

ECM transcriptome dynamics during aging

Mazalov Aleksei (PhD 1), Malygina Alexandra (MSc 2), Shipulina Eva (MSc 2)

Bulk RNA-seq Analysis (reproducing the main results reported in [1])

Data Source & Metadata

- RNA-seq data for analysis were obtained from
https://www.gtexportal.org/home/downloads/adult-gtex/bulk_tissue_expression
(the v8 subsection)
- Sample and donor metadata were obtained from
<https://storage.googleapis.com/adult-gtex/annotations/v8/metadata-files>
- Genes were classified by type (*core matrisome, matrisome associated, other*) and category as per [1]

| | | | |
|---|-------------|-------|-----------|
| > as.matrix(colnames(metadata_article)) | | | |
| [,1] | "SAMPID" | | |
| [1,] | "SMATSSCR" | | |
| [2,] | "SMCENTER" | | |
| [3,] | "SMPHTNNTS" | | |
| [4,] | "SMRIN" | | |
| [5,] | "SMTS" | | |
| [6,] | "SMTSD" | | |
| [7,] | "SMURRTD" | | |
| [8,] | | | |
| > colnames(subj_metadata) | | | |
| [1] | "SUBJID" | "SEX" | "AGE" |
| | | | "DTHHRDY" |
| Categories | | | |
| Core matrisome | | | |
| Collagens | | | |
| ECM Glycoproteins | | | |
| Proteoglycans | | | |
| Matrisome associated | | | |
| ECM-affiliated Proteins | | | |
| ECM Regulators | | | |
| Secreted Factors | | | |
| | | | [,1] |
| | Blood | 755 | |
| | Heart | 705 | |
| | Muscle | 641 | |
| | Brain | 432 | |
| | Lung | 431 | |
| | Liver | 177 | |
| | Kidney | 73 | |

[1]. MatrisomeDB: 2023 updates of the ECM protein knowledge database. Shao X, Gomez CD, Kapoor N, Considine JM, Gao Y, Naba A. *Nucleic Acids Research*, 2022, gkac1009. doi.org/10.1093/nar/gkac1009

Data Filtering & Analysis Setup

- Gene expression analyzed only for samples collected from the ***blood, brain, heart, kidney, liver, lungs, muscle***
- Analysis performed separately for samples collected from males and females
- No detailed information provided by the authors on the filters used (or the source data was modified lately) => discrepancies in the cohort sizes between our analysis and the one reported.

Our filtering result

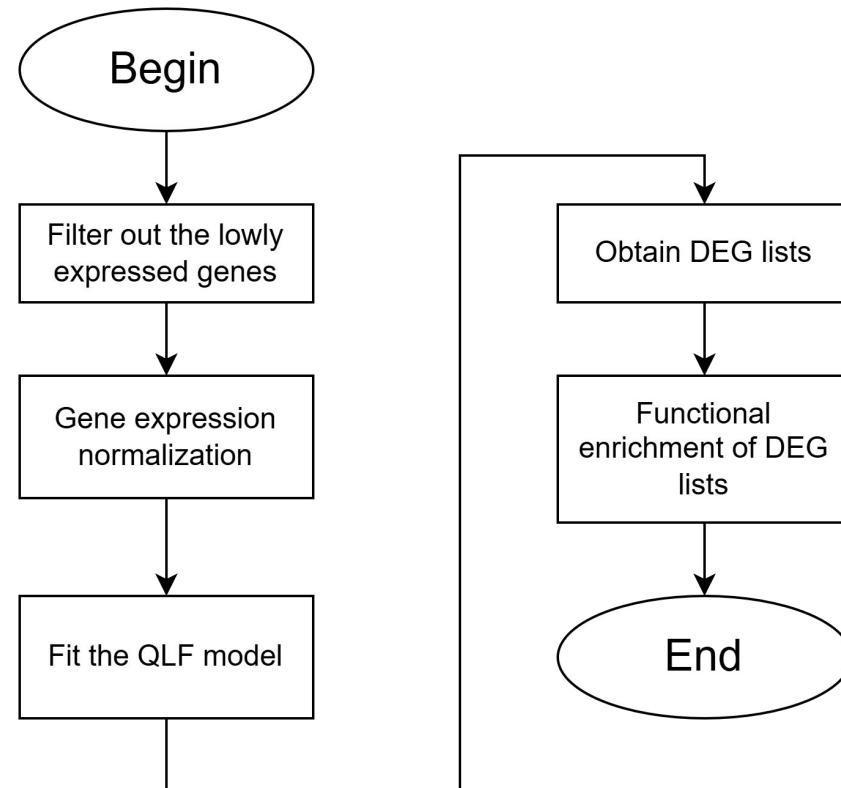
| Sex | Age Group | Blood | Brain | Heart | Kidney | Liver | Lungs | Muscle |
|--------|-----------|-------|-------|-------|--------|-------|-------|--------|
| Male | 20-39 | 92 | 14 | 29 | 4 | 12 | 35 | 69 |
| | 40-59 | 226 | 130 | 229 | 24 | 64 | 143 | 191 |
| | 60-79 | 183 | 175 | 221 | 28 | 51 | 123 | 180 |
| Female | 20-39 | 44 | 6 | 24 | 1 | 4 | 15 | 28 |
| | 40-59 | 121 | 43 | 110 | 8 | 25 | 59 | 102 |
| | 60-79 | 89 | 64 | 92 | 8 | 21 | 56 | 71 |

Supplementary Table S1. GTEx cohort characteristics by tissue.

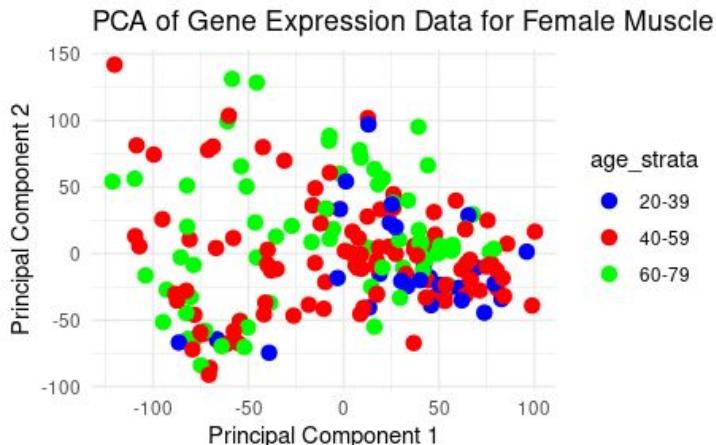
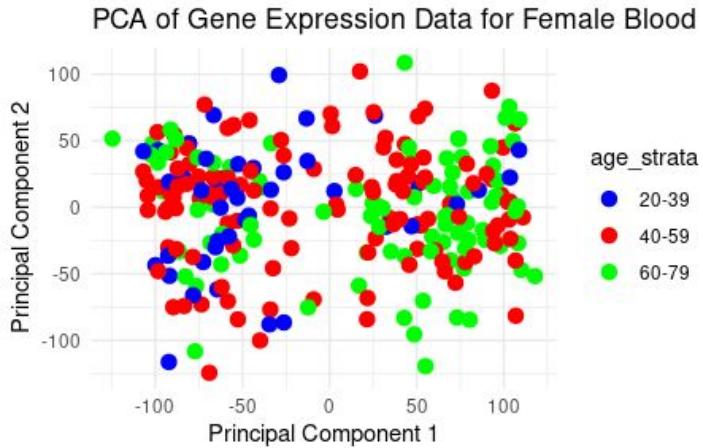
| Sex | Age group | Blood | Brain | Heart | Kidney | Liver | Lungs | Muscle |
|--------|-----------|-------|-------|-------|--------|-------|-------|--------|
| Male | 20-39 | 92 | 2 | 17 | 7 | 18 | 50 | 93 |
| | 40-59 | 226 | 20 | 103 | 29 | 84 | 105 | 245 |
| | 60-79 | 183 | 27 | 83 | 30 | 59 | 140 | 205 |
| Female | 20-39 | 44 | 3 | 11 | 2 | 6 | 23 | 39 |
| | 40-59 | 121 | 10 | 52 | 8 | 34 | 88 | 134 |
| | 60-79 | 89 | 8 | 34 | 9 | 25 | 72 | 87 |

Data Analysis Tools & Pipeline

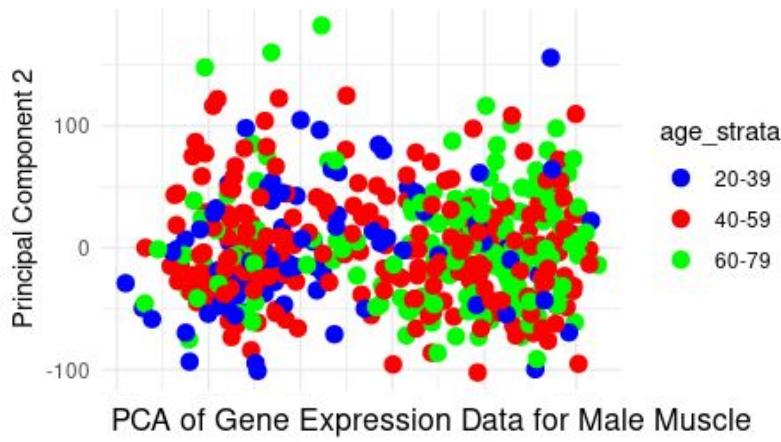
- EdgeR used for RNA-seq analysis: determining the DEGs between age 2 and 3 age groups + PCA analysis
- StringDB used for functional enrichment and exploring interprotein interactions



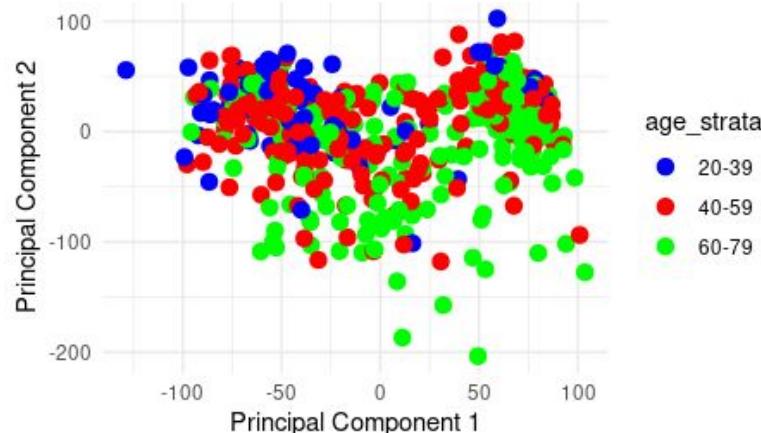
PCA analysis



PCA of Gene Expression Data for Male Blood

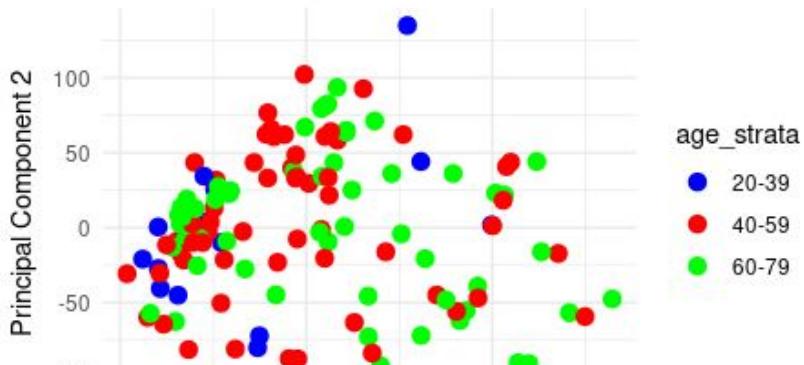


PCA of Gene Expression Data for Male Muscle

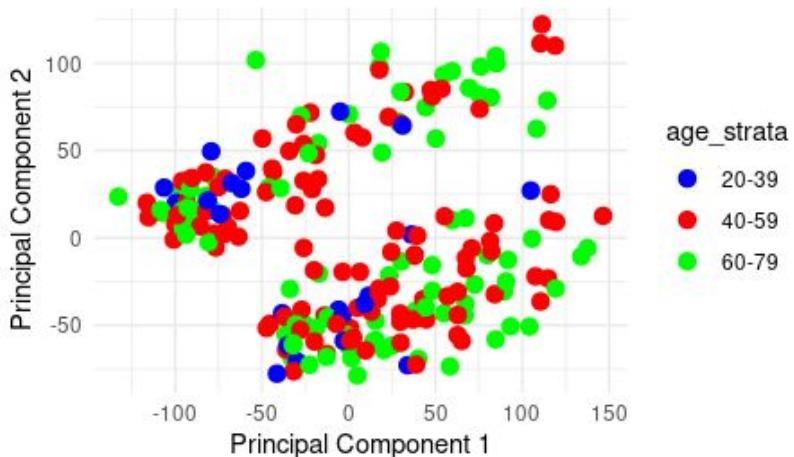


PCA analysis (all genes)

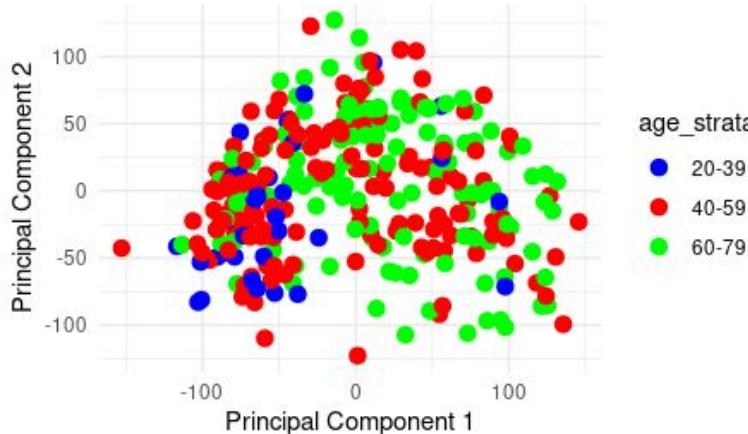
PCA of Gene Expression Data for Female Lung



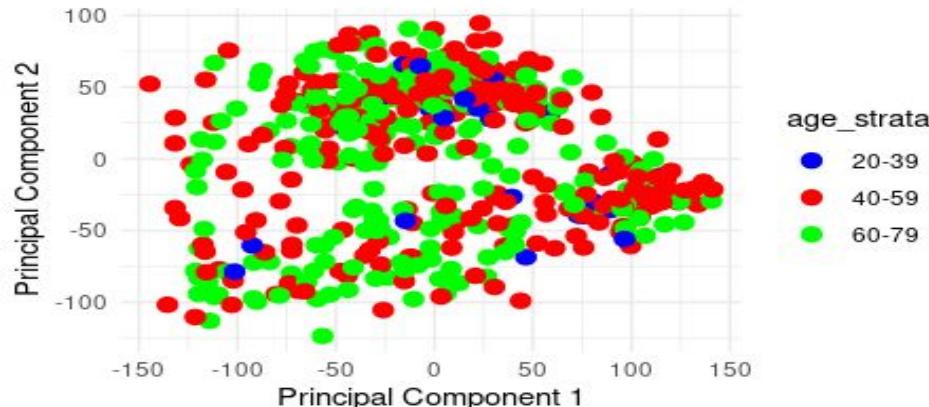
PCA of Gene Expression Data for Female Heart



PCA of Gene Expression Data for Male Lung



PCA of Gene Expression Data for Male Heart



DEG Analysis Results

| Tissue/Organ | Female | | | | Male | | | |
|--------------|-------------|----|---------------|----|-------------|-----|---------------|----|
| | Upregulated | | Downregulated | | Upregulated | | Downregulated | |
| | C | MA | C | MA | C | MA | C | MA |
| Blood | 32 | 98 | 16 | 73 | 34 | 119 | 22 | 92 |
| Heart | 54 | 71 | 1 | 16 | 53 | 129 | 11 | 47 |
| Brain | 15 | 37 | 10 | 17 | 19 | 77 | 1 | 19 |
| Liver | 6 | 16 | 5 | 22 | 4 | 11 | 9 | 11 |
| Kidneys | 0 | 2 | 8 | 23 | 6 | 25 | 5 | 17 |
| Lung | 57 | 85 | 2 | 26 | 87 | 172 | 11 | 93 |
| Muscle | 46 | 74 | 3 | 33 | 100 | 169 | 4 | 36 |

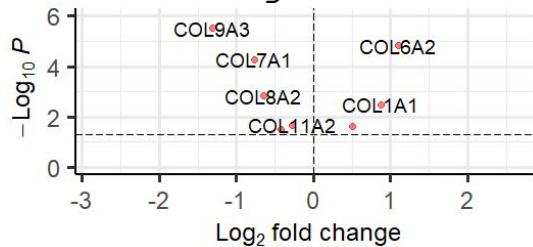
Our analysis,
 $\log\text{CPM} > 0$,
 $p\text{-value} < 0.05$

| Tissue/Organ | Female | | | | Male | | | |
|--------------|-------------|----|---------------|----|-------------|-----|---------------|-----|
| | Upregulated | | Downregulated | | Upregulated | | Downregulated | |
| | C | MA | C | MA | C | MA | C | MA |
| Blood | 29 | 99 | 17 | 79 | 33 | 129 | 27 | 98 |
| Heart | 27 | 21 | 0 | 15 | 43 | 87 | 10 | 50 |
| Brain | 19 | 35 | 6 | 25 | 7 | 19 | 13 | 25 |
| Liver | 8 | 23 | 7 | 36 | 4 | 15 | 16 | 55 |
| Kidneys | 0 | 15 | 1 | 8 | 11 | 26 | 10 | 18 |
| Lung | 78 | 99 | 7 | 49 | 106 | 194 | 10 | 107 |
| Muscle | 54 | 84 | 4 | 34 | 106 | 164 | 19 | 38 |

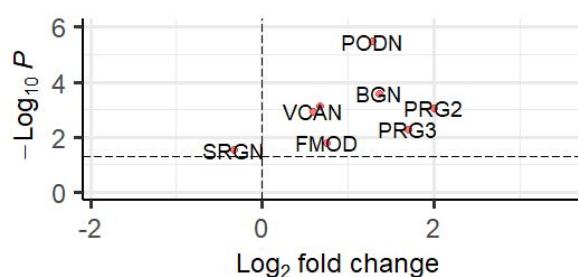
Reported results,
 $\log\text{CPM} > 0$,
 $p\text{-value} < 0.05$

DEG Analysis Results. ECM genes in Female Blood

Collagens



Proteoglycans



ECM glycoproteins



collagens

COL6A2
COL9A3
COL7A1

FGG
FGB
FGA
SPP1
CRISPLD2
TNFAIP6
AEBP1
MFAP3

ECM glycoproteins

SPON2
THBS1
LTBP4
COLQ
TNXB
LAMB3
LRG1
FGL2
FBN2
LAMC1
THBS3
MFAP1

proteoglycans

PRG2
PRG3
SPOCK2
VCAN
SRGN

$\log\text{CPM}>2$, $p\text{-value} < 0.05$

DEG Analysis Results. ECM genes in Male Blood

| Description | logFC | FDR | gene_subtype |
|-------------|------------|--------------|--------------|
| COL6A2 | 1.0972975 | 1.716062e-04 | Collagen |
| COL1A1 | 0.8768284 | 1.041375e-02 | Collagen |
| COL18A1 | -0.4165368 | 6.516823e-02 | Collagen |
| COL7A1 | -0.7600833 | 4.349153e-04 | Collagen |
| COL9A3 | -1.3141332 | 5.426705e-05 | Collagen |

| Description | logFC | FDR | gene_subtype |
|-------------|------------|--------------|---------------|
| PRG3 | 2.6961919 | 5.752495e-08 | Proteoglycans |
| PRG2 | 2.3969168 | 6.535625e-07 | Proteoglycans |
| SPOCK2 | 0.7392273 | 2.755983e-06 | Proteoglycans |
| BGN | 0.5864127 | 2.288502e-02 | Proteoglycans |
| VCAN | 0.5314022 | 1.212506e-04 | Proteoglycans |
| SRGN | -0.4705252 | 2.628872e-05 | Proteoglycans |

logCPM > 2, p-value < 0.05

| Description | logFC | FDR | gene_subtype |
|-------------|------------|--------------|------------------|
| FGG | 2.4200640 | 9.467534e-07 | ECM_glycoprotein |
| FGA | 2.1350490 | 4.471059e-06 | ECM_glycoprotein |
| FGB | 1.8538164 | 6.640801e-05 | ECM_glycoprotein |
| VWCE | 1.7856452 | 5.893204e-12 | ECM_glycoprotein |
| THBS1 | 1.3948098 | 7.517501e-08 | ECM_glycoprotein |
| COLQ | 0.9301354 | 6.332205e-08 | ECM_glycoprotein |
| LTBP4 | 0.8758652 | 4.399225e-10 | ECM_glycoprotein |
| SPP1 | 0.6301921 | 6.311077e-02 | ECM_glycoprotein |
| FN1 | 0.6219741 | 1.167423e-02 | ECM_glycoprotein |
| LTBP3 | 0.4362056 | 4.965415e-04 | ECM_glycoprotein |
| CRISPLD2 | -1.0885172 | 4.868042e-10 | ECM_glycoprotein |
| PCOLCE2 | -0.8799335 | 7.845014e-05 | ECM_glycoprotein |
| LRG1 | -0.7308963 | 3.713276e-04 | ECM_glycoprotein |
| VWA5A | -0.7120060 | 6.903215e-08 | ECM_glycoprotein |
| EFEMP2 | -0.5037302 | 7.614181e-06 | ECM_glycoprotein |
| MFAP1 | -0.4678938 | 1.090727e-07 | ECM_glycoprotein |
| THBS3 | -0.4138677 | 1.444837e-06 | ECM_glycoprotein |

Reported DEG Analysis Results. ECM genes in Blood

| Tissue/Organ | Category | Upregulated | | Downregulated | |
|--------------|-------------------|---------------|----------------|----------------|---|
| | | Female | Male | Female | Male |
| Blood | ECM glycoproteins | collagens | <i>COL6A2</i> | <i>COL6A2</i> | <i>COL9A3</i> <i>COL7A1</i> <i>COL9A2</i> <i>COL18A1</i> |
| | | | <i>FGG</i> | <i>FGG</i> | <i>CRISPLD2</i> |
| | | | <i>FGB</i> | <i>VWCE</i> | <i>TNFAIP6</i> |
| | | | <i>FGA</i> | <i>THBS1</i> | <i>CRISPLD2</i> |
| | | | <i>SPP1</i> | <i>FGA</i> | <i>AEBP1</i> |
| | | | <i>SPON2</i> | <i>SPON2</i> | <i>MFAP3</i> |
| | | | <i>THBS1</i> | <i>FGB</i> | <i>LRG1</i> |
| | | | <i>LTBP4</i> | <i>COLQ</i> | <i>VWA5A</i> |
| | | | <i>COLQ</i> | <i>LTBP4</i> | <i>EFEMP2</i> |
| | | | <i>TNXB</i> | <i>IGFBP4</i> | <i>LAMC1</i> |
| | | | <i>LAMB3</i> | <i>LTBP3</i> | <i>MFAP1</i> |
| | | proteoglycans | <i>PRG2</i> | <i>PRG3</i> | <i>THBS3</i> |
| | | | <i>PRG3</i> | <i>PRG2</i> | <i>THBS3</i> |
| | | | <i>SPOCK2</i> | <i>SPOCK2</i> | <i>GAS6</i> |
| | | | <i>VCAN</i> | <i>VCAN</i> | |
| | | collagens | <i>COL6A3</i> | | <i>SRGN</i> |
| | | | <i>COL1A1</i> | | <i>SRGN</i> |
| | | | <i>COL8A2</i> | | |
| | | | <i>COL1A2</i> | | |
| | | | <i>COL16A1</i> | | |
| | | | | <i>COL27A1</i> | <i>COL26A1</i> |

logCPM > 2, p-value < 0.05

DEG Analysis Results. Matr. Assoc. genes in Male Blood

| Description | logFC | FDR | gene_subtype | Description | logFC | FDR | gene_subtype | Description | logFC | FDR | gene_subtype |
|-------------|------------|--------------|------------------|-------------|------------|--------------|----------------|-------------|------------|--------------|----------------|
| IGF2 | 2.5008682 | 1.748333e-10 | Secreted factors | HRG | 3.3727695 | 1.965179e-09 | ECM regulators | SDC2 | 1.1559916 | 1.356018e-06 | ECM affiliated |
| CCL3 | 2.2004194 | 2.864790e-09 | Secreted factors | AMBP | 2.5797060 | 3.084958e-07 | ECM regulators | ITLN1 | 1.0621543 | 1.678314e-03 | ECM affiliated |
| CCL4L2 | 2.0848667 | 2.345728e-10 | Secreted factors | SERPINH1 | 1.7772279 | 4.964479e-08 | ECM regulators | PLXDC1 | 1.0332113 | 3.926363e-08 | ECM affiliated |
| TNFSF9 | 1.9090413 | 1.854429e-08 | Secreted factors | SERPINA3 | 1.6466851 | 7.128530e-06 | ECM regulators | SDC4 | 0.9070199 | 1.058836e-07 | ECM affiliated |
| AREG | 1.6918613 | 2.689206e-08 | Secreted factors | PLAU | 1.5911112 | 7.769653e-06 | ECM regulators | PLXNA3 | 0.8694726 | 2.365615e-08 | ECM affiliated |
| LIF | 1.5909113 | 3.477499e-04 | Secreted factors | TGM3 | 1.3852067 | 5.509492e-06 | ECM regulators | SEMA4C | 0.8080364 | 4.646054e-09 | ECM affiliated |
| VEGFA | 1.4370014 | 8.894309e-08 | Secreted factors | CTSW | 1.3239998 | 1.749443e-09 | ECM regulators | CLEC2B | 0.7605047 | 1.233481e-08 | ECM affiliated |
| HBEGF | 1.2931964 | 6.832705e-06 | Secreted factors | SERPINB2 | 1.2371238 | 1.491554e-06 | ECM regulators | CLEC11A | 0.7513596 | 6.527319e-05 | ECM affiliated |
| IL1RN | -0.8175009 | 4.412686e-08 | Secreted factors | CTSG | 1.1614011 | 1.022408e-03 | ECM regulators | C1QC | 0.7459376 | 4.566718e-03 | ECM affiliated |
| MEGF9 | -0.8207596 | 6.262274e-08 | Secreted factors | SERPINE1 | 1.1533815 | 2.503985e-05 | ECM regulators | C1QA | 0.7339886 | 9.782455e-04 | ECM affiliated |
| S100A4 | -0.8472878 | 3.715264e-09 | Secreted factors | SERPINB8 | -0.7100229 | 9.279917e-12 | ECM regulators | SEMA4B | -0.6299925 | 1.895395e-07 | ECM affiliated |
| TNFSF14 | -0.8759828 | 1.521736e-07 | Secreted factors | MMP8 | -0.7228314 | 7.458185e-03 | ECM regulators | PLXNC1 | -0.7496036 | 7.207472e-07 | ECM affiliated |
| S100A8 | -0.8784909 | 6.325390e-05 | Secreted factors | MMP25 | -0.7379472 | 7.561588e-05 | ECM regulators | CLEC1B | -0.7811976 | 2.100740e-05 | ECM affiliated |
| TGFA | -0.8944810 | 2.519802e-07 | Secreted factors | CTSK | -0.7429314 | 2.519705e-13 | ECM regulators | ANXA1 | -0.7982271 | 1.364615e-07 | ECM affiliated |
| S100A9 | -0.9054147 | 6.723969e-07 | Secreted factors | CSTA | -0.7500154 | 7.244520e-08 | ECM regulators | CLEC4A | -0.8516600 | 4.720722e-11 | ECM affiliated |
| S100A12 | -0.9662648 | 1.434318e-05 | Secreted factors | ADAMTS2 | -0.7505984 | 1.789789e-02 | ECM regulators | ANXA3 | -0.9369413 | 6.802901e-06 | ECM affiliated |
| PDGFC | -1.0365963 | 3.546775e-06 | Secreted factors | ADAM17 | -0.7508632 | 1.283152e-08 | ECM regulators | CLEC12B | -0.9895588 | 2.282027e-07 | ECM affiliated |
| TNFSF10 | -1.2624050 | 4.963350e-09 | Secreted factors | EGLN1 | -0.7649804 | 8.623554e-08 | ECM regulators | CLEC4D | -1.0608045 | 9.194342e-06 | ECM affiliated |
| | | | | SERPINB1 | -0.8477462 | 1.678013e-06 | ECM regulators | ANXA9 | -1.0711524 | 2.281853e-12 | ECM affiliated |
| | | | | HPSE | -1.0798299 | 3.854136e-14 | ECM regulators | LGALSL | -1.2334179 | 1.097280e-12 | ECM affiliated |

Reported DEG Analysis Results. Matr. Assoc. genes in Blood

| Tissue/ Organ | Category | Upregulated ↑ | | Downregulated ↓ | |
|-------------------------|----------|------------------------|------------------------|-------------------------|-------------------------|
| | | Female | Male | Female | Male |
| ECM Regulators | Blood | <i>AMBP</i> (1.79) | <i>HRG</i> (1.47) | <i>PRSS2</i> (-0.9) | <i>HPSE</i> (-0.55) |
| | | <i>SERPINA3</i> (1.5) | <i>SERPINH1</i> (0.91) | <i>HPSE</i> (-0.43) | <i>HYAL2</i> (-0.46) |
| | | <i>HRG</i> (1.34) | <i>AMBP</i> (0.89) | <i>MMP25</i> (-0.41) | <i>SERPINB1</i> (-0.43) |
| | | <i>SERPINH1</i> (0.97) | <i>PLAU</i> (0.82) | <i>EGLN1</i> (-0.38) | <i>ADAMTS2</i> (-0.42) |
| | | <i>CTSW</i> (0.81) | <i>TGM3</i> (0.7) | <i>HYAL2</i> (-0.31) | <i>MMP25</i> (-0.41) |
| | | <i>SERPINB2</i> (0.77) | <i>CTSW</i> (0.66) | <i>ST14</i> (-0.27) | <i>EGLN1</i> (-0.4) |
| | | <i>PLAU</i> (0.62) | <i>CTSG</i> (0.64) | <i>CTSK</i> (-0.26) | <i>CTSK</i> (-0.37) |
| | | <i>P4HA1</i> (0.57) | <i>SERPINB2</i> (0.63) | <i>ADAM17</i> (-0.26) | <i>ADAM17</i> (-0.37) |
| | | <i>CTSG</i> (0.57) | <i>ELANE</i> (0.53) | <i>SERPINB8</i> (-0.25) | <i>CSTA</i> (-0.37) |
| | | <i>ADAMTS10</i> (0.51) | <i>MMP19</i> (0.52) | <i>ADAM19</i> (-0.24) | <i>SERPINB8</i> (-0.36) |
| ECM-affiliated Proteins | Blood | <i>SDC2</i> (0.86) | <i>CLC</i> (0.71) | <i>SFTPB</i> (-1.02) | <i>CLEC4E</i> (-0.67) |
| | | <i>C1QC</i> (0.54) | <i>SDC2</i> (0.59) | <i>ANXA9</i> (-0.49) | <i>ANXA9</i> (-0.54) |
| | | <i>C1QB</i> (0.47) | <i>PLXDC1</i> (0.54) | <i>CLEC4A</i> (-0.36) | <i>CLEC4D</i> (-0.53) |
| | | <i>C1QA</i> (0.46) | <i>SDC4</i> (0.46) | <i>CLEC4E</i> (-0.34) | <i>CLEC12A</i> (-0.49) |
| | | <i>CLC</i> (0.44) | <i>PLXNA3</i> (0.45) | <i>SEMA4B</i> (-0.34) | <i>ANXA3</i> (-0.48) |
| | | <i>PLXNA3</i> (0.43) | <i>CLEC11A</i> (0.42) | <i>ANXA3</i> (-0.32) | <i>CLEC12B</i> (-0.48) |
| | | <i>CLEC2B</i> (0.42) | <i>SEMA4C</i> (0.4) | <i>PLXNA2</i> (-0.29) | <i>ANXA1</i> (-0.41) |
| | | <i>CLEC11A</i> (0.42) | <i>C1QC</i> (0.36) | <i>CLEC12A</i> (-0.29) | <i>CLEC4A</i> (-0.41) |
| | | <i>SEMA4C</i> (0.34) | <i>CLEC2B</i> (0.36) | <i>CLEC4D</i> (-0.28) | <i>CLEC1B</i> (-0.39) |
| | | <i>ANXA6</i> (0.31) | <i>C1QA</i> (0.35) | <i>PLXNC1</i> (-0.27) | <i>PLXNC1</i> (-0.38) |
| Secreted Factors | Blood | <i>IGF2</i> (1.29) | <i>CCL3L3</i> (1.13) | <i>TNFSF10</i> (-0.53) | <i>TNFSF10</i> (-0.66) |
| | | <i>IFNG</i> (1.24) | <i>IGF2</i> (1.12) | <i>INHBB</i> (-0.46) | <i>PDGFC</i> (-0.53) |
| | | <i>CCL3</i> (1.15) | <i>CCL3</i> (1.08) | <i>IL1B</i> (-0.44) | <i>S100A12</i> (-0.5) |
| | | <i>CCL4</i> (1.03) | <i>CCL4L2</i> (1.05) | <i>IL1RN</i> (-0.43) | <i>S100A9</i> (-0.47) |
| | | <i>LEP</i> (1.02) | <i>CCL4</i> (0.93) | <i>TNFSF14</i> (-0.42) | <i>S100A8</i> (-0.47) |
| | | <i>LIF</i> (0.98) | <i>TNFSF9</i> (0.93) | <i>S100A4</i> (-0.34) | <i>TNFSF14</i> (-0.46) |
| | | <i>CCL4L2</i> (0.92) | <i>EREG</i> (0.84) | <i>INSL3</i> (-0.34) | <i>TGFA</i> (-0.45) |
| | | <i>CCL5</i> (0.81) | <i>XCL2</i> (0.81) | <i>TGFA</i> (-0.33) | <i>S100A4</i> (-0.44) |
| | | <i>AREG</i> (0.8) | <i>AREG</i> (0.81) | <i>MEGF9</i> (-0.32) | <i>MEGF9</i> (-0.41) |
| | | <i>VEGFA</i> (0.8) | <i>PRL</i> (0.76) | <i>PPBP</i> (-0.29) | <i>IL1RN</i> (-0.4) |

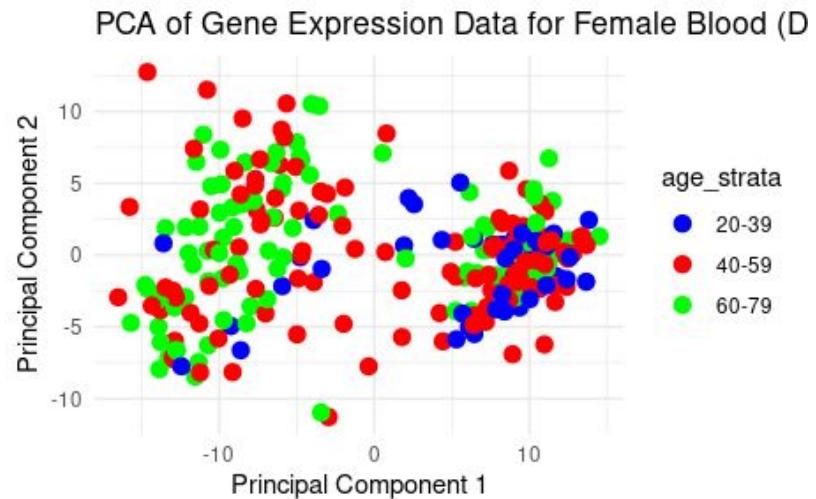
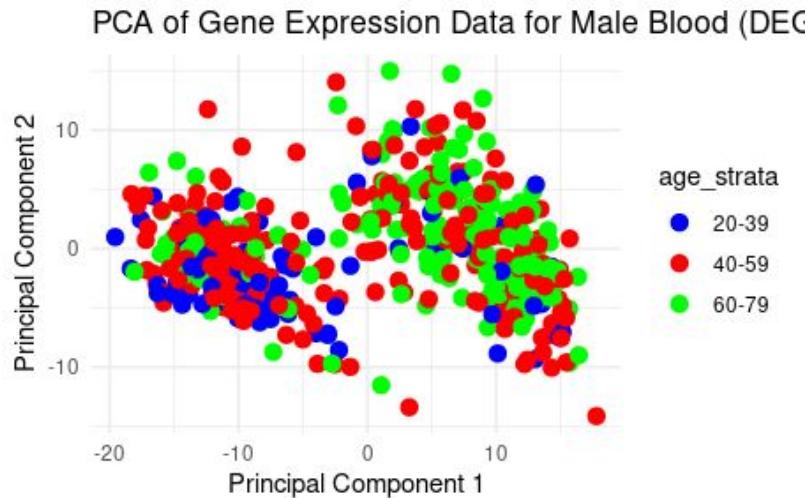
logCPM > 2, p-value < 0.05

DEG Analysis Results. Matr. Assoc. genes in Female Blood

| Description | logFC | FDR | gene_subtype |
|-------------|------------|--------------|------------------|
| IGF2 | 2.9887340 | 3.676168e-06 | Secreted factors |
| CCL3 | 2.4949868 | 7.974830e-05 | Secreted factors |
| LIF | 2.3596570 | 1.129358e-03 | Secreted factors |
| CCL4L2 | 2.2200385 | 4.594902e-05 | Secreted factors |
| VEGFA | 1.7803708 | 6.561193e-05 | Secreted factors |
| AREG | 1.7383353 | 3.792461e-04 | Secreted factors |
| CCL5 | 1.7306954 | 5.979227e-08 | Secreted factors |
| HBEGF | 1.7295252 | 2.556154e-04 | Secreted factors |
| FGFBP2 | 1.6402980 | 9.036047e-06 | Secreted factors |
| PDGFC | -0.6153285 | 6.792712e-02 | Secreted factors |
| TGFA | -0.6706868 | 6.760612e-03 | Secreted factors |
| S100A4 | -0.7102221 | 5.548460e-04 | Secreted factors |
| INSL3 | -0.7392880 | 1.989358e-03 | Secreted factors |
| IL1B | -0.8025321 | 8.238446e-04 | Secreted factors |
| IL1RN | -0.8415156 | 9.158334e-05 | Secreted factors |
| TNFSF14 | -0.8460055 | 1.233192e-03 | Secreted factors |
| INHBB | -0.9576619 | 1.542016e-02 | Secreted factors |
| TNFSF10 | -1.0690030 | 5.946457e-04 | Secreted factors |

| Description | logFC | FDR | gene_subtype |
|-------------|------------|--------------|----------------|
| AMBP | 3.9791284 | 4.581579e-06 | ECM regulators |
| HRG | 3.5054283 | 7.480164e-05 | ECM regulators |
| SERPINA3 | 3.2553723 | 1.468284e-06 | ECM regulators |
| SERPINH1 | 2.2025614 | 5.365416e-05 | ECM regulators |
| SERPINB2 | 1.7660984 | 1.968972e-05 | ECM regulators |
| CTSW | 1.7081052 | 8.097528e-06 | ECM regulators |
| PLAU | 1.4278838 | 1.310080e-02 | ECM regulators |
| P4HA1 | 1.1968426 | 3.604021e-06 | ECM regulators |
| SERPING1 | 1.1942335 | 2.012673e-03 | ECM regulators |
| CTSL | 1.1668899 | 5.928092e-04 | ECM regulators |
| SERPINE1 | 1.1324895 | 2.495722e-03 | ECM regulators |
| ADAM19 | -0.4911141 | 2.557666e-03 | ECM regulators |
| SERPINB8 | -0.5129647 | 5.525374e-04 | ECM regulators |
| CSTA | -0.5133750 | 1.127372e-02 | ECM regulators |
| ST14 | -0.5266343 | 3.521507e-03 | ECM regulators |
| ADAM17 | -0.5399849 | 6.760612e-03 | ECM regulators |
| CTSK | -0.5618765 | 2.341715e-04 | ECM regulators |
| EGLN1 | -0.7575184 | 2.983619e-04 | ECM regulators |
| HPSE | -0.8573201 | 1.243439e-04 | ECM regulators |
| MMP25 | -0.8637453 | 1.507610e-03 | ECM regulators |
| PRSS2 | -1.9613242 | 1.125871e-03 | ECM regulators |
| SDC2 | 1.9488576 | 2.030401e-06 | ECM affiliated |
| C1QC | 1.2959030 | 2.269876e-03 | ECM affiliated |
| C1QB | 1.1283732 | 4.069469e-03 | ECM affiliated |
| C1QA | 1.0231171 | 4.390234e-03 | ECM affiliated |
| PLXNA3 | 0.9019938 | 2.174400e-04 | ECM affiliated |
| CLEC2B | 0.8322105 | 2.463298e-04 | ECM affiliated |
| CLEC11A | 0.7701702 | 7.428039e-03 | ECM affiliated |
| SEMA4C | 0.7365973 | 6.572554e-04 | ECM affiliated |
| PLXNC1 | -0.5425531 | 1.533909e-02 | ECM affiliated |
| CLEC1B | -0.5433505 | 4.078531e-02 | ECM affiliated |
| CLEC4D | -0.6085946 | 5.951738e-02 | ECM affiliated |
| PLXNA2 | -0.6269029 | 9.240206e-04 | ECM affiliated |
| SEMA4B | -0.6615662 | 1.728616e-04 | ECM affiliated |
| ANXA3 | -0.6796452 | 1.479927e-02 | ECM affiliated |
| CLEC4A | -0.7433130 | 4.170954e-05 | ECM affiliated |
| ANXA9 | -0.9741165 | 4.829923e-05 | ECM affiliated |
| LGALSL | -1.1418744 | 4.343717e-05 | ECM affiliated |
| SFTPB | -1.9793875 | 5.309454e-04 | ECM affiliated |

PCA Analysis of the DEGs. Samples can be split by the difference in the DEG expression profile



Functional Enrichment of the upregulated DEGs list in all Female Tissues

| color | cluster Id | gene count | description |
|-----------------|------------|------------|--|
| reddish-pink | Cluster 1 | 92 | + Extracellular matrix organization |
| light red | Cluster 2 | 7 | Semaphorin-plexin signaling pathway |
| brown | Cluster 3 | 7 | + Molecules associated with elastic fibres |
| yellow | Cluster 4 | 6 | + Complement activation, classical pathway |
| greenish-yellow | Cluster 5 | 5 | + Canonical Wnt signaling pathway |
| light green | Cluster 6 | 3 | + LGI-ADAM interactions |
| dark green | Cluster 7 | 3 | + Antagonism of Activin by Follistatin |
| light green | Cluster 8 | 3 | + Antigen processing and presentation of exogenous peptide antigen via MHC clas... |
| dark green | Cluster 9 | 3 | Defective GALNT3 causes HFTC |
| light green | Cluster 10 | 2 | CTSG, SERPINB9 |
| blue | Cluster 11 | 2 | Mixed, incl. Cytolysis, and Regulation of extrathymic T cell differentiation |
| purple | Cluster 12 | 2 | Von Willebrand factor (vWF) type C domain |
| purple | Cluster 13 | 2 | Alcoholic pancreatitis, and Typhus |
| purple | Cluster 14 | 2 | Extracellular matrix structural constituent conferring compression resistance |
| purple | Cluster 15 | 2 | S100/CaBP-9k-type, calcium binding, subdomain, and Annexin |
| purple | Cluster 16 | 2 | S-100/CaBP type calcium binding domain |
| purple | Cluster 17 | 2 | Platelet-derived and vascular endothelial growth factors (PDGF, VEGF) family |

Network Stats

number of nodes: 167
 number of edges: 1086
 average node degree: 13
 avg. local clustering coefficient: 0.537

expected number of edges: 173
 PPI enrichment p-value: < 1.0e-16

Interaction Enrichment

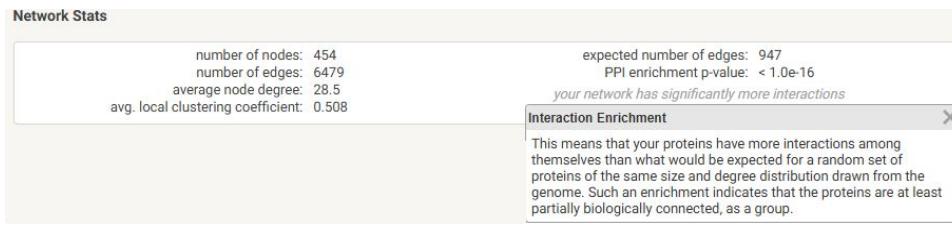
This means that your proteins have more interactions among themselves than what would be expected for a random set of proteins of the same size and degree distribution drawn from the genome. Such an enrichment indicates that the proteins are at least partially biologically connected, as a group.

KEGG Pathways

| description | count in network | strength | signal | false discovery rate |
|---|------------------|----------|--------|----------------------|
| ECM-receptor interaction | 18 of 88 | 1.38 | 3.28 | 2.96e-16 |
| Complement and coagulation cascades | 15 of 82 | 1.33 | 2.72 | 2.80e-13 |
| PI3K-Akt signaling pathway | 28 of 349 | 0.98 | 2.35 | 2.96e-16 |
| Focal adhesion | 20 of 195 | 1.08 | 2.31 | 1.98e-13 |
| Proteoglycans in cancer | 17 of 194 | 1.01 | 1.84 | 1.79e-10 |
| AGE-RAGE signaling pathway in diabetic complications | 12 of 96 | 1.17 | 1.8 | 5.57e-09 |
| Amoebiasis | 12 of 101 | 1.15 | 1.75 | 8.33e-09 |
| Human papillomavirus infection | 20 of 324 | 0.86 | 1.56 | 6.75e-10 |
| Protein digestion and absorption | 11 of 100 | 1.11 | 1.55 | 9.10e-08 |
| Chagas disease | 10 of 97 | 1.08 | 1.37 | 7.87e-07 |
| Malaria | 7 of 46 | 1.25 | 1.25 | 8.89e-06 |
| Cytokine-cytokine receptor interaction | 13 of 282 | 0.74 | 0.88 | 3.44e-05 |
| Rheumatoid arthritis | 7 of 83 | 1.0 | 0.86 | 0.00024 |
| Pathways in cancer | 18 of 515 | 0.62 | 0.84 | 1.67e-05 |
| MAPK signaling pathway | 12 of 286 | 0.69 | 0.75 | 0.00021 |
| Pertussis | 6 of 73 | 0.99 | 0.72 | 0.0011 |
| Relaxin signaling pathway | 7 of 126 | 0.82 | 0.61 | 0.0026 |
| Small cell lung cancer | 6 of 92 | 0.89 | 0.61 | 0.0031 |
| TGF-beta signaling pathway | 6 of 91 | 0.89 | 0.61 | 0.0031 |
| Viral protein interaction with cytokine and cytokine receptor | 6 of 96 | 0.87 | 0.59 | 0.0034 |
| Toll-like receptor signaling pathway | 6 of 100 | 0.85 | 0.58 | 0.0040 |

Functional Enrichment of the Upregulated DEGs List in all Male Tissues

| color | cluster id | gene count | description |
|--------------|------------|------------|--|
| ● Cluster 1 | 189 | 189 | + Extracellular matrix organization |
| ● Cluster 2 | 43 | 43 | + Viral protein interaction with cytokine and cytokine receptor |
| ● Cluster 3 | 30 | 30 | + Semaphorin-plexin signaling pathway |
| ● Cluster 4 | 13 | 13 | + Lysosome |
| ● Cluster 5 | 12 | 12 | + Molecules associated with elastic fibres |
| ● Cluster 6 | 9 | 9 | + Canonical Wnt signaling pathway |
| ● Cluster 7 | 9 | 9 | Complement activation |
| ● Cluster 8 | 8 | 8 | + Calcium-dependent protein binding |
| ● Cluster 9 | 8 | 8 | Regulation of pathway-restricted SMAD protein phosphorylation, and Signaling by B... |
| ● Cluster 10 | 5 | 5 | + Regulation of pathway-restricted SMAD protein phosphorylation, and Signaling b... |
| ● Cluster 11 | 5 | 5 | + LGI-ADAM interactions |
| ● Cluster 12 | 5 | 5 | + Myeloid leukocyte mediated immunity |
| ● Cluster 13 | 5 | 5 | + Osteogenesis |
| ● Cluster 14 | 5 | 5 | + Basement membrane |
| ● Cluster 15 | 4 | 4 | Defective B3GALTL causes PpS |
| ● Cluster 16 | 4 | 4 | Defective GALNT3 causes HFTC |
| ● Cluster 17 | 3 | 3 | Transforming growth factor beta complex |
| ● Cluster 18 | 3 | 3 | Galactoside-binding lectin |



| KEGG Pathways | description | count in network | strength | signal | false discovery rate |
|---|-------------|------------------|----------|----------|----------------------|
| ECM-receptor interaction | 34 of 88 | 1.22 | 4.11 | 2.83e-25 | |
| Protein digestion and absorption | 29 of 100 | 1.1 | 3.08 | 4.45e-19 | |
| Cytokine-cytokine receptor interaction | 49 of 282 | 0.88 | 2.72 | 6.27e-24 | |
| Amoebiasis | 26 of 101 | 1.05 | 2.64 | 3.98e-16 | |
| PI3K-Akt signaling pathway | 55 of 349 | 0.83 | 2.62 | 5.53e-25 | |
| Focal adhesion | 37 of 195 | 0.92 | 2.56 | 4.45e-19 | |
| TGF-beta signaling pathway | 24 of 91 | 1.06 | 2.55 | 4.04e-15 | |
| Axon guidance | 30 of 176 | 0.87 | 2.09 | 1.73e-14 | |
| Malaria | 14 of 46 | 1.12 | 1.89 | 1.15e-09 | |
| Complement and coagulation cascades | 18 of 82 | 0.98 | 1.83 | 2.51e-10 | |
| Viral protein interaction with cytokine and cytokine receptor | 19 of 96 | 0.93 | 1.76 | 2.97e-10 | |
| Proteoglycans in cancer | 28 of 194 | 0.8 | 1.73 | 6.10e-12 | |
| Rheumatoid arthritis | 16 of 83 | 0.92 | 1.53 | 1.45e-08 | |
| Human papillomavirus infection | 35 of 324 | 0.67 | 1.49 | 1.20e-11 | |
| AGE-RAGE signaling pathway in diabetic complications | 17 of 96 | 0.89 | 1.49 | 1.43e-08 | |
| Pathways in cancer | 42 of 515 | 0.55 | 1.21 | 2.48e-10 | |
| Chagas disease | 14 of 97 | 0.8 | 1.09 | 3.42e-06 | |
| Hippo signaling pathway | 18 of 154 | 0.71 | 1.07 | 1.26e-06 | |
| MAPK signaling pathway | 26 of 286 | 0.6 | 1.05 | 2.28e-07 | |
| Ras signaling pathway | 22 of 225 | 0.63 | 1.02 | 9.00e-07 | |
| Rap1 signaling pathway | 19 of 201 | 0.61 | 0.89 | 1.02e-05 | |
| EGFR tyrosine kinase inhibitor resistance | 11 of 77 | 0.79 | 0.89 | 6.60e-05 | |
| IL-17 signaling pathway | 12 of 91 | 0.76 | 0.88 | 5.41e-05 | |

Conclusions

- The reported results were reproduced using the same bioinformatic tools as employed in the considered analysis
- The results for the blood samples are almost fully consistent with the ones reported in the article
- The rest of the results were reproduced *grosso modo*, with the causes of imperfection lying in the lack of detailed information on the used filtering and/or the possibly changed data annotations
- The results for DEG list functional enrichment and network analysis of the corresponding proteins interactions are consistent with the ones reported and indicate the importance of studying ECM involvement in aging.

Weighted Gene Co-expression Network Analysis (WGCNA)



Workflow: Heart, Liver, Lung

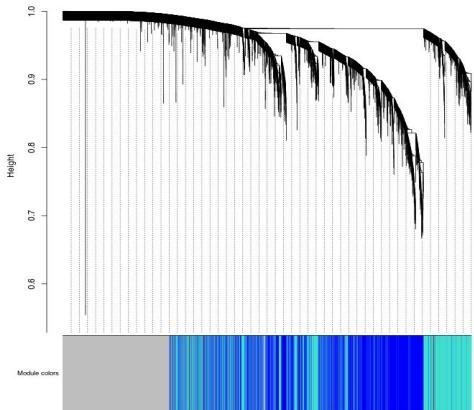
Data for 7 tissues (blood, brain, heart, kidneys, liver, lungs, and skeletal muscle) were used

Data were preprocessed (metadata, gene filtration, normalization, variance stabilizing transformation)

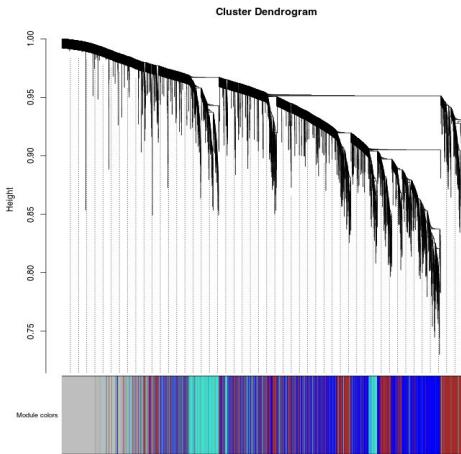
Power of correlation was chosen

Gene modules were identified [due to technical reasons] only for Heart, Liver and Lung samples; corr < 0.7

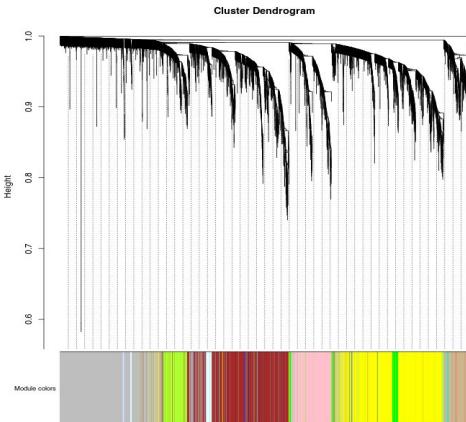
Analysis of genes in the most significant modules



Heart



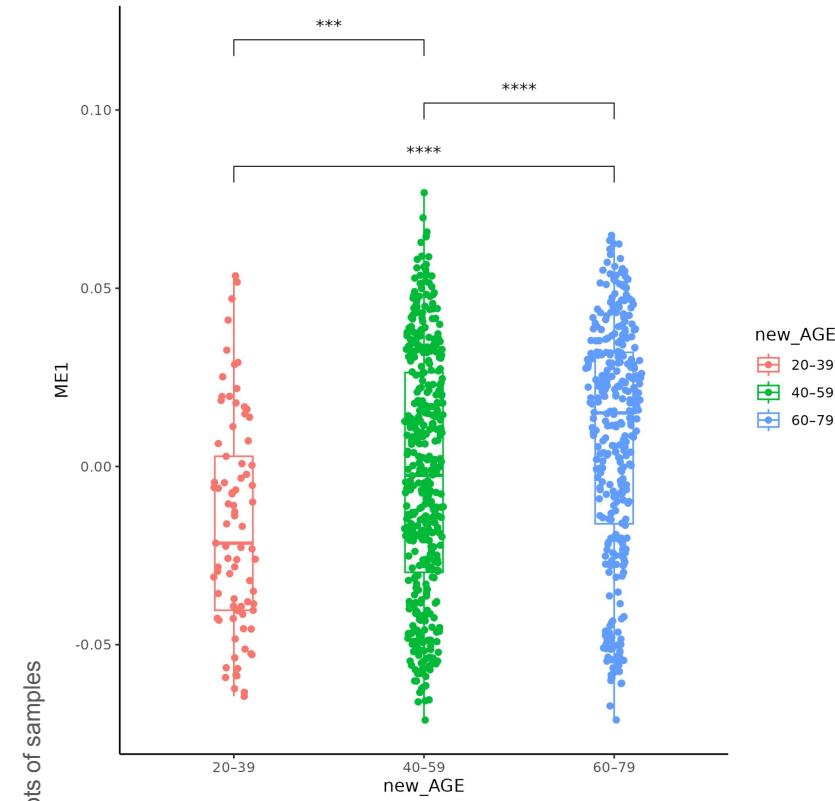
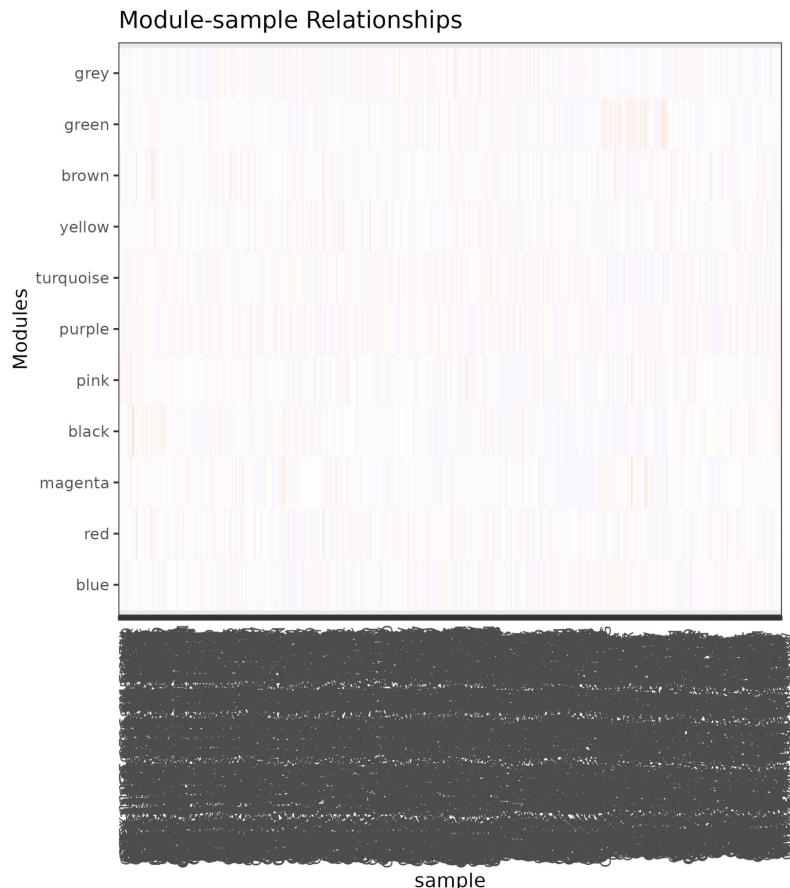
Lung



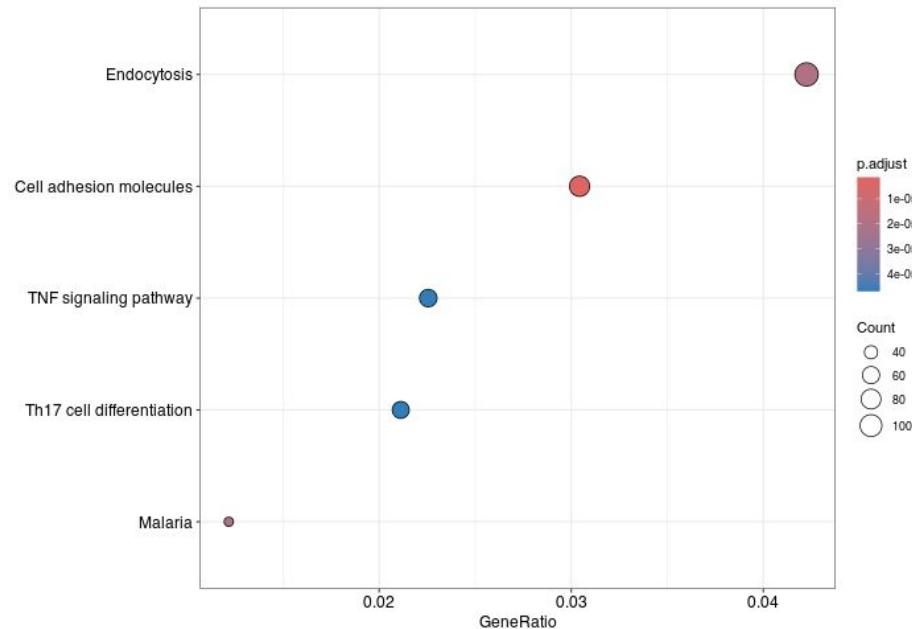
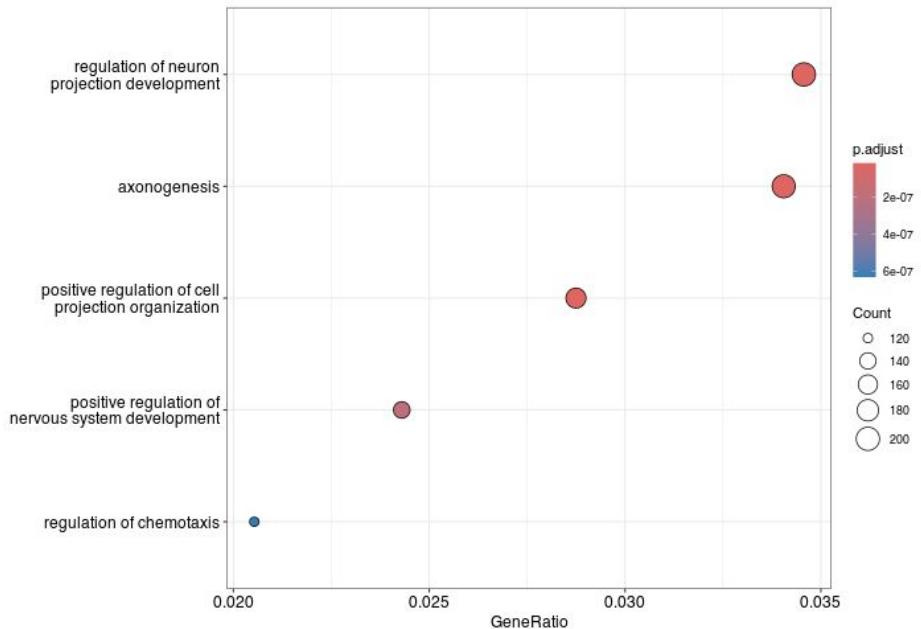
Liver



Heart: significant but not really bright



GO- and KEGG-enrichment for Heart



As it was expected or not?

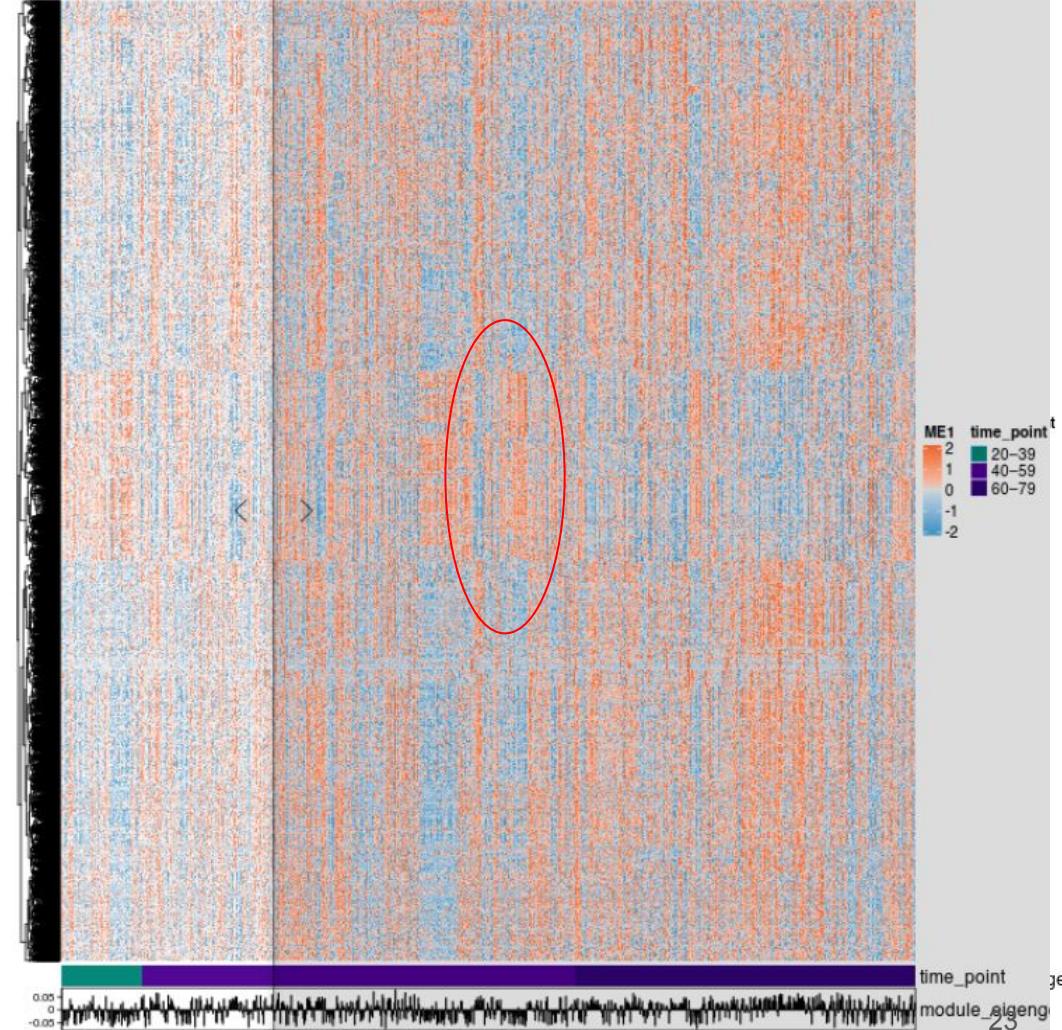
Heart

Are samples homogeneous?

Heart samples consist of left ventricle and atrial appendage tissues meaning cardiomyocytes, blood vessels, autonomic nerves, and connective tissues.

What is about donors?

They're post-mortem. Did they have any illnesses?



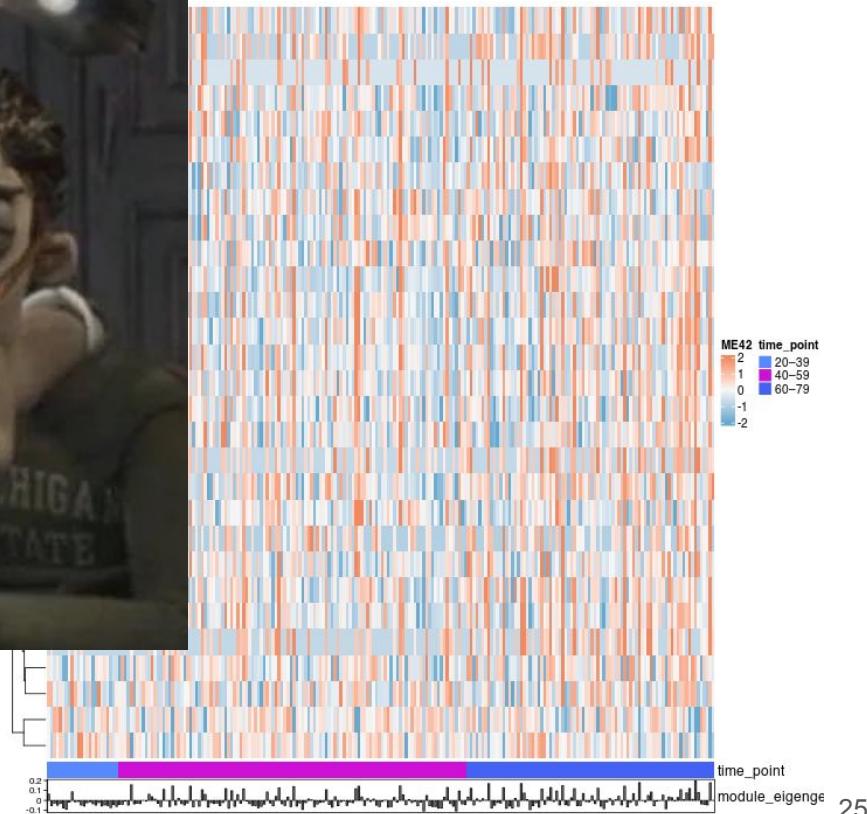
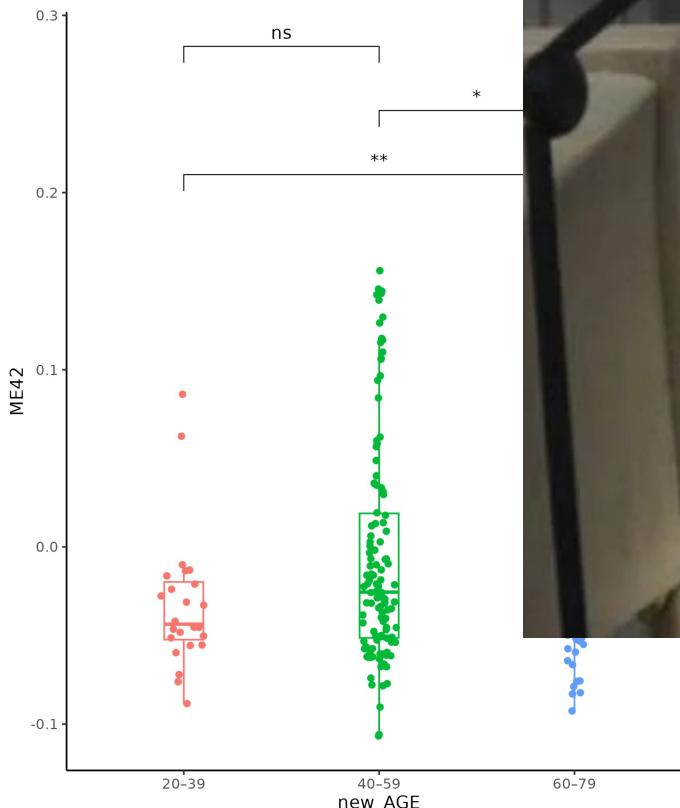
The most significant Heart module genes \cap matrisome genes

DCN, VCAN, COL9A2, ELN, COL11A1, NTN1, FBLN1, COL4A4, EPYC, COL16A1, LTBP4, COL9A3, PAPLN, MXRA5, SRPX, COMP, PCOLCE, AEBP1, OGN, ASPN, ECM2, LGI1, COL1A1, TECTA, VWA5A, MGP, COL12A1, COL9A1, SMOC2, IMPG1, SPARC, THBS4, EFEMP1, FN1, PRG4, MFAP2, LTBP2, TNN, TGFB1, CRISPLD1, FMOD, SRGN, TNFAIP6, MATN4, COL21A1, FGL2, COL5A1, LAMA5, EMILIN2, MATN2, POSTN, LAMC1, CHAD, IGFBPL1, EMILIN1, CIIP, MMRN1, FRAS1, LUM, KERA, FBLN5, MFAP1, MGGE8, IGFBP4, COL5A1, NTN5, TINAGL1, FNDC7, DPT, HMCN1, ECM1, FBLN7, COL8A1, SLIT2, VWA5B2, HAPLN1, RSPO3, VWDE, IGFBP3, IGFBP1, RSPO2, CRIM1, SPOCK1, IGSF10, LGI2, LGI4, ABI3BP, ACAN, SPON2, CILP2, NTN3, MATN1, NTNG1, SNED1, COL6A3, IGFBP7, FBLN2, PCOLCE2, ESM1, BMPER, COL1A2, FNDC1, SBSPON, CTHRC1, SVEP1, FBN1, MFAP4, IGFBP6, LTBP3, TNXB, COL3A1, COL4A3, RSPO1, THBS3, COL22A1, COL24A1, COL8A2, MMRN2, PODN, BGN, TSKU, COL18A1, SLIT3, NELL2, THBS2, EMID1, HAPLN4, COL14A1, COL4A5, AGRN, PRELP, RELN, NTNG2, COL27A1, COL13A1, MFAP5, SMOC1, COL5A2, COL15A1, COL6A6, COL28A1, BGLAP, SPON1

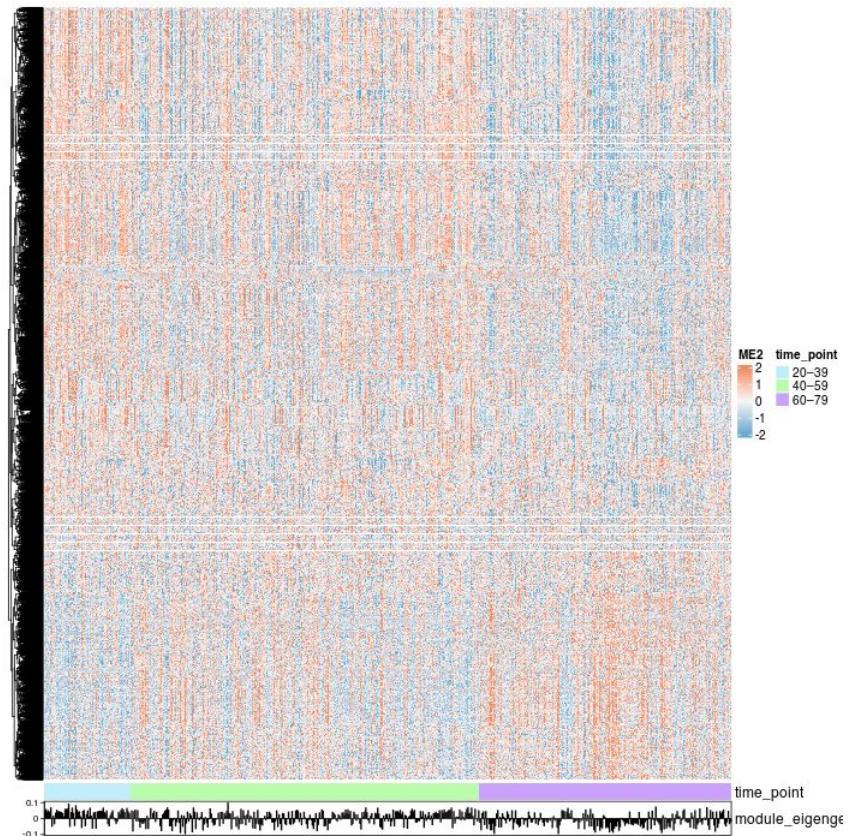
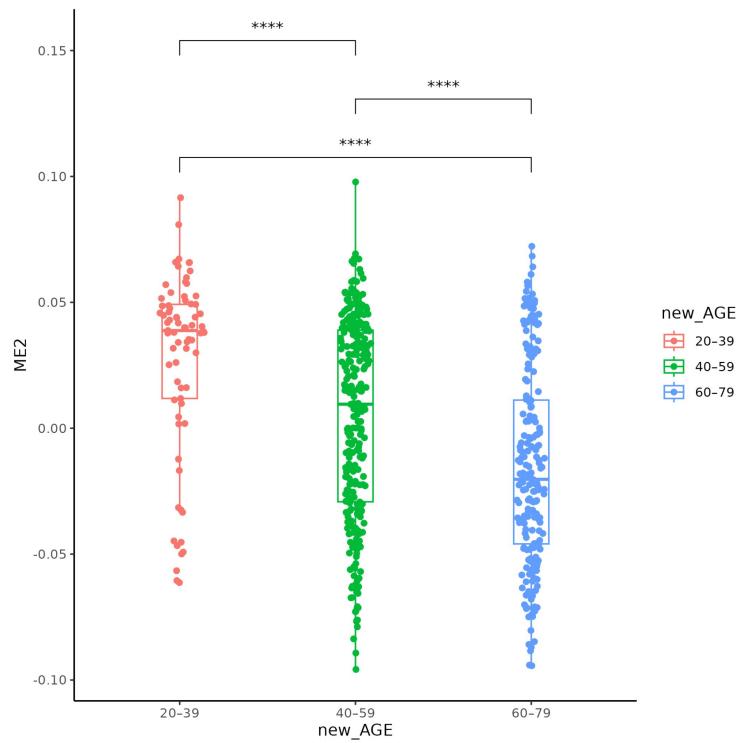
143

Size of a module is huge (9127 genes).
Intersection is significant at $p < 0.05$
(Chi-Square)

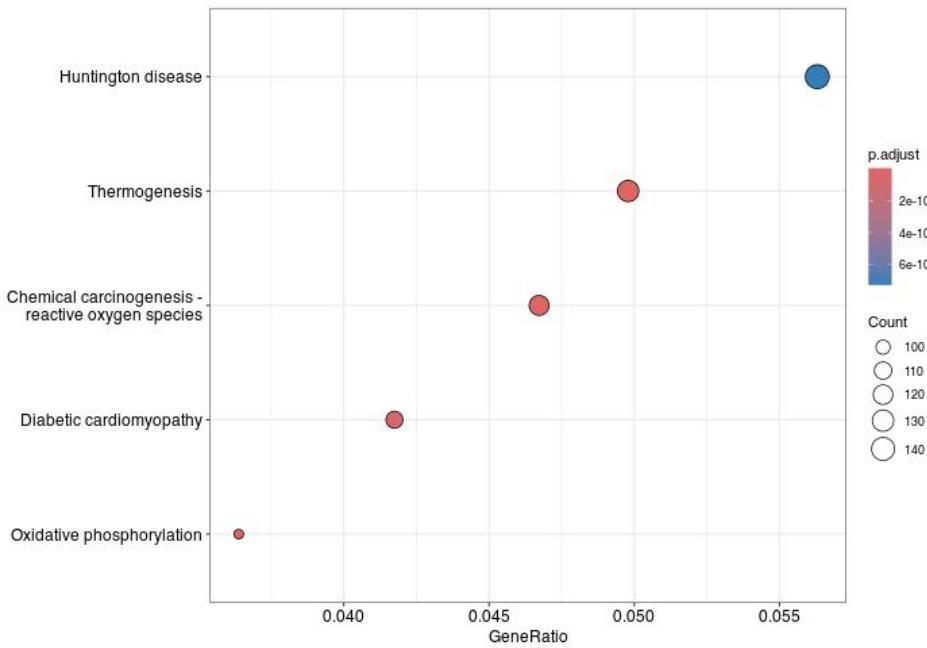
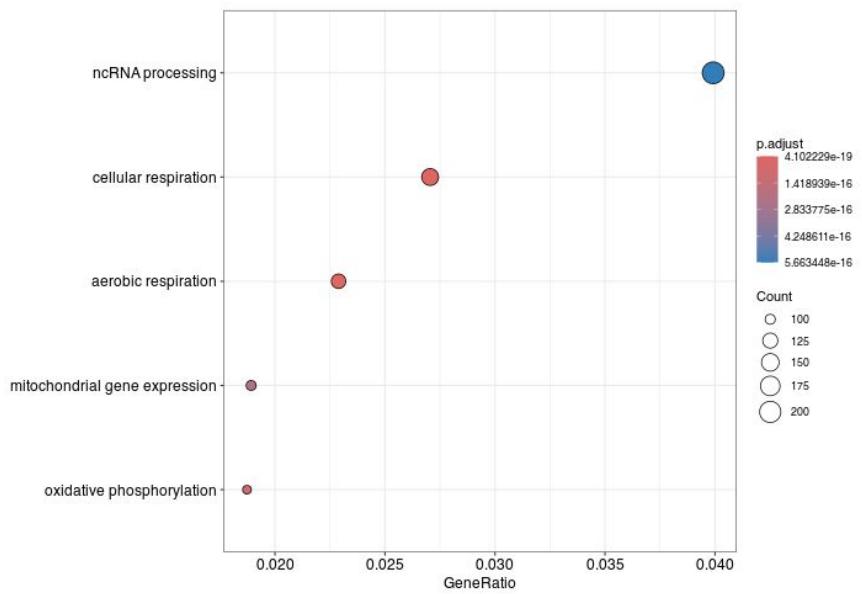
Liver: no GO and KEGG, no intersection



Lung: the same idea of homogeneity



Lung: GO and KEGG



Intersection: 40 matrisome genes

What can be done further?

Check for other modules and for other samples from this database

Repeat of the same workflow with more homogeneous samples (filtration by expression patterns?)

Intersect matrisome genes with categories in GO/KEGG enrichment

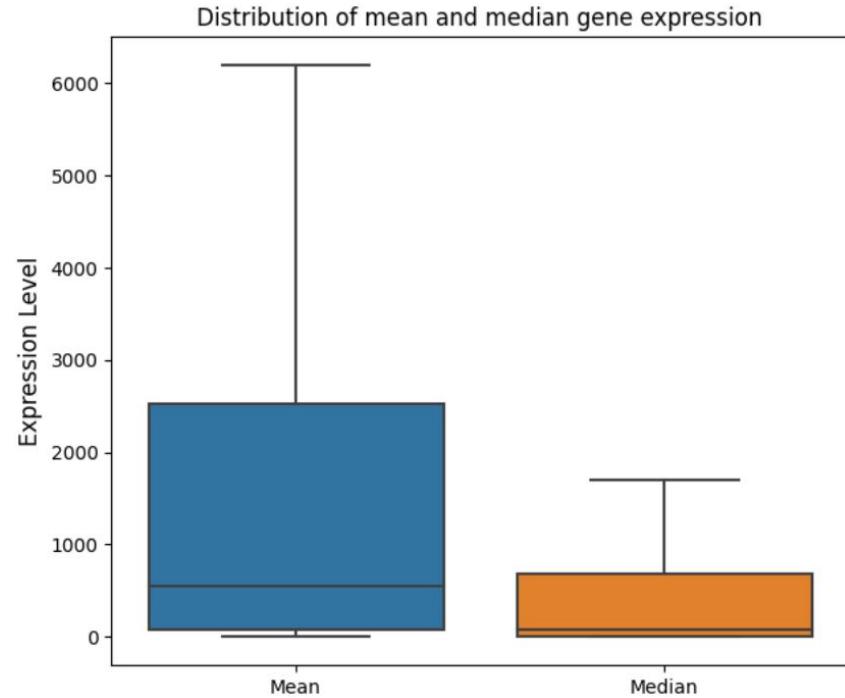
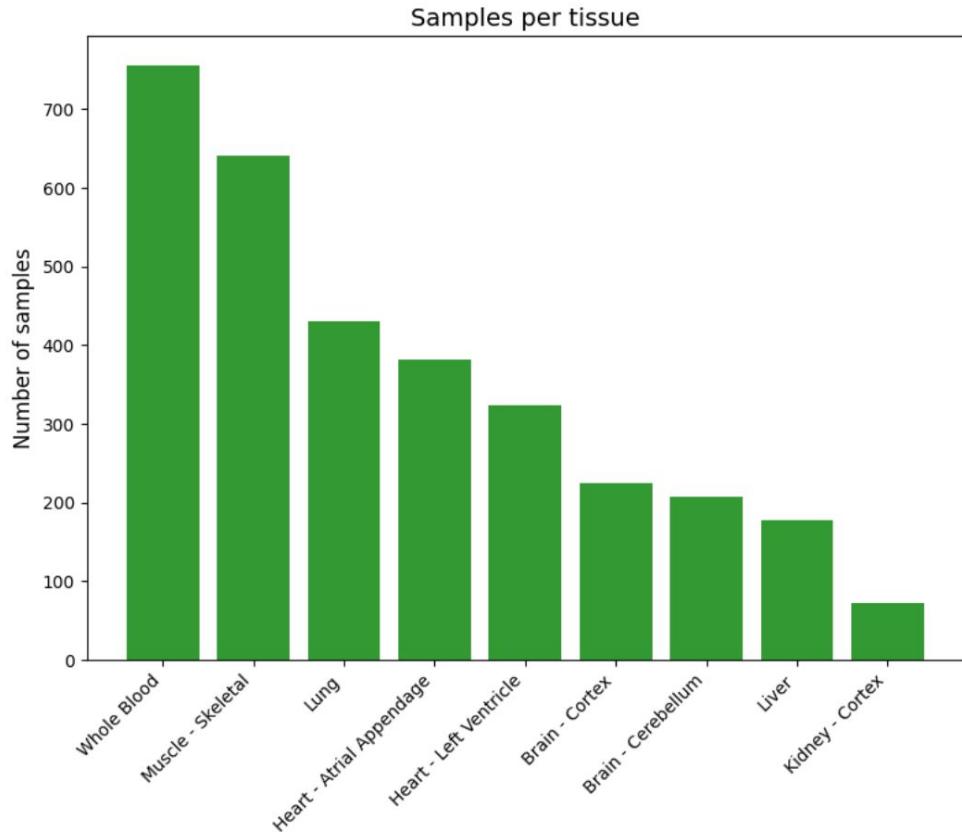
Use GO/KEGG groups to compare particular gene expression dynamics through different age (for now it's not really informative)

Construct regulatory networks to restore molecular mechanisms of agents interactions in modules

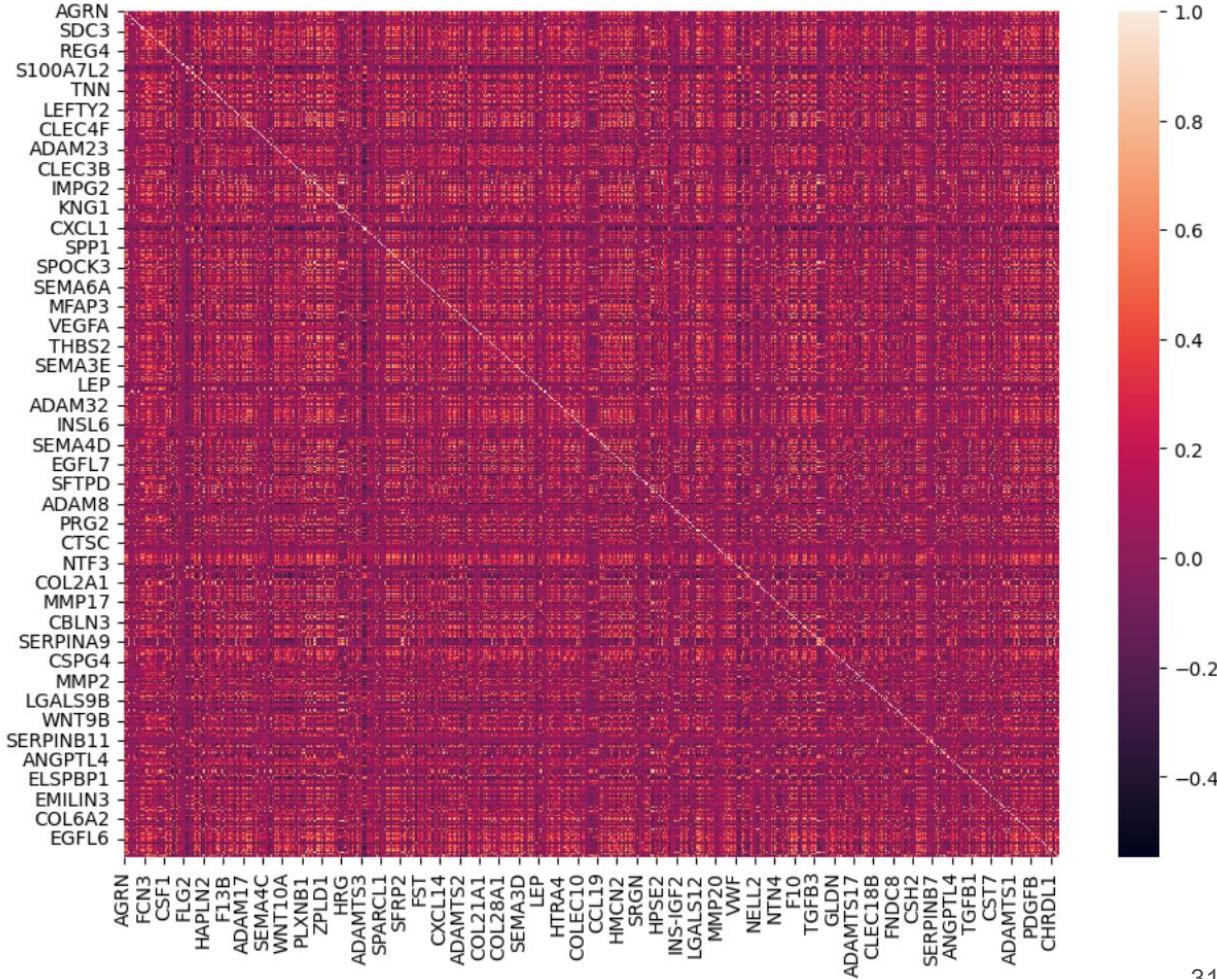
Check for found matrisome gene functions

Matrisome gene expression prediction

3415 samples and 902 matrisome genes

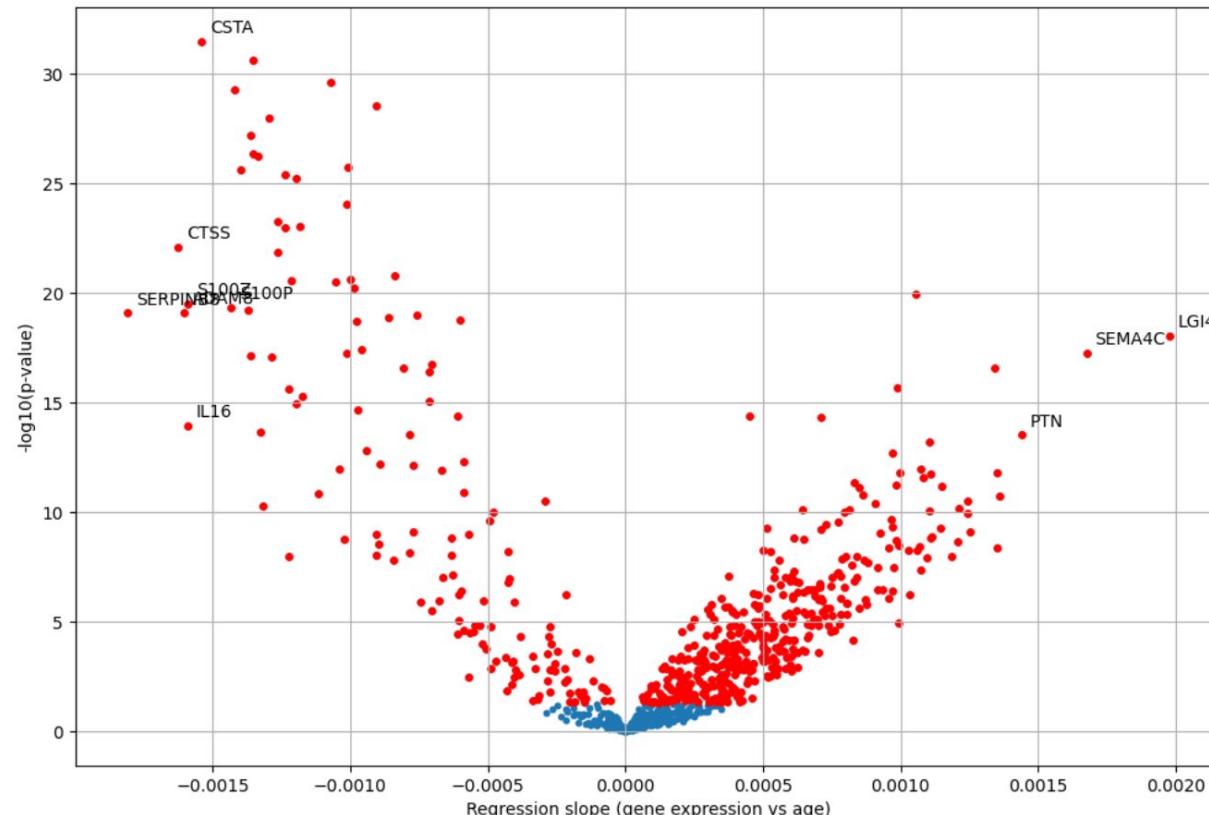


5808 unique gene
pairs with:
 $0.7 < \text{correlation} < 1$

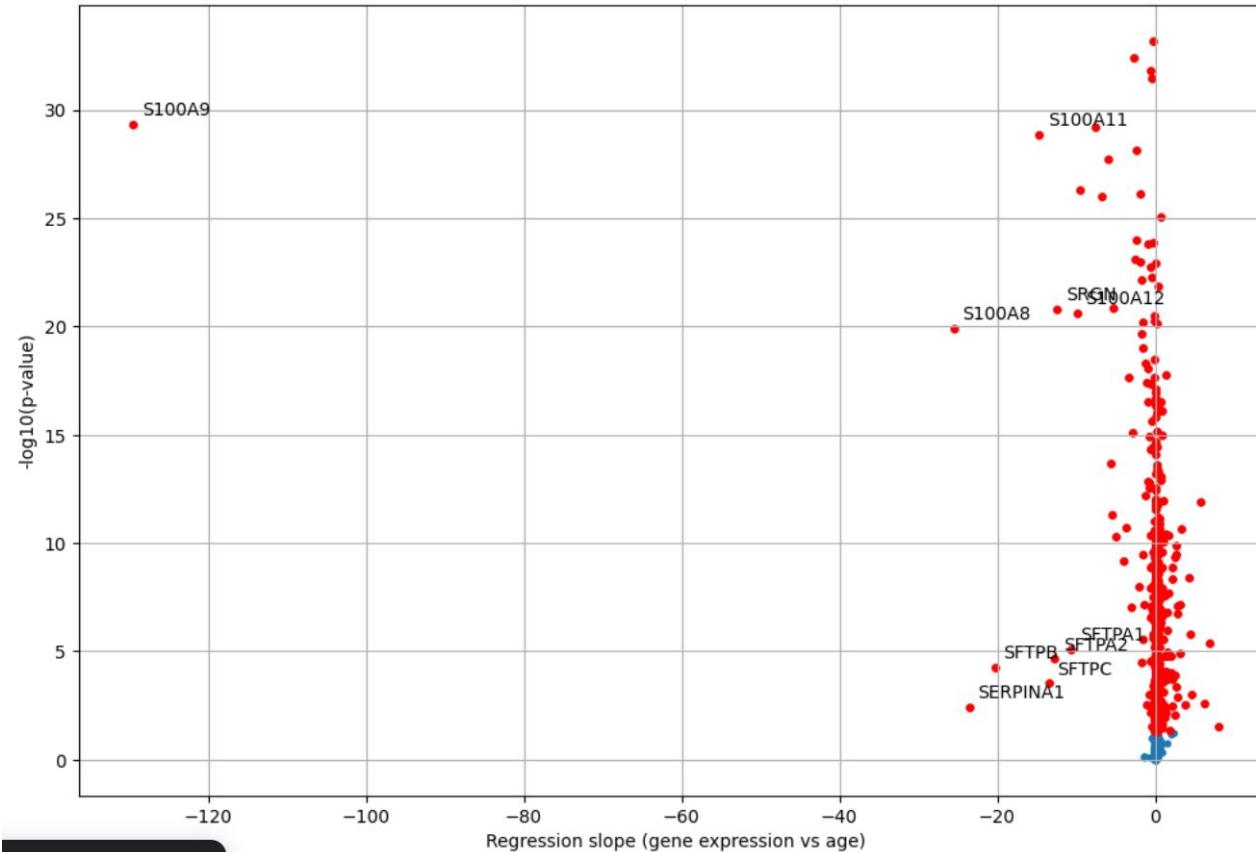


LGI4
SERPINB8
SEMA4C
CTSS
ADAM8
IL16
S100Z
CSTA
PTN
S100P

Regression slope (gene expression vs age) and p-value

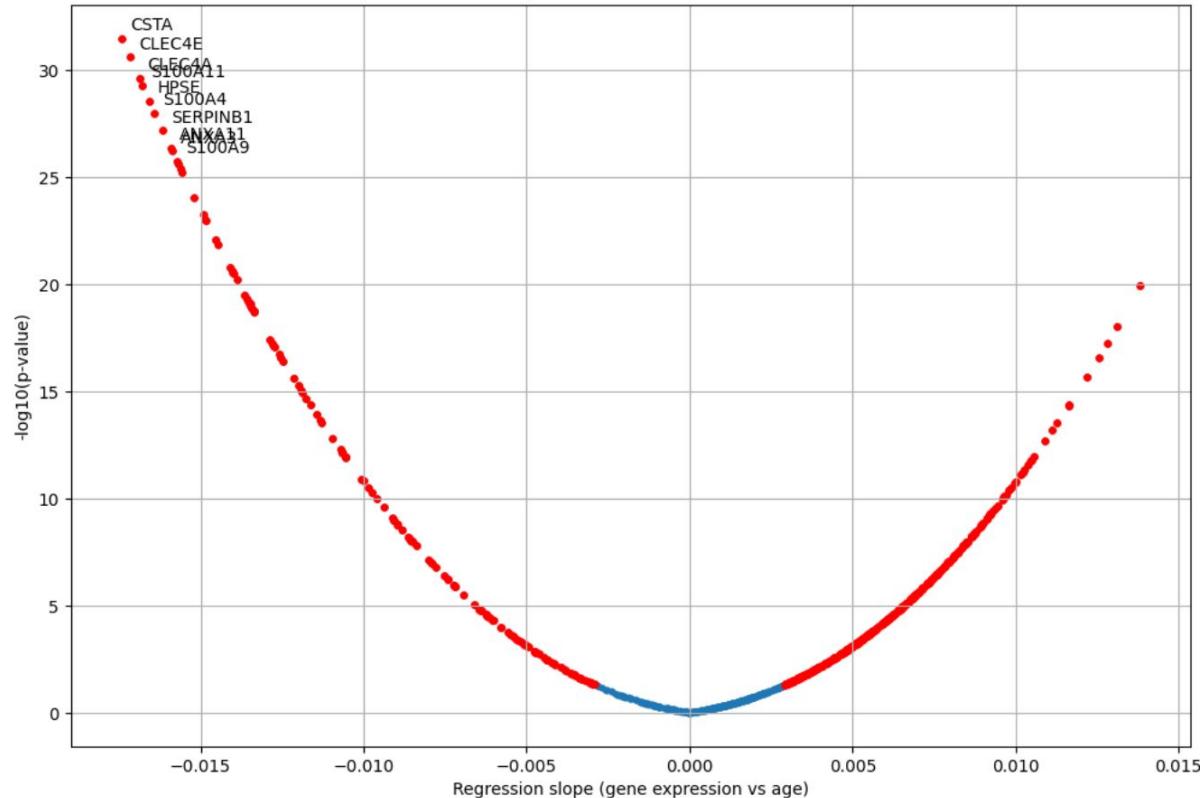


CPM normalisation



CSTA
CLEC4E
CLEC4A
S100A11
HPSE
S100A4
SERPINB1
ANXA11
ANXA3
S100A9

StandartScaler



Workflow

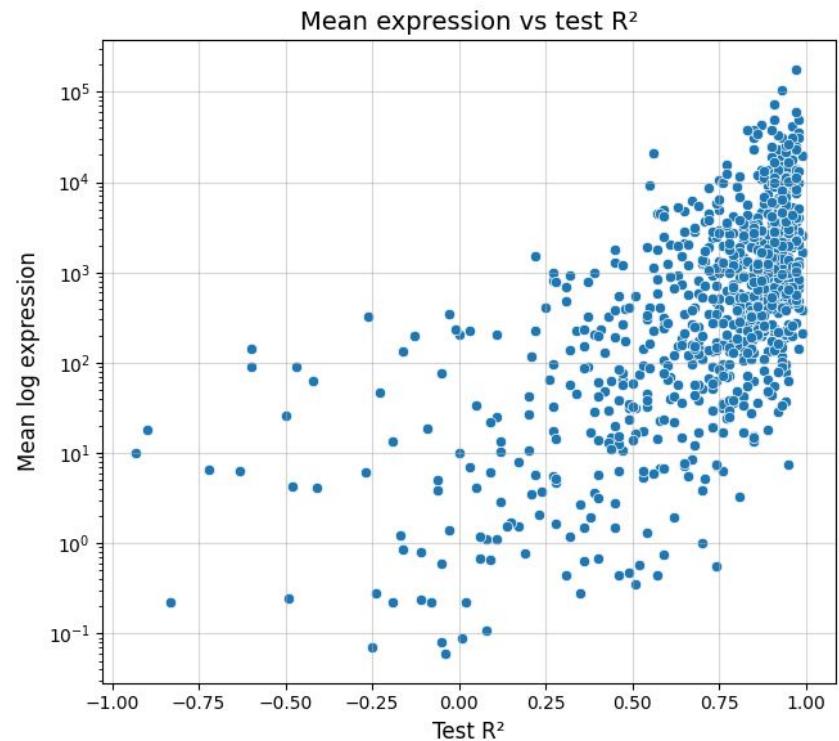
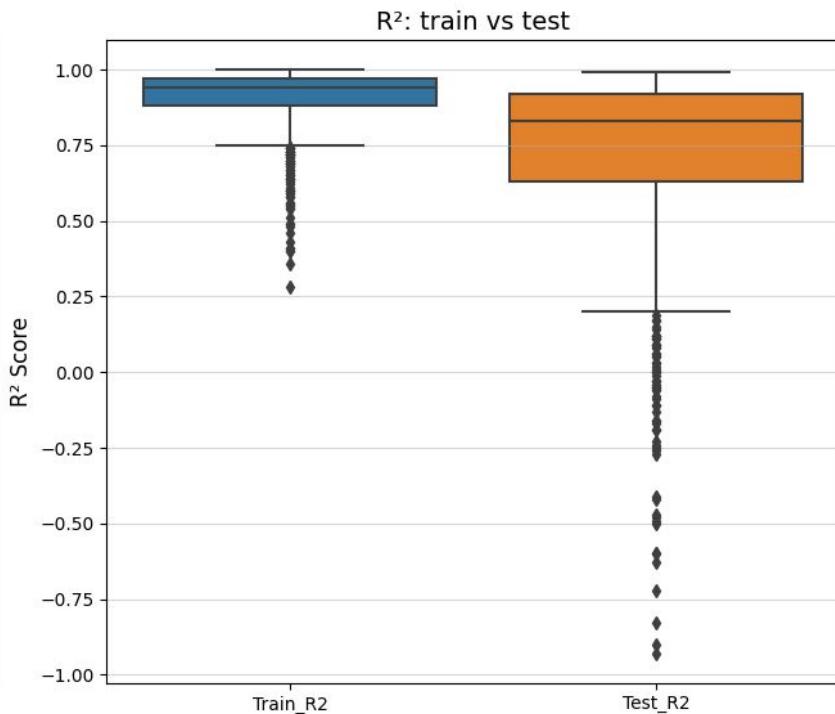
- 1) Keep only **matrisome genes** from the gene expression table (**902 genes**)
- 2) Encode categorical features (Age, Sex, Tissue)
- 3) **Divide** dataset on **train and test** sets
- 4) On train set obtain **correlations**
- 5) Choose model and go through all 902 genes:
 - remove redundant genes (correlation > 0.7)
 - optimise hyperparameters with optuna (if needed)
 - make prediction on train and test sets

Results

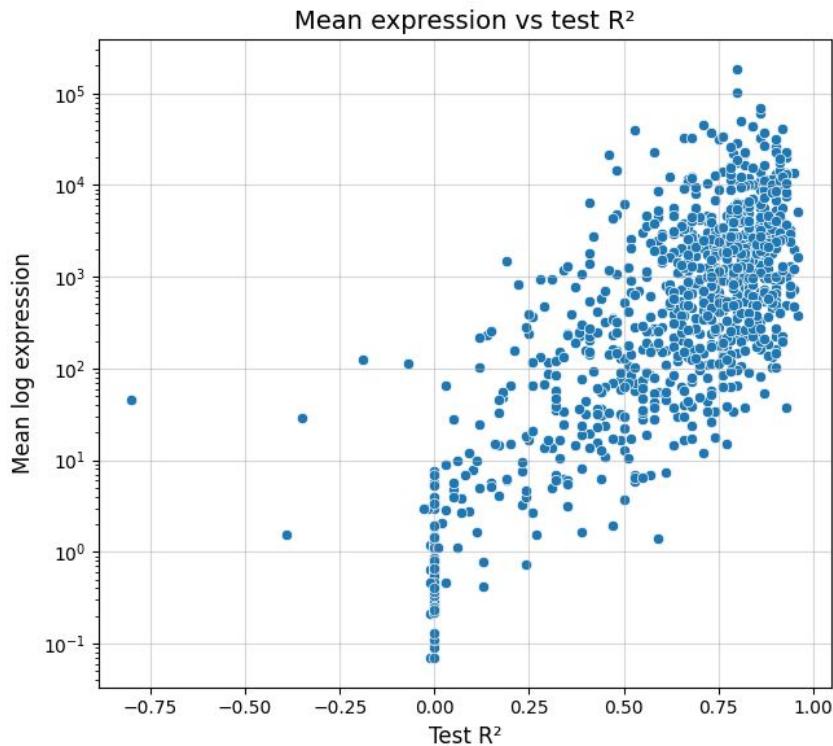
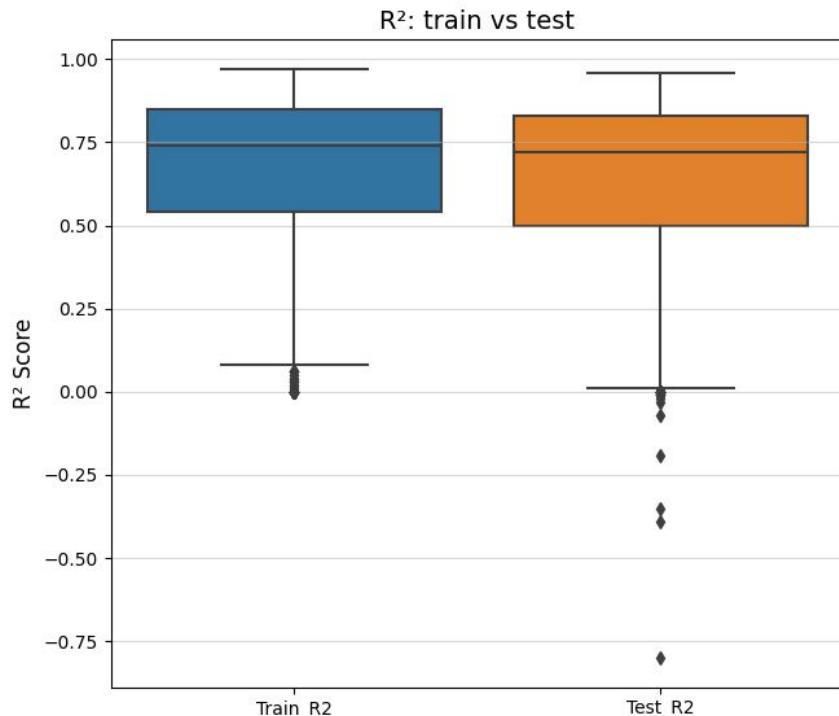
| | Linear Regression | Linear Regression (without corr genes) | Ridge (alpha = 100) | Ridge (optuna) | Lasso (optuna) | Elasticnet (optuna) | KNN | DecisionTree |
|----------------|-------------------|--|---------------------|----------------|----------------|---------------------|-------------|--------------|
| Mean R2 test | -0,21 | 0,09 | 0,32 | 0,39 | 0,41 | 0,57 | 0,59 | -0,46 |
| Median R2 test | 0,72 | 0,74 | 0,82 | 0,79 | 0,79 | 0,72 | 0,7 | 0,6 |
| Smallest R2 | -217,72 | -182,19 | -102,48 | -91,46 | -88,87 | -36,48 | -17,63 | -390,82 |
| Number of R2<0 | 126 | 89 | 64 | 34 | 22 | 17 | 23 | 101 |

*XGBoost and random forest were computed for 500+ minutes each, so I stopped them

Ridge (alpha = 100)



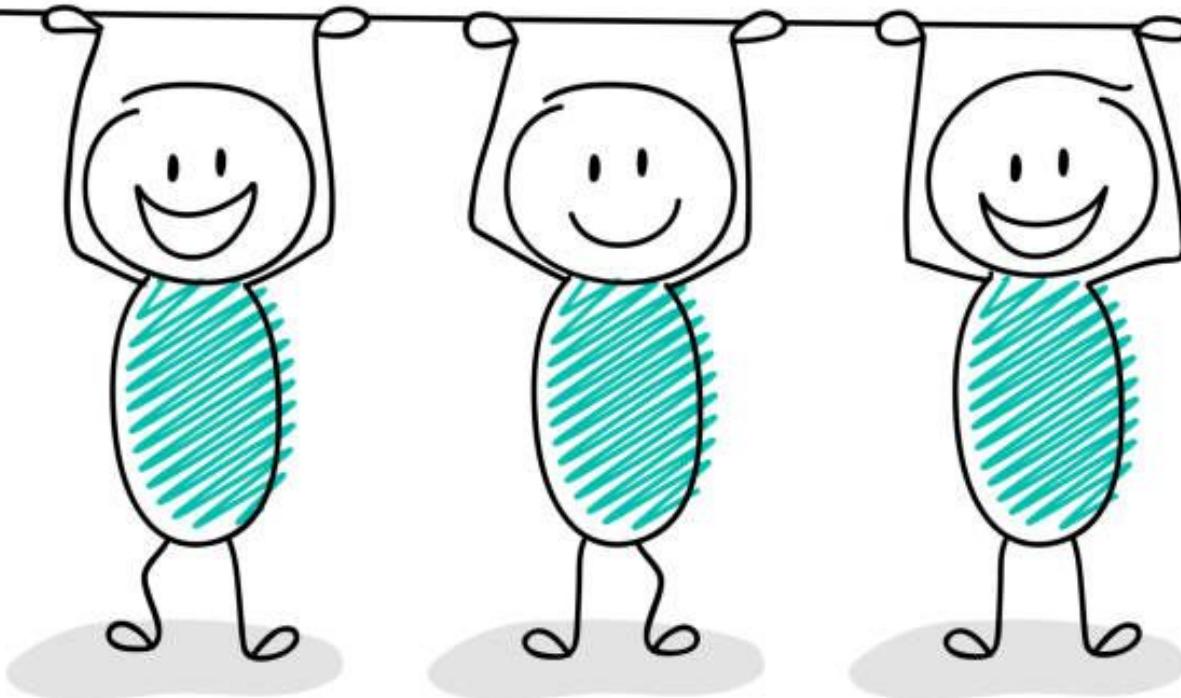
ElasticNet



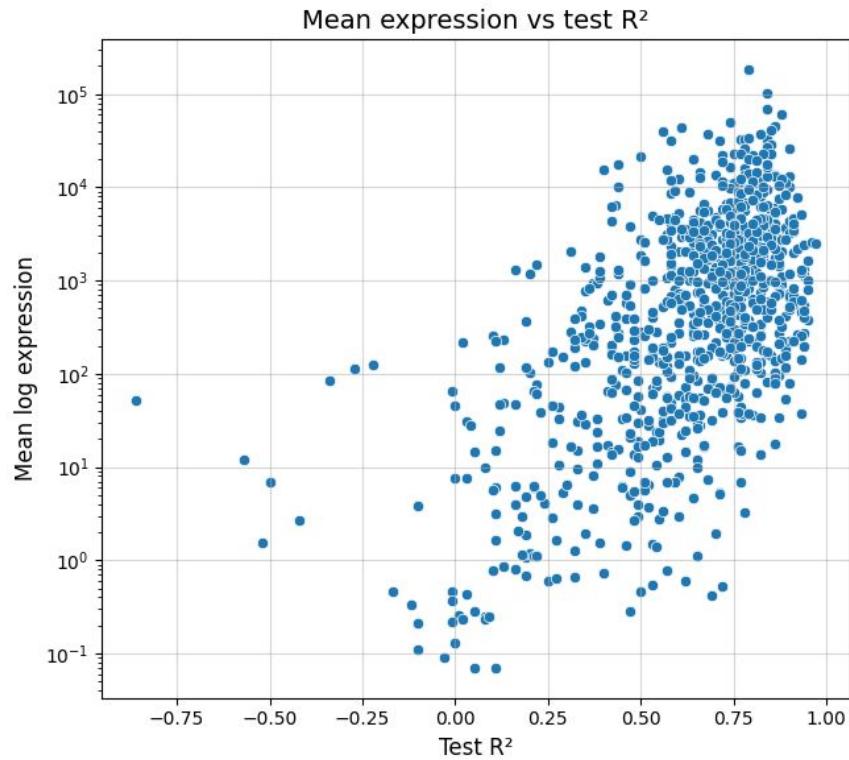
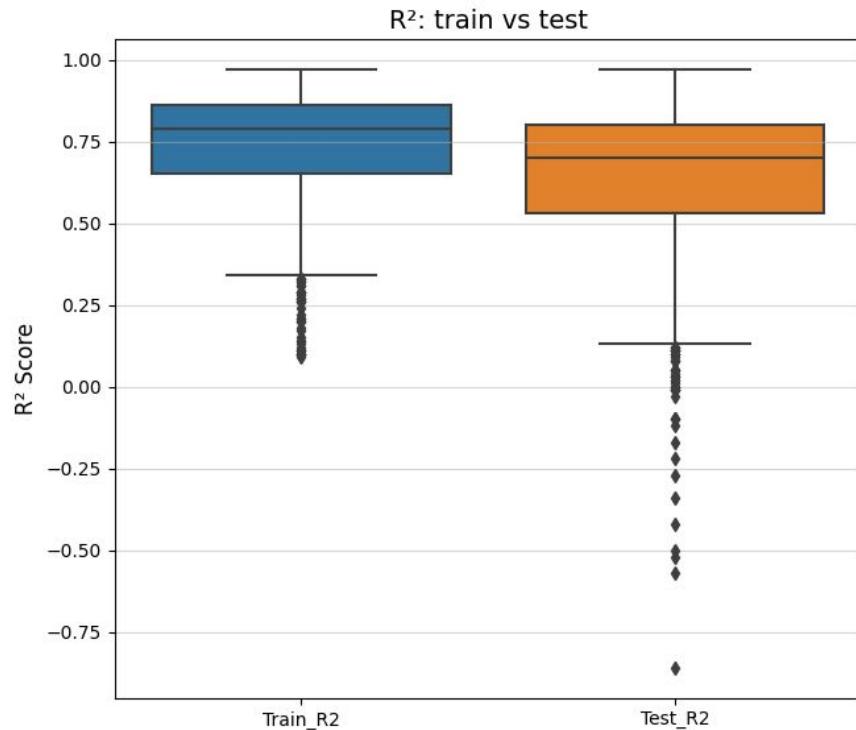
What else can be done?

1. Optimize code
2. Maybe normalise data with another approach
3. Check other regression models from sklearn
4. Use more data (not only tissues from article)
5. Combine with WGCNA analysis
6. Take age coefficients from regression and check, for which genes they are the highest + compare them with volcano plot results

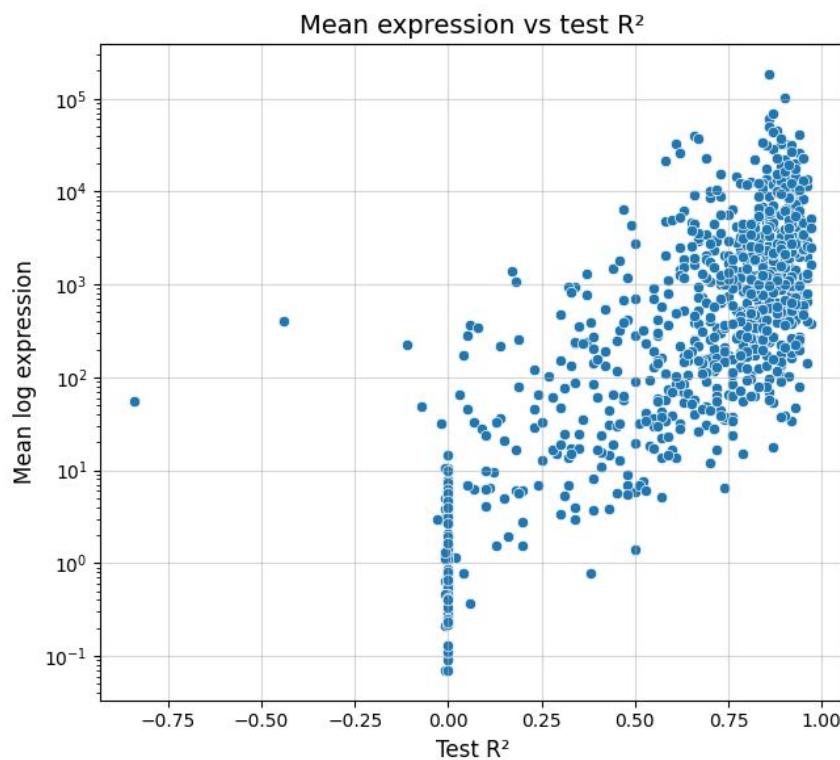
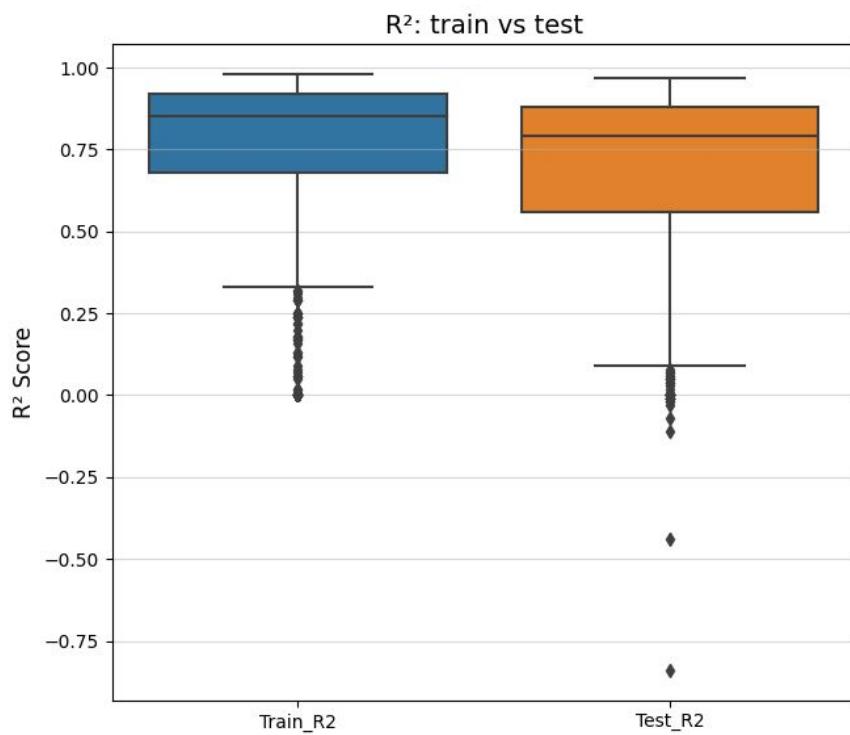
THANK YOU



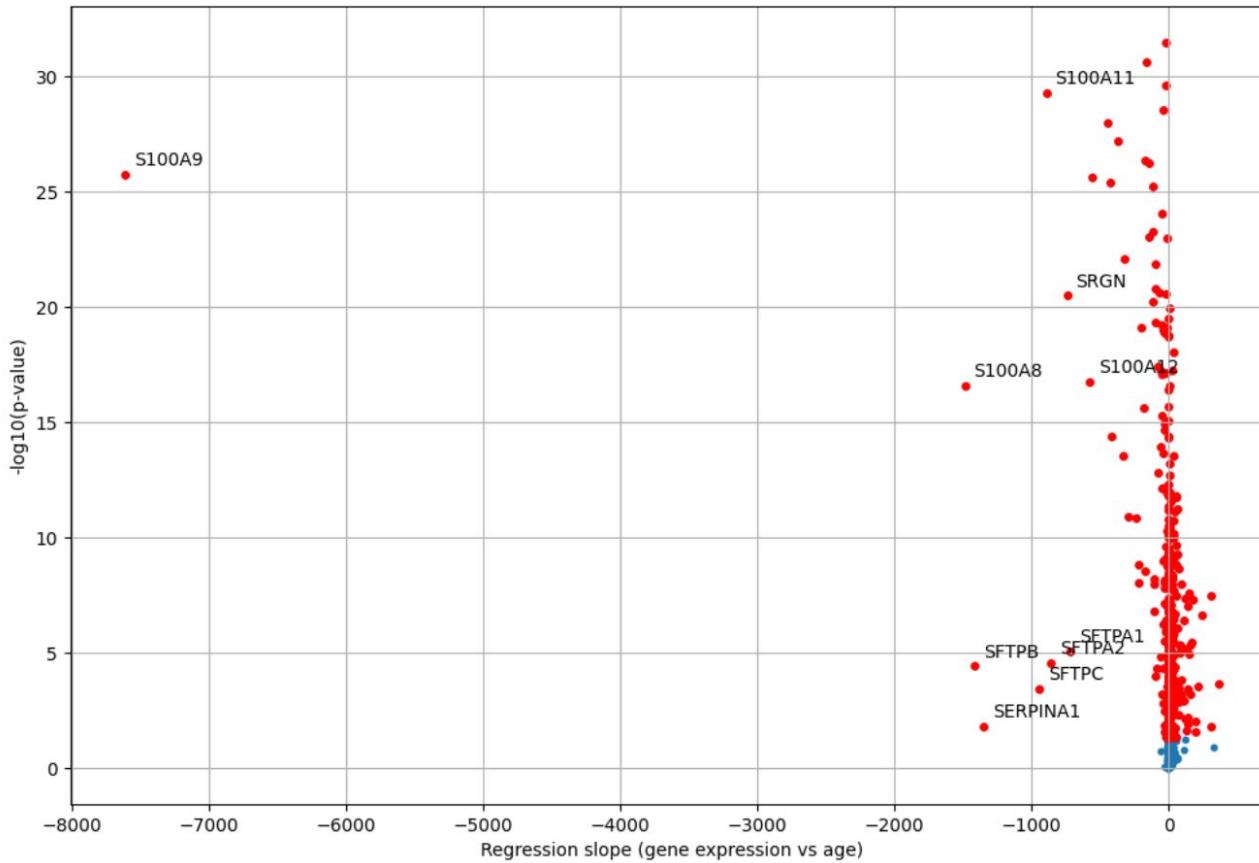
KNN Regressor

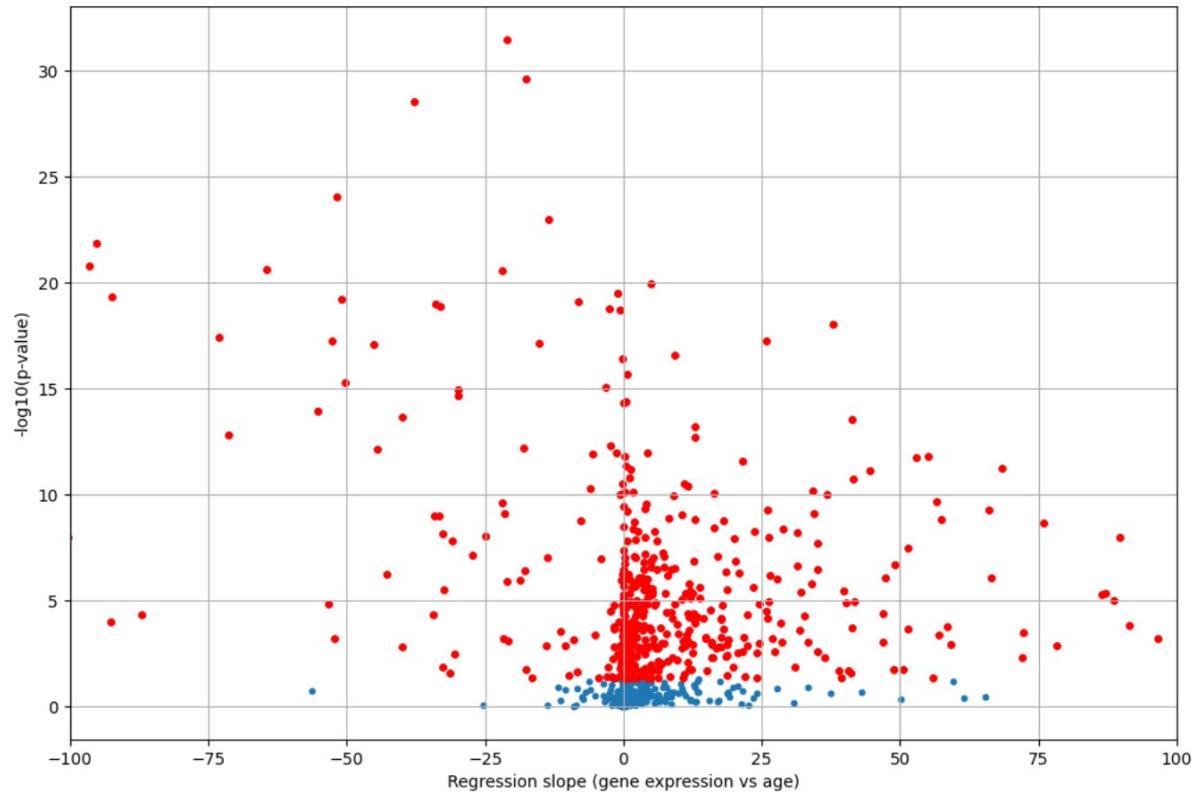


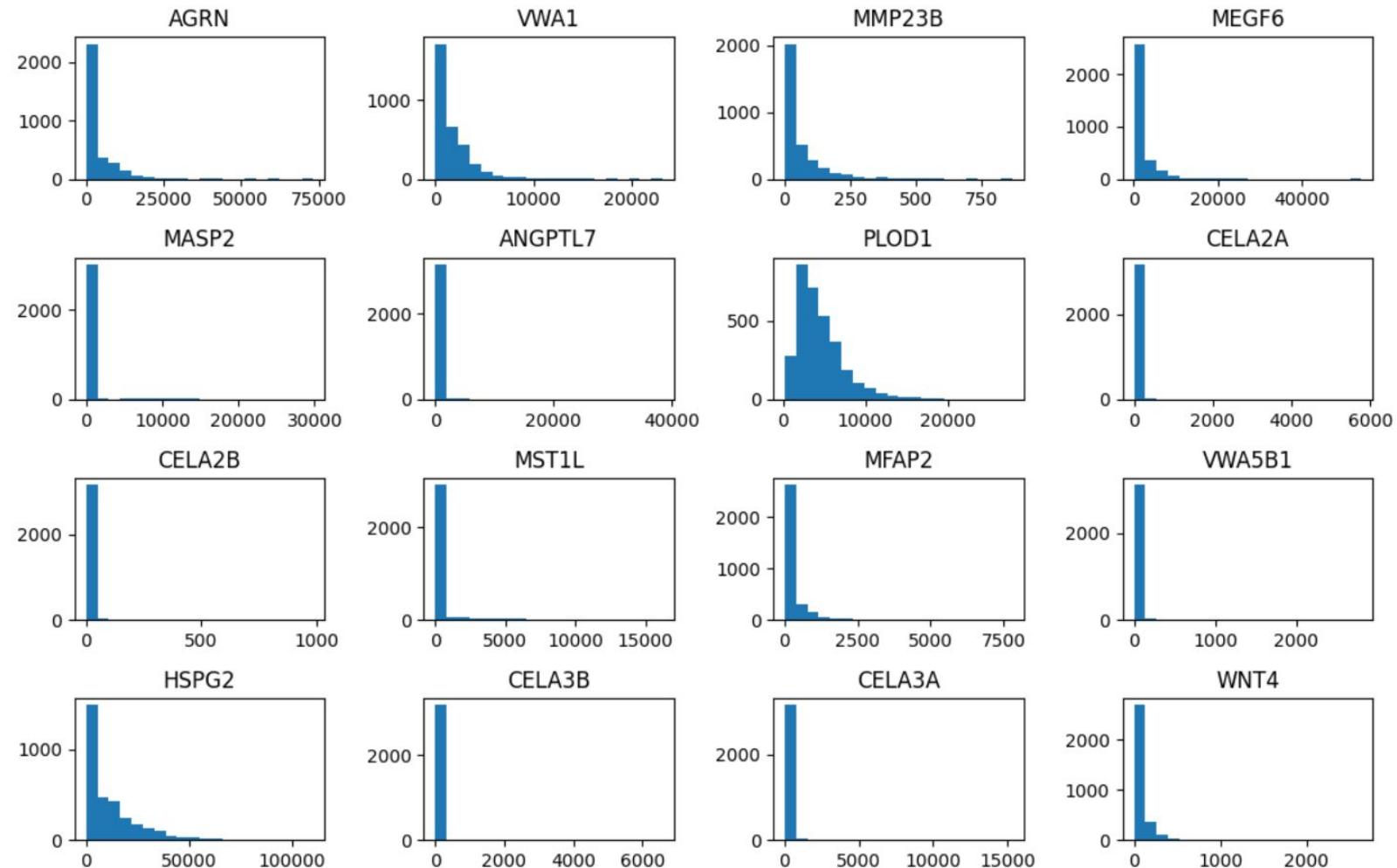
Lasso



Without normalisation







$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}$$