

Dec 27, 2024

🌐 **Comparative proteomic analysis of the composition of decellularized extracellular matrix (dECM) and dECM-based inks as compared to the native tissue**

DOI

<https://dx.doi.org/10.17504/protocols.io.8epv52zr5v1b/v1>

Ainhua Irastorza Lorenzo^{1,2}, Paula Vazquez-Aristizabal³, Lore Zumeta Olaskoaga¹, Maider Mateo¹, Pedro Guerrero^{4,5,6}, Koro de la Caba^{4,5}, Ander Izeta⁷

¹BioGipuzkoa HRI; ²University of the Basque Country (UPV/EHU); ³CIC biomaGUNE; ⁴UPV/EHU; ⁵BCMaterials; ⁶Proteinmat Materials SL; ⁷Biodonostia HRI



Paula Vazquez-Aristizabal

CIC biomaGUNE

Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN  ACCESS



DOI: <https://dx.doi.org/10.17504/protocols.io.8epv52zr5v1b/v1>

Protocol Citation: Ainhua Irastorza Lorenzo, Paula Vazquez-Aristizabal, Lore Zumeta Olaskoaga, Maider Mateo, Pedro Guerrero, Koro de la Caba, Ander Izeta 2024. Comparative proteomic analysis of the composition of decellularized extracellular matrix (dECM) and dECM-based inks as compared to the native tissue. **protocols.io**

<https://dx.doi.org/10.17504/protocols.io.8epv52zr5v1b/v1>

License: This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working

We use this protocol and it's working

Created: November 21, 2024

Last Modified: December 27, 2024

Protocol Integer ID: 112505

Keywords: bioprinted decellularized extracellular matrix, decellularized extracellular matrix, biocompatible ink, source biomaterials for 3d printing, tissue engineering approach, comparative proteomic analysis of the composition, proteomic analysis, comparative proteomic analysis, native tissue regenerative medicine, porcine tissue, native tissue, source biomaterial, biomolecule, tissue layer, mass spectrometry, 3d printing, tissue, regenerative medicine, based ink, liquid chromatography, standard processing of decm, tandem mass spectrometry, comparative liquid chromatography

Funders Acknowledgements:

Instituto de Salud Carlos III (ISCIII) and co-funded by the European Union

Grant ID: PI19/01621, PI22/01247, PT23/00142 and DTS24/00167

Department of Economic Development, Sustainability and Environment of the Basque Government - Elkartek

Grant ID: bMG24; KK-2024/00041

Department of Economic Development, Sustainability and Environment of the Basque Government - Hazitek

Grant ID: ITEAS; ZE-2022/00021

Department of Education of the Basque Government

Grant ID: PRE_2019_1_0031

Asociación Katxalin

Abstract

Regenerative medicine and tissue engineering approaches based on the use of 3D-bioprinted decellularized extracellular matrix (dECM) present the advantage of a relatively biomolecule-rich matrix, which directs cell growth and differentiation in a tissue-specific manner. However, little is known about the composition changes that occur with standard processing of dECM-based inks. To characterize this process, six porcine tissues/tissue layers (artery, breast, dermis, epidermis, muscle and nerve) were independently decellularized via chemical, mechanical and enzymatic processes and the resulting dECMs formulated into biocompatible inks, to serve as source biomaterials for 3D printing. A comparative liquid chromatography–tandem mass spectrometry (LC–MS/MS)-based proteomic analysis was carried out for native tissue, decellularized and formulated ECMs, and the resulting complexity of the matrisome analyzed.

En este protocolo se encuentra descrito la información relativa a las muestras (information and classification), al análisis mediante LC-MS/MS y posterior identificación de las muestras, el procesamiento y curado de los resultados así como la generación de las figuras publicadas.

Samples

1 Sample information and classification

Species: *Sus scrofa domestica*

- **Strain:** Large White
- **Age:** 2 months
- **Sex:** Female
- **Type of samples (by tissue type):**
 - o Aortic artery
 - o Breast
 - o Dorsal dermis
 - o Dorsal epidermis
 - o Biceps femoris muscle
 - o Sciatic nerve

Porcine tissue collection was sourced from cadaveric animals according to Royal Decrees 118/2021 and 53/2013 and the three Rs principles, to ensure the protection of animals used in experiments and other scientific purposes.

1.1 Native, unprocessed tissue: ECM (data set acronym 'E')
N=3 (biological replicates)

1.2 Decellularized tissue: dECM (data set acronym 'D')
N=3 (biological replicates); n=3 (technical replicates); total samples analyzed=9

1.3 Digested dECMs: Ink (I) (data set acronym 'I')

2 Decellularization protocols

Each tissue decellularization protocols are described in the following references:

- Artery, muscle and nerve: <https://doi.org/10.1101/2024.09.23.614437>
- Breast: <https://doi.org/10.1016/j.mtbio.2023.100826>
- Dermis and epidermis: <https://doi.org/10.1002/adhm.202401136>

3 Digestion protocols

Digestion protocols for the generation of dECM-derived inks are described in the work:

<https://doi.org/10.1101/2024.09.23.614437>

LC-MS/MS analysis and data generation

4 LC-MS/MS analysis was performed by CIC bioGUNE Proteomics Platform (<https://www.cicbiogune.es/research/platforms/proteomic>) (Bilbao,

Spain).

- 4.1
 - ECM and dECM samples: 100 mg of cryopreserved (-80 °C) tissues
 - Ink samples: 200 µl of neutralized digestions (4 °C) were sent for their analysis

4.2 In solution digestion

1. Samples were incubated in a sample containing 7 M urea 2 M Thiourea 4 % CHAPS and 5mM DTT for 30 min at RT under agitation
2. Samples were digested following the filter-aided FASP protocol described by Wisniewski *et al.* (<https://doi.org/10.1038/nmeth.1322>) with minor modifications. Trypsin was added to a trypsin:protein ratio of 1:50, and the mixture was incubated overnight at 37 °C
3. Samples were dried out in a RVC2 25 speedvac concentrator (Christ), and resuspended in 0.1 % FA
4. Peptides were desalted and resuspended in 0.1 % FA using C18 stage tips (Millipore).

4.3 Mass spectrometry analysis

Samples were analyzed in a hybrid trapped ion mobility spectrometry (timsTOF Pro with PASEF, Bruker Daltonics) coupled online to a nanoElute liquid chromatograph (Bruker). 200ng of samples were directly loaded in a 15 cm Bruker nanoelute FIFTEEN C18 analytical column (Bruker) and resolved at 400 nL/min with a 100 min gradient. Column was heated to 50 °C using an oven.

4.4 Protein identification and quantification

MaxQuant software was used with default settings (<https://doi.org/10.1038/nbt.1511>). Searches were carried out against a database consisting of pig protein entries (UniProt/Swissprot+TrEMBL). Precursor and fragment tolerances were set to 20 ppm and 0.05 Da.

All the data regarding the analysis is contained in the **"ALL INFO"** sheet of **"RAW"** data file (.xlsx). **"SEL INFO"** sheet of **"RAW"** data file contains selected information about the identifications. Available at **10.5281/zenodo.14195914**. For more information "

⇒ go to step #12 "

Dataset

Raw and processed searches files

NAME

[10.5281/zenodo.14195914](https://doi.org/10.5281/zenodo.14195914)

LINK

Data processing and curation

- 5 Only proteins identified with at least 2 peptides at FDR<1 % were considered for further analysis.
- 6 In the case of uncharacterized proteins, BLAST tool was used (<https://www.UniProt.org/blast>) to complete the information.
 - UniProtKB reference proteomes + Swiss-Prot were used as target databases.
 - Only *Sus scrofa* or *Homo sapiens* results with an identity >98 % were considered.
- 7 Protein lists were completed with the corresponding encoding gene name for each accession number.
- 8 Matrisome-specific proteins were classified by using the MatrisomeDB 2.0 database (<https://matrisomedb.org/>). For that, MatrisomeDB protein list was downloaded (available at <https://sites.google.com/uic.edu/matrisome/home?authuser=0>) and compared to the generated data for each tissue using Excel (Office 2016).
- 9 The LFQ intensity value of each identification was normalized against the total LFQ intensity.

$$Relative\ abundance\ (\%) = \frac{LFQ(x)}{\sum LFQ(x)_n} * 100$$

- 10 Average and standard deviation were calculated for technical replicates (only dECM) from intensity (LFQ) data.
i.e.: $AVG(D1A;D1B;D1C)=D1$
- 11 A file resulting from steps 8-10 is generated and named as '**PROCESSED**'. Available at Zenodo ([doi:10.5281/zenodo.14195914](https://doi.org/10.5281/zenodo.14195914)).



Dataset

Raw and processed searches files

NAME

[10.5281/zenodo.14195914](https://doi.org/10.5281/zenodo.14195914)

LINK

Data information

12 "RAW" DATA FILE

12.1 ALL INFO

■ COLUMNS

- *Protein IDs*: refers to the unique identifiers for the proteins or protein groups based on the UniProt/TrEMBL databases.
- *Majority protein IDs*: protein or proteins within a protein group that have the highest number of unique peptides identified. The protein with the most evidence (peptides) is labelled as the "majority" protein.
- *Peptide counts (all)*: total number of peptides identified for a given protein.
- *Peptide counts (razor+unique)*: number of unique and shared (razor) peptides identified for a given protein.
- *Peptide counts (unique)*: number of unique peptides identified for a given protein.
- *Fasta headers*: contain detailed information from the FASTA database used for the identification. Provides protein description, accession number and information about the species and encoding genes.
- *Number of proteins*: refers to the distinct number of proteins (or protein groups) identified in a given dataset.
- *Peptides*: represents the total number of peptides identified for a given protein or protein group.
- *Razor + unique peptides*: represents the sum of razor peptides and unique peptides used for protein identification and quantification.
- *Unique peptides*: peptides that are specific to that protein and not shared with others.
- *Sequence coverage [%]*: percentage of the protein sequence that is covered by the identified proteins. Provides a measure of how much of the total protein sequence has been matched with peptides.
- *Unique + razor sequence coverage [%]*: the percentage of the protein sequence

that is covered by the sum of razor and unique peptides.

- *Unique sequence coverage [%]*: the percentage of the protein sequence that is covered only by unique peptides.
- *Mol. Weight*: mass of the protein in kDa.
- *Sequence length(s)*: refers to the total number of amino acids in a given protein or peptide sequence
- *Q-value*: is the False Discovery Rate (FDR) adjusted p-value for a particular peptide or protein identification. It gives a measure of confidence in the identification, representing the minimum FDR at which a particular identification is considered significant.
- *Score*: the confidence score assigned to a peptide or protein identification by the search algorithm. It is based on the quality of the match between the measured MS/MS spectra and the theoretical spectra for a peptide.
- *Identification type*: specifies the method of identification. Commonly, a direct match to an MS/MS spectrum identifies the peptide.
- *Intensity*: the peak intensity of the identified peptides or protein in the MS1 (precursor ion) spectrum. It is used as a proxy for the relative abundance of the protein.
- *LFQ intensity*: stands for Label-Free Quantification Intensity and reflects the relative abundance of a protein in a sample based on the intensity of the detected precursor ions (MS1 data) across multiple samples.
- *MS/MS count*: (or Spectral Counts) represents the number of times MS/MS spectra were collected for a given peptide or protein during the LC-MS/MS analysis. It represents the confidence in identification.
- o *Only identified by site*: indicates that the protein was only identified by site-specific modification peptides, such as phosphorylation or oxidation.
- *Reverse*: is assigned to decoy hits, proteins from a reversed or randomized version of the database used to estimate the FDR. Should be excluded from final analysis.
- *Potential contaminant*: marks proteins that are identified as common contaminants (e.g., keratins, trypsin or other proteins that may come from lab material or sample handling).
- *ID*: unique identifier for each protein in the MaxQuant output.
- *Peptide IDs*: refers to the unique identifier assigned to each peptide in the MaxQuant peptides table.
- *Peptide is razor*: indicated whether the peptide is a razor peptide, meaning it is assigned to a protein group but is not unique to a single protein in that group.
- *Mod. peptide IDs*: unique identifiers for modified peptides.
- *Evidence IDs*: links to the protein or peptide to the evidence table in MaxQuant, provides information on the spectra used to identify peptides and proteins.
- *MS/MS IDs*: link the identification to the specific MS/MS spectra used in the analysis.
- *Best MS/MS*: refers to the best-scoring MS/MS

spectrum used to identify a particular peptide. The best-scoring one is considered the most reliable for identification.

- *Oxidation (M) site IDs*: refers to the unique identifier for peptides that have oxidation on methionine residues (M).
- *Oxidation (M) site positions*: specifies the amino acid position where the oxidation modification occurred on the methionine residue in the peptide or protein sequence.
- *Taxonomy IDs*: unique numerical identifiers that refer to the species or organism from which the protein or peptide originates. These are sourced from the NCBI taxonomy database.

- **ROWS**: Each row represents an identified protein.

12.2 SEL INFO

- **COLUMNS**

- *Fasta headers*: used database for the identification Swissprot (sp) or TrEMBL (tr).
- *Accession*: unique identifier of each protein based on UniProt.
- *Description*: gives information about the accession number, organism name (OS), organism taxonomy (OX), protein and gene name (GN), protein evidence level (PE) and sequence version (SV).
- *Mol. Weight*: mass of the protein in kDa.
- *Number of proteins*: number of distinct protein groups identified in the analysis.
- *Peptides*: total number of peptides identified for a particular protein.
- *Unique peptides*: number of peptides that are specific to a single protein.
- *Razor + unique peptides*: the sum of the razor peptides, which are peptides that are shared among proteins; and unique peptides.
- *LFQ intensity (Label-Free Quantification)*: is a measure of the relative abundance of proteins. LFQ is calculated by comparing the signal intensities of peptides across different experimental conditions or samples.
- *MS/MS Counts*: refers to spectral counts, which represent the number of MS/MS spectra associated with peptides from a given protein.
- *SAF (Spectral Abundance Factor)*: is used to normalize spectral counts based on protein length. For that, the count are divided by the length of the protein.
- *NSAF (Normalized Spectral Abundance Factor)*: a further refinement of SAF that normalizes across all proteins in the dataset, allowing comparison of abundances across different samples.

- **ROWS**: Each row represents an identified protein.

Figure generation

13 **FIGURE 2.** Representation of the contribution as well as conservation of the matrisome and its different categories for the different tissues and samples.

13.1 The total number of identifications (Matrisome and non-Matrisome) in 'SEL INFO' was considered for this figure.
i.e.: LFQ intensity results for each biological replicate.

	E1	E2	E3	HIT SCORE
	100	25000	0	1
	0	0	0	0

1: GENE 1
2: GENE 2

13.2 The lists of ECM-related genes (see step 5) for each tissue (artery, breast, dermis, epidermis, muscle and nerve) and step (ECM, dECM, Ink) were submitted to Flaski 3.12.2 (<https://flaski.age.mpg.de/home/>) for the generation of the Venn diagrams.

14 **FIGURE 3.** Analysis of the similarities between native-ECM matrisomal protein profiles.

14.1 The lists of genes with a hit score=1 described in step 13.1 for each tissue ECM were submitted to InteractiVenn (<https://www.interactivenn.net>) for the generation of Figure 3A.

14.2 A matrix was built with the data explained in steps 8-11, comparing the intensity of the genes between ECM tissue samples, and named as '**ALL TISSUES PROCESSED**' file available at Zenodo (doi:10.5281/zenodo.14195914).

Dataset

Raw and processed searches files

NAME

[10.5281/zenodo.14195914](https://doi.org/10.5281/zenodo.14195914)

LINK

i.e.: Processed file

	TISSUE1 E1	TISSUE1 E2	TISSUE1 E3	TISSUE2 E1	TISSUE2 E2	TISSUE2 E3
	0,1	0,3	0,2	10,1	5,0	0,7
	0,005	2,5	1,1	0,0	0,0001	25,0

1: GENE1
2: GENE2

14.3 The matrix (available at **10.5281/zenodo.14195915**) was submitted to MetaboAnalyst 6.0 and analyzed according to 'Statistical analysis [one factor]' module.

<https://dev.metaboanalyst.ca/MetaboAnalyst/upload/StatUploadView.xhtml>

14.4 Parameters:

- Concentrations
- Samples in columns (unpaired)

Submit and proceed

14.5 Data filtering

- Reliability filter: RSD greater than 10%
- Variance filter: IQR 0%
- Abundance filter: Mean intensity value 0%

Proceed

14.6 Normalization

- Sample normalization: none
- Data transformation: none
- Data scaling: auto scaling

Normalize and proceed

14.7 Chemometric analysis **Partial Least Squares - Discriminant Analysis (PLS-DA)** was selected. 'Group order matters' option was deselect.

A cross-validation was carried out with 4 maximum components to search, 5-fold CV method and Q2 as performance measure, to obtain the number of components (**Figure S6**).

14.8 Within the PLS-DA analysis tool, 2D scores plot were used to represent the covariance between the components selected in step 14.7.

14.9 Variable importance in projection (VIP) were also plotted for each component using the top 30 features. The colored boxes on the right indicate the relative concentrations of the corresponding metabolite in each group under study.

- 15 **FIGURE 4.** Proportions of the different matrisomal protein categories in each tissue at the ECM, dECM and ink processing stages.

A matrix was built with the data explained in steps 8-11, comparing the intensity of the genes between step-samples (ECM, dECM and ink) within each tissue, and named as '[TISSUE] PROCESSED'. Files are available at Zenodo (doi:10.5281/zenodo.14195914).

These results were plotted using the categorized data according to MatrisomeDB.

- 16 **FIGURE 5.** Heatmap of matrisome proteins of each tissue ECM.

The 'ALL TISSUES PROCESSED' file generated in step 14.2 was submitted to MetaboAnalyst 6.0, choosing 'Heatmap' in this step.

- 16.1
- Settings were left as default
 - Data source: Normalized data
 - Standarization: Autoscale features
 - Distance measure: Euclidean
 - Clustering method: Ward
 - Other view options: Samples, T-test/ANOVA
 - Top 100 features were represented.

- 17 **FIGURE 6.** PLS-DA analysis of ECM, dECM and ink samples for each studied tissue.

Steps 14.3-14.8 were replicated for '[TISSUE] PROCESSED' datasets.

- 18 **FIGURE 7.** Representation of proportion of the most abundant matrisomal proteins for each tissue and sample.

These results were plotted using the identified proteins with an abundance >1% for each tissue and sample based on the data generated in step 11 ('[TISSUE] PROCESSED' datasets).

- 19 **FIGURE S2.** Heatmap of the matrisome proteins coverage shared between native tissues.

The coverage between tissues was calculated based on the dataset 'ALL TISSUES PROCESSED', by using the number of common and total identifications.

- 20 **FIGURE S3.** Adjusted p-values of biological processes gene ontology (GO) common to every tissue.

'ALL TISSUES PROCESSED' file generated in step 14.2 was used for this figure.

- 20.1 The list of identifications for each tissue ECM was submitted to <https://maayanlab.cloud/Enrichr/>. Parameters were set in order to include the top 500

most relevant proteins encoding genes.

20.2 'GO Biological Process 2023' was selected among the 'Ontologies' tab

20.3 The results from 'Table' were sorted according to the adjusted *p*-value, and the common concepts among the tissues were selected. Then, the adjusted *p*-value for each tissue and concept was plotted.

21 **FIGURE S4.** Collagen types present within the collagen family for each tissue ECM, dECM and ink.

The proportion of collagen isoforms for each tissue and sample was calculated and plotted based on the '[TISSUE] PROCESSED' datasets.

22 **FIGURE S5.** PLSDA-based VIP-scores of artery, breast, dermis, epidermis, muscle and nerve ECM, dECM and ink for each representative component.

According to step 17, steps 14.3-14.9 were replicated for '[TISSUE] PROCESSED' datasets. In the case of step 14.9, only the top 15 features were represented.

23 **FIGURE S6.** Cross validation of the PLS-DA results of ECM, artery, breast, dermis, epidermis, muscle and nerve.

As explained in 14.7, cross-validation was carried out with 4 maximum components to search, 5-fold CV method and Q2 as performance measure, to obtain the number of components for 'ALL TISSUES PROCESSED' and '[TISSUE] PROCESSED' datasets.

24 **TABLE S3.** Quantification of the number and proportion of the total as well as matrisome proteins for each tissue and sample.

The matrix generated in step 13.1 was used for the calculation of the number of identifications and proportions (of the total) of matrisome and non-matrisome proteins for each tissue and sample type. Median, range and standard deviation (SD) of the sample types were also included in the table.

25 **TABLE S4.** GO terms, adjusted *p*-values and related gene names for each tissue ECM.

Based on the results obtained in the FIGURE 5, the proteins identified for each tissue cluster were analysed according to steps 20.1 and 20.2. GO terms, adjusted *p*-values are related genes are indicated.