Label-free plasma proteomics profiling of healthy individuals across lifespan in a Sicilian Cohort

Valentina Siino, Giulia Accardi, Ashfaq Ali, Anna Aiello, Sergio Mosquim Junior, Giuseppina Candore, Calogero Caruso, Fredrik Levander and Sonya Vasto

**Abstract**

**Introduction**

Ageing, defined as a time-dependent functional decline of living organisms (Lopez-Otin C, Cell 2013), is characterized by a progressive deterioration of physiological functions, often leading to development of age-related diseases. Major age-related disorders include atherosclerosis, neurodegenerative disorders and diabetes (Vasto S et al, Front Biosci 2010).

Several factors have been characterized as risks for the development of age-related diseases such as: genomic predispositions, telomeric and epigenetic alterations, mitochondrial dysfunction and cellular senescence (Lopez-Otin, Cell 2013). Also, chronic systemic inflammation might lead to higher risk of developing cardiovascular disease, both representing a major cause of death in people over 65 years old (Ferrucci and Fabbri 2018).

Long-lived individuals (LLs) represent a good model of healthy ageing since, during the years, they have escaped the development of age-related diseases and show good health. Investigating ageing mechanisms and, even more, how LLIs have developed a healthy ageing process, is of great interest to unravel potential biomarkers that could prolong human lifespan and/or promote healthy ageing. Even though plasma proteomics is very challenging due to the presence of very high abundant proteins (Anderson NL, Anderson NG. 2002) and the lack of high-throughput approaches, the study of circulating proteins from plasma can contribute to the identification of unique protein signatures in the elderly population.

In the presented study, we analyzed the plasma proteome of a Sicilian cohort of healthy individuals in the age range 20-100+ with the aim of identifying unique proteins across lifespan and that are specifically related to healthy ageing. Specifically we have used a high-throughput automated protein digestion approach to reduce the variability between samples combined with a data-independent acquisition (DIA) method to deepen protein coverage. Using this novel plasma proteomic approach in combination with statistical analysis we have confirmed some of the findings from other ageing studies and we have also identified new proteins that could play an important role in healthy ageing.

**Material and methods**

**Sample preparation for mass spectrometry analysis**

Plasma was lysed in 5% SDS in 100 mM Tris (pH=7.55) under ice by ultrasound probe sonication with Branson Digital Sonifier® 250-D (Branson Ultrasonics Corporation, Danbury, USA), at amplitude 10%, with 10s pulse on and 20s off, for a total of 36 cycles. Lysate was centrifuged at 13,000 rpm for 8 minutes to remove debris and the supernatant, containing proteins, was recovered.

50 μg of plasma proteins were digested into peptides using on-bead digestion on HILIC Microspheres (ReSyn Bioscience, Gauteng, South Africa). The process has been fully automated using a 96-well plate in the robotic system King-Fisher Flex (Thermo Fisher Scientific, Bremen). The automated procedure consisted on the following steps: magnetic microspheres (1:10 protein:beads ratio) were incubated and equilibrated in equilibration buffer (15% ACN, 100 mM NH4Ac, pH=4.5); protein sample was incubated in binding buffer (30% ACN, 200 mM NH4Ac, pH=4.5) where proteins would bind to HILIC beads. Beads were then washed twice in 95% ACN to remove unspecific and low binding proteins. Beads-binding proteins were then incubated for 1h at 47°C with trypsin (20:1 protein:trypsin ratio) dissolved in 50 mM AMBIC. Peptides were recovered from the plate and dried in a Speedvac (Thermo Fisher Scientific, Germany) prior to C18 desalting procedure.

Desalting procedure was performed using BioPureSPN Mini, PROTO 300 C18 (The Nest Group, Inc., MA, USA). Briefly, columns were equilibrated with 100 μl 70% ACN, 5% FA and conditioned using 100 μl 5% FA. Samples were resuspended in 100 ul 5% FA and loaded into the column. Column was washed in 5% FA and cleaned peptides were eluted using 100 μl 50% ACN, 5% FA. All the steps were performed using an Eppendorf bench top centrifuge at 50xg for 2 minutes.

Cleaned peptides were dried and stored at -20°C prior to quantification and injection into the mass spectrometer.

**nanoLC mass spectrometry**

300 ng of each sample was resuspended in 0.1% FA loaded onto an EASY-nano LC system (Thermo Fisher Scientific, Germany). The analytical column was 15 cm long fused silica capillary (75 μm\* 16 cm Pico Tip Emitter, New Objective, USA) packed in house with C18 material ReproSil-Pur 1.9 μm (Dr. Maisch GmbH, Germany). Peptides were separated using an 80 min gradient from 5% to 90% solvent B (80% ACN, 0.1% FA) at a constant flow rate of 250 nl/min. The nanoLC system was coupled to Q-Exactive HF-X Mass Spectrometer (Thermo Fisher Scientific, Germany). Data were acquired using either data-dependent acquisition (DDA) or data-independent acquisition (DIA).

**DDA acquisition**

The Orbitrap was set in a positive mode. Precursor spectra (375 to 1500 m/z) were acquired at 120,000 resolution with a automatic gain control (AGC) of 3x106 and a miximum injection time of 50 ms. The 20 most abundant ion peptides were selected for fragmentation. Fragmented spectra were acquired at 15,000 resolution with an AGC of 1x105 and a maximum injection time of 20 ms. Isolation width was set to 1.2 m/z.

**DIA acquisition**

To generate the library, the MS was set to acquire six chromatogram library acquisitions with 4 m/z DIA spectra (4 m/z precursor isolation windows at 30,000 resolution, AGC target 1e6, maximum inject time 60 ms) using an overlapping window pattern from narrow mass ranges (402.43-496.47, 502.47-596.52, 602.52-696.56, 702.56-796.61, 802.61-896.65 and 902.66-996.70 m/z). Maybe have a supplementary with the window range? A full window spectrum (395-1005 m/z at 30,000 resolution using an AGC target value of 3x106 ions and a maximum injection time of 55 ms) was scanned every 10 samples.

For quantitative samples, the MS was set to acquire DIA spectra at 15,000 with AGC target of 1x106, maximum injection time of 20 ms and using windows from 404.43 to 992.70 m/z with isolation window between them of 8 m/z.

DDA and DIA spectra were set with normalized collision energy (NCE) of 27. The Xcalibur software v3.0 (Thermo Fisher Scientific, Germany) controlled the nanoLC system, the mass spectrometer and was used to acquire and visualize the RAW data.

**Data analyses**:

Data missingnes was anayzed using "mice" package in R. We conisdered proteins present in at least 50% of the samples for further analyses. Principle component analyses was used to perform exploratory analyses of the proteomics data. Explained variance by each principle component was calculated using the output from ´prcomp´function in R.

Correlation between principle components and phenotype (clinical) variables was calculated using ´lm´ function in R and "adj.r.squared" values were extracted from the summary statistics produced by the ´lm´ function.

**Differetial expression analyses**

For differential expression analyses we used "limma" package in R. Following models were used to analyze deferentially abundant proteins.

- \*Basic model\* protein ~ Age

- \*Gender and BMI adjusted model\* protein ~ gender + BMI + Age

- \*Interaction with environmental variables\* protein ~ gender + bmi + Age + ENV + Age\*ENV

**Age related protein clustering**

To identify patterns associated with age we performed cluster analyses of on proteins associated with age. To visualize trends we create age related intervals as following.

1. Plot for groups between 0-40, 61-60, 61-80 and 81-112.

2. Plots for groups for age span of 10 years each

Mean protein intensity was calculated for proteins in each cluster followed by average over each age intervals.

For clustering, protein data was scaled followed by calculating dsitance between protein observation using Euclidean distance. Clustering was then performed by using complete linkage between observations.

**SWANDe analyses**

Differential expression Sliding window analysis distinguishes (DEswan) was used to identify waves of aging plasma proteins.

Age span between 20 years to 120 years with interval size of 10 years was slected for sliding window analyses in the ´DEswan´ function implemented i ´DEswan´ package. Sex was used as covariate in the DEswan analyses.

**Mass spectrometry data processing**

**Results**

The present study aimed to use an automated high-throughput plasma protein digestion method in combination with DIA to identify differentially expressed proteins in plasma among different age groups from a healthy Sicilian cohort.

**Plasma ageing proteome overview**

The current study included a total of 86 plasma samples from healthy men and women aged 20 to 95+ years. A schematic overview of the samples and the workflow employed is displayed in **Figure 1**. Further individuals’ information regarding clinical and biochemical parameters as well as BMI, age and location (origin of the samples) are reported in **Supplementary table 1**.

By using an automated sample processing and a label-free approach on undepleted plasma combined with data independent acquisition (DIA) a total amount of 435 proteins were identified. After inspection of the data, some proteins were removed due to missing values and the final protein list contains 410 proteins.

**Figure 2A** displays the overall sample distribution in the principal component analysis (PCA) plot. Percent variance explained in the principal components (PCs), shows a good distribution of the data in the first 10 PCs (**Figure 2B**). Thus, the association between the 10 PCs and the clinical variables collected from each individual was further investigated. Results show that parameters like age, category (young, middle age, old, very old and centenaries), location (city or village), uric acid (UA2) and gender play a role in proteins distribution, while biochemical factors, such as LDL, HDL and BMI, do not significantly contribute to the distribution of proteins (**Figure 2C**).

**Protein correlation with age**

The correlation between proteins and age was further investigated.

From this analysis we could identify 106 proteins that significantly (q-value<0.05) correlate with age (**Supplementary figure 1**). To further discriminate between up and down regulated proteins, a cut-off on effect size >/< 0.01 was applied. 20 age-associated proteins are negatively regulated, while 27 are positively regulated as shown in the volcano plot (**Figure 3A**). Among these last group of proteins, fibulin-1 (FBLN1, P23142), lysozyme C (LYZ, P61626), dystroglycan (DAG1, Q14118) and gamma glutamyl hydrolase (GGH, Q92820) were already reported in another study as related to aging (Lehallier B et al. 2019). Indeed, FBLN1 and DAG1 have been shown to play a crucial role in ageing since both have a positive effect on brain homeostasis and neurotrophic activities (Moore SA et al. Nature 2002; Ohsawa I et al. J Neurochemistry 2001). Interestingly, also extracellular superoxide dismutase (SOD3, P08240) is increasing with age, suggesting a higher protection from oxygen reactive species in older individuals.

Most of the age-associated proteins that are negatively correlated with age are immunoglobulins. It has been reported that accumulation of immunoglobulins activates the immune system and results in chronic inflammation during ageing (Aprahamian T et al Clin Exp Immunol 2008). The reduced expression of immunoglobulins in our older cohort suggests a better response to inflammation and cellular stress.

Furthermore, IGFBP-3 and IGFBP-5 are less expressed in our elderly cohort pointing towards reduced cell senescence in these individuals as well as reduced cellular inflammation. Indeed, both proteins have been associated with cellular senescence and they have been reported to induce fibroblast activation and inflammatory response (Elzi DJ et al PNAS 2012; Kojima H et al. Cell cycle 2012; Sanada F et al. Front. Endocrinol. 2018).

To identify proteomic changes during individuals’ lifespan we used the differential expression-sliding window analysis (DE-SWAN) algorithm (http://lehallib.github.io/DEswan/), which has been demonstrated to provide useful information about protein changes at particular stages of life (Lehallier B et al. 2019). The algorithm employed here analyses protein levels within a window of 10 years. Using this approach we were able to identify significant proteins changing mainly in two age waves with a peak at age 40 and then a consistent peak at age 80 (**Figure 3B**). The upset plot (**Figure 3C**) is also showing the number of significant proteins that are unique or shared between age waves. Most of the significant proteins in the age-waves are unique to long lived individuals (LLI). The age range X80 comprises of 16 proteins, while X80 and X90 share 22 proteins (**Supplementary table 2**). These results might reveal a unique protein signature for LLIs. Among these proteins, many of them are related to extracellular matrix (ECM) structures and functions. The ECM comprises of proteins providing structural and biochemical support to cells. Since the cellular behavior is highly affected by the surrounding environment, age-related changes of the ECM impact the ability to give support to the cells and might influence most processes in the body (Birch HL 2018).

On the other hand, 7 proteins that are unique to age range 40 are related to the immune response and they result to be (up or down regulated compared to old? Ashfaq could you look into that?).

**Protein correlation with other parameters**

In the present dataset, most of the plasma samples were collected from the city (Palermo), while some of the LLIs and their relatives plasma were coming from a village in the Sicilian mountains. 9 proteins correlate to the location at a set q-value<0.05 (**Supplementary figure 2**). **Figure 4A** shows the intensity of 3 representative proteins (B2MG, LYSC and CYTC) and their distribution in regard to age (x-axis) and to the location (blue village and red city). These proteins are more represented in the elderly population in the village than in the city, their expression is increasing with age and they are related to the cellular protein metabolic process.

Furthermore, the correlation of age and location was investigated in parallel, and 9 proteins were significantly correlating to both variables. The heatmap (**Figure 4B**)shows the differential expression of these proteins. Matrix metalloprotease 9 (MMP9) was negatively regulated across samples, confirming what has been previously shown in correlation to LLIs (Cancemi P et al. 2020).

Among the significant proteins, 11 of them are correlated to gender. As expected, 2 proteins are hormonal proteins (SHBG and PZP). Can we have a plot for the proteins related to gender?

UA2 is the end product of purine metabolism. The role of UA is controversial since it has been reported to enhance oxidative stress (Yu MA et al 2010) while other studies suggest that UA is a reactive oxygen species (ROS) scavenger, thus playing an antioxidant effect (Ridi and Tallima 2017; Sautin and Johnson 2008; Glantzounis GK et al 2005). Even UA correlation with age is controversial since it has been shown to physiologically increase during adulthood and it has been associated to endothelial dysfunctions and hypertension (Kuzuya M et al 2002; Kawamoto R et al 2013; Zhou H et al 2017). On the contrary, increased levels of UA in elderly populations have proved a greater muscle strength (Lee J et al 2019), while in *Caenorhabditis elegans* high levels of UA seem to increase life span (Quan QL et al 2020).

In the present cohort, 29 proteins are significantly correlating with UA2, 18 of them are positively correlated to UA2 while 11 proteins are correlating negatively to UA2, as show in **Figure 4C**. Of the 18 proteins positively correlated to UA2, 9 of them are part of the cytoskeleton machinery such as ACT, CFL1, FLN-A, PFN1, TLN1, TPM4 and VCL. The fact that most of these cytoskeleton-associated proteins are mostly expressed in older individuals, could imply a higher need to recruit proteins to the cytoskeleton to support and maintain cellular homeostasis (Amberg D et al 2012).

**Cluster of significant proteins and their trajectories**

Normalized significant proteins were divided into clusters with an age step size of 10 years. As depicted in **Figure 5**, the first 7 clusters show high intensity and were further investigated. Cluster 1 and 3 are the most represented ones (41 and 18 proteins respectively) and show a linear increase with age; while cluster 2, 4, 5, 6 and 7 (19, 15, 5, 13 and 15 proteins respectively) show a step-wise decrease trajectory with age. Most of the proteins in cluster 1 and 3 are related to coagulation and clotting such as fibrinogen, von Willebrand factor as well as coagulation factor V and IX. These proteins are well known to increase with the physiological process of aging and are higher in centenarians, where a state of high coagulation enzyme activity has been reported (Mari D et al 1995; Manucci PM et al 1997; Mari D et al 2008). Among the proteins that are representing the decreasing trajectories, there are several immunoglobulins and serpins. Serpins are known to be highly expressed in senescent cells, providing a fine balance between thrombosis and thrombolysis cascades, and have been identified as plasma biomarkers of ageing (Basisty N et al 2020; Tanaka T et al 2018). We could speculate that the lower expression of many serpins in our elderly population could indicate an escaping mechanism from cell senescence.