High-throughput image analysis of quantitative immunohistochemistry data reveals age-dependent changes in mitochondrial and extracellular matrix proteins in multiple human tissues

Dora Marcec[1], Jarek Byrk[1] and **Yun-Wah Lam**[1,2]

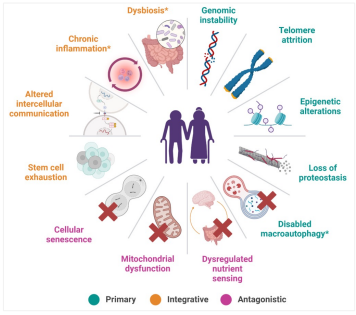
[1]School of Applied Sciences, University of Huddersfield, Huddersfield HD1 3DH, UK

[2]Department of Diagnostic Sciences, School of Nursing and Health Sciences, Hong Kong Metropolitan University, Hong Kong, China

AIM

To develop a high-throughput computational workflow for quantitative analysis of Human Protein Atlas immunohistochemistry images and identify proteins whose expression changes significantly during human ageing.

BACKGROUND



Ageing is associated with progressive molecular and structural alterations that contribute to tissue dysfunction, frailty, and age-related disease. Although transcriptomic studies have generated extensive knowledge regarding ageing biology, protein-level investigations remain comparatively limited. This is particularly true for extracellular matrix (ECM) proteins and other structurally complex proteins, whose abundance cannot always be accurately inferred from mRNA expression because of post-transcriptional regulation, protein turnover, and extensive post-translational modification.

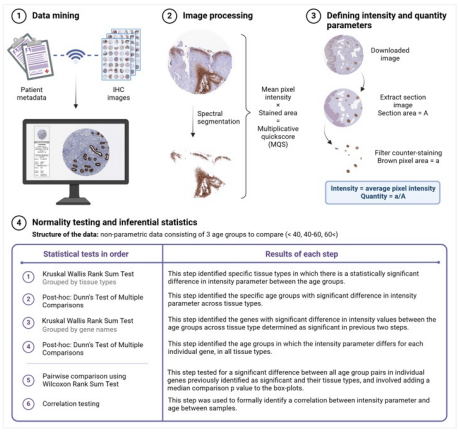
Mass spectrometry (MS)-based proteomics has become the dominant approach for large-scale protein profiling. However, many proteins remain difficult to analyse using conventional MS workflows. ECM proteins are often highly cross-linked, insoluble, and heavily glycosylated, making extraction and detection challenging. Furthermore, MS typically requires tissue homogenisation, resulting in the loss of spatial information regarding cellular localisation and tissue architecture.

The Human Protein Atlas (HPA) provides an alternative and largely underutilised resource for proteomic ageing research. Established as one of the world's largest antibody-based proteomics initiatives, the HPA contains millions of immunohistochemistry (IHC) images generated from human tissues using validated antibodies and linked to detailed tissue metadata. Unlike conventional proteomic datasets, HPA images preserve the spatial distribution of proteins at cellular resolution while simultaneously capturing tissue-level organisation.

Despite its scale and biological richness, the HPA has primarily been used as a qualitative reference atlas, with protein expression assessed by visual inspection and pathologist annotation. The enormous volume of image data has historically limited its use for quantitative biological discovery.

Recent advances in image analytics and computational pathology now create an opportunity to transform these archived images into a quantitative resource. By combining automated image processing with statistical analysis, large-scale IHC repositories can be repurposed as digital platforms for ageing biomarker discovery. We therefore hypothesised that high-throughput analysis of HPA images could identify age-dependent changes in mitochondrial and extracellular matrix proteins across multiple human tissues, providing a scalable and cost-effective approach for protein-level ageing research.

DIGITAL AGEING BIOMARKER DISCOVERY PIPELINE



Patient metadata and immunohistochemistry images were mined from the Human Protein Atlas database. Images were subjected to automated colour deconvolution and image processing to quantify antibody staining intensity. Statistical analysis was then performed to identify proteins demonstrating significant age-dependent expression patterns across human tissues.

Dataset analysed:

- 153,919 immunohistochemistry images
- 44 tissue types
- 734 proteins
- 1,268 antibodies

KEY FINDINGS

The workflow identified age-associated expression changes in multiple proteins across diverse human tissues.

Nine proteins demonstrated statistically significant correlations between protein expression and age.

Four proteins demonstrated significant differences between age groups.

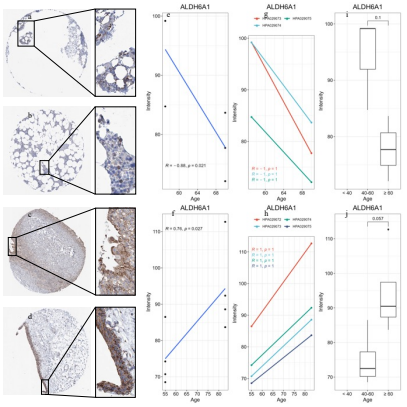
Three proteins satisfied both criteria:

- ALDH6A1
- MMP9
- ITGAM

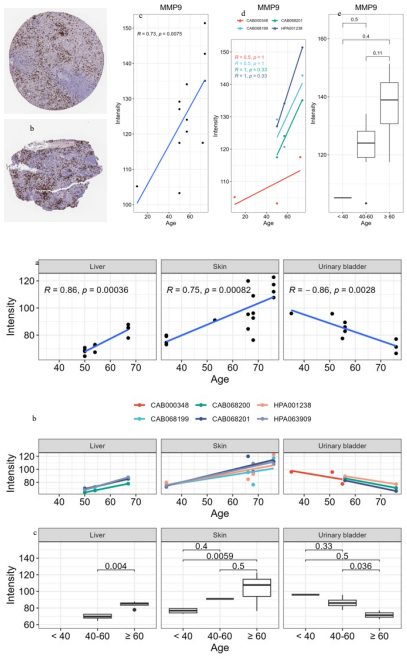
These proteins emerged as the strongest candidates for age-associated biomarker signatures.

Tissue	*Gene*	*Parameter*	*group1*	*group2*	*p*
Spleen	ALDH6A1	Intensity	< 40	= 60	0.004
Liver	MMP9	Intensity	40-60	= 60	0.004
Skin	MMP9	Intensity	< 40	= 60	0.006
Cervix and uterine	ALDH6A1	Intensity	< 40	40-60	0.024
Heart muscle	ITGAM	Intensity	40-60	= 60	0.024
Spleen	COASY	Intensity	< 40	= 60	0.024
Liver	ALDH6A1	Intensity	40-60	= 60	0.028
Skeletal muscle	ALDH6A1	Intensity	< 40	= 60	0.029
Urinary bladder	MMP9	Intensity	40-60	= 60	0.036
Urinary bladder	ALDH6A1	Intensity	40-60	= 60	0.057
Bone marrow	ALDH6A1	Intensity	40-60	= 60	0.100
Spleen	MMP9	Intensity	40-60	= 60	0.110
Breast	HDGF	Intensity	< 40	40-60	0.133
Skeletal muscle	COASY	Intensity	< 40	= 60	0.133
Endometrium	CCBE1	Intensity	< 40	40-60	0.200
Thyroid gland	ADAMTS13	Intensity	< 40	40-60	0.267
Urinary bladder	MMP9	Intensity	< 40	40-60	0.333
Endometrium	F13A1	Intensity	< 40	= 60	0.333
Endometrium	F13A1	Intensity	40-60	= 60	0.333
Skin	MMP9	Intensity	< 40	40-60	0.400
Skin	MMP9	Intensity	40-60	= 60	0.500
Urinary bladder	MMP9	Intensity	< 40	= 60	0.500
Endometrium	CCBE1	Intensity	< 40	= 60	0.500
Endometrium	CCBE1	Intensity	40-60	= 60	0.667
Endometrium	F13A1	Intensity	< 40	40-60	0.667

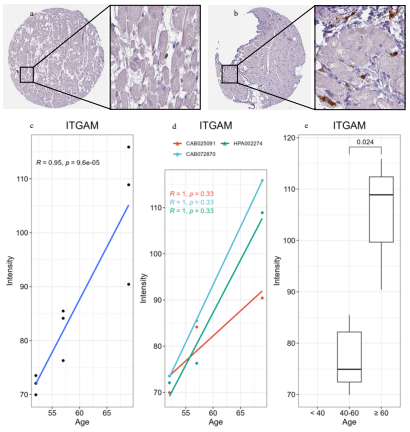
AGEING SIGNATURE 1: MITOCHONDRIAL DYSFUNCTION ALDH6A1



AGEING SIGNATURE 2: EXTRACELLULAR MATRIX REMODELLING MMP9



AGEING SIGNATURE 3: INFLAMMAGING ITGAM



GERONTECHNOLOGY IMPLICATIONS

This study demonstrates that existing digital pathology archives can be repurposed as large-scale ageing research infrastructure.

The workflow enables:

- Quantitative protein analysis without generating new experimental data
- Preservation of spatial protein localisation
- Discovery of tissue-specific ageing biomarkers
- Analysis of proteins that are difficult to study using conventional proteomics
- Future integration with artificial intelligence and computational pathology platforms

By transforming archived pathology images into quantitative datasets, this approach establishes a foundation for next-generation digital biomarker discovery and precision gerontology.

FUTURE DIRECTIONS

- Expansion to the entire Human Protein Atlas proteome.
- Integration with transcriptomic, clinical, and imaging datasets.
- Development of tissue-specific biological age signatures.
- Machine-learning prediction of biological age from digital pathology data.
- Creation of a digital atlas of human protein ageing.