P1816

Development of a protein-clock for estimating age in the epidermis using machine learning: insights from the INSPIRE-T cohort

EA DV CONGRESS PARIS 17-20 SEPT 2025

Patrick Bogdanowicz[1], Amaury Alves[2], Pedro Vásquez-Ocmín[1], Pascale Bianchi[1], Aimée Simcic-Mori[1], Eleonore Gravier[1], Gwendal Josse[1], Luciana Bostan[3], Anna Kychygina[3], Michel Simon[3], Nicolas Gaudenzio[3], Nicola Coley[4], Sophie Guyonnet[4], Bruno Vellas[4], Emma Guillou[5], Gautier Doat[5], Katia Ravard[1], Hélène Duplan[1], Sandrine Bessou-Touya[1]

[1]Pierre Fabre Dermo-Cosmetics & Personal Care, R&D Department, Toulouse, France, [2]Ippon Innovation, Toulouse, France, [3]INSERM U1291, INFINITY, Toulouse, France, [4]CHU Toulouse, IHU HealthAge, Toulouse, France, [5]Pierre Fabre Dermo-Cosmetics & Personal Care, Laboratoire Dermatologique Eau Thermale Avène, Lavaur, France.

INTRODUCTION

Aging is characterized by the progressive accumulation of molecular and cellular damage, leading to functional decline and increased disease risk. In contrast to chronological age, biological age reflects an individual's physiological state and is shaped by complex cellular, molecular, and biochemical processes—particularly within dynamic tissues such as the skin. As skin aging is influenced by both intrinsic and extrinsic factors, it serves as a valuable model for investigating age-related biological changes.

Recent advances in machine learning have enabled the integration of multi-omics data to predict biological age and identify molecular drivers of aging. In this context, the translational INSPIRE-T cohort (n=1,200; ages 20–102) was established to explore trajectories of healthy aging through the integration of clinical and molecular markers. Here, we applied machine learning to epidermal proteomic profiles from the INSPIRE-T cohort, developing predictive models to estimate biological skin age based on a defined protein signature.

METHODOLOGY

The INSPIRE-T cohort included 77 individuals, from whom skin biopsies were obtained from the inner arm. Participants were stratified into three age groups: Young (n = 19), Middle-aged (n = 21), and Aged (n = 37). Epidermis and dermis were then separated by heat-induced dissociation. Epidermal samples were then processed and analyzed using shotgun proteomics by LC–MS/MS (Fig. 1).

For LC–MS/MS analysis, 250 ng of peptides were injected into a Vanquish Neo system equipped with a PepMap100 C18 column. Data acquisition was performed on an Exploris 480 mass spectrometer (Thermo Scientific), and raw data were processed using Proteome Discoverer 3.0. Protein identification was performed with the SEQUEST-HT algorithm.

Proteins with more than 80% missing values or labeled as "contaminants" were excluded. Principal component analysis (PCA) was performed to show a separation of age classes. Feature selection was conducted using Recursive Feature Elimination (RFE) on a training set of 62 subjects, resulting in a subset of 28 proteins. An XGBoost machine learning model was then trained on this set and evaluated on an independent test set of 15 subjects. Model hyperparameters were optimized using four-fold cross-validation

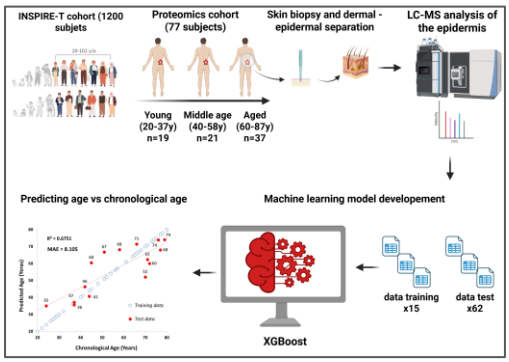

Figure 1 *: Machine learning workflow from INSPIRE-T cohort*

RESULTS

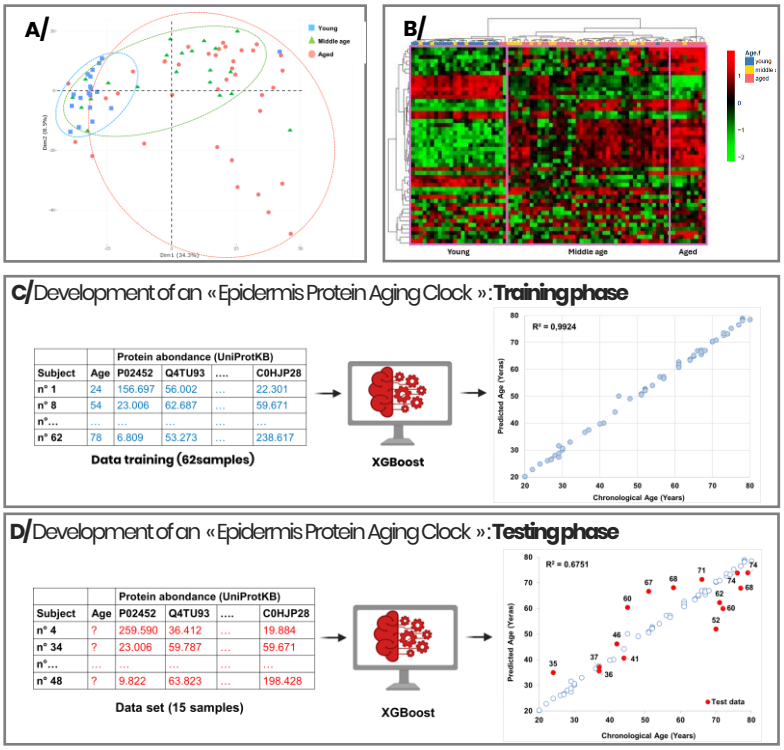

Figure 2 *: **Construction of an « Epidermis Protein Aging Clock »** A/ Principal Componant Analyse (PCA) from 3 age groups.. B/ Heatmap of the top 50 most variable proteins across samples.. Development of « Epidermis Protein Aging Clock » : Training step (C) and testing step. (D).*

From pre-cleaned proteomic dataset comprising 4,054 proteins, principal component analysis (PCA) was performed (Fig. 2A). The first two principal components accounted 42.8% of the total variance. The PCA plot revealed a clear separation between groups: younger individuals formed a distinct cluster, while the middle-aged group exhibited substantial overlap and dispersion between the younger and older groups. To complement findings, a heatmap analysis was conducted using the 50 most variable proteins (Fig. 2B). Consistent with the PCA results, younger individuals clustered separately from other age groups, with the exception of one 29-year-old participant who clustered with the aged group. Notably, over 60% of the middle-aged participants showed dispersed clustering between the young and aged clusters. Interestingly, two individuals from the aged group (61 and 80 years old) clustered with the young group (Fig. 2B).

Then, we used an artificial intelligence model called XGBoost to construct a model that could determine the biological age of a sample by using the abundance of 28 selected proteins. Initially, we integrated into XGBoost the abundance of proteins present in 62 samples while providing the subject's age (Training phase, Fig. 2C). The training phase enabled the identification of a set of 28 proteins, out of 4,054, that were most significant in age calculation (Fig. 3). Following this training phase, we conducted a testing phase of the model using proteomic data from 15 samples (Fig. 2D). The results demonstrate our capability to determine age with a mean absolute error (MAE) of 8.1 years and an R^2 of 0.6751 on an independent test set.

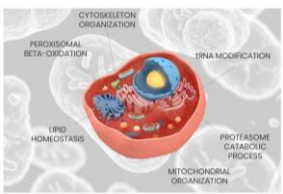

Figure 3 *: Set of 28 proteins identified using the XGboost model*

Pathway analysis of the machine learning-selected 28 proteins identified five canonical pathways: biotin-carboxyl carrier protein assembly, carnitine metabolism, fatty acid activation, phenylalanine degradation IV, and the unfolded protein response. These canonical pathways are involved in various processes that may play a role in aging, including proteasomal degradation, lipid homeostasis, tRNA modification, cytoskeletal organization, and mitochondrial maintenance (Fig. 3).

CONCLUSION

While gold-standard aging clocks, such as Hannum's and Horvath's DNA methylation-based models, are well established, recent advances in artificial intelligence applied to proteomic data have facilitated the development of protein-based aging clocks, primarily from plasma samples. Our protein clock approach presents strong potential for applications in the dermo-cosmetic field, offering new avenues to decipher the molecular mechanisms underlying epidermal aging and to inform targeted, personalized anti-aging solutions.