Gerontology

Clinical Section: Research Article

Gerontology 2024;70:269–278
DOI: 10.1159/000534753

Received: April 30, 2023
Accepted: October 9, 2023
Published online: January 12, 2024

Serum Proteomic Signatures of Common Health Outcomes among Older Adults

Jackson A. Roberts[a, b] Sayantani Basu-Roy[a] Jong Shin[a] Vijay R. Varma[a] Andrew Williamson[a] Chad Blackshear[c] Michael E. Griswold[c] Julián Candia[d] Palchamy Elango[d] Ajoy C. Karikkineth[e] Toshiko Tanaka[d] Luigi Ferrucci[d] Madhav Thambisetty[a]

[a]Clinical and Translational Neuroscience Section, Laboratory of Behavioral Neuroscience, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA; [b]Columbia University Vagelos College of Physicians and Surgeons, New York, NY, USA; [c]University of Mississippi Medical Center, Jackson, MS, USA; [d]Longitudinal Studies Section, Translational Gerontology Branch, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA; [e]Clinical Research Core, National Institute on Aging, National Institutes of Health Intramural Research Program, Baltimore, MD, USA

Keywords
Proteomics · Aging · Diabetes · Health · Physical activity

Abstract

Introduction: In aging populations, the coexistence of multiple health comorbidities represents a significant challenge for clinicians and researchers. Leveraging advances in omics techniques to characterize these health conditions may provide insight into disease pathogenesis as well as reveal biomarkers for monitoring, prognostication, and diagnosis. Researchers have previously established the utility of big data approaches with respect to comprehensive health outcome measurements in younger populations, identifying protein markers that may provide significant health information with a single blood sample. ***Methods:*** Here, we employed a similar approach in two cohorts of older adults, the Baltimore Longitudinal Study of Aging (mean age = 76.12 years) and InCHIANTI Study (mean age = 66.05 years), examining the relationship between levels of serum proteins and 5 key health outcomes: kidney function, fasting glucose, physical activity, lean body mass, and percent body fat. ***Results:*** Correlations between proteins and health outcomes were primarily shared across both older adult cohorts. We further identified that most proteins associated with health outcomes in the older adult cohorts were not associated with the same outcomes in a prior study of a younger population. A subset of proteins, adiponectin, MIC-1, and NCAM-120, were associated with at least three health outcomes in both older adult cohorts but not in the previously published younger cohort, suggesting that they may represent plausible markers of general health in older adult populations. ***Conclusion:*** Taken together, these findings suggest that comprehensive protein health markers have utility in aging populations and are distinct from those identified in younger adults, indicating unique mechanisms of disease with aging.

Jackson A. Roberts and Sayantani Basu-Roy contributed equally to this work.

karger@karger.com
www.karger.com/ger

Correspondence to:
Madhav Thambisetty, thambisettym@mail.nih.gov

Karger

Introduction

Multimorbidity, the coexistence of multiple chronic illnesses, is a significant challenge for primary health care [1], often representing long-term and expensive health conditions which require complex and ongoing care. As aging populations experience chronic diseases, there is an urgent global need for an integrated approach to prevent illness, distinct from the more prevalent disease-specific outlook [2].

The use of big data collected from numerous sources including genomic sequencing, health records, and pharmaceutical research may improve healthcare services and the early identification of chronic diseases [3]. This approach has resulted in advances in data-driven disease monitoring, prevention, and diagnosis [4], as well as broad advancements in the field of personalized medicine [5].

A recent proof-of-concept study [6] examined the relationship between plasma proteins and chronic health indicators in a midlife cohort. The study identified a subset of proteins among ~5,000 quantified using the SOMAscan platform that were significantly associated with 11 health outcomes of current health state, modifiable behavioral factors, and future metabolic health risk. These results suggest that protein measurements from a single blood draw may provide a personalized, cost-efficient assessment of key health indicators at midlife.

Chronic diseases in older adults (above the age of 65 years) account for significant health expenditure in the USA and globally as populations continue to age [7]. Additionally, life course studies have indicated that risk factors at midlife may differ from those at older ages for a number of chronic diseases including Alzheimer's disease and cardiovascular disease [8, 9]. We, therefore, were interested in identifying proteins unique to an older adult cohort that were associated with the current health status and modifiable behavioral factors. The identification of proteins specific to older adults may provide greater insight into the physiologic processes of aging and is critical to developing noninvasive markers of chronic disease.

Using the 1.3k SOMAscan proteomic platform, we first identified serum proteins associated with health outcomes in older adults from the Baltimore Longitudinal Study of Aging (BLSA) (average age: 76.1 years; n = 148). Second, we examined the associations between plasma levels of these proteins and health outcomes in an independent older adult cohort, the Invecchiare in Chianti (InCHIANTI) study (average age: 66.1 years; n = 918). These shared proteins constituted the older adult protein health signature. Finally, we compared the significant proteins identified in both older adult cohorts with the proteins reported by Williams et al. [6] in a midlife cohort (average age: 51.6 years) to better understand those that were unique to an older adult cohort and those that were shared across the life course.

Materials and Methods

Older Adult Participants

The National Institute on Aging's BLSA is one of the longest-running scientific studies of human aging in the USA [10]. This prospective study was launched in 1958 and performs longitudinal, clinical, radiological, and laboratory evaluations on community-dwelling volunteer participants. The individuals in our study were participants (n = 148) in the neuroimaging substudy of the BLSA [11]. Written informed consent was obtained at each visit from all participants. The study protocol has ongoing approval from the Institutional Review Board (IRB) of the National Institute of Environmental Health Science, National Institutes of Health ("Early Markers of Alzheimer's Disease [BLSA]," IRB No. 2009-074). The neuroimaging substudy began in 1994. At the time of entry, participants were free of central nervous system disease, severe cardiac disease, pulmonary disease, or metastatic cancer.

The InCHIANTI study is a population-based epidemiological study aimed at evaluating factors that influence mobility in the older population living in the Chianti region in Tuscany, Italy. Details of the study have been previously reported [12]. For this analysis, data from the baseline visit conducted between 1998 and 2000 were used. All participants provided written informed consent, and the study protocol was approved by the Italian National Institute of Research and Care of Aging Institutional Review and approved by the Internal Review Board of the Intramural Research Program of the National Institutes of Health.

Blood Samples

Serum blood samples from BLSA participants were collected at the National Institute on Aging Clinical Research Unit in Harbor Hospital, Baltimore. Plasma blood samples from the InCHIANTI cohort were collected at site-specific study clinics at Greve in Chianti and Bagno a Ripoli. Additional details on sample collection and processing have been reported previously (BLSA – [13, 14]; InCHIANTI – [12, 15]). Blood samples in both BLSA and InCHIANTI were collected after an overnight fast. The midlife cohort in Williams et al. [6] used plasma blood samples, as described in the index publication. Because the two older adult studies used different tissue matrices (i.e., serum in BLSA and plasma in InCHIANTI), we examined correlations between plasma and serum samples from a subset (n = 10) of BLSA participants with proteomic measures in both matrices (described in the Statistical Analysis section).

Proteomic Quantification

Protein concentrations for 1,322 SOMAmers were assessed using the 1.3K SOMAscan assay at the Trans-NIH Center for Human Immunology, Autoimmunity, and Inflammation, National Institute of Allergy and Infectious Disease, National Institutes of Health (Bethesda, MD, USA). The SOMAscan assay

platform includes 1,322 SOMAmer reagents, of which 12 are hybridization controls, 5 are viral proteins, and 5 are nonspecifically targeted SOMAmers. We additionally excluded 17 proteins that were not measured in the Williams et al. [6] study. As a result, our analysis included 1,283 SOMAmers. SOMAmers are single-stranded deoxynucleotides that are selected from large random libraries due to their binding of specific molecular targets, including proteins. As described previously [16], the 1.3K SOMAscan platform was developed by selecting against proteins across broad biological subgroups, including 47 percent secreted proteins, 28 percent extracellular domains, and 25 percent intracellular proteins. The assay, therefore, serves as a broad, general-purpose assay for systems-based proteomics without specific a priori inclusion criteria.

The experimental procedure for assessment of protein concentrations and normalization of SOMAscan assay data has been previously reported [15–17]. In brief, targets were generated by Selected Evolution of Ligands by Exponential Enrichment, a method of identifying high-affinity binding targets from much larger sequence libraries. This method detects proteins across a dynamic range of eight orders of magnitude [18]. The SOMAscan assay uses SOMAmers to translate protein concentrations into measurable DNA signals, which are then quantified using the standard DNA detection procedures. This is achieved by affinity binding and biotin capture on streptavidin beads. The DNA concentrations obtained from this method are reported as relative fluorescence units, resulting from fluorescent SOMAmer hybridized to its complimentary probe on an Agilent array, and are directly proportional to the reported relative abundance of SOMAmer reagents. The process of normalizing the data includes hybridization, control normalization, median signal normalization, and calibration normalization [16].

Data collected in the BLSA and InCHIANTI samples used the 1.3K SOMAscan assay platform that has since been expanded to include ~5,000 proteins in the 5K SOMAscan assay platform, the platform used in Williams et al. [6]. 1,283/1,300 proteins measured by the 1.3K SOMAscan platform were also measured on the 5K SOMAscan platform. A full list of proteins measured across all three cohorts (BLSA, InCHIANTI, and Williams et al. [6]) – i.e., 1,283 proteins – is included in online supplementary Table 1 (for all online suppl. material, see https://doi.org/10.1159/000534753).

Health Outcomes

In the BLSA and InCHIANTI studies, we sought to examine health outcomes also assessed in Williams et al. [6]. In BLSA, 4 current health state outcomes were measured. The included health outcomes span multiple health domains and have low barriers to assessment in the outpatient setting. Kidney filtration (mL/min) was estimated from the glomerular filtration rate (eGFR) (above/below 60). Body composition, including percentage body fat (%) and lean body mass (kg), was measured using dual-energy X-ray absorption (DEXA). Fasting glucose was measured using the oral glucose tolerance test (OGTT). In BLSA, 1 additional modifiable behavioral factor was measured. Physical activity was estimated using a self-report questionnaire of the amount of time spent performing activities including exercise, housework, and social activity over the prior 2 years [19]. Energy expenditure in METs (kilocalories/kg/min) for all activities (normalized to a 24-h – i.e., daily value) was used as the physical activity outcome.

In InCHIANTI, the same 4 current health state outcomes were measured. Kidney filtration was measured using serum creatine assessed by a kinetic-colorimetric assay (Roche Diagnostics GmbH, Mannheim, Germany) and used to estimate eGFR [20]. Body composition – percentage body fat and lean body mass – were assessed using peripheral quantitative computed tomography (pQCT) using an XCT 2000 device (Stratec Medizintechnik, Pforzheim, Germany). The cross-sectional images obtained by pQCT were analyzed using BonAlyse software (BonAlyse Oy, Jyvaskyla, Finland). Fat and muscle area (mm^2) was assessed by the transverse scans at 66% of tibia length. Fasting glucose was measured using enzymatic colorimetric methods (Roche Diagnostics GmbH, Mannheim, Germany). In InCHIANTI, physical activity was obtained during a structured interview in the past year and assessed as a binary variable of sedentary (less than light exercise 2–4 h/week) versus non-sedentary.

Williams et al. [6] included all health outcomes included in BLSA and InCHIANTI including 4 current health state outcomes: kidney filtration (mL/min estimated glomerular filtration rate [eGFR], above/below 60), body composition – percentage body fat (DEXA), lean body mass (DEXA), and fasting glucose (OGTT). All outcomes were measured identically to BLSA. Fasting glucose in Williams et al. [6] was used to predict conversion to diabetes within 10 (see Statistical Analysis section). Williams et al. [6] also included 1 modifiable behavioral factor included in BLSA and InCHIANTI: physical activity (actigraphy and individually calibrated heart rates as kJ/kg/day). A comparison of outcomes used in BLSA, InCHIANTI, and Williams et al. [6] studies is included in online supplementary Table 2.

Statistical Analysis

The primary goal of this study was to examine plasma/serum proteins associated with health outcomes in two older adult cohorts and compare those to a midlife cohort to determine blood-based indicators of health status that are either unique to an older adult sample or shared across the life course (i.e., shared with the midlife cohort). To identify proteins in the older adult cohorts, we first used a univariate approach to identify proteins significantly associated with each health outcome in both BLSA and InCHIANTI. We then tested correlations between proteins assayed in both plasma and serum in a subset of the BLSA cohort in order to assess differences in protein levels driven by the tissue matrix (i.e., serum or plasma) used for the assays. We then determined significant proteins shared between both older adult cohorts, defined as the older adult protein signature. Finally, in order to identify proteins that were unique to an older adult cohort or shared across the life course, we compared the older adult protein signature to proteins identified by Williams et al. [6] in their midlife cohort.

BLSA: Univariate Analyses of Association between Serum Proteins and Health Outcomes

We identified serum proteins associated with 4 current health state outcomes and 1 modifiable behavioral factor for 1,283 proteins included in the analyses of samples from BLSA. For continuous outcomes, the relationship between the proteins and the health outcome of interest was tested using the Spearman's rank correlation test. For binary outcomes, we used logistic regression models. Statistical significance of proteins from models

was corrected for multiple comparisons using a Benjamini-Hochberg [21] false discovery rate (FDR)-adjusted $p < 0.10$. This threshold was used to be consistent with the Williams et al. [6] study, which used the same FDR threshold to rank proteins in univariate analyses and reduce dimensionality. Additionally, the resulting probability of false positivity when assessing overlapping proteins in two (BLSA and InCHIANTI) and three (BLSA, InCHIANTI, and Williams et al. [6]) independent cohorts (see shared protein analyses below) each with an FDR-adjusted $p < 0.10$ is 1 and 0.1%, respectively. In sensitivity analyses, we included age, gender, and serum sample storage time as covariates.

Correlation Analyses between Serum and Plasma Proteins

To assess the relationship between protein concentrations from two different tissue matrices, i.e., serum and plasma, we examined the correlation using the Spearman nonparametric test between concentrations of proteins assayed in both the serum and plasma. We used a subset of 10 BLSA participants with SOMAscan protein data in both plasma and serum from the same follow-up visit blood draw. We reported the number and percentage of proteins that were correlated at $r > 0.50$.

InCHIANTI: Univariate Analyses of Association between Plasma Proteins and Health Outcomes

We identified plasma proteins associated with 4 current health state outcomes and 1 modifiable behavioral factor for 1,283 proteins included in the analyses of samples from the InCHIANTI cohort. Analytic methods were identical to those used for BLSA univariate analyses. For continuous outcomes, the relationship between the proteins and the health outcome of interest was tested using the Spearman's rank correlation test. For binary outcomes, we used logistic regression models. Similar to BLSA and the Williams et al. [6] study, statistical significance of proteins from models was corrected for multiple comparisons using a Benjamini-Hochberg FDR-adjusted $p < 0.10$. In sensitivity analyses, we included age, gender, and serum sample storage time as covariates.

BLSA and InCHIANTI: Shared Proteins Defining the Older Adult Protein Signature

We identified significant proteins shared between BLSA and InCHIANTI across all 4 current health state outcomes and 1 modifiable behavioral factor. To determine the significance of shared proteins between BLSA and InCHIANTI, we used the R package SuperExactTest [22] and reported p values and fold enrichment (FE) similar to prior work [23]. The SuperExact Test computes the statistical distributions of a set of intersections using combinatorial theory to calculate efficiently exact probabilities. This method determines the number of intersecting proteins that would be expected if the proteins were picked at random and compares the null "expected intersection" to the "observed intersection" to generate a p value and FE (observed intersection/expected intersection). A significant p value suggests that the intersection is likely suggestive of a shared older adult protein signature versus a chance intersection. We additionally performed over-representation enrichment analysis for the proteins included in each older adult protein signature in WebGestalt [24], utilizing the Gene Ontology Biological Process. We applied Benjamini-Hochberg FDR correction and extracted the top 10 most significant gene sets for each health outcome, considering FDR-adjusted $p < 0.05$, the default setting for WebGestalt, as our significance threshold.

BLSA, InCHIANTI, and the Williams et al. Study: Unique and Shared Proteins Comparing the Older Adult Protein Signature to Midlife Proteins

To identify the proteins that may be unique and shared by comparing the older adult cohorts to the midlife cohort, we compared the older adult protein signature to the significant proteins identified in the Williams et al. [6] study. The Williams et al. cohort assayed 5,000 serum proteins including 1,283 proteins measured in the BLSA and InCHIANTI. We restricted analyses to only the 1,283 shared proteins.

We reported the number and percentage (significantly associated proteins/total proteins included in this analysis × 100) of proteins that were unique to the older adult cohorts or shared across all 3 studies for the shared health outcomes: 4 current health state outcomes and 1 modifiable behavioral factor. To determine the significance of the multiset interaction, we used the SuperExactTest [22] and reported p values and FE. For proteins unique to the older adult cohorts, we also reported the number and percentage of proteins that were shared across health outcomes. We additionally performed over-representation enrichment analysis for the proteins overlapping across all 3 cohorts for each health outcome in WebGestalt [24], utilizing the Gene Ontology Biological Process. We applied Benjamini-Hochberg FDR correction and extracted the top 10 most significant gene sets for each health outcome, considering FDR-adjusted $p < 0.05$, the default setting for WebGestalt, as our significance threshold. We used R Studio 1.2.5033 for all data analyses.

Results

BLSA Participants

The demographic characteristics of the BLSA participants ($n = 148$) are summarized in Table 1. The mean age of participants was 76.1 years (range: 67.6–84.6), 52.0% were female, and 77.8% were white.

InCHIANTI Participants

The demographic characteristics of the InCHIANTI participants ($n = 918$) are summarized in Table 1. The mean age of participants was 66.1 years (range: 21.0–94.0), 54.3% were female, and all individuals were white, with participants from 2 study sites, Greve and Bagno a Ripoli.

Williams et al. Participants

The demographic characteristics of participants included in the Williams et al. [6] study ($n = 16,984$) are summarized in extended data figures in the original study publication. The average age of participants across the 5 midlife cohorts contributing data to the 5 health outcomes included in this study was 51.6 years.

Table 1. Participant demographics

Demographic variable	BLSA (n = 148)	InCHIANTI (n = 918)
Female, n (%)	77 (52.02)	499 (54.36)
White, n (%)	115 (77.77)	918 (100)
Storage time, years, mean (SD)	6.63 (4.02)	18.39 (0.44)
Age, mean (SD), years	76.12 (8.52)	66.05 (15.27)
Age range, years	67.6–84.6	21.0–94.0

BLSA: Univariate Analyses of Associations between Serum Proteins and Health Outcomes

For the current health state outcomes, at an FDR threshold of $p < 0.10$, 306 proteins were significantly associated with kidney function, 252 with percentage body fat, 144 with lean body mass, and 366 with fasting glucose. For the modifiable behavioral factor, 15 proteins were associated with physical activity (online suppl. Table 3). In sensitivity analyses including covariates, the number of significant (FDR <0.10) proteins reduced from 252 to 53 for percent body fat, 366 to 358 for fasting glucose, 306 to 144 for kidney function, 144 to 3 for lean body mass, and 15 to 0 for physical activity (online suppl. Table 4).

Correlation Analyses between Serum and Plasma Proteins for BLSA Participants

Across the 1,283 proteins in 10 BLSA individuals providing serum and plasma samples at the same time point, 850/1,283 (66%) had a correlation >0.50. The distribution of correlations across proteins is included in online supplementary Figure 1.

InCHIANTI: Univariate Analyses of Association between Plasma Proteins and Health Outcomes

For the current health state outcomes, at an FDR threshold of $p < 0.10$, 353 serum proteins were found to be significantly associated with kidney function, 144 with percentage body fat, 234 with lean body mass, and 153 with fasting glucose. For the modifiable behavioral factor, 180 proteins were significantly associated with physical activity (online suppl. Table 3). In sensitivity analyses including covariates, the significant (FDR $p < 0.10$) number of significant proteins reduced from 252 to 45 for percent body fat, 153 to 102 for fasting glucose, 353 to 309 for kidney function, 234 to 87 for lean body mass, and 180 to 48 for physical activity (online suppl. Table 4).

BLSA and InCHIANTI: Defining the Older Adult Protein Signature via Protein Set Intersection Analysis

Across 5 health outcomes, the following number of proteins overlapped across BLSA and InCHIANTI (Table 2): 151 with kidney function, 56 with percentage body fat, 48 with lean body mass, 39 with fasting glucose, and 9 with physical activity. Online supplementary Table 5 lists the significant proteins included in the older adult signature for each health outcome. We tested the statistical significance of this overlap using the SuperExactTest [22]. This procedure computes the statistical distributions of multiset intersections using combinatorial theory and efficiently calculates their exact probabilities. We specified the background population of 1,283 proteins, representing the total number of proteins included in models. The overlap for kidney function was significant ($p < 0.00001$; FE = 1.79), indicating that the overlap of 151 proteins significantly exceeded the expected or null overlap of 84.2. The overlap of percent body fat was significant ($p < 0.0001$; FE = 1.98) indicating the overlap of 56 proteins significantly exceeded the expected overlap of 28.3. The overlap of lean body mass was significant ($p < 0.0001$; FE = 1.83), indicating that the overlap of 48 proteins significantly exceeded the expect overlap of 26.3. The overlap for fasting glucose was not significant ($p = 0.837$; FE = 0.89), indicating that the overlap of 39 proteins did not significantly exceed the expected overlap of 43.6. Lastly, the overlap of physical activity was significant ($p < 0.0001$; FE = 4.3), indicating that the overlap of 9 proteins significantly exceeded the expected overlap of 2.10 proteins.

We also performed over-representation analysis for overlapping proteins in each of the older adult proteomic signatures. For kidney function, over-representation analysis identified enrichment of biological processes including biological adhesion, locomotion, and signaling receptor activity (online suppl. Table 7a; FDR $p < 0.0001$). For percent body fat, over-representation analysis identified enrichment of multiple processes including the regulation of endopeptidase activity, inflammatory and defense responses, protein processing, and response to stimuli (online suppl. Table 7d, FDR $p < 0.0005$). For lean body mass, over-representation analysis identified enrichment of biological processes including the regulation of protein metabolism, regulation of peptidase activity, inflammatory response, and response to stimuli (online suppl. Table 7c; FDR $p < 0.0001$). For fasting glucose,

Table 2. SuperExactTest results for 2-way intersection of significant model proteins in BLSA and InCHIANTI

	BLSA proteins	InCHIANTI proteins	2-way intersection	Expected overlap	Fold enrichment	*p* value
Kidney function	306	353	151	84.2	1.79	<0.0001
Lean body mass	144	234	48	26.3	1.83	<0.0001
Fasting glucose	366	153	39	43.6	0.837	0.837
Percent body fat	252	144	56	28.3	1.98	<0.0001
Physical activity	15	180	9	2.10	4.3	<0.001

over-representation analysis identified enrichment of the MAPK cascade, regulation of stress response, and signal transduction regulation (online suppl. Table 7b; FDR $p < 0.010$). Lastly, for physical activity, over-representation analysis did not identify significantly enriched gene sets for the physical activity signature (online suppl. Table 7e).

BLSA, InCHIANTI, and Williams et al.: Unique and Shared Proteins Comparing the Older Adult Protein Signature to Midlife Proteins

Across 5 health outcomes, the majority of proteins significantly associated with health outcomes in both BLSA and InCHIANTI were unique to these two older adult cohorts and were not significantly associated with the same outcomes in Williams et al. [6] (online suppl. Fig. 2) study. For kidney function, 95% of proteins significant in both BLSA and InCHIANTI were unique to the older adult cohort, 65% for percentage body fat, 69% for lean body mass, 69% for fasting glucose, and 89% for physical activity.

Among the unique (i.e., not significantly associated with the same outcomes in Williams et al. [6]) older adult proteins, the majority of significant proteins for each health outcome were specific to that outcome. 218 unique proteins were significantly associated with at least one health outcome in the older adult cohorts only. Of these, 29 (13.3%) were associated with at least 2 health outcomes, 3 (1.4%) were associated with at least 3 health outcomes, and 1 (0.5%) was associated with 4 health outcomes (Table 3). The three proteins that were most frequently shared across health outcomes models (i.e., shared in at least 3 health outcomes) were (1) adiponectin, which was significantly associated with kidney function, lean body mass, percent body fat, and fasting glucose; (2) MIC-1, which was significantly associated with kidney function, percent body fat, and physical activity; and (3) NCAM-120, which was significantly associated with kidney function, percent body fat, and fasting glucose.

The shared proteins (i.e., multiset interaction) across all three studies were significant based on the SuperExactTest for all health outcomes (Table 4). We specified the background population of 1,283 proteins, representing the total number of proteins included in models across all three studies. Across all 3 cohorts, the overlap of kidney function proteins was significant ($p = 0.002$; FE = 5.33), with the 7-protein overlap significantly exceeding the expected null overlap of 1.31. The overlap of lean body mass was also significant ($p < 0.001$; FE = 15.9), with the 15-protein overlap exceeding the expected overlap of 0.94. The overlap of percent body fat was significant ($p < 0.0001$; FE = 8.89), with the 19-protein overlap significantly exceeding the expected overlap of 2.14. The overlap of fasting glucose was significant ($p = 0.0009$; FE = 2.99), with the 11-protein overlap significantly exceeding the expected overlap of 3.67. Lastly, the overlap of physical activity was significant ($p = 0.036$, FE = 27.7), with the 1-protein overlap significantly exceeding the expected overlap of 0.036. online supplementary Table 6 lists the significant proteins overlapping across all 3 cohorts. Over-representation analysis did not identify any significantly enriched gene sets at the FDR $p < 0.05$ level for any of the 5 health outcomes (online suppl. Table 8a–e). A description of the known biological functions and described associations with health outcomes for each protein is included in online supplementary Table 9.

Among the proteins shared between the older adult protein signature and the midlife proteins for the health outcomes (i.e., shared across all three studies) that were also significant based on the SuperExactTest, 4 were significant across more than 1 health outcome: cystatin C, FSH, myoglobin, and troponin I. Cystatin C was significant for kidney function and lean body mass, and FSH, myoglobin, and troponin I were significant for both lean body mass and percent body fat. The majority of overlapping proteins were specific to only one health outcome.

Discussion

The use of blood-based protein and metabolite data to improve healthcare services and identify preclinical risk factors has become an important part of the success of

Table 3. Shared significant proteins across health outcomes among unique older adult proteins

	Kidney function	Lean body mass	Fasting glucose	Percent body fat	Physical activity
Kidney function	100 (151)	**29**	17	**20**	**56**
Lean body mass	9.2	100 (48)	11	**28**	0
Fasting glucose	4.0	8.3	100 (39)	9.3	11
Percent body fat	7.3	**31**	14	100 (56)	**22**
Physical activity	3.3	0	2.8	3.7	100 (9)

Cell values indicate the percent overlap across outcomes between proteins identified as significant (FDR $p < 0.10$) only in both BLSA and InCHIANTI, i.e., significant proteins exclusive to older adult cohorts. Columns indicate the percentage of significant results for that outcome that were also significant for the outcomes listed in each row, i.e., column 2 row 3 indicates the significant proteins overlapping between lean body mass and kidney function, divided by the number of proteins significant in kidney function alone (so 9.2% of the 151 kidney proteins are also associated with lean body mass?). Percentage overlaps >20% are indicated in bold.

Table 4. SuperExactTest results for 3-way intersection of significant model proteins across all cohorts

	BLSA proteins	InCHIANTI proteins	Williams proteins	3-way intersection	Expected overlap	Fold enrichment	*p* value
Kidney function	306	353	20	7	1.31	5.33	0.0002
Lean body mass	144	234	46	15	0.94	15.9	<0.0001
Fasting glucose	366	153	108	11	3.67	2.99	0.0009
Percent body fat	252	144	97	19	2.14	8.89	<0.0001
Physical activity	15	180	22	1	0.036	27.7	0.036

preventive health efforts during aging [25]. Our study examined which proteins were significantly associated with health outcomes and physical activity in two independent older adult studies in comparison to a previously published midlife cohort. We also assessed proteins that were unique or shared in their associations with specific outcomes between the two cohorts of older adults versus individuals at midlife. Unique proteins likely reflect pathophysiological processes specific to aging, while shared proteins are likely important markers of health across the midlife to older adult life course.

We identified significant correspondence in protein associations across the two older adult cohorts for all health outcomes examined in this study, and most associations were unique to older adults and not shared with the midlife cohort. The majority of proteins unique to the older adult cohorts were significantly associated with one health outcome only, suggesting that they may reflect biological processes specific to these health outcomes during aging.

Identifying blood biomarkers unique to older adult cohorts allows for an improved understanding of biological processes that may specifically impact health in older adults. In this study, the outcome with the largest set of concordant protein associations across the older adult cohorts was kidney function, measured by eGFR. Interestingly, prior work has identified that proteomic alterations associated with kidney function occur to a greater extent than mRNA transcriptomic changes during aging [26]. It is unsurprising that kidney function as an outcome displayed the highest proportion of proteins unique to the older adult cohorts, as kidney function begins to decline more rapidly in elderly populations compared to those in middle age [27]. The high proportion of unique proteins in these older cohorts in other outcomes similarly may reflect that the perturbations associated with age-related decline are distinct from the mechanisms mediating health earlier in the lifespan.

We also identified a set of proteins that may represent more general markers of health in older populations. Three proteins (adiponectin, MIC-1, and NCAM-120) were

associated with at least 3 health outcomes in both older cohorts but not in the midlife cohort. Adiponectin, which was associated with kidney function, lean body mass, percent body mass, and fasting glucose, is an adipokine with central roles in fat metabolism and insulin sensitivity. In its canonical role, adiponectin works via two receptors to reduce atherogenesis, decrease inflammation, and sensitize cells to the effects of insulin [28]. Prior studies have identified a role for adiponectin in chronic kidney disease, as levels of this adipokine have been found to be elevated in chronic kidney disease and are related to the progression of disease [29]. Current evidence suggests that adiponectin, which offers an indication of metabolic and adiposity status, may therefore represent a link between obesity and kidney disease, most likely mediated by oxidative stress [30].

MIC-1, which was associated with kidney function, percent body fat, and physical activity in the older adult cohorts, is a member of the transforming growth factor-β superfamily and has been implicated in appetite regulation, metabolism, cellular survival, and immune tolerance [31]. Interestingly, MIC-1 has been found to be elevated significantly in some pathological states, such as cancer, but also has a potent ability to induce anorexia and cachexia [32]. Pertinent for aging populations, MIC-1 has previously been associated with reduced cognitive performance and cognitive decline and has shown an inverse relationship with white matter integrity in older adults [33, 34]. In the context of aging health, MIC-1, therefore, may play a role at the convergence of obesity, chronic inflammation, and energy homeostasis. Furthermore, it has already been proposed as a relevant biomarker for several diseases and a predictor of all-cause mortality [35].

Lastly, NCAM-120 was associated with kidney function, percent body fat, and fasting glucose in both older adult cohorts and is involved in neuron-neuron adhesion and neurite outgrowth, additionally serving as a receptor for viruses including rabies and Zika. The loss of NCAM-120 has previously been shown to result in long-term memory impairment [36], and it additionally has been determined to be a marker of skeletal muscle fiber denervation that increases with aging [37]. One study has shown that a 5-month resistance training regimen reduces NCAM-120 positivity in skeletal muscle in obese older adults, suggesting it may represent a theragnostic biomarker in older adults [38].

While most proteomic associations identified in this study were unique to the older adult cohorts, there was a subset of proteins that were shared with multiple health outcomes in the midlife cohort and therefore may represent markers of health that are consistent across the midlife to older adult life course. Cystatin C, associated with kidney function and lean body mass, is included in several novel GFR estimation calculations [39]. FSH, known primarily for its role in female reproduction, rises significantly in both males and females with aging and plays a role in fat accumulation and redistribution, leading to alterations in leptin and adiponectin signaling [40]. Myoglobin, with its role in oxygenation of cardiac and skeletal muscle, has previously been shown to demonstrate similar intra-individual variability in both young and elderly populations and is proportional to muscle mass in these groups [41, 42]. Troponin I, associated with lean body mass and percent body fat, is a commonly used biomarker during acute coronary events but has further been suggested as a marker of all-cause mortality regardless of age in prior large cohort studies [43]. Given their association with multiple health outcomes in two independent cohorts of older adults and in a cohort of midlife adults, these proteins may have utility as markers of general health status independent of adult age.

There are a number of limitations that are important to consider when interpreting the results of this study. We were not able to compare all proteins included in the Williams et al. [6] with the older adult cohorts because of differences in the number of proteins quantified by the SOMAscan assay used in the BLSA and InCHIANTI cohorts (1.3K SOMAscan) compared to the Williams et al. [6] study (5K SOMAscan). Additionally, there were important across study differences: (1) not all health outcomes were measured identically across studies; (2) while the BLSA and InCHIANTI cohorts were on average 25 years older than the Williams et al. [6] participants, the age ranges for these cohorts overlapped; (3) there were storage time differences between BLSA and InCHIANTI that could impact study results; however, prior work has identified only minimal effects of storage time on SOMAscan proteins and a lack of an interaction effect with other participant characteristics [44, 45]; and (4) the sample matrix varied across the older adult cohorts (i.e., serum in BLSA and plasma in InCHIANTI). We attempted to address serum and plasma differences in protein measurement by correlating protein levels in the serum and plasma of the same individuals in a subset of the BLSA study. These correlations suggested that while protein levels were highly correlated between serum and plasma samples for the majority of proteins, there are differences that may impact defining an older adult protein signature. Finally, this study is limited in our ability to quantify only a subset of more than 20,000 known human proteins, suggesting that there are likely additional unquantified proteins relevant to the health outcomes reported here. Additional work with completely unbiased proteomics techniques would strengthen and improve upon our findings. Importantly, though our analysis identifies small sets of proteins associated with multiple domains of health

status, it remains to be seen how these protein signatures relate to clinical endpoints of disease. Future work examining the utility of these proteins in relation to the existing clinical risk estimators would better characterize their capacity to serve as clinical biomarkers.

The strengths of this study include analysis of 1,283 proteins quantified with the same proteomics platform in 3 independent cohorts at two different stages of age. We quantified these proteins in two well-characterized cohorts of older adults, BLSA and InCHIANTI, to derive unique signatures of proteins associated with 5 different health outcomes measured in both cohorts. Comparing these associations to those identified in Williams et al. [6], we derived protein signatures unique to older individuals as well as proteins that may represent more general markers of health across the adult lifespan. In our analyses, we applied a conservative FDR threshold of $p < 0.10$, and we additionally tested the significance of overlapping proteins across cohorts using the multiset intersection testing from SuperExactTest [22]. This approach allowed for a parsimonious selection of a small number of high-confidence proteins that may serve as the basis for further research.

In summary, we have compared the findings of Williams et al.'s [6] study of comprehensive markers of health in a middle-aged population to those of two independent cohorts of older adults, deriving protein signatures both unique to older populations and shared across age groups. The finding that most associated proteins shared across the older adult cohorts were not identified in the midlife cohort suggests that many of the proteomic alterations impacting health at older ages possibly reflect distinct pathophysiological mechanisms associated with aging.

Acknowledgments

We are grateful to BLSA and ROS participants for their invaluable contributions. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Statement of Ethics

All BLSA participants provided written informed consent. The BLSA study protocol has ongoing approval from the Institutional Review Board (IRB) of the National Institute of Environmental Health Science, National Institutes of Health ("Early Markers of Alzheimer's Disease (BLSA)", IRB No. 2009-074). All InCHIANTI participants provided written informed consent, and the study protocol was approved by the Italian National Institute of Research and Care of Aging Institutional Review and approved by the Internal Review Board of the Intramural Research Program of the National Institutes of Health.

Conflict of Interest Statement

The authors report no relevant conflicts of interest.

Funding Sources

This work was supported by the Andrew and Lillian A. Posey Foundation (M.T.) and the Intramural Research Program of the NIH, National Institute on Aging (M.T.).

Author Contributions

Conceptualization: V.R.V., M.T., J.A.R., and Y.A. Methodology: Y.A., S.V., M.E.B., C.B., V.R.V., J.S., J.A.R., M.T., T.T., and P.E. Investigation: S.B.-R., J.A.R., P.E., and A.C.K. Visualization: J.A.R., A.W., S.B., and V.R.V. Funding acquisition: M.T. Project administration: M.T., L.F., P.E., V.R.V., and T.T. Supervision: M.T., L.F., and V.R.V. Writing – original draft: J.A.R., S.B., V.R.V., M.T., J.C., and Y.A. Writing – review and editing: M.T., L.F., T.T., P.E., M.E.G., and C.B.

Data Availability Statement

The full proteomic dataset results associated with this study from the Baltimore Longitudinal Study of Aging (BLSA) are available to researchers and can be requested at https://blsa.nih.gov/researchers, in accordance with BLSA and NIH policy.

References

1 Diederichs C, Berger K, Bartels DB. The measurement of multiple chronic diseases: a systematic review on existing multimorbidity indices. J Gerontol A Biol Sci Med Sci. 2011;66(3):301–11.
2 Bahler C, Huber CA, Brungger B, Reich O. Multimorbidity, health care utilization and costs in an elderly community-dwelling population: a claims data based observational study. BMC Health Serv Res. 2015;15:23.
3 Chen B, Butte AJ. Leveraging big data to transform target selection and drug discovery. Clin Pharmacol Ther. 2016;99(3):285–97.
4 Song C, Kong Y, Huang L, Luo H, Zhu X. Big data-driven precision medicine: starting the custom-made era of iatrology. Biomed Pharmacother. 2020;129:110445.
5 Bauer C, Stec K, Glintschert A, Gruden K, Schichor C, Or-Guil M, et al. BioMiner: paving the way for personalized medicine. Cancer Inform. 2015;14:55–63.
6 Williams SA, Kivimaki M, Langenberg C, Hingorani AD, Casas JP, Bouchard C, et al. Plasma protein patterns as comprehensive indicators of health. Nat Med. 2019;25(12):1851–7.

7 Hung WW, Ross JS, Boockvar KS, Siu AL. Recent trends in chronic disease, impairment and disability among older adults in the United States. BMC Geriatr. 2011;11(47):47.
8 Barnes DE, Yaffe K, Byers AL, McCormick M, Schaefer C, Whitmer RA. Midlife vs late-life depressive symptoms and risk of dementia: differential effects for Alzheimer disease and vascular dementia. Arch Gen Psychiatry. 2012;69(5):493–8.
9 Power MC, Tingle JV, Reid RI, Huang J, Sharrett AR, Coresh J, et al. Midlife and late-life vascular risk factors and white matter microstructural integrity: the atherosclerosis risk in communities neurocognitive study. J Am Heart Assoc. 2017;6(5):e005608.
10 Ferrucci L. The Baltimore Longitudinal Study of Aging (BLSA): a 50-year-long journey and plans for the future. J Gerontol A Biol Sci Med Sci. 2008;63(12):1416–9.
11 Resnick SM, Pham DL, Kraut MA, Zonderman AB, Davatzikos C. Longitudinal magnetic resonance imaging studies of older adults: a shrinking brain. J Neurosci. 2003;23(8):3295–301.
12 Ferrucci L, Bandinelli S, Benvenuti E, Di Iorio A, Macchi C, Harris TB, et al. Subsystems contributing to the decline in ability to walk: bridging the gap between epidemiology and geriatric practice in the InCHIANTI study. J Am Geriatr Soc. 2000;48(12):1618–25.
13 Casanova R, Varma S, Simpson B, Kim M, An Y, Saldana S, et al. Blood metabolite markers of preclinical Alzheimer's disease in two longitudinally followed cohorts of older individuals. Alzheimers Dement. 2016;12(7):815–22.
14 Westwood S, Leoni E, Hye A, Lynham S, Khondoker MR, Ashton NJ, et al. Blood-based biomarker candidates of cerebral amyloid using PiB PET in non-demented elderly. J Alzheimers Dis. 2016;52(2):561–72.
15 Tanaka T, Lavery R, Varma V, Fantoni G, Colpo M, Thambisetty M, et al. Plasma proteomic signatures predict dementia and cognitive impairment. Alzheimers Dement. 2020;6(1):e12018.
16 Candia J, Cheung F, Kotliarov Y, Fantoni G, Sellers B, Griesman T, et al. Assessment of variability in the SOMAscan assay. Sci Rep. 2017;7(1):14248.
17 Balasko M, Soos S, Szekely M, Petervari E. Leptin and aging: review and questions with particular emphasis on its role in the central regulation of energy balance. J Chem Neuroanat. 2014;61-62:248–55.
18 Gold L, Ayers D, Bertino J, Bock C, Bock A, Brody EN, et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. PLoS One. 2010;5(12):e15004.
19 Verbrugge LM, Gruber-Baldini AL, Fozard JL. Age differences and age changes in activities: Baltimore Longitudinal Study of Aging. J Gerontol B Psychol Sci Soc Sci. 1996;51(1):S30–41.
20 Kusek JW, Caggiula AW, Williams GW, Klahr S. An overview of the modification of diet in renal disease study. Contrib Nephrol. 1990;81:50–60.
21 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B. 1995;57(1):289–300.
22 Wang M, Zhao Y, Zhang B. Efficient test and visualization of multi-set intersections. Sci Rep. 2015;5:16923.
23 Roberts JA, Varma VR, An Y, Varma S, Candia J, Fantoni G, et al. A brain proteomic signature of incipient Alzheimer's disease in young *APOE* ε4 carriers identifies novel drug targets. Sci Adv. 2021;7(46):eabi8178.
24 Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res. 2019;47(W1):W199–205.
25 Justice JN, Ferrucci L, Newman AB, Aroda VR, Bahnson JL, Divers J, et al. A framework for selection of blood-based biomarkers for geroscience-guided clinical trials: report from the TAME Biomarkers Workgroup. Geroscience. 2018;40(5–6):419–36.
26 Takemon Y, Chick JM, Gerdes Gyuricza I, Skelly DA, Devuyst O, Gygi SP, et al. Proteomic and transcriptomic profiling reveal different aspects of aging in the kidney. Elife. 2021;10:e62585.
27 Fehrman-Ekholm I, Skeppholm L. Renal function in the elderly (>70 years old) measured by means of iohexol clearance, serum creatinine, serum urea and estimated clearance. Scand J Urol Nephrol. 2004;38(1):73–7.
28 Kadowaki T, Yamauchi T. Adiponectin and adiponectin receptors. Endocr Rev. 2005;26(3):439–51.
29 Sweiss N, Sharma K. Adiponectin effects on the kidney. Best Pract Res Clin Endocrinol Metab. 2014;28(1):71–9.
30 Susztak K, Raff AC, Schiffer M, Böttinger EP. Glucose-induced reactive oxygen species cause apoptosis of podocytes and podocyte depletion at the onset of diabetic nephropathy. Diabetes. 2006;55(1):225–33.
31 Wischhusen J, Melero I, Fridman WH. Growth/differentiation factor-15 (GDF-15): from biomarker to novel targetable immune checkpoint. Front Immunol. 2020;11:951.
32 Tsai VWW, Husaini Y, Sainsbury A, Brown DA, Breit SN. The MIC-1/GDF15-GFRAL pathway in energy homeostasis: implications for obesity, cachexia, and other associated diseases. Cell Metab. 2018;28(3):353–68.
33 Fuchs T, Trollor JN, Crawford J, Brown DA, Baune BT, Samaras K, et al. Macrophage inhibitory cytokine-1 is associated with cognitive impairment and predicts cognitive decline: the Sydney Memory and Aging Study. Aging Cell. 2013;12(5):882–9.
34 Jiang J, Trollor JN, Brown DA, Crawford JD, Thalamuthu A, Smith E, et al. An inverse relationship between serum macrophage inhibitory cytokine-1 levels and brain white matter integrity in community-dwelling older individuals. Psychoneuroendocrinology. 2015;62:80–8.
35 Wiklund FE, Bennet AM, Magnusson PK, Eriksson UK, Lindmark F, Wu L, et al. Macrophage inhibitory cytokine-1 (MIC-1/GDF15): a new marker of all-cause mortality. Aging Cell. 2010;9(6):1057–64.
36 Vukojevic V, Mastrandreas P, Arnold A, Peter F, Kolassa IT, Wilker S, et al. Evolutionary conserved role of neural cell adhesion molecule-1 in memory. Transl Psychiatry. 2020;10(1):217.
37 Gosztonyi G, Naschold U, Grozdanovic Z, Stoltenburg-Didinger G, Gossrau R. Expression of Leu-19 (CD56, N-CAM) and Nitric Oxide Synthase (NOS) I in denervated and reinnervated human skeletal muscle. Microsc Res Tech. 2001;55(3):187–97.
38 Messi ML, Li T, Wang ZM, Marsh AP, Nicklas B, Delbono O. Resistance training enhances skeletal muscle innervation without modifying the number of satellite cells or their myofiber association in obese older adults. J Gerontol A Biol Sci Med Sci. 2016;71(10):1273–80.
39 Murty MS, Sharma UK, Pandey VB, Kankare SB. Serum cystatin C as a marker of renal function in detection of early acute kidney injury. Indian J Nephrol. 2013;23(3):180–3.
40 Liu XM, Chan HC, Ding GL, Cai J, Song Y, Wang TT, et al. FSH regulates fat accumulation and redistribution in aging through the Gαi/Ca(2+)/CREB pathway. Aging Cell. 2015;14(3):409–20.
41 Anesi A, Rondanelli M, Trotti R, Melzi d'Eril GV. Biological variability of myoglobin in healthy elderly and younger subjects. Aging. 2000;12(3):168–72.
42 Weber MA, Kinscherf R, Krakowski-Roosen H, Aulmann M, Renk H, Kunkele A, et al. Myoglobin plasma level related to muscle mass and fiber composition: a clinical marker of muscle wasting? J Mol Med Berl. 2007;85(8):887–96.
43 Sedighi SM, Nguyen M, Khalil A, Fulop T. The impact of cardiac troponin in elderly patients in the absence of acute coronary syndrome: a systematic review. Int J Cardiol Heart Vasc. 2020;31:100629.
44 Liu RX, Thiessen-Philbrook HR, Vasan RS, Coresh J, Ganz P, Bonventre JV, et al. Comparison of proteomic methods in evaluating biomarker-AKI associations in cardiac surgery patients. Transl Res. 2021;238:49–62.
45 Kim CH, Tworoger SS, Stampfer MJ, Dillon ST, Gu X, Sawyer SJ, et al. Stability and reproducibility of proteomic profiles measured with an aptamer-based platform. Sci Rep. 2018;8(1):8382.