Rakhan Aimbetov

# The Extracellular Matrix Aging Atlas

## A knowledgebase of time-resolved matrisome signatures extracted from public proteomic datasets

The extracellular matrix is a complex substance localized in the extracellular space, serving as a medium where cells reside. Besides providing anchoring support, the extracellular matrix is a mechanical and biochemical environment that directs cellular functions and processes via a collection of various stimuli, prompting gene expression profiles to reflect developmental and physiological contexts. As a dynamic structure, the extracellular matrix composition is subject to change as a function of age. Presently, there is a lack of a unified consensual understanding of the qualitative and quantitative aspects of these changes. A few recent publications look into proteomic alterations that happen in the extracellular matrix of some tissues with aging. Consequently, aggregating the published datasets into a database of the extracellular matrix aging signatures is proposed.

# Introduction

The **extracellular matrix** (ECM) is a complex biological scaffold that gives tissues and organs their structural foundation (1). The ECM and the cells that synthesize and assemble it communicate reciprocally to modulate tissue homeostasis (2). Consequently, any alteration, physiological or not, in either the cell or its microenvironment, represented by the ECM, launches an array of mechanisms to restore balance.

The ECM is mainly protein in nature. The sum total of proteins that constitute the ECM is called the **matrisome** (3). The matrisome comprises 1027 genes for the human genome and 1110 genes for the mouse genome (4, 5) **(Fig. 1)**. Any given tissue consists of over 150 ECM and ECM-associated proteins; with characteristic differences in the ECM composition of different tissues (6, 7). Furthermore, qualitative and quantitative matrisomic alterations in response to insult can be used as a biomarker for undercurrent pathology (8).

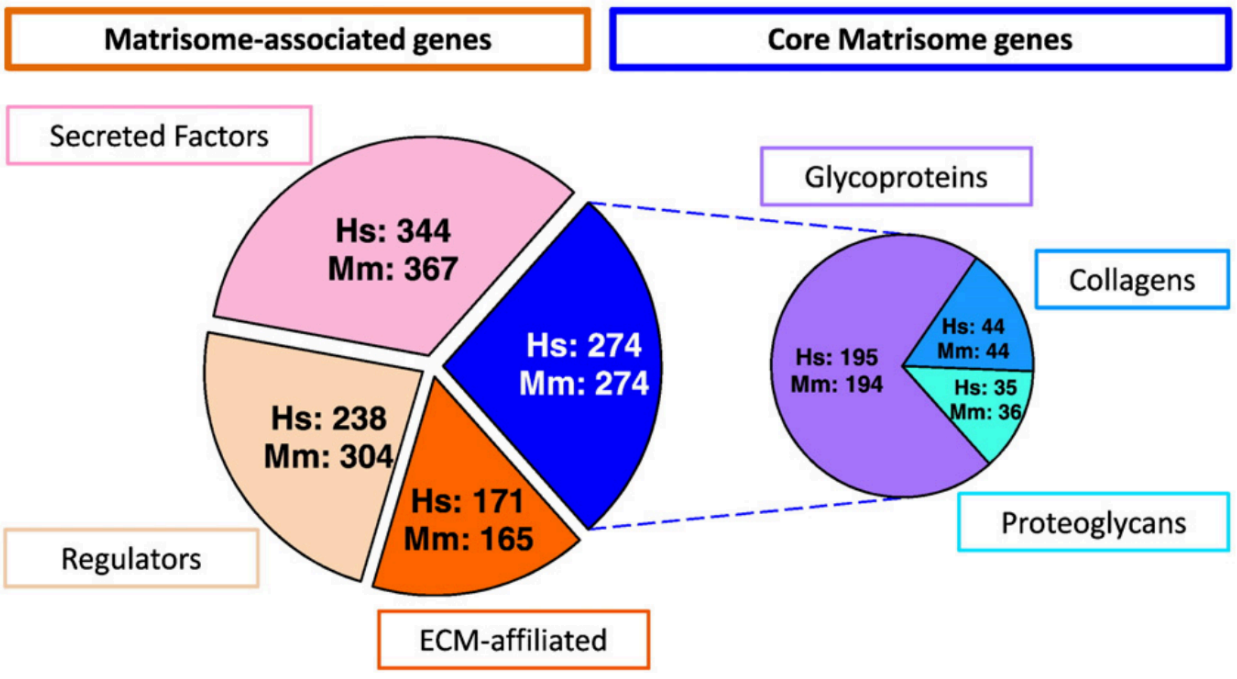

**Fig. 1. The matrisome** (5)**.** The core matrisome comprises ECM glycoproteins, collagens, and proteoglycans. Matrisome-associated proteins include ECM-affiliated proteins, ECM regulators, and secreted factors. Hs, *Homo sapiens*; Mm, *Mus musculus*.

Matrisomes, being a dynamic structure, are subject to compositional variation as a function of age (9). Although we know the cross-sectional structural make-up of different tissues in the norm (10), little is known about how the matrisomes change temporally – at different stages of life. Of note, it is important to describe the ECM on the protein level since transcriptomic signatures might not adequately reflect proteomic changes (11). In the past few years, several papers describing such time-resolved shifts in matrisomic composition for select tissues in humans and mice using protein mass spectrometry have been published (12–24). The creation of a database that would integrate the public datasets on age-associated ECM characteristics consistently and comparably is a natural next step. A database like that would help the matrix biologists to better understand the temporal ECM dynamics, and identify potential targets for interventions.

## Aims

I propose to create such a database from publicly available datasets for human and murine tissues – **The ECM Aging Atlas**. As a result, a curated database, where each subsequent new dataset is processed according to a standardized schema, complementing and enriching the whole body of information in a harmonized fashion, will be deployed.

# Materials/methods

## Study type

- Other

This is a bioinformatics and computational project to integrate the available knowledge on age-related extracellular matrix dynamics into a standardized database format, aiming to enhance research accessibility and facilitate new discoveries in the field.

## Data

- Registration prior to analysis of the data

The data selected for the project are described in the spreadsheet. These include processed datasets and associated raw data files from peer-reviewed research on tissue- or organ-level, time-resolved matrisomic signatures in mammals (12–24).

## Processing and analysis

To efficiently compile and standardize the **ECM Aging Atlas** database, focusing on the essence and eliminating redundancies, the workflow includes:

- **Database and metadata standards.** A relational database schema organizes data into interrelated tables, ensuring efficient storage and query execution. Metadata standards specify the format and content of additional information about data (e.g., experimental

conditions, sample preparation), which is crucial for interpreting and reproducing research findings.

- **Data normalization and analysis protocols.** Data normalization adjusts measurements to a common scale or reference, compensating for variations in experimental methods. Standardized analysis workflows use consistent procedures for processing and analyzing data, enhancing comparability across studies.
- **Quality control and integrating data.** Quality control involves checks to ensure data accuracy and consistency. Data integration merges information from various sources into a unified database, maintaining logical connections among different data elements.
- **Database access and community interaction.** A user-friendly web interface facilitates access to the database, allowing researchers to easily explore and extract information. Engaging with the scientific community ensures the database remains relevant and up-to-date, encouraging contributions and feedback.

This streamlined approach to database creation supports the harmonization of diverse datasets into a singular, valuable resource for the scientific community, fostering advancements in understanding ECM changes with aging.

To reanalyze the raw data, I will use one of the open-source solutions tailored for proteomics workflows. For example,

- OpenMS is a versatile environment for proteomic analysis (25, 26), compatible with such workflow systems as Nextflow, KNIME, and Galaxy.
  **Nextflow** is an open-source workflow system for automating bioinformatics pipelines. It is designed to be scalable, reproducible, and portable, making it suitable for large-scale analyses on cloud platforms, such as **Google Cloud** (doc).
- Deployed on **Nextflow**, quantms is an open-source cloud-based pipeline for massively parallel proteomic data analysis (27). Currently, the workflow supports three major MS-based analytical methods: (i) data-dependent acquisition (DDA) label-free and isobaric quantitation (e.g. TMT, iTRAQ); (ii) data-independent acquisition (DIA) label-free quantification.
- MaxQuant is an open-source software platform for analyzing large-scale quantitative proteomics data; it provides a comprehensive set of tools for protein identification, quantification, and statistical analysis (28, 29). **MaxQuant** on **Galaxy** offers seamless data analysis, eliminating software installation needs (30) (doc).
- CloudProteoAnalyzer is a cloud-computing platform designed to offer a user-friendly interface and accurate analysis of comprehensive proteomics data (31). This platform harnesses the computational capacity of multiple computing nodes within a supercomputer, thereby ensuring scalability for large datasets.

# Budget

## Costs

| | |
|---|---|
| **Google Cloud compute (VM + storage):** | 250 USD/mo. * 6 mo. = **1500 USD** |
| **ChatGPT subscription:** | 20 USD/mo. * 12 mo. = **240 USD** |
| **Researcher hourly time:** | 25 USD/hr. * 130 hrs. = **3250 USD** |
| | **TOTAL: 4990 USD** |

# Links

- **PROJECT DATAROOM**
- **PROJECT UPDATES**

## Databases

### ECM

GAG-DB; https://gagdb.glycopedia.eu/
MatriNet; https://www.matrinet.org/
Matrisome AnalyzeR; https://matrinet.shinyapps.io/MatrisomeAnalyzer/
MatrisomeDB 2.0; https://matrisomedb.org/
MatrixDB; http://matrixdb.univ-lyon1.fr/
The Human Protein Atlas; https://www.proteinatlas.org/
The Matrisome Project; http://matrisome.org/

### Datasets

Google Dataset Search; https://datasetsearch.research.google.com/
Mendeley Data; https://data.mendeley.com/
OmicsDI; https://www.omicsdi.org/

# References


1. Frantz, C., Stewart, K. M., and Weaver, V. M. (2010) The extracellular matrix at a glance. *J. Cell Sci.* **123**, 4195–4200
2. Humphrey, J. D., Dufresne, E. R., and Schwartz, M. A. (2014) Mechanotransduction and extracellular matrix homeostasis. *Nat. Rev. Mol. Cell Biol.* **15**, 802–812
3. Hynes, R. O., and Naba, A. (2012) Overview of the matrisome--an inventory of extracellular matrix constituents and functions. *Cold Spring Harb. Perspect. Biol.* **4**, a004903
4. Naba, A., Clauser, K. R., Hoersch, S., Liu, H., Carr, S. A., and Hynes, R. O. (2012) The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. *Mol. Cell. Proteomics MCP.* **11**, M111.014647
5. Naba, A., Clauser, K. R., Ding, H., Whittaker, C. A., Carr, S. A., and Hynes, R. O. (2016) The extracellular matrix: tools and insights for the “omics” era. *Matrix Biol. J. Int. Soc. Matrix Biol.* **49**, 10–24
6. Shao, X., Taha, I. N., Clauser, K. R., Gao, Y. (Tom), and Naba, A. (2020) MatrisomeDB: the ECM-protein knowledge database. *Nucleic Acids Res.* **48**, D1136–D1144
7. Sacher, F., Feregrino, C., Tschopp, P., and Ewald, C. Y. (2021) Extracellular matrix gene expression signatures as cell type and cell state identifiers. *Matrix Biol. Plus.* **10**, 100069
8. Taha, I. N., and Naba, A. (2019) Exploring the extracellular matrix in health and disease using proteomics. *Essays Biochem.* **63**, 417–432
9. Ewald, C. Y. (2020) The matrisome during aging and longevity: a systems-level approach toward defining matreotypes promoting healthy aging. *Gerontology*. **66**, 266–274
10. Shao, X., Gomez, C. D., Kapoor, N., Considine, J. M., Grams, C., Gao, Y. T., and Naba, A. (2022) MatrisomeDB 2.0: 2023 updates to the ECM-protein knowledge database. *Nucleic Acids Res.* 10.1093/nar/gkac1009
11. Jiang, L., Wang, M., Lin, S., Jian, R., Li, X., Chan, J., Dong, G., Fang, H., Robinson, A. E., GTEx Consortium, and Snyder, M. P. (2020) A quantitative proteome map of the human body. *Cell*. **183**,

269-283.e19
12. Caldeira, J., Santa, C., Osório, H., Molinos, M., Manadas, B., Gonçalves, R., and Barbosa, M. (2017) Matrisome profiling during intervertebral disc development and ageing. *Sci. Rep.* **7**, 11629
13. Angelidis, I., Simon, L. M., Fernandez, I. E., Strunz, M., Mayr, C. H., Greiffo, F. R., Tsitsiridis, G., Ansari, M., Graf, E., Strom, T.-M., Nagendran, M., Desai, T., Eickelberg, O., Mann, M., Theis, F. J., and Schiller, H. B. (2019) An atlas of the aging lung mapped by single cell transcriptomics and deep tissue proteomics. *Nat. Commun.* **10**, 963
14. Tam, V., Chen, P., Yee, A., Solis, N., Klein, T., Kudelko, M., Sharma, R., Chan, W. C., Overall, C. M., Haglund, L., Sham, P. C., Cheah, K. S. E., and Chan, D. (2020) DIPPER, a spatiotemporal proteomics atlas of human intervertebral discs for exploring ageing and degeneration dynamics. *eLife.* **9**, e64940
15. McCabe, M. C., Hill, R. C., Calderone, K., Cui, Y., Yan, Y., Quan, T., Fisher, G. J., and Hansen, K. C. (2020) Alterations in extracellular matrix composition during aging and photoaging of the skin. *Matrix Biol. Plus.* **8**, 100041
16. Li, M., Li, X., Liu, B., Lv, L., Wang, W., Gao, D., Zhang, Q., Jiang, J., Chai, M., Yun, Z., Tan, Y., Gong, F., Wu, Z., Zhu, Y., Ma, J., and Leng, L. (2021) Time-resolved extracellular matrix atlas of the developing human skin dermis. *Front. Cell Dev. Biol.* **9**, 783456
17. Li, Z., Tremmel, D. M., Ma, F., Yu, Q., Ma, M., Delafield, D. G., Shi, Y., Wang, B., Mitchell, S. A., Feeney, A. K., Jain, V. S., Sackett, S. D., Odorico, J. S., and Li, L. (2021) Proteome-wide and matrisome-specific alterations during human pancreas development and maturation. *Nat. Commun.* **12**, 1020
18. Randles, M., Lausecker, F., Kong, Q., Suleiman, H., Reid, G., Kolatsi-Joannou, M., Tian, P., Falcone, S., Davenport, B., Potter, P., Van Agtmael, T., Norman, J., Long, D., Humphries, M., Miner, J., and Lennon, R. (2021) Identification of an altered matrix signature in kidney aging and disease. *J. Am. Soc. Nephrol. JASN*. 10.1681/ASN.2020101442
19. Lofaro, F. D., Cisterna, B., Lacavalla, M. A., Boschi, F., Malatesta, M., Quaglino, D., Zancanaro, C., and Boraldi, F. (2021) Age-related changes in the matrisome of the mouse skeletal muscle. *Int. J. Mol. Sci.* **22**, 10564
20. Ariosa-Morejon, Y., Santos, A., Fischer, R., Davis, S., Charles, P., Thakker, R., Wann, A. K., and Vincent, T. L. (2021) Age-dependent changes in protein incorporation into collagen-rich tissues of mice by in vivo pulsed SILAC labelling. *eLife.* **10**, e66635
21. Ouni, E., Nedbal, V., Da Pian, M., Cao, H., Haas, K. T., Peaucelle, A., Van Kerk, O., Herinckx, G., Marbaix, E., Dolmans, M.-M., Tuuri, T., Otala, M., Amorim, C. A., and Vertommen, D. (2022) Proteome-wide and matrisome-specific atlas of the human ovary computes fertility biomarker candidates and open the way for precision oncofertility. *Matrix Biol. J. Int. Soc. Matrix Biol.* **109**, 91–120
22. Tsumagari, K., Sato, Y., Aoyagi, H., Okano, H., and Kuromitsu, J. (2023) Proteomic characterization of aging-driven changes in the mouse brain by co-expression network analysis. *Sci. Rep.* **13**, 18191
23. Chmelova, M., Androvic, P., Kirdajova, D., Tureckova, J., Kriska, J., Valihrach, L., Anderova, M., and Vargova, L. (2023) A view of the genetic and proteomic profile of extracellular matrix molecules in aging and stroke. *Front. Cell. Neurosci.* 10.3389/fncel.2023.1296455
24. Dipali, S. S., King, C. D., Rose, J. P., Burdette, J. E., Campisi, J., Schilling, B., and Duncan, F. E. (2023) Proteomic quantification of native and ECM-enriched mouse ovaries reveals an age-dependent fibro-inflammatory signature. *Aging.* **15**, 10821–10855
25. Röst, H. L., Sachsenberg, T., Aiche, S., Bielow, C., Weisser, H., Aicheler, F., Andreotti, S., Ehrlich, H.-C., Gutenbrunner, P., Kenar, E., Liang, X., Nahnsen, S., Nilse, L., Pfeuffer, J., Rosenberger, G., Rurik, M., Schmitt, U., Veit, J., Walzer, M., Wojnar, D., Wolski, W. E., Schilling, O., Choudhary, J. S., Malmström, L., Aebersold, R., Reinert, K., and Kohlbacher, O. (2016) OpenMS: a flexible open-source software platform for mass spectrometry data analysis. *Nat. Methods.* **13**, 741–748
26. Pfeuffer, J., Bielow, C., Wein, S., Jeong, K., Netz, E., Walter, A., Alka, O., Nilse, L., Colaianni, P. D., McCloskey, D., Kim, J., Rosenberger, G., Bichmann, L., Walzer, M., Veit, J., Boudaud, B., Bernt, M., Patikas, N., Pilz, M., Startek, M. P., Kutuzova, S., Heumos, L., Charkow, J., Sing, J. C., Feroz, A., Siraj, A., Weisser, H., Dijkstra, T. M. H., Perez-Riverol, Y., Röst, H., Kohlbacher, O., and Sachsenberg, T. (2024) OpenMS 3 enables reproducible analysis of large-scale mass

spectrometry data. *Nat. Methods.* **21**, 365–367
27. Dai, C., Pfeuffer, J., Wang, H., Zheng, P., Käll, L., Sachsenberg, T., Demichev, V., Bai, M., Kohlbacher, O., and Perez-Riverol, Y. (2024) quantms: a cloud-based pipeline for quantitative proteomics enables the reanalysis of public proteomics data. *Nat. Methods.* **21**, 1603–1607
28. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372
29. Tyanova, S., Temu, T., and Cox, J. (2016) The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11**, 2301–2319
30. Pinter, N., Glätzer, D., Fahrner, M., Fröhlich, K., Johnson, J., Grüning, B. A., Warscheid, B., Drepper, F., Schilling, O., and Föll, M. C. (2022) MaxQuant and MSstats in Galaxy enable reproducible cloud-based analysis of quantitative proteomics experiments for everyone. *J. Proteome Res.* **21**, 1558–1565
31. Li, J., Xiong, Y., Feng, S., Pan, C., and Guo, X. (2024) CloudProteoAnalyzer: scalable processing of big data from proteomics using cloud computing. *Bioinforma. Adv.* **4**, vbae024