisons involving the AD group (at least 20 proteins being altered compared to CT against a maximum of 14 altered proteins in PD vs. CT, Supplementary Figure 4a-b) and only a small subset of proteins being altered in only one comparison (15/41 in the whole serum and 11/39 in HMW serum, Figure 4a and 4b, respectively). Besides those

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

similarities, different profiles of altered proteins were observed depending on the approach used. Hence, it can also be highlighted that, while the whole serum strategy (Figure 4a) mainly found proteins altered between the two disease groups (39 proteins in AD vs. PD compared to 20 and 12 proteins in AD vs. CT and PD vs. CT, respectively), the HMW approach (Figure 4b) captured more differences between AD and CT samples (36 out of the 39 altered proteins). Due to this complementarity, the combination of results from both approaches resulted in a more comprehensive profile (Supplementary Figure 4a-c), with a general increase in the number of altered proteins per group (total of 69 altered proteins, Figure 4c). This improvement is particularly evident in the case of proteins altered between PD and CT samples, for which only one protein was considered commonly altered in the two approaches (Figure 2c and Supplementary Figure 4b), thus resulting in the duplication of the list of proteins with the potential to serve as biomarkers for PD versus CT individuals. Common to all mapped profiles was the low number of proteins altered between all three groups (4, 1, and 6 for Whole serum, HMW serum and the combination of both, Figure 4a-c, respectively) and the absence of proteins altered exclusively between PD and CT samples.

Figure 4 – Identification of potential circulating biomarkers of AD and PD. **(a-c)** Venn diagrams representing the distribution of the altered proteins among the different comparisons (AD vs. CT; PD vs. CT; AD vs. PD) considering the whole serum, HMW serum and the combination of the two (Whole + HMW serum) types of samples, respectively. **(d-f)** LDA using all the altered proteins per sample type or the combination of the two. The number of proteins used in each model is indicated above the graphic, and for each model it is indicated their specificity (percentage of healthy individuals correctly identified; indicated in black) and the sensitivity per disease condition (percentage of AD or PD patients correctly identified; indicated in red and blue, respectively). Specificity and sensitivity values were also summarized in Supplementary Table 7, and the LDA discriminant functions generated and their

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

respective statistical confidence, were summarized in Figure 5a. **(g-h)** Comparative ROC curves of the discriminant functions generated using all the altered proteins per sample type or the combination of the two. An independent evaluation was performed for each disease group being studied. The AUC of the ROC curves, their 95% CI and the pairwise comparisons are summarized in **(i)**. CI was calculated as follows AUC ± 1.96 SE. n.s., non-significant alterations. * and # indicate a $p<0.05$ for statistically significant differences in comparison to the Whole Serum or HMW+Whole Serum in comparison to HMW Serum, respectively, using the method of Delong et al. (1988) [21].

To further confirm the biomarker potential of the altered proteins from the three strategies presented above, they were used as input to build discriminant models that could differentiate between the studied groups (Figure 4d-f). From each dataset, candidates whose combination resulted in the best possible discriminant model were automatically selected (detailed information regarding the methods in Table 2), resulting in three distinct and statistically valid models (all with $p < 0.0032$) capable of discriminating the three groups being studied. The whole serum dataset resulted in a reasonable model composed of 4 proteins (Figure 4d), with a median sensitivity (the capacity to classify the individuals from each disease group correctly) of 86.95% (sensitivity and specificity are summarized in Supplementary Table 8). On the other hand, the model created with six proteins from the HMW approach (Figure 4e) had lower performance, with a median sensitivity of only 80.3% (corresponding to 83.3% predicting capacity for PD samples and 77.3% for AD samples). Moreover, neither of the models was particularly good in the classification of CT samples, resulting in a specificity of only 66.7% and 75% in the whole serum and HMW serum models, respectively. Remarkably, the combined model (created from the dataset containing the altered proteins from both approaches - Figure 4f) clearly outperformed the two models based only on proteins from a single approach. For this combined model, a total of 10 proteins were selected and

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

integrated, creating a discriminant model capable of correctly classifying more than 96% of all tested samples (93% in a cross-validation test; Table 2), including 100% correct classification of PD samples (Figure 4f). Altogether, this combined method presented a median sensitivity of 97.75% and a specificity of 91.7%.

The diagnostic capacity of these models was further evaluated by ROC curves of the capacity to positively classify the AD and PD patients against all the remaining samples (Figure 4g-i). This analysis confirmed that the best model is the one created with the combination of proteins from both approaches and that, in general, the model using only proteins from the whole serum approach is better than the model from the HMW approach. The respective statistics (Figure 4i) further support that the whole serum model performed better than the HMW serum model but without statistically significant differences between the two ROC curves. Additionally, the statistical analysis also confirmed that the combined model (AUC = 0.999 for the classification of AD and PD patients) is the best model, and that it performed significantly better ($p < 0.05$) than both other models for PD classification (HMW serum, AUC = 0.888; whole serum, AUC = 0.960) and better than the HMW serum model (AUC = 0.919) in the case of AD classification. The robustness of the combined model is further evidenced by the confidence interval (CI) of the AUC, which has a lower limit above 0.93 for both diseases, in contrast with the values 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.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

By looking at the proteins selected to build the different methods (Table 2), it was observed that the combined method is not the simple combination of the proteins previously selected from each of the individual methods. The combined model is built by the combination of 10 proteins, five from each dataset, including three [forkhead box protein M1 (*FOXM1*), immunoglobulin lambda variable 3-19 (*IGLV3-19*) and pigment epithelium-derived factor (*SERPINF1*)] that were not selected on the database-specific models. On the other hand, some previously selected proteins [serotransferrin (*TF*) and platelet basic protein (*PPBP*)] were not included in the combined model. Finally, the Hemoglobin subunit beta (*HBB*) was selected in both approach-specific models, although only data from the whole serum dataset was used in the combined model. These results demonstrate that the increase in the initial amount of data provided for the discriminant analysis has an important impact on the generated model by making it possible to test different combinations of proteins and, thus, allowing for the identification of better combinations than those highlighted in the analysis of individual datasets. Interestingly, all ten proteins selected in the combined model have a MW below 90 kDa (Table 2), confirming that all the proteins selected from the HMW approach have a MW below the theoretical cut-off of the filters used for fractionation, which supports the hypothesis that this approach may be capable of evaluating the remodeling of molecular complexes. By plotting the individual values of each of the ten proteins selected in the combined model (Figure 5a), it is possible to observe that, as expected, those values present some variation characteristic of the individuality of each patient. Nevertheless, considering that the model was able to correctly classify more than 90% of all the patients (Figure 4f), it is possible to infer that the combinations performed in the model could diminish the impact of the biological variability, proving that the combination of different markers can overcome their individual weaknesses. The analysis of these plots immediately reveals that: i) only three proteins from the model (the proteins encoded by the genes *FOXM1*, *HBB*, and *SRRD*) are significantly altered between

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

all three groups; and that ii) only one protein, the apolipoprotein C1 (Apo-CI, encoded by the *APOC1* gene), is altered between a single comparison, in this case between AD and PD which may indicate that this protein may have a particularly important role in this model for distinguishing AD from PD patients. Among the remaining 6 proteins: i) three are altered between both disease groups and control sample (all three found in the HMW fraction); ii) two are altered in AD in comparison to both PD and CT, and one, the Vitamin K-dependent protein C (encoded by the PROC gene), is altered in PD patients in comparison to the other two groups. Moreover, it is noteworthy that while there was a tendency to incorporate proteins that were increased in AD compared to CT from the whole serum approach (three out of the five proteins from the whole serum model encoded by the genes *FOXM1*, *HBB*, and *APOA1*), the opposite trend was observed in the case of proteins from the HMW approach. Specifically, three out of the five proteins (the proteins encoded by the genes *SRRD*, *APOE*, and *SERPINF1*) were found to be decreased in the AD vs. CT comparison. On the contrary, for the PD vs. CT comparison there were no major differences in terms of tendencies when considering the proteins captured in whole serum or the HMW serum.

Finally, all proteins from the whole serum dataset, in addition to SRR1-like protein (encoded by the *SRRD* gene) and Apo-CI from the HMW dataset, were significantly altered between both disease groups. From these, three proteins [the protein Forkhead box protein M1 (encoded by the *FOXM1* gene), hemoglobin subunit beta (encoded by *HBB* gene), and apolipoprotein A1 (Apo-AI, encoded by the *APOA1* gene)], are less abundant in PD samples than in AD samples, while the remaining four are increased.

Figure 5 – Characterization of the proteins selected by the LDA model created by the combination of Whole Serum and HMW Serum analyses and their correlation. (a) Violin plots representing the group's distribution of the levels of each of the ten proteins from the model. The dashed lines inside the violin plots indicate the first, second (median) and third quartiles. *,

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

, and * indicate a $p<0.05$, $p<0.01$, and $p<0.001$ for statistically significant differences in comparison to the control group. #, ##, and ### indicate a $p<0.05$, $p<0.01$, and $p<0.001$ for statistically significant differences in comparison between disease groups. Statistical analysis was performed using the Kruskal–Wallis H test followed by the Dunn's Test for pairwise comparison. **(b)** Person's correlation analysis between the overall regulation profile of the ten proteins included in the model. **(c)** Heatmap and hierarchical clustering analysis of the 10 proteins from the model. Clustering was performed for both the proteins and the individuals analyzed in this study. Three different clusters (Cluster PD, CT and AD) containing the large majority of the individuals from a given group can be highlighted from the analysis. The average profile of each cluster is indicated on the right and can be considered as the profile of expression of those ten proteins within the groups considered in this study. **(d)** The interaction network of the ten proteins included in the model was carried out with STRING with medium confidence (0.4) score. The color of the edges indicates the type of evidence that supports a given interaction, while the color of the nodes represents the categorization of the proteins considering UniProt Keywords enriched in this dataset (complete functional enrichment analysis in the Supplementary Table 8). The calculated PPI enrichment p-value is 2.07e-05. Three clusters (Cluster 1 to 3) can be identified within the network, with the dashed edges indicating the separation between the two clusters. Cluster 1 corresponds to proteins whose interactions are experimentally confirmed, while cluster 2 is composed of proteins that are theoretically related, and finally, cluster 3 corresponds to non-related proteins. **(e)** Reactome pathways enrichment analysis using the ten proteins included in the diagnostic model. The analysis was performed by FunRich functional enrichment analysis. The red line indicates Bonferroni corrected p-value with the corrected $p < 0.05$, meaning a significant enrichment. The grey dashed line indicates the reference line ($p = 0.05$). The complete analysis can be found in Supplementary Table 9.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

The correlation analysis of the protein abundances among groups, confirmed that, in general, there is no particularly evident correlation between the profiles and the magnitude of regulation of these proteins (Figure 5b). There were, however, some observed exceptions, including a strong positive correlation between the proteins encoded by the genes *APOA1* and *FOXM1* (r=0.8) and, to a lesser extent, the proteins encoded by *APOC1*, *APOE and SERPINF1* (r=0.5-0.6). Interestingly, the proteins exhibiting a positive correlation originated from the same approach: i) the gene products of *APOA1* and *FOXM1* were both highlighted in the whole serum approach, indicating that the total levels of these two proteins were modified in the same way: while, ii) the products of the *APOC1*, *APOE*, and *SERPINF1* genes were found to be altered in the HMW fractionation strategy, which may indicate that these proteins could be involved in the same complex and consequently regulated similarly. No remarkable negative correlations were found, with the strongest being observed between the proteins encoded by the *APOA1* and *SRRD* genes, which indicates that none of the proteins in the model present a completely opposite regulation profile. Additionally, an unsupervised clustering analysis using these ten proteins (Figure 5c) confirmed their capacity to partially distinguish the three groups being studied, revealing that, besides the existence of individual variability, it was possible to identify three independent clusters composed exclusively or mainly of samples from one of the three groups. This analysis also demonstrates that this set of proteins is particularly efficient in isolating the AD patients from the remaining individuals from the study: the AD cluster was composed exclusively of AD patients and only 6 out of the 22 AD patients were not included in this cluster. On the other hand, a slightly lower separation capacity was observed for both PD and CT samples. These two clusters contained few samples that did not belong to their respective groups, resulting in a higher percentage of individuals not properly grouped (10 out of 24 and 4 out of 12 samples for PD and CT, respectively). The discrepancies observed between the clustering analysis and the discriminant model results, where the latter

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

correctly classified over 90% of the samples, can be attributed to the fact that the clustering analysis relied solely on the individual protein distribution profiles across the samples. In contrast, the discriminant analysis employed equations with different weightings for each protein, resulting in a single model that effectively reduces the intragroup variability while promoting a better separation between the analyzed groups. Despite that, the clustering analysis remains an important approach for understanding how the proteins are modulated within the samples. Thus, from the three different clusters highlighted in the analysis, it was possible to infer the median protein abundance profile of these proteins among the three groups. For instance, the gene products of *SERPINF1*, *APOE* and *APOC1* tend to be less abundant in both disease groups compared to CT samples. Furthermore, some proteins are more abundant in each disease group, namely the gene products of *PROC* and *KRT9* in PD samples and the gene products of *HBB*, *APOA1* and *FOXM1* in AD samples. Another disease-specific observation was the smaller amount of immunoglobulin lambda variable 3-19 (*IGLV3-19*) in AD samples compared to both other groups. Overall, these tendencies characterize the unique profiles determined for each disease group, which may be a precursor to a potential future biomarker panel, more informative than the analysis based on any single protein.

Finally, STRING analysis (Figure 5d and Supplementary Table 9) revealed that these ten proteins have more interactions among themselves than what would be expected for a random set of proteins of the same size and degree of distribution, indicating that this set of proteins is, at least partially, biologically connected (PPI enrichment $p = 2.07\times10^{-5}$). This result may be mainly due to the strong network involving apolipoproteins and Hemoglobin subunit beta (cluster 1). Again, two out of the ten proteins selected for the discriminant method revealed to be associated with high-density lipoproteins (HDL) and chylomicron (ultra-low-density lipoproteins particles) remodeling and assembly (Figure 5e and Supplementary Table 10),

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

158 highlighting the potential importance of these mechanisms in the
159 neurodegenerative process, and confirming that the proteins related with
160 these mechanisms could be good biomarker candidates for their diagnosis.

161 Given the central role that apolipoproteins appear to play in this model, a
162 discriminant analysis was performed using data from ten altered proteins
163 involved in apolipoprotein-related mechanisms to investigate if the
164 diagnostic model could be limited to this set of functionally related
165 proteins (Supplementary Figure 5). However, the generated model exhibited
166 lower diagnostic capacity compared to the combined approach, with only 82.75%
167 of the samples being correctly classified. The model showed a specificity of
168 66.7%, and a sensitivity of 86.4% and 87.5% for AD and PD, respectively,
169 resulting in ROC curves with AUCs equal to or below 0.955. Thus, besides the
170 importance of apolipoproteins, the results from these proteins alone are not
171 enough to distinguish the three groups, which emphasizes the importance of
172 having diagnostic models based on several complementary candidates instead
173 of a single or just a few candidates. Nonetheless, the identification of
174 this robust core of functionally related proteins underscores the
175 significance of the combined approach for identifying new potential
176 biomarkers, since the dysregulation of Apo-C1 and Apolipoprotein E (Apo-E,
177 encoded by the *APOE* gene) was discovered using the HMW fractionation approach
178 while the dysregulation of Apo-AI was identified using the whole serum
179 approach.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Discussion

The present study presents a proof of concept of a novel two-pronged approach to biomarker discovery in complex peripheral biological fluids. More specifically, it was demonstrated that through the combination of two complementary proteomics strategies, the direct analysis of whole serum and the analysis of serum HMW fraction (above 300 kDa) in non-denaturing conditions, another level of proteome characterization of the samples could be achieved resulting in more robust diagnostic models. In this sense, when applied to serum samples from a cohort of control individuals and individuals afflicted by neurodegenerative diseases, this strategy allowed for a strong discriminant model to be built, able to distinguish all studied groups more effectively than the models generated from a single proteomics analysis. The most noticeable findings from this model showed that several, otherwise overlooked, proteins may yet serve as potential biomarkers of disease, in this case, AD and PD, particularly when analyzed together in a model created using the two different approaches. Thus, these results confirm the importance of having a panel of potential candidates rather than a single protein biomarker, and besides that, it also demonstrates that the biomarker discovery field will also benefit from combining data from the sample obtained through different sample processing strategies. Although not sufficient to be considered as a biomarker by itself, the substantial influence of apolipoproteins, namely Apo-AI, Apo-CI, and Apo-E, in the aforementioned discriminant model points out for a possible disease-specific dysregulation of the lipoprotein metabolism in AD and PD patients.

HMW fractionation may reveal a potentially altered macromolecular and macromolecular-complex organization

In this work it was demonstrated that interrogating serum samples with the HMW fractionation method adds an extra layer of information capable of bringing new insight into the behavior of the serum proteins, particularly

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

regarding their potential macromolecular organization. Because the fractionation procedure took place under non-denaturing conditions and since aberrant protein aggregation [8] is a common hallmark of both AD and PD, it was hypothesized that macromolecular complexes, potentially altered between the studied groups, could be captured through the HMW fractionation approach.

The present results confirmed this premise, as evidenced by the fact that although no variation was observed in the overall serum protein captured by both strategies, proteins exclusively altered in the HMW fraction accounted for 40% of the total list of altered proteins. Furthermore, with the exception of one protein (Centrosome-associated protein CEP250), all proteins had a MW below 90 kDa, which is considerably lower than the 300 kDa cut-off filter used. It is worth noting that 72% of proteins altered in the HMW fraction did not exhibit alterations in their total levels. This supports the possibility that different regulatory mechanisms of these proteins, apart from expression and degradation, are being revealed and studied using this approach.

Moreover, the results show that most of these proteins have several reported interactors and thus may be involved in the formation of large complexes. An example is the protein clusterin (*CLU* gene), also known as apolipoprotein J, which has been reported to be involved in the metabolism of aggregation-prone proteins, such as those involved in NDs [26-28]. For instance, the interaction of clusterin with Aβ42 has been shown to increase its clearance from the brain through the blood-brain barrier (BBB) [26]. Moreover, clusterin has already been pointed out as being related to different stages of PD disease, including a potential neuroprotective role arising from its interaction with α-synuclein aggregates [27]. Additionally, the interaction of clusterin with α-synuclein has already been detected in plasma samples [28]. Besides this specific example, the generic functional analysis of the altered proteins, particularly those from the HMW serum strategy, reveals

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

that those protein are highly related with the amyloids and clot formation, which can be large structures.

On the other hand, some of these proteins may instead, or additionally, be present in large biological structures not composed exclusively of proteins, like exosomes or lipoproteins, which would not only likewise justify their presence in this HMW serum fraction but also give further understanding of the potentially altered mechanisms related to the diseases being studied. Such may be the case of the proteins clusterin and serum amyloid A-4 (*SAA4* gene), which were found to be altered in serum neuron-derived exosomes of AD patients [29]. Additionally, exosomal clusterin was found to be altered in patients at different stages of PD when compared to controls [30]. Thus, although the presence of exosomes in the HMW fraction was not confirmed, given the MWCO of the filters used in this work, it is feasible that some of the proteins being analyzed in the HMW fraction may correspond to proteins linked to the extracellular vesicles.

Altogether, these findings support the notion that the HMW fractionation approach can provide a new level of information that may provide new insights into how proteins are organized within a given sample.

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

Importantly, the combination of the two approaches in this study led to a robust and promising potential biomarker panel composed of ten proteins, quantified in whole serum or HMW serum. A major finding revealed by this model was the involvement of several lipoproteins in discriminating the studied groups. Among the ten proteins used in the best discriminant model, three are apolipoproteins: Apo-AI from whole serum, and Apo-CI and Apo-E from HMW serum. Besides those three proteins, other altered apolipoproteins were observed in this study but not included in the model, namely: i) the Apo-AII, Apo-LI and clusterin highlighted in the HMW serum strategy; ii) the Apo-AIV from the whole serum; and, iii) beta-2-glycoprotein 1 in both

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

approaches. Furthermore, a lipoprotein-related enzyme, lecithin-cholesterol acyltransferase, was also altered in HMW serum. This is further supported by the functional enrichment analysis of the altered proteins discovered in both strategies that highlight the involvement of those proteins in lipoprotein metabolism and HDL-mediated lipid transport pathways. Cumulatively, all these findings point to the relevance of lipoproteins in the context of NDs and, although not absolutely clear, the link between these diseases, in particularly AD, and apolipoproteins has been the focus of many studies [26-28, 31-40].

Both Apo-AI and Apo-E have an established relation to toxic species clearance from the brain in the context of AD and PD [26, 28, 31, 41-44]. Additionally, regarding AD, our findings for both proteins are contrary to what can be found in the literature [32, 37]. For Apo-AI, we found an increase in abundance in AD patients as opposed to the decrease reported for most studies [32]. However, in another study where no significant alterations in total serum Apo-AI content of AD patients were reported, further investigation revealed that some proteoforms of this protein were significantly increased compared to the levels observed in the controls [33]. This has been suggested as a possible explanation for the different observations regarding this protein in the context of AD, which might be related to the use of different detection methods within different studies [32]. Regarding this protein's connection to clearance mechanisms, evidence suggests that for HDLs containing Apo-AI (APOA1-HDL), the structure seems to influence not only the disaggregation of Aβ fibrils but also its ability to cross the BBB, with lipid-poor discoidal APOA1-HDL having the best performance when compared to APOA1-HDL in other lipidation states [31]. Moreover, in the present study, phosphatidylcholine-sterol acyltransferase, an enzyme known to affect HDL structure through lipidation of Apo-AI in plasma [39], was also found to be altered in AD patients. As this enzyme participates in HDL maturation in plasma [45], this result could reflect the dysregulation of lipoproteins in

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

AD. For Apo-E, we found a decrease in abundance in AD patients as opposed to the increase reported for MCI patients in a previous study [37]. Isoform and lipidation status of Apo-E is also crucial for the Aβ clearance, with the Apo-E4 isoform, the genetic risk factor most associated with the onset of AD [26, 42], and higher lipidation having detrimental effects on the process [26, 42]. Further investigation using the combined approach presented in this study, particularly in the context of AD, should also involve a lipid profile analysis and APOE genotyping of the participants to enable a more comprehensive characterization of serum HDLs and Apo-E content, respectively.

Although systemic lipid abnormalities have also been implicated in PD, there are much fewer findings connecting it to HDL-related proteins, as compared to AD [27]. Nonetheless, a previous report observed significantly decreased values of Apo-AI in mild PD patients when compared to healthy controls, but much like what we observed for this protein, a less impactful and non-significant decrease was observed in moderate/severe PD patients [46]. In fact, most research indicates that Apo-AI may have a protective role in PD [27] and it has been hypothesized that APOA1-HDL could take part in the efflux of α-synuclein from the brain [28]. Additionally, both Apo-AI and Apo-E have been reported to interact with α-synuclein [28].

Finally, HDL size and plasma levels have been shown to be dependent on the levels of Apo-CI [47], another apolipoprotein included in the best diagnostic model generated in this study. Besides that, a previous study also shows that the dysregulation of this protein can lead to impaired memory processes in mice [48]. This suggests that the regulation of Apo-CI can be pivotal in the brain and that a systemic disruption of this process could have effects detectable beyond the CNS, particularly in lipoprotein metabolism, that could be observable in peripheral biofluids. Interestingly, in this study, it was found that Apo-CI was significantly decreased in the HMW serum fraction of AD in comparison to PD patients, but only a slight and non-significant decrease was observed when compared to CT patients. Similarly, only a small

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

non-significant decrease in HMW serum Apo-CI was observed for controls in comparison to PD patients, which is in accordance to what was already reported for whole plasma [34]. Since impairment of memory is a hallmark of AD and is not a predominant feature among the PD patients [49], these results may be impacted, at least in part, by the age discrepancies observed between the patients from the PD group and the other two groups. In fact, the PD patients are on average younger than the other two groups that have a similar age-distribution among them. In this sense, the lower levels of Apo-CI in comparison to PD may be in part linked with some memory impairment associated to natural aging [50]. On the other hand, the results of Apo-CI may also be influenced by the individual Apo-CI and Apo-E isoforms. In fact, the *APOE* and *APOC1* genes are in linkage disequilibrium [51], and carriers of the *APOE(ε4)* and *APOC1(H2)* alleles have been shown to have an increased risk of developing AD [40]. This was further confirmed in an study using human *APOE*-carrying mice, which demonstrated that those animals carrying the *APOE(ε4)* allele were found to have decreased serum Apo-CI content when compared to those carrying the *APOE(ε3)* allele [52]. However, Apo-CI has also been suggested to potentially play a modulatory role in the development of AD, with reported effects on mice cognitive function independent of Apo-E expression [53]. Again, these observations strengthen the importance of combining these results with further characterization of the individuals, including genotyping of the apolipoproteins' isoforms.

Despite of the connection between these three apolipoproteins to AD and PD, as evidenced by previously mentioned studies, and their relevance for the discriminant model, the use of these proteins alone or in combination with other proteins associated with apolipoprotein-related mechanisms did not result in a robust diagnostic model capable of effectively distinguishing between the studied groups (as shown in Supplementary Figure 5). This indicates that these three apolipoproteins had to be combined with other seemingly unrelated proteins to be used as potential biomarkers. Further

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

studies should be directed towards elucidating this potential relationship to understand: i) the importance of the identified proteins/mechanisms for the pathophysiology of the studied NDs, and ii) to which extent these mechanisms are differently altered between the two diseases.

In summary, in this work it was demonstrated that the combination of two complementary sample processing approaches is a more effective strategy to reach potential biomarkers rather than a single approach. Besides that, the strategies used here, that combine the analysis of the whole serum and the HMW fractionation of non- denaturing serum, can also identifying proteins being differentially modulated besides the conventional alteration in their total levels. In this work, this new strategy was applied to a cohort of NDs patients and respective CT individuals, being able to build a good predictive model capable of distinguishing all the three groups studied (AD, PD and CT). This predictive model highlighted the linkage of the apolipoprotein family and NDs, with three out of the ten proteins included in this model being apolipoproteins. Nevertheless, further validation in a larger and independent cohort is needed to confirm the soundness of the model, as well as more studies to link the alterations observed and these pathologies. Controlling the lipid profile of each individual included in future studies is also advised, as altered lipid metabolism was a major finding of the present work. Another interesting aspect to be further explored would be the identification of protein complexes in the HMW fraction to better understand the origin of the protein alterations observed. This is particularly relevant since many of the proteins captured in this fraction have a molecular weight that would normally exclude them from analysis through the 300kDa cut off fractionation approach. Overall, the results of this study demonstrate that HMW fractionation under non-denaturing conditions could be a valuable addition to routine biofluid analysis.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Acknowledgments

The authors would like to thank the financial support of the European Regional Development Fund (ERDF), through the COMPETE 2020 - Operational Programme for Competitiveness and Internationalisation and Portuguese national funds via FCT – Fundação para a Ciência e a Tecnologia, I.P., under projects: POCI-01-0145-FEDER-016428 (ref.: SAICTPAC/0010/2015), POCI-01-0145-FEDER-30943 (ref.: PTDC/MEC-PSQ/30943/2017), PTDC/MED-NEU/27946/2017, POCI-01-0145-FEDER-016795 (ref.: PTDC/NEU-SCC/7051/2014), POCI-01-0145-FEDER-029311 (ref.: PTDC/BTM-TEC/29311/2017); EXPL/BTM-TEC/1407/2021, and, UIDB/04539/2020, UIDP/04539/2020 and LA/P/0058/2020, and the National Mass Spectrometry Network (RNEM) [POCI-01-0145-FEDER-402-022125 Ref. ROTEIRO/0028/2013). SIA was supported by the CEEC grant 2021.04378.CEECIND and MR was supported by Ph.D. fellowship 2020.07749.BD.

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

References

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

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

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

[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: 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;73:637-48.

[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;92:329-37.

[9] Association GAotWM. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. The Journal of the American College of Dentists. 2014;81:14.

[10] 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;55:181-4.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

[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;7:263-9.

[12] Anjo SI, Simoes I, Castanheira P, Graos M, Manadas B. Use of recombinant proteins as a simple and robust normalization method for untargeted proteomics screening: exhaustive performance assessment. Talanta. 2019;205:120163.

[13] Anjo SI, Santa C, Manadas B. Short GeLC-SWATH: a fast and reliable quantitative approach for proteomic screenings. Proteomics. 2015;15:757-62.

[14] Santa C, Anjo SI, Manadas B. Protein precipitation of diluted samples in SDS-containing buffer with acetone leads to higher protein recovery and reproducibility in comparison with TCA/acetone approach. Proteomics. 2016;16:1847-51.

[15] Anjo SI, Melo MN, Loureiro LR, Sabala L, Castanheira P, Grãos M, et al. oxSWATH: An integrative method for a comprehensive redox-centered analysis combined with a generic differential proteomics screening. Redox biology. 2019;22:101130.

[16] Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, et al. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Molecular & Cellular Proteomics. 2012;11.

[17] Tang WH, Shilov IV, Seymour SL. Nonlinear fitting method for determining local false discovery rates from decoy database searches. Journal of proteome research. 2008;7:3661-7.

[18] Sennels L, Bukowski-Wills JC, Rappsilber J. Improved results in proteomics by use of local and peptide-class specific false discovery rates. BMC bioinformatics. 2009;10:179.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

[19] Lambert J-P, Ivosev G, Couzens AL, Larsen B, Taipale M, Lin Z-Y, et al. Mapping differential interactomes by affinity purification coupled with data-independent mass spectrometry acquisition. Nature methods. 2013.

[20] Collins BC, Gillet LC, Rosenberger G, Röst HL, Vichalkovski A, Gstaiger M, et al. Quantifying protein interaction dynamics by SWATH mass spectrometry: application to the 14-3-3 system. Nature methods. 2013.

[21] DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988:837-45.

[22] Caraux G, Pinloche S. PermutMatrix: a graphical environment to arrange gene expression profiles in optimal linear order. Bioinformatics. 2005;21:1280-1.

[23] Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Research. 2010;38:W214-W20.

[24] Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. 2017;45:D362-d8.

[25] Pathan M, Keerthikumar S, Ang CS, Gangoda L, Quek CY, Williamson NA, et al. FunRich: An open access standalone functional enrichment and interaction network analysis tool. Proteomics. 2015;15:2597-601.

[26] Bell RD, Sagare AP, Friedman AE, Bedi GS, Holtzman DM, Deane R, et al. Transport pathways for clearance of human Alzheimer's amyloid beta-peptide and apolipoproteins E and J in the mouse central nervous system. Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism. 2007;27:909-18.

[27] Berdowska I, Matusiewicz M, Krzystek-Korpacka M. HDL Accessory Proteins in Parkinson's Disease—Focusing on Clusterin (Apolipoprotein J) in Regard to Its Involvement in Pathology and Diagnostics—A Review. Antioxidants. 2022;11.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

[28] Emamzadeh FN, Allsop D. α-Synuclein Interacts with Lipoproteins in Plasma. Journal of Molecular Neuroscience. 2017;63:165-72.

[29] Arioz BI, Tufekci KU, Olcum M, Durur DY, Akarlar BA, Ozlu N, et al. Proteome profiling of neuron-derived exosomes in Alzheimer's disease reveals hemoglobin as a potential biomarker. Neuroscience Letters. 2021;755:135914.

[30] Kitamura Y, Kojima M, Kurosawa T, Sasaki R, Ichihara S, Hiraku Y, et al. Proteomic Profiling of Exosomal Proteins for Blood-based Biomarkers in Parkinson's Disease. Neuroscience. 2018;392:121-8.

[31] Dal Magro R, Simonelli S, Cox A, Formicola B, Corti R, Cassina V, et al. The Extent of Human Apolipoprotein A-I Lipidation Strongly Affects the β-Amyloid Efflux Across the Blood-Brain Barrier in vitro. Frontiers in neuroscience. 2019;13:419.

[32] Endres K. Apolipoprotein A1, the neglected relative of Apolipoprotein E and its potential role in Alzheimer's disease. Neural Regeneration Research. 2021;16.

[33] Fania C, Arosio B, Capitanio D, Torretta E, Gussago C, Ferri E, et al. Protein signature in cerebrospinal fluid and serum of Alzheimer's disease patients: The case of apolipoprotein A-1 proteoforms. PLOS ONE. 2017;12:e0179280.

[34] Paslawski W, Zareba-Paslawska J, Zhang X, Hölzl K, Wadensten H, Shariatgorji M, et al. α-synuclein–lipoprotein interactions and elevated ApoE level in cerebrospinal fluid from Parkinson's disease patients. Proceedings of the National Academy of Sciences. 2019;116:15226-35.

[35] Paula-Lima AC, Tricerri MA, Brito-Moreira J, Bomfim TR, Oliveira FF, Magdesian MH, et al. Human apolipoprotein A–I binds amyloid-β and prevents Aβ-induced neurotoxicity. The International Journal of Biochemistry & Cell Biology. 2009;41:1361-70.

[36] Qiang JK, Wong YC, Siderowf A, Hurtig HI, Xie SX, Lee VM-Y, et al. Plasma apolipoprotein A1 as a biomarker for Parkinson disease. Annals of Neurology. 2013;74:119-27.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

[37] Song F, Poljak A, Crawford J, Kochan NA, Wen W, Cameron B, et al. Plasma Apolipoprotein Levels Are Associated with Cognitive Status and Decline in a Community Cohort of Older Individuals. PLOS ONE. 2012;7:e34078.

[38] Swanson CR, Berlyand Y, Xie SX, Alcalay RN, Chahine LM, Chen-Plotkin AS. Plasma apolipoprotein A1 associates with age at onset and motor severity in early Parkinson's disease patients. Movement Disorders. 2015;30:1648-56.

[39] van der Vorst EPC. High-Density Lipoproteins and Apolipoprotein A1. In: Hoeger U, Harris JR, editors. Vertebrate and Invertebrate Respiratory Proteins, Lipoproteins and other Body Fluid Proteins. Cham: Springer International Publishing; 2020. p. 399-420.

[40] Zhou Q, Zhao F, Lv Z-p, Zheng C-g, Zheng W-d, Sun L, et al. Association between APOC1 Polymorphism and Alzheimer's Disease: A Case-Control Study and Meta-Analysis. PLOS ONE. 2014;9:e87017.

[41] Castellano JM, Kim J, Stewart FR, Jiang H, DeMattos RB, Patterson BW, et al. Human apoE Isoforms Differentially Regulate Brain Amyloid-β Peptide Clearance. Science Translational Medicine. 2011;3:89ra57-89ra57.

[42] Deane R, Sagare A, Hamm K, Parisi M, Lane S, Finn MB, et al. apoE isoform-specific disruption of amyloid beta peptide clearance from mouse brain. The Journal of clinical investigation. 2008;118:4002-13.

[43] Merino-Zamorano C, Fernández-de Retana S, Montañola A, Batlle A, Saint-Pol J, Mysiorek C, et al. Modulation of Amyloid-β 1-40 Transport by ApoA1 and ApoJ Across an in vitro Model of the Blood-Brain Barrier. Journal of Alzheimer's Disease. 2016;53:677-91.

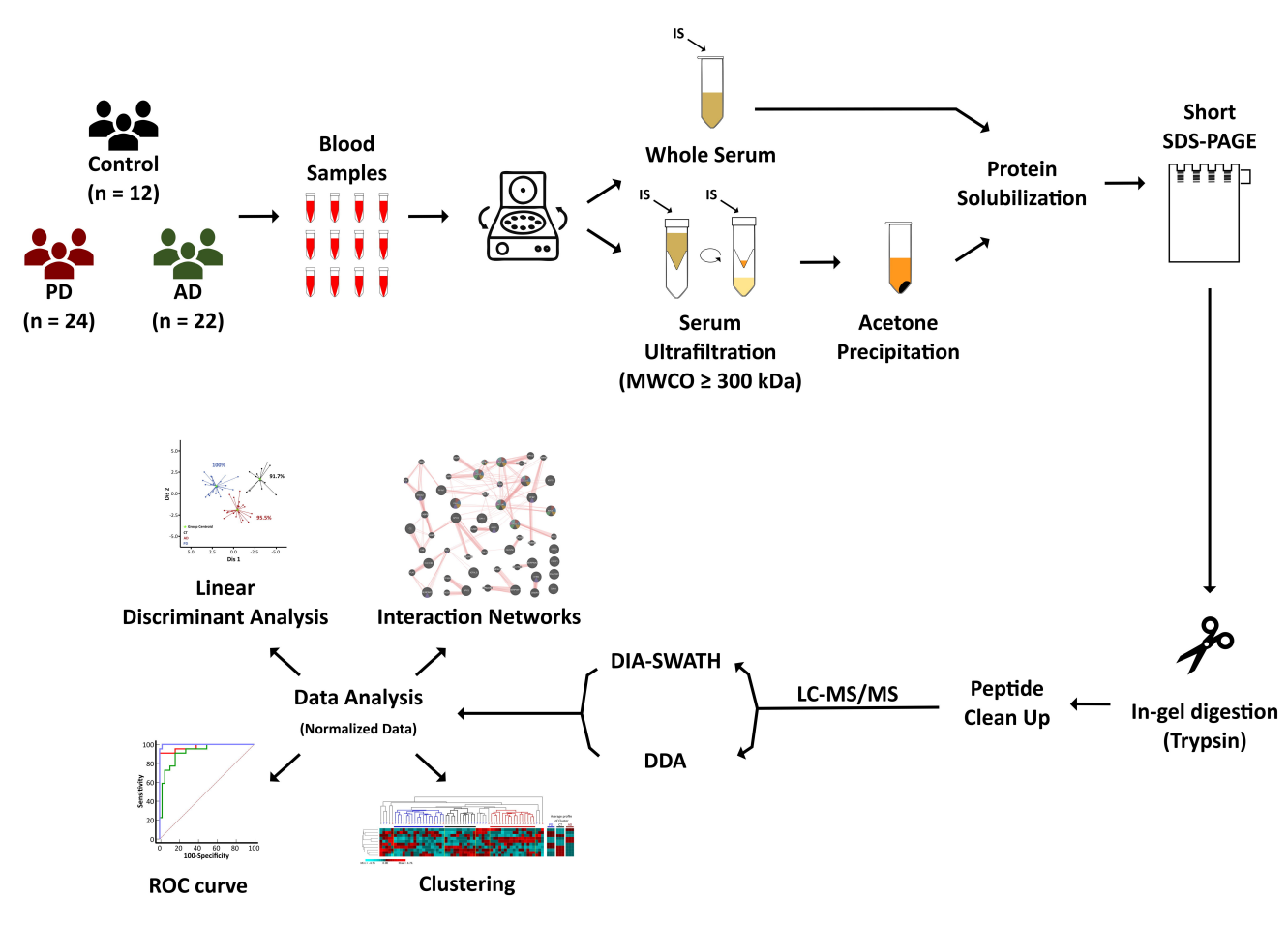
[44] Verghese PB, Castellano JM, Garai K, Wang Y, Jiang H, Shah A, et al. ApoE influences amyloid-β (Aβ) clearance despite minimal apoE/Aβ association in physiological conditions. Proceedings of the National Academy of Sciences. 2013;110:E1807-E16.

[45] Rye K-A, Clay MA, Barter PJ. Remodelling of high density lipoproteins by plasma factors. Atherosclerosis. 1999;145:227-38.

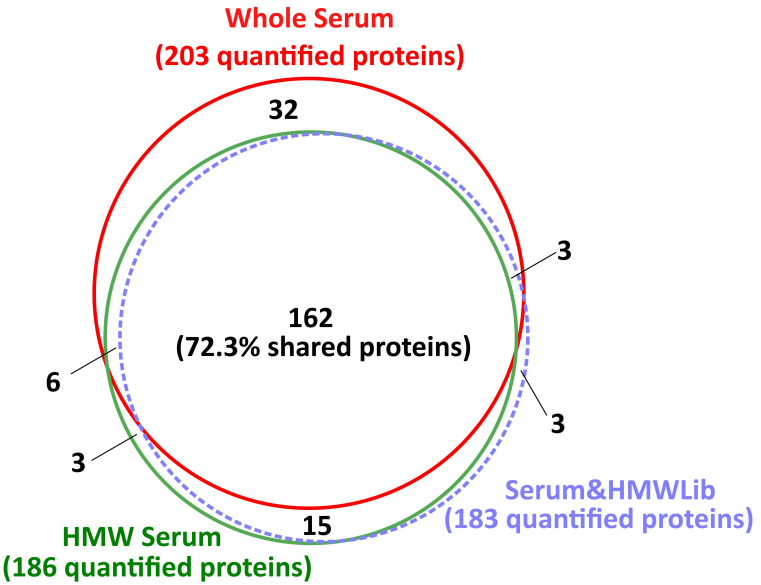
[46] Zhang X, Yin X, Yu H, Liu X, Yang F, Yao J, et al. Quantitative proteomic analysis of serum proteins in patients with Parkinson's disease using an

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.23.568435; this version posted November 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

552 isobaric tag for relative and absolute quantification labeling, two-
553 dimensional liquid chromatography, and tandem mass spectrometry. Analyst.
554 2012;137:490-5.
555 [47] de Haan W, Out R, Berbée JFP, van der Hoogt CC, Dijk KWv, van Berkel
556 TJC, et al. Apolipoprotein CI inhibits scavenger receptor BI and increases
557 plasma HDL levels in vivo. Biochemical and Biophysical Research
558 Communications. 2008;377:1294-8.
559 [48] Berbée JFP, Vanmierlo T, Abildayeva K, Blokland A, Jansen PJ, Lütjohann
560 D, et al. Apolipoprotein CI Knock-Out Mice Display Impaired Memory Functions.
561 Journal of Alzheimer's Disease. 2011;23:737-47.
562 [49] Albert AD, Brad R. Parkinson disease and cognitive impairment.
563 Neurology: Clinical Practice. 2016;6:452.
564 [50] Vance DE, Graham MA, Fazeli PL, Heaton K, Moneyham L. An Overview of
565 Nonpathological Geroneuropsychology: Implications for Nursing Practice and
566 Research. Journal of Neuroscience Nursing. 2012;44.
567 [51] Fuior EV, Gafencu AV. Apolipoprotein C1: Its Pleiotropic Effects in
568 Lipid Metabolism and Beyond. International Journal of Molecular Sciences.
569 2019;20:5939.
570 [52] Cudaback E, Li X, Yang Y, Yoo T, Montine KS, Craft S, et al.
571 Apolipoprotein C-I is an APOE genotype-dependent suppressor of glial
572 activation. Journal of Neuroinflammation. 2012;9:192.
573 [53] Abildayeva K, Berbée JFP, Blokland A, Jansen PJ, Hoek FJ, Meijer O, et
574 al. Human apolipoprotein C-I expression in mice impairs learning and memory
575 functions. Journal of Lipid Research. 2008;49:856-69.
576 [54] McDonnell T, Wincup C, Buchholz I, Pericleous C, Giles I, Ripoll V, et
577 al. The role of beta-2-glycoprotein I in health and disease associating
578 structure with function: More than just APS. Blood Rev. 2020;39:100610-.
579 [55] Gupta A, Watkins A, Thomas P, Majer R, Habubi N, Morris G, et al.
580 Coagulation and inflammatory markers in Alzheimer's and vascular dementia.
581 International Journal of Clinical Practice. 2005;59:52-7.

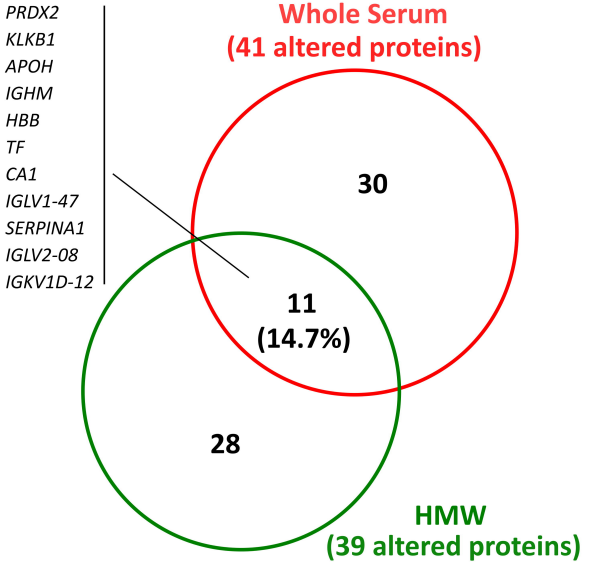

Control
(n = 12)
PD
(n = 24)
AD
(n = 22)
Blood
Samples
IS
Whole Serum
IS
IS
Serum
Ultrafiltration
(MWCO ≥ 300 kDa)
Acetone
Precipitation
Protein
Solubilization
Short
SDS-PAGE
In-gel digestion
(Trypsin)
Peptide
Clean Up
LC-MS/MS
DIA-SWATH
DDA
Data Analysis
(Normalized Data)
Linear
Discriminant Analysis
Interaction Networks
ROC curve
Clustering

(a)



Total: 224 quantified proteins

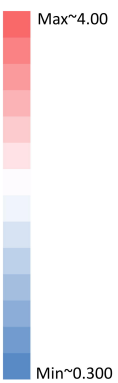
(b)

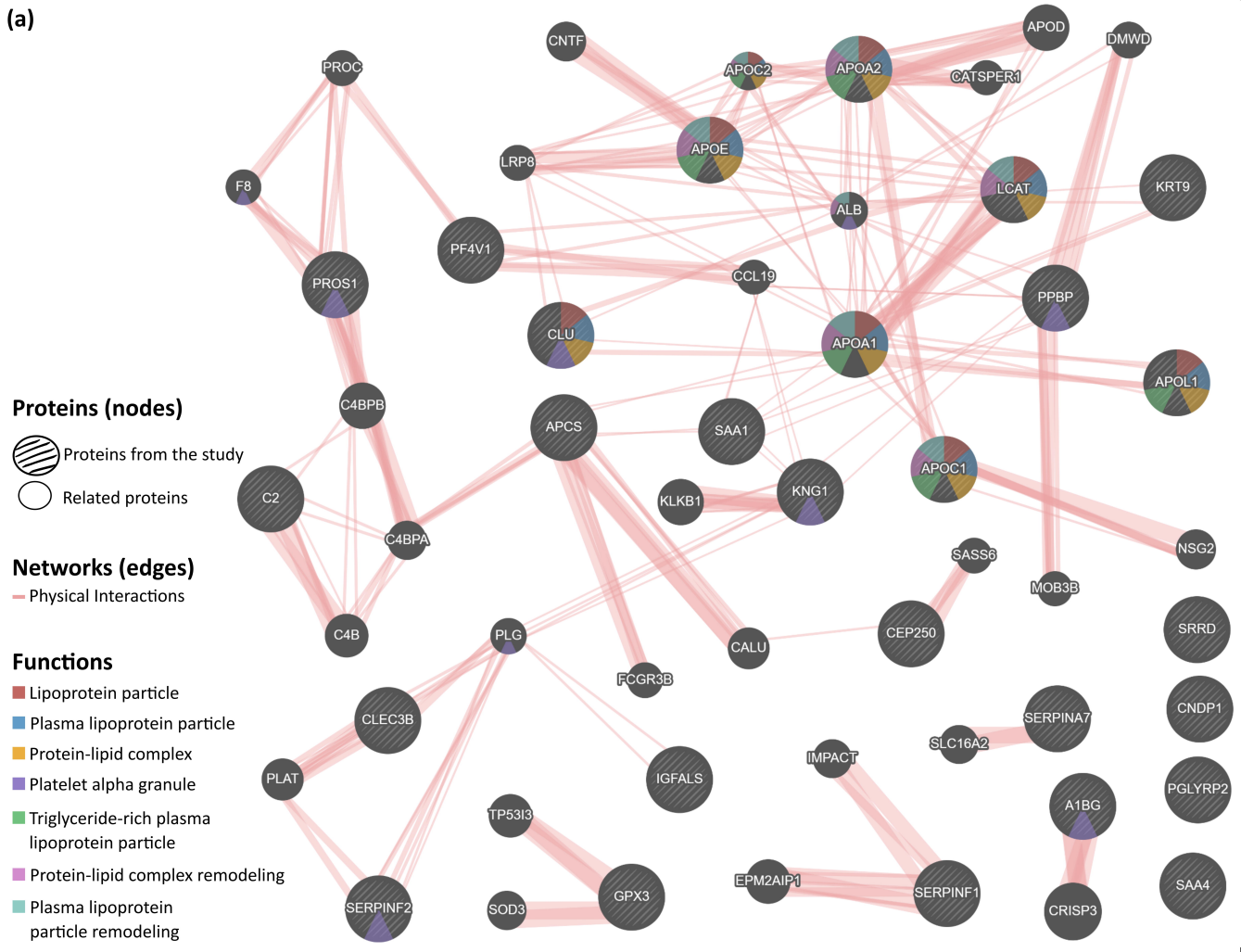


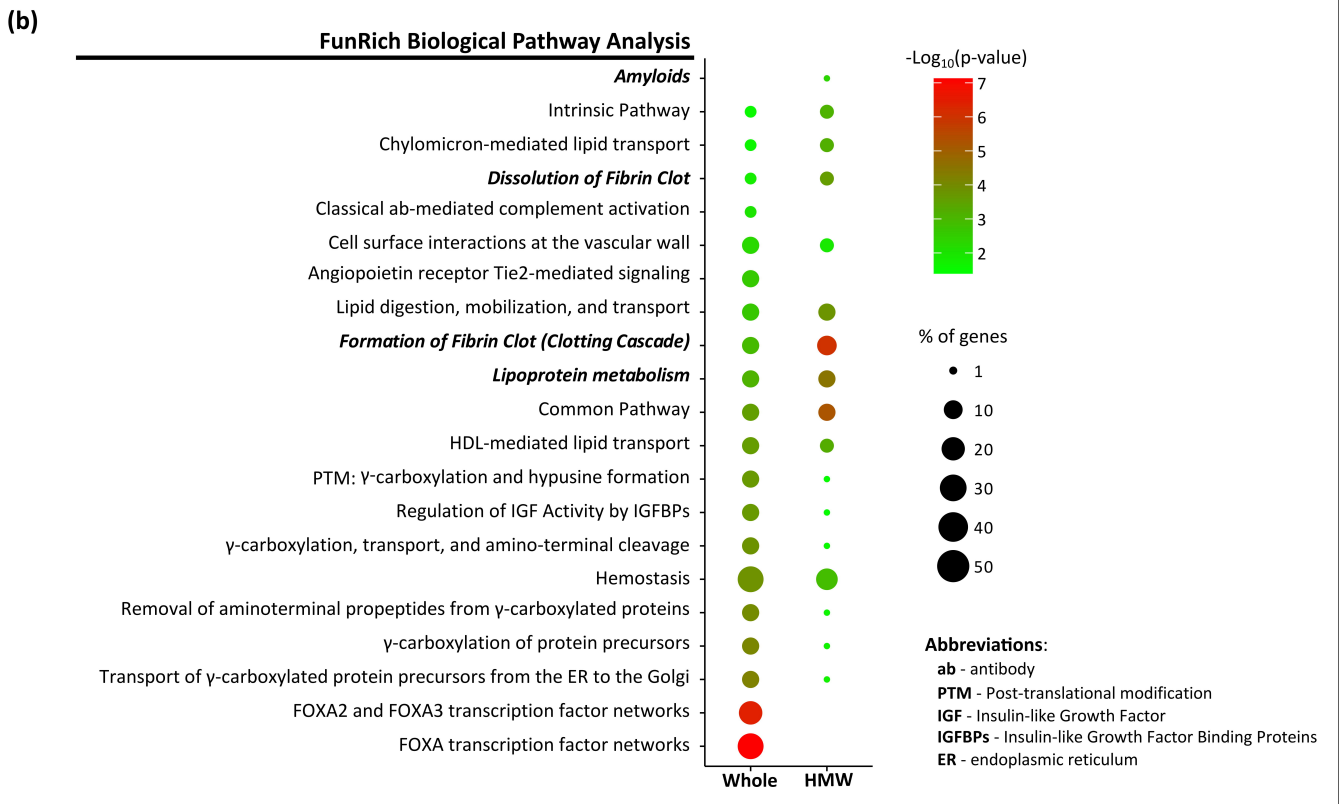
Total: 69 altered proteins

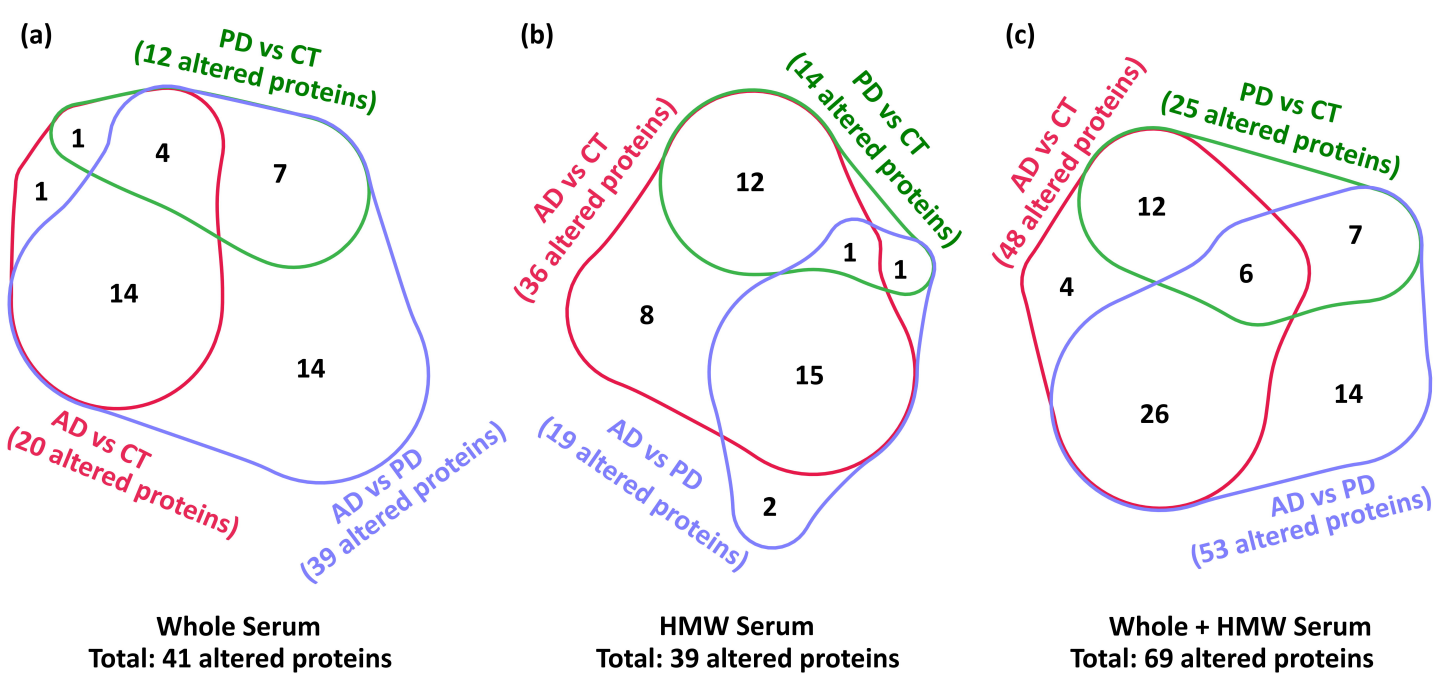
(c)

Group comparison	Gene Name	Protein Name	Fold Change Serum	Fold Change 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




(a)
APOD
DMWD
CNTF
PROC
APOC2
APOA2
CATSPER1
APOE
LRP8
F8
LCAT
KRT9
ALB
PF4V1
PROS1
CCL19
PPBP
CLU
APOA1
APOL1
C4BPB
APCS
SAA1
APOC1
C2
KLKB1
KNG1
NSG2
C4BPA
SASS6
MOB3B
C4B
PLG
CALU
CEP250
SRRD
FCGR3B
CNDP1
CLEC3B
SERPINA7
SLC16A2
IMPACT
PLAT
IGFALS
PGLYRP2
A1BG
TP53I3
EPM2AIP1
SERPINF1
SAA4
SERPINF2
SOD3
GPX3
CRISP3
Proteins (nodes)
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

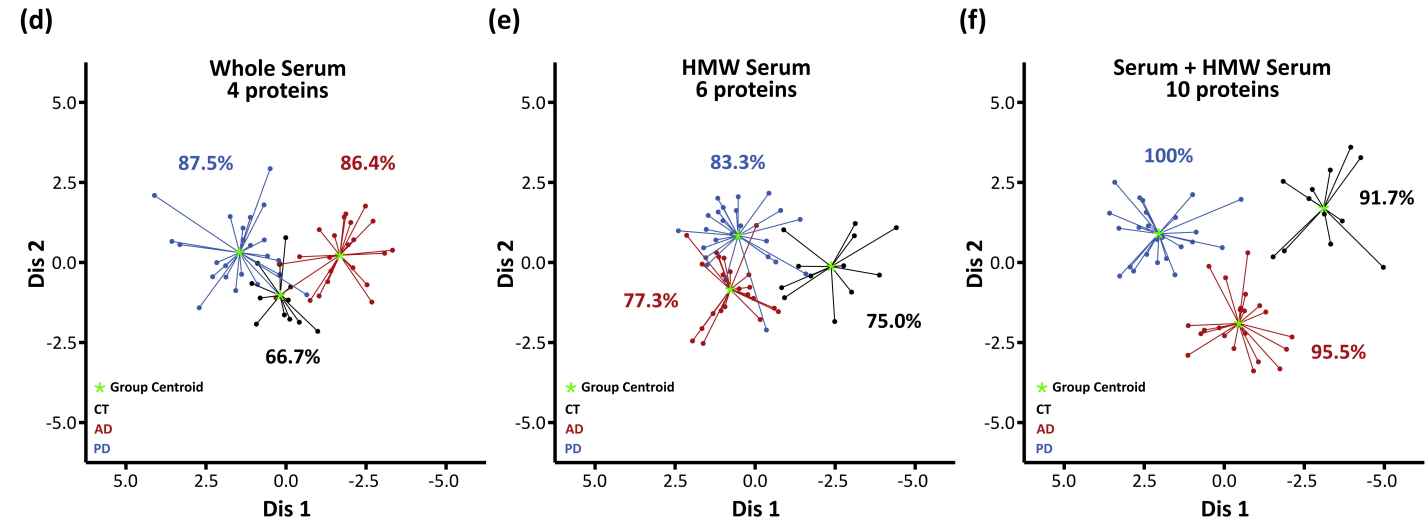

(b)
FunRich Biological Pathway Analysis
Amyloids
Intrinsic Pathway
Chylomicron-mediated lipid transport
Dissolution of Fibrin Clot
Classical ab-mediated complement activation
Cell surface interactions at the vascular wall
Angiopoietin receptor Tie2-mediated signaling
Lipid digestion, mobilization, and 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
Whole
HMW
-Log10(p-value)
7
6
5
4
3
2
% of genes
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

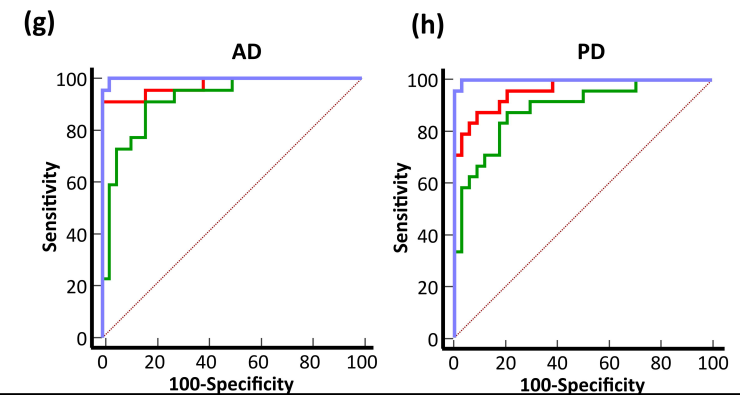


Whole Serum
Total: 41 altered proteins

HMW Serum
Total: 39 altered proteins

Whole + HMW Serum
Total: 69 altered proteins

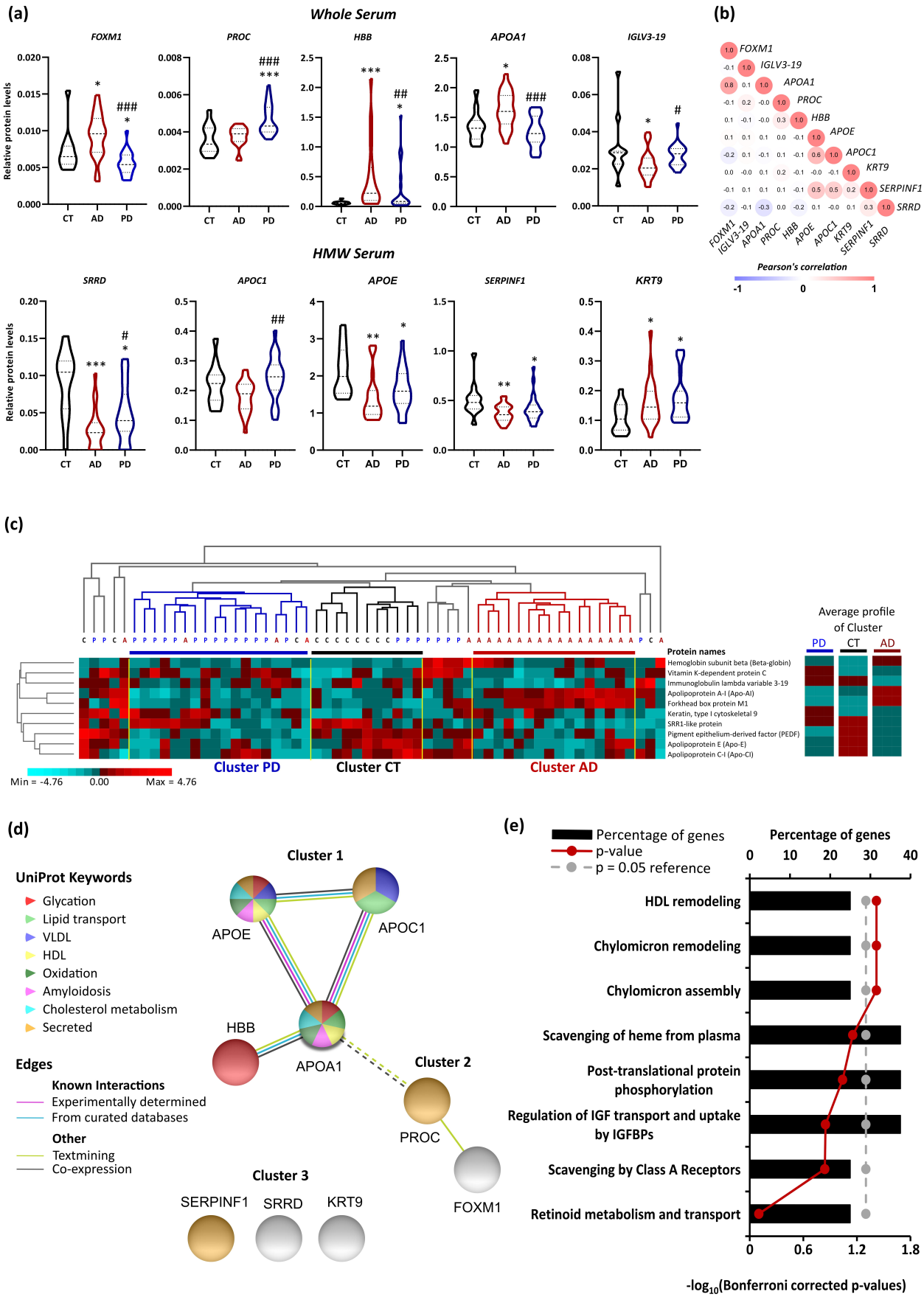




(i)

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
FOXM1
PROC
HBB
APOA1
IGLV3-19
HMW Serum
SRRD
APOC1
APOE
SERPINF1
KRT9
Relative protein levels
CT
AD
PD
(b)
Pearson's correlation
(c)
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
Cluster PD
Cluster CT
Cluster 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
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
-log10(Bonferroni corrected p-values)