**RESEARCH ARTICLE**

# voyAGEr: free web interface for the analysis of age-related gene expression alterations in human tissues

*Short running title:*

Age-related gene changes profiling app

Arthur L. Schneider[1,2]
arthur.schneider@centraliens-nantes.org

Nuno Saraiva-Agostinho[1,3]
nunodanielagostinho@gmail.com

Nuno L. Barbosa-Morais[1*]
nmorais@medicina.ulisboa.pt

[1] Instituto de Medicina Molecular João Lobo Antunes, Faculdade de Medicina, Universidade de Lisboa, Av. Prof. Egas Moniz, 1649-028 Lisbon, Portugal

[2] *Present address:* Sia Partners, 4 Rue Voltaire, 44000 Nantes, France

[3] *Present address:* European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, CB10 1SD, Cambridge, UK

* Corresponding author

## Abstract

We herein introduce voyAGEr, an online graphical interface to explore age-related gene expression alterations in 48 human tissues. voyAGEr offers a visualization and statistical toolkit for the finding and functional exploration of sex- and tissue-specific transcriptomic changes with age. In its conception, we developed a novel bioinformatics pipeline leveraging RNA sequencing data, from the GTEx project, for more than 700 individuals.

voyAGEr reveals transcriptomic signatures of the known asynchronous aging between tissues, allowing the observation of tissue-specific age-periods of major transcriptional changes, that likely reflect so-called digital aging, associated with alterations in different biological pathways, cellular composition, and disease conditions.

voyAGEr therefore supports researchers with no expertise in bioinformatics in elaborating, testing and refining their hypotheses on the molecular nature of human aging and its association with pathologies, thereby also assisting in the discovery of novel therapeutic targets. voyAGEr is freely available at https://compbio.imm.medicina.ulisboa.pt/voyAGEr

## Keywords

Humans; Aging; Computational Biology; Transcriptome; Gene Expression Profiling

# Main text

## Introduction

The aging-associated progressive loss of proper tissue homeostasis maintenance makes age a prevalent risk factor for many human pathologies, including cancer, neurodegenerative and cardiovascular diseases [1–3]. A better comprehension of the molecular mechanisms of human aging is thus required for the development and effective application of therapies targeting its associated morbidities.

Dynamic transcriptional alterations accompany most physiological processes occurring in human tissues [4]. Transcriptomic analyses of tissue samples can thus provide snapshots of cellular states therein and insights into how their modifications over time impact tissue physiology. A small proportion of transcripts has indeed been shown to vary with age in tissue- [5] and sex-specific [6–11] manners. Such variations reflect deregulations of gene expression that underlie cellular disfunctions [5].

Many studies analyzed the age-related changes in gene expression in rodent tissues [12–18], emphasizing the role in aging of genes related to inflammatory responses, cell cycle or the electron transport chain. However, while it is possible to monitor the modifications in gene expression over time in those species by sequencing transcriptomes of organs of littermates at different ages, as a close surrogate of longitudinality, such studies cannot be conducted in humans for obvious ethical reasons. Indeed, most studies aimed at profiling aging-related gene expression changes in human tissues focused on a single tissue (e.g. muscle [19–21], kidney [22], brain [23–27], skin [28,29], blood [30,31], liver [32], retina [33]) and/or limited to a comparison between young and old individuals [20,21,28,32,33], failing to fully capture the changes of tissue-specific gene expression landscape along aging [5]. A few studies were nonetheless led on more than one tissue in humans, from post-mortem samples [34,35] and biopsies [36,37], and in mice [18,36] and rats [38]. The age-related transcriptional profiles derived therein, either from regression [18,34,35,37] or comparison between age groups [36,38], highlight an asynchronous aging of tissues (discussed in [39]), with some of them more affected by age-related gene expression changes associated with biological mechanisms known to be

impacted by aging such as mitochondrial activity or metabolic homeostasis. In particular, tissue-specific periods of major transcriptional changes in the fifth and eighth decades of human lifespan have been revealed [34], reflecting the so-called digital aging [39] and consistently with what is observed in mice [17,18]. Furthermore, despite outlining the tissue specificity of the transcriptomic signatures of human aging, some synchronization was found between tissues like lung, heart and whole blood, which exhibit a co-aging pattern [35]. Nevertheless, as each study followed its own specific procedures, from sample collection to data processing, results from their analyses are hardly comparable across studies.

Processed data from those studies have not been made easily accessible and interpretable to researchers lacking computational proficiency but aiming to use them to test their novel hypotheses. To fill this void, we have developed voyAGEr, a web application providing flexible visualization of comprehensive functional analyses of gene expression alterations occurring in 48 human tissues with age in each sex. We leverage the large RNA-seq dataset from the Genotype-Tissue Expression (GTEx) project [40], encompassing post-mortem tissue samples from hundreds of donors aged from 20 to 70 years, with a pipeline for gene expression profiling with an optimized temporal resolution. voyAGEr allows to investigate aging from two perspectives: (i) gene-centric – how each gene's tissue-specific expression progresses with age; (ii) tissue-centric – how tissue-specific transcriptomes change with age. Additionally, voyAGEr enables the examination of modules of co-expressed genes altered with age in 4 tissues (brain, skeletal muscle, heart (left ventricle), whole blood), namely their enrichment in specific cell types, biological pathways, and association with diseases. We therefore expect voyAGEr to become a valuable support tool for researchers aiming to uncover the molecular mechanisms underlying human aging. Moreover, being open-source, voyAGEr can be adapted by fellow developers to be used with alternative datasets (e.g. from other species) or to incorporate other specific functionalities.

voyAGEr is freely available at https://compbio.imm.medicina.ulisboa.pt/voyAGEr

# Results

voyAGEr's interactive exploration of tissue-specific gene expression landscapes in aging is based on the sequential fitting of linear models (v. Materials and Methods) to estimate, for each gene in each tissue (**Figure 1**):

(i) the *Age* effect, i.e., how the age-associated changes in gene expression evolve with age itself;

(ii) the *Sex* effect, i.e., how the differences in gene expression between sexes evolve with age;

(iii) the *Age&Sex* interaction effect, i.e., how the *Sex* effect changes with age, equivalent to how the *Age* effect differs between sexes; in other words, how the age-associated changes in differences in gene expression between sexes (or how the differences between sexes of age-associated changes in gene expression) evolve with age.

We named our approach Shifting Age Range Pipeline for Linear Modelling (ShARP-LM). Briefly, this method carries out differential gene expression analyses (with gene expression as a function of the donors' *Age*, *Sex* and *Age&Sex* interaction) in age windows spanning 16 years centered in consecutive years of age. By considering the percentage of genes differentially expressed in each age range, we highlight age periods of major tissue-specific transcriptomic alterations (**Figure 1**).

## *Gene*-centric analyses of human tissue-specific expression changes across age

The progression of tissue-specific expression of a specific gene across age can be examined in votAGEr's *Gene* tab. By entering its symbol in the Gene selector, the user has access to graphical summaries of the gene's tissue-specific expression (sub-tab *Profile*) (**Figure 2A**) and of the significance age-related changes in its expression due to the *Age*, *Sex*, and *Age&Sex* (sub-tab *Alteration*) (**Figure 2B**) across age. Results can be displayed in a heatmap for all tissues or in a scatter plot for a chosen individual tissue. When the gene is studied in a single tissue, the user can graphically and statistically profile the association of the donors' sex and reported conditions (e.g., history of heart attack or pneumonia) with the gene's expression profile (**Figure 2C**). A table summarizing the donors' metadata is also shown (**Figure 2C**). The

user can interactively select donors of interest on the scatter plot and further examine their information in the automatically subsetted table.

Cellular senescence, a stress-induced cell cycle arrest limiting proliferation of potentially oncogenic cells but progressively creating an inflammatory environment in tissues as they age, is an example of a process whose molecular mechanisms are of particular interest to aging researchers [41,42]. *CDKN2A*, encoding cell cycle regulatory protein $p16^{INK4A}$ that accumulates in senescent cells [43,44], has its expression increased with age in the vast majority of tissues profiled (**Supplementary Figure 1A**). Similarly, reduced levels of proliferation markers, such as *PCNA* [45] and *MKI67* [46], can be studied as putative markers of aging of certain tissues. These genes have their expression significantly altered with age in transverse colon and display a similar expression profile (decreasing from 20 to 30 years old (y.o.), constant until 55 y.o. and decreasing in older ages) (**Figure 2C** and **Supplementary Figure 1B**). However, these trends appear to vary according to the donor's history of pneumonia (**Supplementary Figure 1B**), illustrating voyAGEr's use in helping to find associations between gene expression and age-related diseases.

On a different note, sex biases have been reported in for the expression of *SALL1* and *KAL1* in adipose tissue and lung, respectively [10]. voyAGEr allows to not only recapitulate those observations but also estimate that these differences occur mostly in older ages (**Supplementary Figure 2**).

## *Tissue*-specific assessment of gene expression changes across age

### Peaks

The landscape of global tissue-specific gene expression alteration across age can be examined in voyAGEr's *Tissue* main tab. A heatmap displaying, for all tissues, the statistical significance over age (v. Materials and Methods) of the proportion of genes altered with *Age*, *Sex* or *Age&Sex* interaction (depending on the user's interest) is initially shown (**Figure 3A**), illustrating the aforementioned asynchronous aging of tissues observed for humans and rodents [18,32–36,39].

The user can then spot the age periods with the most significant gene expression alterations in a selected tissue (**Figure 3B a**), the so-called *Peaks* naming the sub-tab, and identify the

associated differentially expressed genes (**Figure 3B b**). The user can also plot the expression of a given gene of interest across age together with the significance of its expression modification with *Age*, *Sex* or *Age&Sex* (**Figure 3B c**).

As an example, the transverse colon appears to go through three main periods of transcriptional changes: one minor at around 27 y.o. (~5 % of genes differentially expressed (DEG)) and two major at around 55 y.o. (~21 % DEG) and 62 y.o. (~40 % DEG) (**Figure 3B a**). Interestingly, some of the genes most differentially expressed in the second peak at ~55 y.o. appear to have their expression modified only at this precise age period (e.g. *E2F1*, *KIF24*, *TRMU*, *CRIM1*). Similarly, genes such as *HOXA-AS3*, *NFYA*, *ZDHHC1*, and *MYL6B* (**Figure 3B b**), are specifically differentially expressed in the third peak at ~62 y.o. (**Figure 3B c**). This particularity is also found for the minor first peak. It therefore seems that different sets of genes drive the periods of major transcriptional changes, which begs assessing if they reflect the activation of distinct biological processes.

## Enrichment

The user can thus explore the biological functions of the set of genes underlying each peak of transcriptomic changes by assessing their enrichment in manually curated pathways from the Reactome database [47]. voyAGEr performs Gene Set Enrichment Analysis (GSEA) [48] and the user can visualize heatmaps displaying the evolution over age of the resulting normalized enrichment score (NES, reflecting the degree to which a pathway is over- or under-represented in a subset of genes) for the given tissue, effect (*Age*, *Sex* or *Age&Sex*) and all or some user-specified Reactome pathways (**Figure 4A**). To reduce redundancy and facilitate the understanding of their biological relevance, we clustered those pathways into families that also include Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways [49] and Gene Ontology (GO) Biological Processes of level 3 [50]. Briefly, we clustered gene sets from the 3 sources based on the overlap of their genes (v. Materials and Methods), thereby creating families of highly functionally related pathways. Taking advantage of the complementary and distinct terminology in Reactome, KEGG and GO, the user can interpret each family's broad biological function by looking at the word cloud of its most prevalent terms, still being able to browse the list of its associated pathways (**Figure 4A**).

The peaks of transcriptomic changes can also be examined for enrichment in a user-provided gene set (**Figure 4B**). To illustrate this procedure, we inspected the aforementioned three

aging peaks of the transverse colon for significant enrichment in genes retrieved from Senequest [42] whose link with senescence is supported by at least 4 sources, having observed it in the first (~27 y.o.) and second (~55 y.o.) peaks (**Figure 4B**).

## *Modules* of co-expressed genes

voyAGEr also allows functional analyses of modules of co-expressed genes (i.e. genes with highly correlated expression across samples, defined by weighted correlation network analysis [51]) altered with age. Genes in the same module are likely to be coregulated and share biological functions or associations with phenotypical or pathological traits [52]. Those modules may also act as markers of core transcriptional features of cellular activity and identity [53].

Concretely, voyAGEr enables the user to visualize how the expression of modules of genes that are associated with a specific cell type or disease progresses over age in a specific tissue. In the example of **Figure 5B**, the "midnight blue" module, gathering 11 genes co-expressed in the heart's left ventricle, shows significant enrichment in fibroblast markers, namely collagens. This module's expression in the left ventricle is steady until ~55 years of age and appears to increase later in life, perhaps reflecting the known age-related changes of the collagen network of the human myocardium [54].

After selecting one tissue of interest, the user has, for each module, access to four levels of information: (i) *Expression*: its eigengene expression progression over age; (ii) *Cell types*: its enrichment in specific cell types, based on marker genes found literature, (iii) *Pathways*: its enrichment in Reactome pathways; (iv) *Diseases*: its enrichment in disease markers, based on gene-disease associations from DisGeNET [55,56], calculated with both the `DOSE` package [57] and with Fisher's tests (**Figure 5C**).

By default, for each tissue, results are displayed in the form of heatmaps showing how all modules evolve with age (**Figure 5A**). The user has also the possibility to select a module of interest and see its eigengene progression over age in a scatter plot (**Figure 5B**), lists of its association with diseases and pathways ordered by significance, and a TreeMap for its cell type enrichment (**Figure 5B**). Like in the *Gene* tab, the user can separate donors based on their sex and previous condition in the scatter plot of the eigengene expression progression. When a significant enrichment results from a Fisher's test, the associated contingency table is shown for its clearer interpretation.

## Discussion

voyAGEr provides a framework to examine the progression over age of gene expression in a broad range of human tissues, thereby being a novel valuable resource for the aging research community. In particular, it helps to identify tissue-specific age periods of major transcriptomic alterations. The results of our analyses show the complexity of human biological aging by stressing its tissue specificity [34] and non-linear transcriptional progression throughout lifetime, consistent with previous results from both proteomic [58] and transcriptomic [28,34] analyses. By revealing and annotating the age-specific transcriptional trends in each tissue, voyAGEr aims at assisting researchers in deciphering the cellular and molecular mechanisms associated with the age-related physiological decline across the human body.

Besides, voyAGEr, to our knowledge unprecedently, scrutinizes and visually displays the tissue-specific differences in gene expression between biological sexes across age. Biological sex is an important variable in the prevalence of aging-associated diseases, as well as in their age of onset and progression [6,7,59] and tissue-specific sex-related biases in gene expression had been reported [8–11]. By profiling the age distribution of those biases, voyAGEr will thus help to better understand their influence in the etiology of the sex specificities of human aging. For instance, we were able to corroborate findings on the sex-differential transcriptome of human adults by Gershoni et al. [9], with voyAGEr emphasizing its tissue-specificity and allowing to discriminate the ages at which sex-related biases appear to be more prevalent (**Supplementary Figure 3**).

One of the limitations of voyAGEr is that most GTEx tissue donors had health conditions and their frequency increased with age, preventing us from defining a class of healthy individuals and identifying age-associated transcriptomic changes that could be more confidently proposed to happen independently of any disease progression. On one hand, we expect the large sample sizes and the biological variability between individuals, reflected on the diversity of combinations of conditions, to mitigate major confounding effects. On the other hand, voyAGEr allows users to assess how tissue-specific gene expression trends vary according to the donors' history of the different conditions (**Supplementary Figure 1B**), therefore helping to find associations between gene expression and age-related diseases.

The development of voyAGEr was accompanied by that of a pipeline, ShARP-LM, that facilitates the holistic depiction of the transcriptional landscapes of adult human tissues upon aging with a yearly age resolution. We take advantage of the comprehensiveness of the collection of transcriptomes from post-mortem human tissues from the GTEx project to make our analyses a valid surrogate of a currently undoable longitudinal study. It confers our method enough statistical robustness to mitigate the inter-individual, overcome the imbalance in the numbers of male and female samples and deal with the non-uniform distribution of the donors' ages. Indeed, we found no confounding between the distribution of samples' ages and the trend of gene expression progression over age in any tissue. However, the distribution of donor's ages does affect the statistical power of detection of transcriptional alterations and we are therefore more likely to find significant ones in age ranges that are more abundant in samples (**Supplementary Figure 4)**. This limitation is to be overcome by the accumulation of transcriptomes of human tissues in public databases, promising a gradually increase in accuracy and age resolution with which human transcriptomic aging can be profiled. Similarly, the accumulation of single-cell transcriptomes in public databases is yielding better gene markers for a growing diversity of human cell types, increasing the usefulness of exploiting bulk transcriptomes in studying the impact of aging on the cellular composition of human tissues, for which the co-expression module approach in voyAGEr provides a proof-of-concept.

As voyAGEr is open source, it is envisaged as an ever-evolving resource, that can be adapted by researchers to accommodate the aforementioned new data and expanded to provide new functionalities (e.g. integration with perturbagen data for the inference of molecular causes of observed gene expression alterations or small molecules for therapeutic purposes [60,61]).

As an *in silico* approach with no experimental validation for its results, voyAGEr is meant to be a discovery tool, supporting biologists of aging in the exploration of a large transcriptomic dataset, thereby generating, refining or preliminary testing hypotheses that will pave the way for subsequent experimental work. It can be an entry point for projects aiming at better understanding the tissue- and sex-specific transcriptional alterations underlying human aging, to be followed by more targeted studies focusing on the functional roles of the most promising markers identified therein in the physiology of aging. Those marker genes can

contribute to the development of more robust and cross-tissue gene signatures of aging [62] and the expansion of age-related gene databases [57,63].

Moreover, the observed diverse and asynchronous changes in gene expression between tissues over the human adult life provides potentially relevant information for the design of accurate diagnostic tools and personalized therapies. On one hand, finding those changes' association with specific disorders could have a prognostic value by enabling the identification of their onset before their clinical symptoms' manifestation [36]. On the other hand, computational screening of databases of genetic and pharmacologically induced human transcriptomic changes could help to infer their molecular causes and shed light on candidate drugs to delay their effects [60,61,64,65].

# Methods

## Development's platform

Data analysis was performed in R (version 3.6.3) and the application developed with R package `Shiny` [66]. voyAGEr's outputs are plots and tables, generated with R packages `highcharter` [67] and `DT` [68], respectively, that can easily be downloaded in standard formats (png, jpeg, and pdf for the plots; xls and csv for the tables).

voyAGEr was deployed using Docker Compose and ShinyProxy 2.6.0 in a Linux virtual machine (64GB RAM, 16 CPU threads and 200GB SSD) running in our institutional computing cluster.

The source code for voyAGEr is available at GitHub (https://github.com/nunolbmorais/voyAGEr).

## Read count data pre-processing

We downloaded the matrix with the RNA-seq read counts for each gene in each GTEx v7 sample from the project's data portal (https://www.gtexportal.org/) [40]. We then pre-processed the read count data for each tissue separately. We started by filtering out the uninformative genes with very low expression across samples. For that purpose, we normalised read counts for library size, to account for the differences in sequencing depth between samples, by transforming them into counts per million (CPM) with function `cpm` from R package `edgeR` [69]. Only genes with at least 1 CPM in at least 40% of the samples were kept for analysis (the number of genes analysed for each tissue can be found in **Supplementary Table 1**). Read counts for those kept genes were used to calculate normalization factors to scale the raw library sizes, using function `calcNormFactors` from `edgeR` that implements the trimmed mean of M-values [70], and normalised accordingly and subsequently log-transformed with the `voom` function [71] from package `limma` [72], which also estimates mean-variance relationships to be used in linear modelling (v. ShARP-LM below). Normalized gene expression was thereby ready to be analyzed and featured in the plots in log2 of CPM (logCPM), this log-transformation yielding more manageable nearly normal distributions.

Finally, from the 53 tissues available from GTEx v7, we discarded five (renal cortex, fallopian tube, bladder, ectocervix, endocervix, chronic myeloid leukemia cell line) due to a low (<50) number of respective samples.

## ShARP-LM

To model the changes in gene expression with age, we developed the Shifting Age Range Pipeline for Linear Modelling (ShARP-LM). For each tissue, we fitted linear models to the gene expression of samples from donors with ages within windows with a range of 16 years shifted through consecutive years of age (i.e. in a sliding window with window size = 16 and step size = 1 years of age). As samples at the ends of the dataset's age range would be thereby involved in fewer linear models, we made the window size gradually increase from 11 to 16 years when starting from the "youngest" samples and decrease from 16 to 11 years when reaching the "oldest" (**Supplementary Figure 5)**.

Function `lm` from `limma` was used to fit the following linear model for gene expression (GE):

GE ~ $\alpha$.*Age* + $\beta$.*Sex* + $\chi$.*Age&Sex* + $\varepsilon$

For each gene, $\alpha$, $\beta$ and $\chi$ are the coefficients to be estimated for the respective hypothesized effects and $\varepsilon$ is the residual. For each sample, *Age* in years and *Sex* in binary were centered and *Age&Sex* interaction given by their product.

For each gene in each model (i.e., each age window in each tissue), we retrieved the t-statistics of differential expression associated with the three variables and their respective p-values. We considered the average age of the samples' donors within the age window as the representative age of the observed expression changes.

In summary, for a given tissue and variable (*Age*, *Sex* and *Age&Sex*), ShARP-LM yields t-statistics and p-values over age for all genes, reflecting the magnitude and significance of the changes in their expression throughout adult life.

## *Gene*-centric visualization of tissue-specific expression changes across age

For visualization purposes, the trend of each gene's expression progression over age in each tissue was derived through Local Polynomial Regression Fitting, using R function `loess` on logCPM values (**Figures 1, 2C, 3B**). For summarizing in a heatmap a given gene's expression

across age in multiple tissues (**Figure 2A**) or the expression of several genes across age in a given tissue, each gene's regression values are centered, but not scaled, with R function `scale`.

For summarizing in a heatmap the significance of a given gene's expression changes over age in multiple tissues (**Figure 2B**), cubic smoothing splines are fitted to -log10(p), p being the t-statistic's p-value, with R function `smooth.spline`.

## Tissue-specific quantification of global transcriptomic alterations across age

To assess the global transcriptomic impact of each of the three modelled effects in each tissue across age, we then analyzed the progression over age (i.e. over consecutive linear models) of the percentage of genes whose expression is significantly altered (t-statistic's p-value ≤ 0.01) by each effect (**Figure 3B**). To evaluate the significance of each percentage and see if high ones can be confidently associated with major transcriptomic alterations, we controlled for their false discovery rate (FDR) by randomly permuting the samples' ages within each age window one million times and performing ShARP-LM on each randomised dataset. We were then able to associate an FDR to each percentage of differentially expressed genes by comparing it with the distribution of those randomly generated (**Figure 3B**).

## Gene Set Enrichment Analysis (GSEA)

For each *Peak* of significant gene expression modifications, we performed GSEA [48] on the ordered (from the most positive to the most negative) t-statistics of differential expression for the respective tissue and age, using R package `fgsea` [73] and the Reactome database [47]. The resulting normalized enrichment score (NES) for each pathway was used in subsequent analyses as a metric of its over- or down-representation in the *Peak*.

## Families of pathways

To reduce pathways' redundancy and facilitate the assessment of their biological relevance in results' interpretation, we created an unifying representation of pathways from Reactome [47] and KEGG [49], and level 3 Biological Processes from GO [50], by adapting a published pathway clustering approach [74] to integrate them into families.

The approach relies on the definition of a hierarchy of pathways based on the numbers of genes they have in common. For each two pathways Pi and Pj, respectively containing sets of genes Gi and Gj, we computed their overlap index [74], defined as following:

$Oiij = |Gi \cap Gj| / \min(|Gi|, |Gj|)$

Where $|Gi|$ is the number of genes in set Gi and $|Gi \cap Gj|$ is the number of genes in common between Gi and Gj. Oiij = 1 would therefore indicate that Pi and Pj are identical in gene composition or that one is a subset of the other. On the contrary, Oiij = 0 would mean that Pi and Pj have no genes in common. To ease the computational work, we removed from the analysis pathways that are subsets of larger pathways (i.e. each pathway whose genes are all present in another pathway).

From the Oiij matrix, from which each row is a vector with the gene overlaps of pathway I with each of all the pathways, we computed the matrix of Pearson's correlation between all pathways' overlap indexes with R function `cor`. That matrix was finally transformed into Euclidean distances with R function `dist`, allowing for pathways to be subsequently clustered with the complete linkage method with R function `hclust`. The final dendrogram was empirically cut in 10 clusters (**Supplementary Figure 6**). One of them, not including Reactome pathways, was dismissed from the analysis. Pathways that initially excluded from the computation for being subsets of others were finally added to the clusters of their respective *parent* pathways. Each *daughter* pathway with more than one parent was assigned to the cluster of the parent with the smallest number of genes, thereby maximizing the daughter-parent similarity. The `data.table` R package, for fast handling of large matrices, was used in this analysis [75].

## Gene co-expression modules

Gene co-expression modules were defined with R package `WGCNA` [51].

For each tissue, genes were first filtered based on the variance of their expression across samples (*variable A*), with lowly variant ones being discarded from subsequent co-expression analyses. Then, an adjacency matrix was computed with the `adjacency` function. To this end, we first computed the matrix of Pearson's correlation between all selected genes' expressions and then raised all of its values to a power $\beta$ (soft thresholding) to emphasise stronger correlations. The soft threshold was determined with the

`pickSoftThreshold` function. The thereby generated matrix ($|correlation|^{\beta}$) was then converted into a dissimilarity matrix based on the topological overlap measure, determined with function `TOMsimilarity`. Finally, gene co-expression modules were identified with a dynamic tree-cutting algorithm, implemented with the `cutreeDynamic` function, that adjusts the minimal number of genes per module (*variable B*) and the height at which the tree is cut (*variable C*).

The definition of those modules of co-expressed genes is highly dependent on the three variables *A*, *B*, and *C*. Thus, we comprehensively tested combinations of them and defined as optimal those that maximized the modules' enrichment in known markers for specific cell types, an approach similar to that reported to have been successfully applied in the revelation of the core transcriptional features of cell types in the human central nervous system [53]. In particular, we used human and murine cell type gene signatures, derived from single cell transcriptomic analyses, for brain (cortex) [76–79], skeletal muscle [80–82], left cardiac ventricle [82–84], and whole blood [82,85,86]. Finally, the 40%, 30%, 35%, and 40% most variant genes, the tree cut at a height of 0.96, and a minimum number of 20, 18, 11, and 16, were respectively chosen for the three variables in those four tissues.

Each module is then represented by a set of genes and an eigengene, the first principal component obtained by singular-value decomposition of the expression of the module's genes and representative of its own expression profile.

The enrichment of modules in cell types, Reactome pathways and diseases from DisGeNET [55,56]was quantified with Fisher's tests. Disease enrichment was also calculated using the `DOSE` package [57].

## Data availability

Processed GTEx v7 RNA-seq data (read count tables) were downloaded from the project's data portal (https://www.gtexportal.org/). Donor metadata were obtained from dbGaP - database of Genotypes and Phenotypes (Accession phs000424.v7.p2; project ID 13661). voyAGEr's output tables be directly downloaded in standard xls and csv formats.

## Acknowledgements

We thank iMM colleagues Joana Neves and Luísa Lopes, as well as all members from the Disease Transcriptomics Lab, for providing valuable feedback on the manuscript.
Funding: This work was supported by European Molecular Biology Organization (EMBO Installation Grant 3057), FCT - Fundação para a Ciência e a Tecnologia I.P. (FCT Investigator Starting Grant IF/00595/2014, CEEC Individual Assistant Researcher contract CEECIND/00436/2018, PhD Studentship SFRH/BD/131312/2017, project LA/P/0082/2020), European Union (EU) Horizon 2020 Research and Innovation Programme (RiboMed 857119), and GenomePT project (POCI-01-0145-FEDER-022184), supported by COMPETE 2020 – Operational Programme for Competitiveness and Internationalization (POCI), Lisboa Portugal Regional Operational Programme (Lisboa2020), Algarve Portugal Regional Operational Programme (CRESC Algarve2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF), and by FCT - Fundação para a Ciência e a Tecnologia.

## Authors' contributions

A.S. and N.L.B.-M. designed the study; A.S. performed all the development and bioinformatics work, under the supervision of N.L.B.-M.; N.S.-A. set up the web and database implementation; A.S. and N.L.B.-M. wrote the manuscript, with contributions from N.S.-A..

## Conflict of Interest statement

The authors declare no conflict of interest.

## Figures and Tables

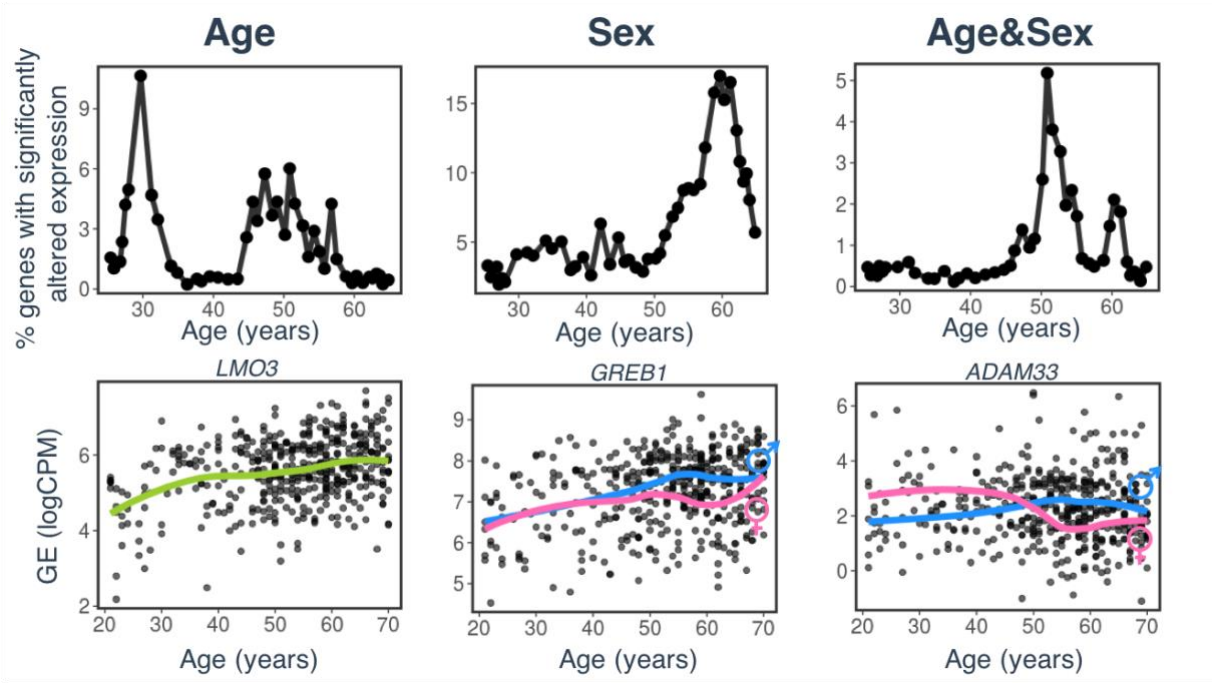



**Figure 1**: *voyAGEr profiles tissue-specific age-associated changes in gene expression and their differences between sexes.*

For each of the 48 human tissues in GTEx, gene expression was linearly modelled in windows spanning 16 years centered in consecutive years of age, to estimate the effects thereon of *Age*, *Sex* and the interaction between them (*Age&Sex*), i.e., how the *Sex* effect changes with age, equivalent to how the *Age* effect differs between sexes (v. Materials and Methods). In each window (i.e., for each age in consecutive years), the percentage of genes with expression significantly altered by of each of those effects gives their respective transcriptomic impact (upper panels). voyAGEr thereby identifies the age periods at which major gene expression changes occur in each tissue. For example, in subcutaneous adipose tissue: major age-related transcriptional alterations are found at 30 and around 50 years of age (upper left panel), illustrated by the behavior of *LMO3* (bottom left panel); major gene expression differences between males and females happen at around 60 years of age (upper center panel), as illustrated by *GREB1* (bottom center panel); major differences between sexes in age-related gene expression alterations happen at around 50 years of age (upper right panel), as illustrated by *ADAM33* (bottom right panel). Solid loess lines in the bottom

panels (green for all donors, pink for females, blue for males). Gene expression (GE) in log2 of counts per million (logCPM).

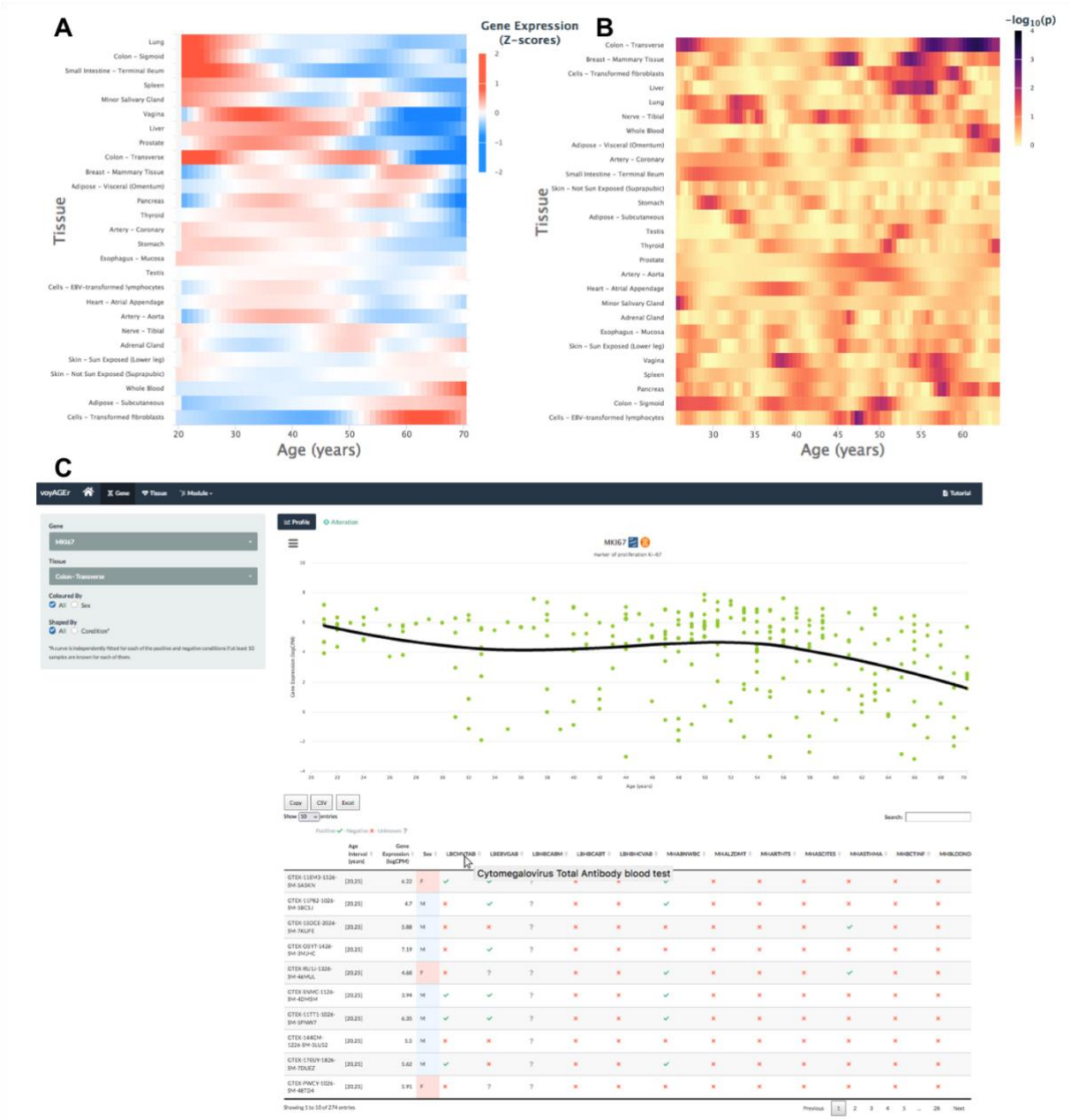



**Figure 2**: *Gene-centric analyses of expression alterations across age*

A. Heatmap of *MKI67* expression across tissues over age.

B. Heatmap of significance of *Age*-associated *MKI67* expression alterations over age. p-values are for the moderated t-statistics of differential gene expression associated with the Age effect (v. the ShARP-LM approach in Materials and Methods). Notably, major transcriptional changes are observed in transverse colon (mid 20's and after 55), breast (mid/late 40's and early/mid 50's), and liver (mid 50's).

C. voyAGEr's *Gene* tab interface. *MKI67* expression in transverse colon is inspected. Donors' information is shown in a table and the scatter plot can be interactively adjusted according to the donors' condition of interest **(Supplementary Figure 1B)**.

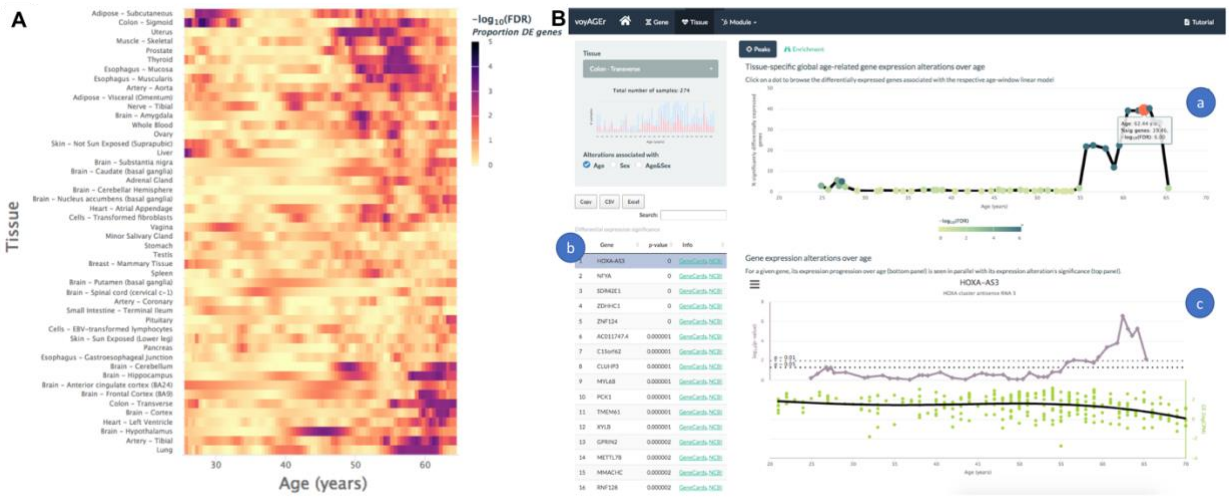



**Figure 3**: *Tissue-specific assessment of gene expression changes across age*

A. Heatmap of significance (FDR based on random permutations of age, v. Materials and Methods) over age of the proportion of genes with expression significantly altered with *Age* in the 48 analyzed tissues.

B. Exploration of gene expression changes across age in transverse colon.

   a. Percentage of genes with significantly altered expression with *Age* over age. Three peaks at 27, 55 and 62 years old are noteworthy.

   b. Clicking on the dot of a specific age (62.44 y.o. in plot *a*) gives access to the list of the most differentially expressed genes at that age, ordered by statistical significance of expression changes (p-value of moderated t-statistic).

   c. Plot of expression of the chosen top gene in the table in *b* across age (bottom) in parallel with the significance of its expression alterations with *Age*. The expression of *HOXA-AS3* significantly decreases at around 62 years old, concomitantly with the third peak of transcriptomic changes with *Age*.

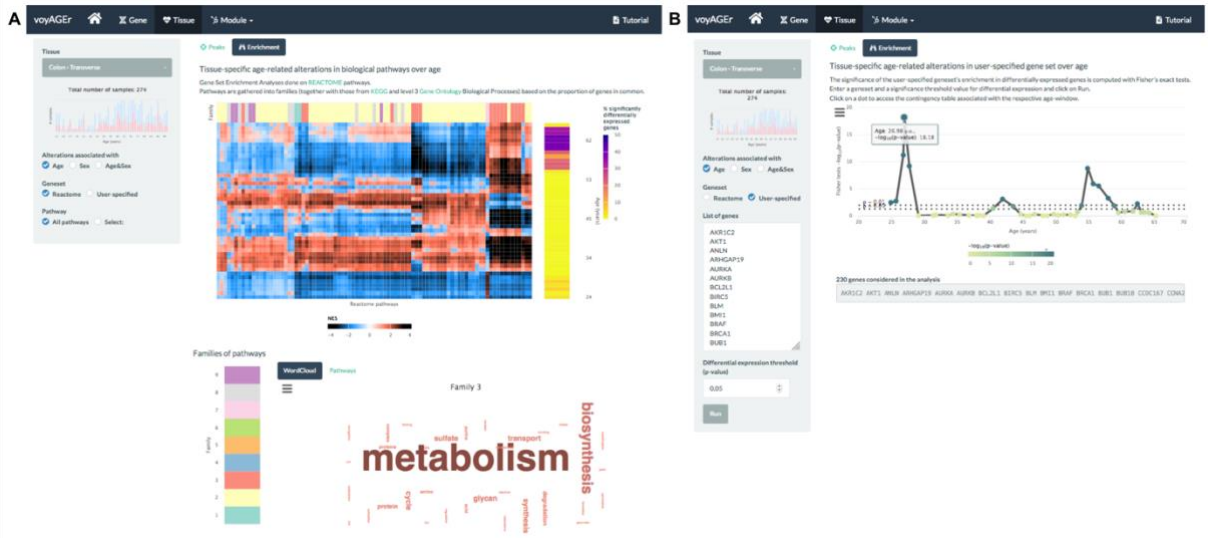



**Figure 4**: *Tissue-specific assessment of pathway expression changes across age*

A. Heatmap of enrichment (above), given by GSEA's normalized enrichment score (NES), of Reactome pathways in genes differentially expressed with *Age* over age. Pathways are classified into broad families (upper bar), whose associated biological processes and functions can be learnt from the word clouds of their most frequent terms (below). Only Reactome pathways found significantly associated (FDR ≤ 0.01) with gene expression alterations in at least one age window are included.

B. Enrichment, given by the significance of Fisher's tests, of a user-provided gene set in genes differentially expressed (based on a user-defined p-value threshold) with *Age* over age. Here, the given gene set is composed of genes from Senequest [42] whose link with senescence is supported by at least 4 sources. The first and second peaks of *Age*-associated transcriptomic changes in the transverse colon appear to be enriched in senescence-related genes.

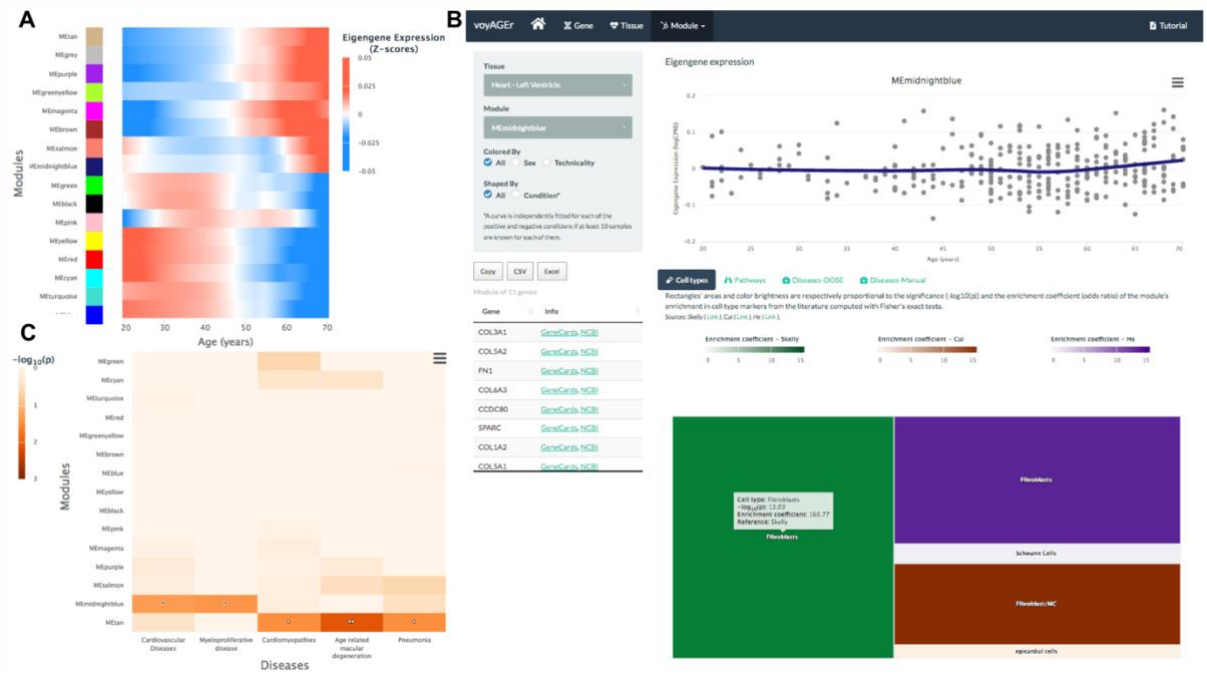



**Figure 5**: *Tissue-specific assessment of age-associated progression of modules of co-expressed genes*

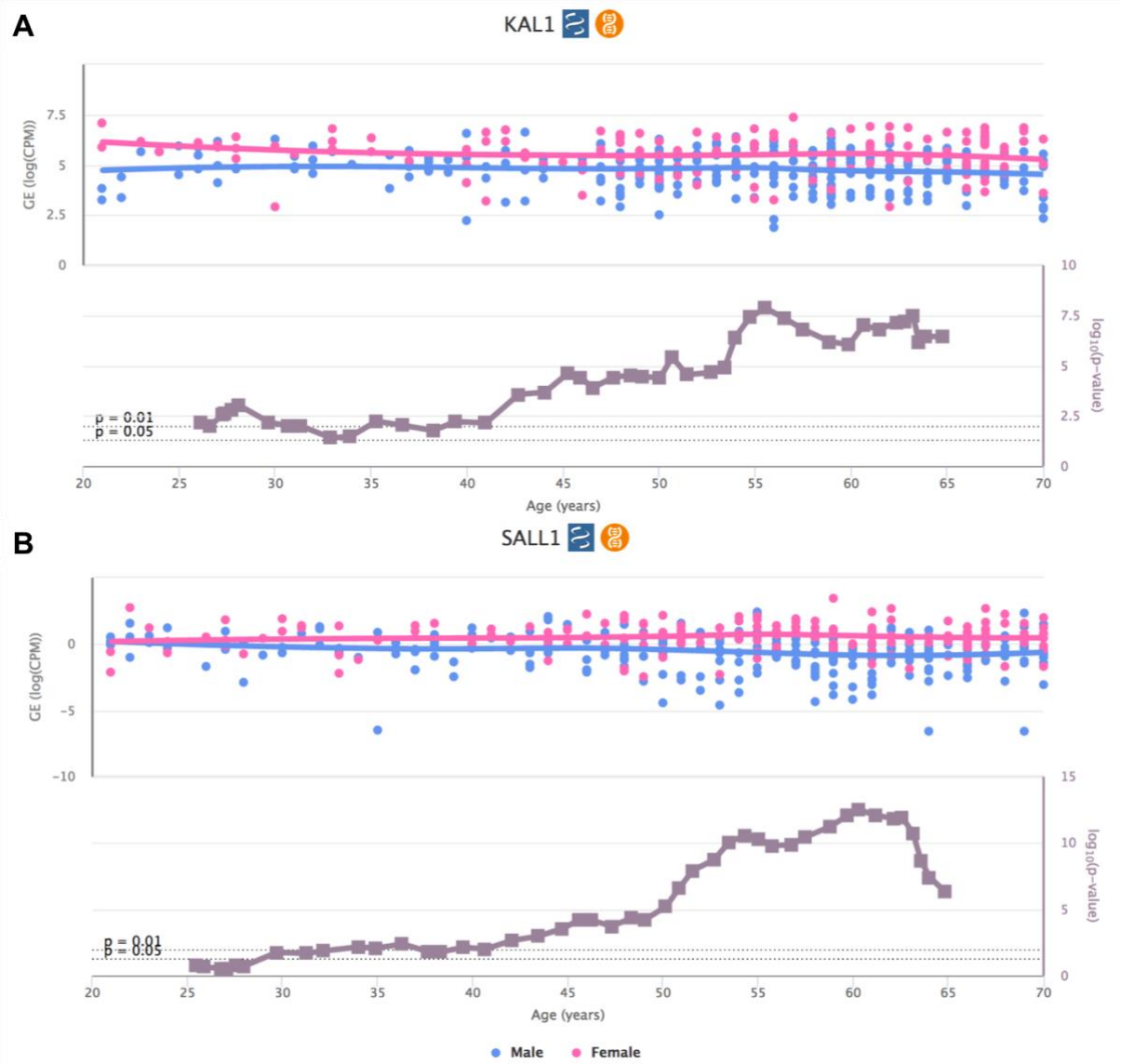
A. Heatmap of eigengene expression for all the inferred modules of co-expressed genes in the heart's left ventricle over age.
B. Scatter plot (above) of eigengene expression over age, in all left ventricle samples, for a selected module of 11 genes. The eigengene expression is derived from the 1st component of the single value decomposition of its genes' expression. This module was analysed, with Fisher's tests, for enrichment in cell types, based on markers from the literature (Skelly [84], Cui [83] and He [82], surnames of the first authors of the studies the markers are retrieved from) and found to be associated with fibroblasts, as revealed by the TreeMap below (where each rectangle's area and darkness are proportional to the significance of its association with a cell type and its color linked to the markers' source study).
C. Heatmap of association of the modules with five selected diseases, computed with the DOSE package [57].

## Supplementary Figures and Tables

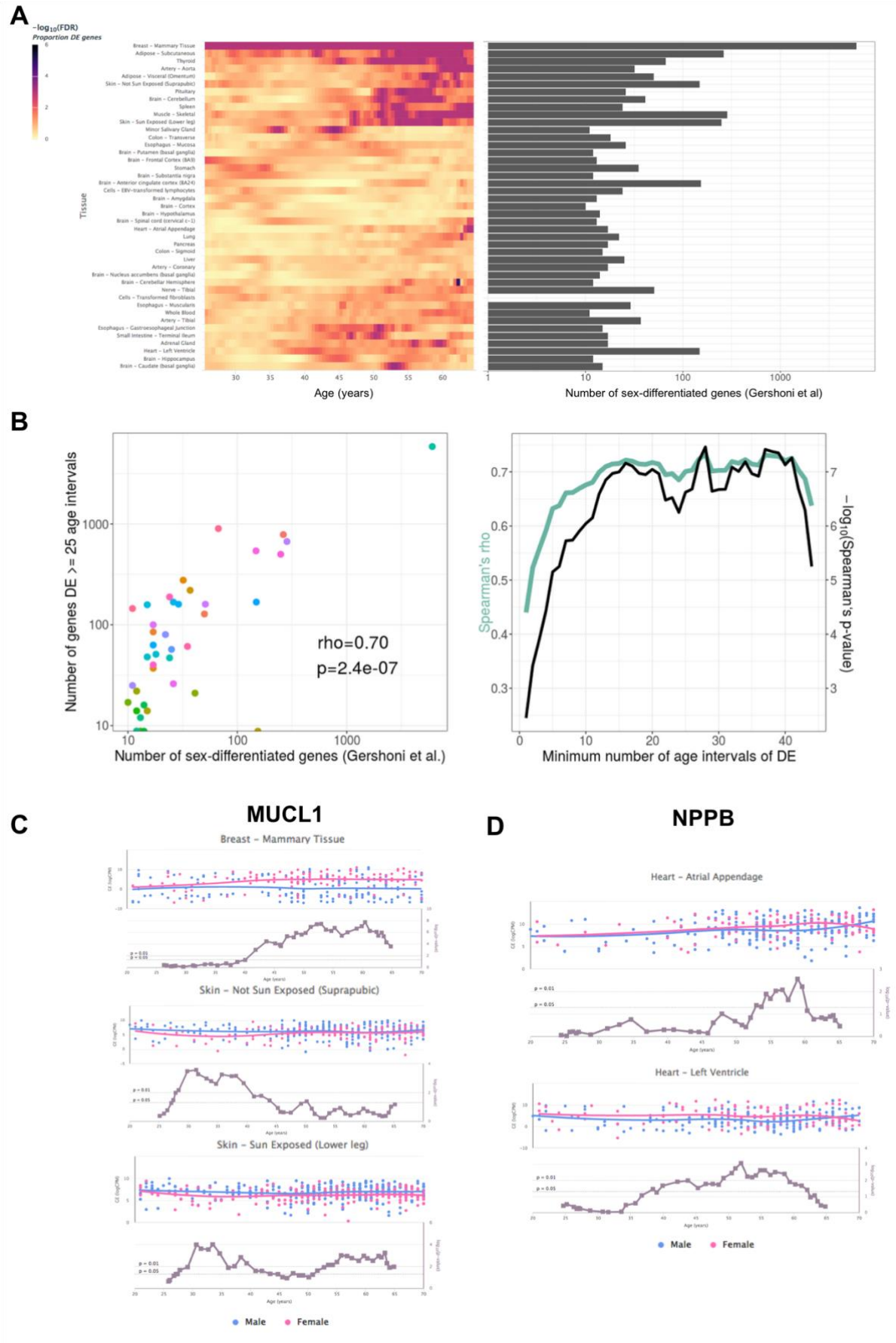
**Supplementary Figure 1**: *voyAGEr's Gene tab interface*.

A. Heatmap of *CDKN2A* expression across tissues over age.

B. *PCNA* expression in transverse colon over age and its association with the donor's history of pneumonia.




**Supplementary Figure 2**: *Sex-specific SALL1 (A) and KAL1 (B) expression alterations over age.* Scatter plots of *SALL1* (A) and *KAL1* (B) expression over age (above), where dots are colored based on the donor's sex (male in blue and female in pink), are shown together with those of the significance of their difference between sexes (below).

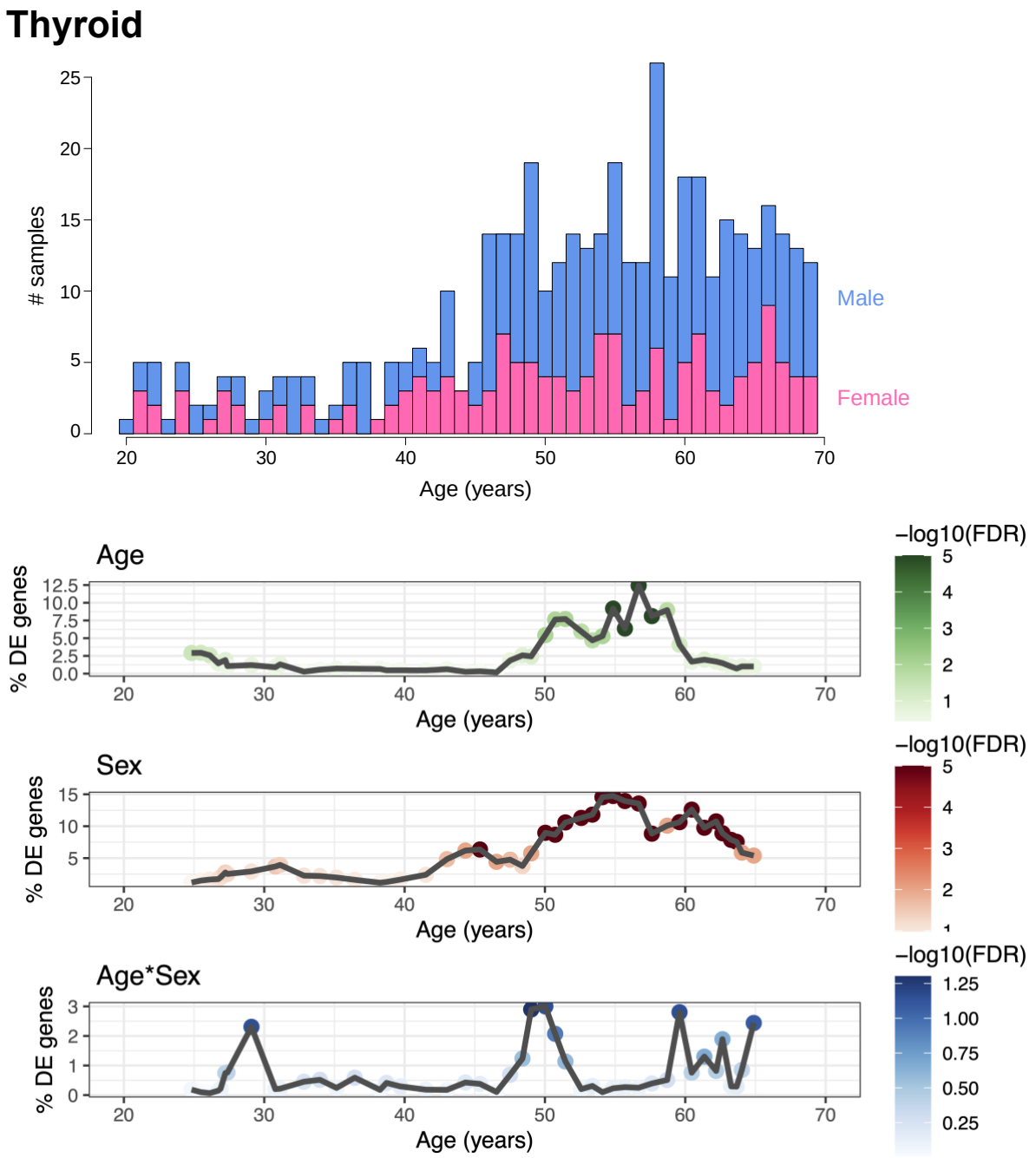



**Supplementary Figure 3**: *Tissue-specific sex-differentiated expression.*

A. Heatmap of significance over age, for each tissue, of the proportion of differentially expressed genes between sexes (left), in parallel with the number of sex-differentiated genes highlighted by Gershoni et al. (v. Figure 1 in [9]) (right). Mammary is corroborated as the tissue with the most sex-differentiated gene expression. For tissues like adipose, skeletal muscle, skin, and heart, sex-related biases in expression mainly arise in late life.

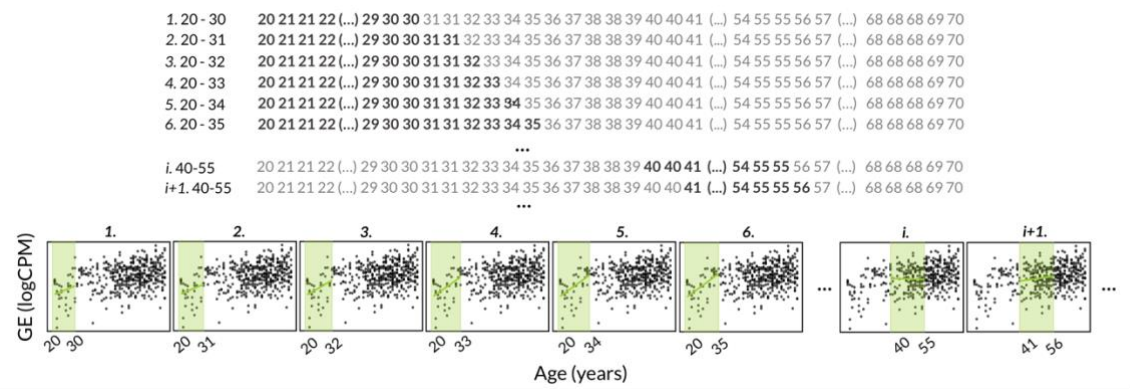
B. Scatter plot relating, for each tissue (each dot), the number of sex-differentiated genes found by Gershoni et al. with the number of genes found by voyAGEr to be differentially expressed between sexes in at least 25 age intervals (left). This correlation is significant regardless of the considered minimum number of age intervals a gene must be differentially expressed in (right). Half visible dots in the left scatter plot represent tissues found to have no sex-differentiated genes by voyAGEr (the log scale would give them a minus infinite value on the Y-axis).

C. and D. Two examples of genes, *MUCL1* (C) and *NPPB* (D), whose expression is known to be sex-differentiated in a tissue-specific manner (Figures S7 and 3 in [9], respectively). Plots of their sex-specific expression across age in parallel with the significance of their differences in expression alterations between sexes are shown for tissues reported in [9] as having sex-differentiated gene expression: breast and skin for *MUCL1* and heart for *NPPB*. voyAGEr helps to refine the definition of the age periods at which those biases occur: late-life in breast and early-life in skin for *MUCL1*, and late-life in heart for *NPPB*.



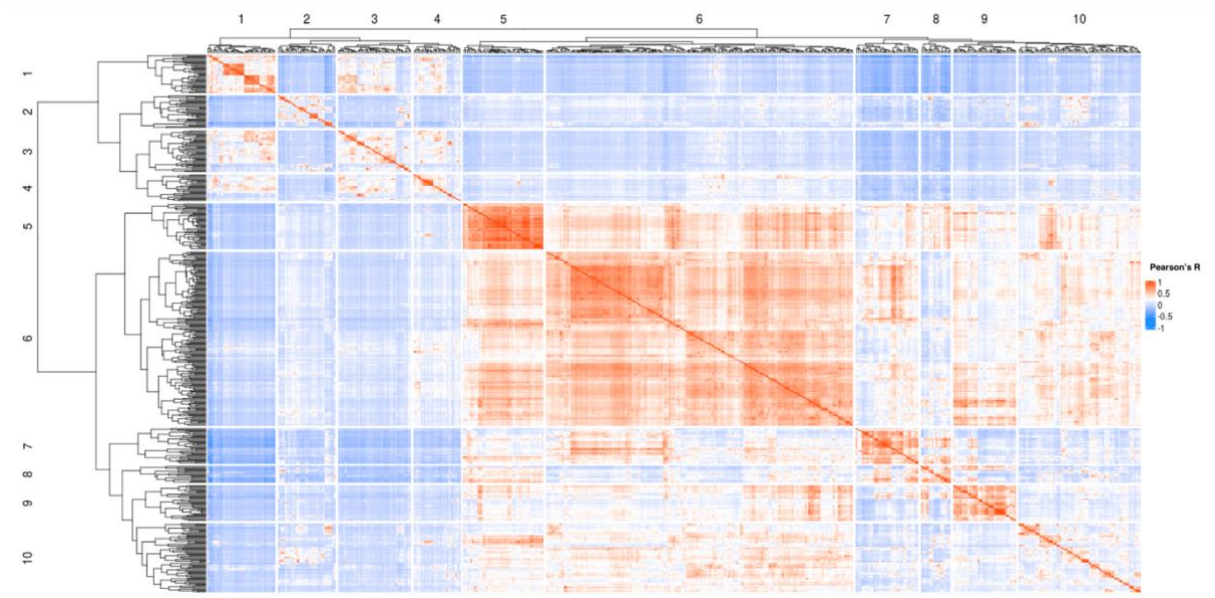

**Supplementary Figure 4**: *Distributions of GTEx donors' ages and its impact on the statistical power of detection of differential gene expression.*

For thyroid as an example tissue, the sex-differentiated distribution of GTEx samples' ages (v7) (above) are shown together with the percentage of differentially expression genes (% DE genes) over age (below) for the three variables (*Age*, *Sex* and *Age&Sex*), as well as their statistical significance, given by the ShARP-LM approach.




**Supplementary Figure 5**: *Principle of the ShARP-LM method.*

ShARP-LM consists of fitting, for each tissue, linear models to the gene expression of samples from donors with ages within windows with a range of 16 years shifted through consecutive years of age, in a sliding window with window size = 16 and step size = 1 years of age. As samples at the ends of the dataset's age range would be thereby involved in fewer linear models, we made the window size gradually increase from 11 to 16 years when starting from the "youngest" samples and decrease from 16 to 11 years when reaching the "oldest". The shown example to illustrate the approach is the expression of *LMO3* in subcutaneous adipose tissue over age, with 8 age windows are shown. Colored rectangles represent the age windows and colored lines the linear modelling of the gene expression over age in each of them.




**Supplementary Figure 6** *Clustering of pathways from Reactome and KEGG, and level 3 Gene Ontology Biological Processes.*

Based on their common genes (v. Materials and Methods), pathways from curated databases were clustered in 10 families.

| Tissue | #Genes |
|---|---|
| Adipose - Subcutaneous | 14151 |
| Adipose - Visceral (Omentum) | 14305 |
| Adrenal Gland | 14075 |
| Artery - Aorta | 13798 |
| Artery - Coronary | 14121 |
| Artery - Tibial | 13510 |
| Bladder | 14696 |
| Brain - Amygdala | 14575 |
| Brain - Anterior cingulate cortex (BA24) | 14588 |
| Brain - Caudate (basal ganglia) | 14778 |
| Brain - Cerebellar Hemisphere | 14291 |
| Brain - Cerebellum | 14497 |
| Brain - Cortex | 14736 |
| Brain - Frontal Cortex (BA9) | 14665 |
| Brain - Hippocampus | 14627 |
| Brain - Hypothalamus | 15030 |
| Brain - Nucleus accumbens (basal ganglia) | 14756 |
| Brain - Putamen (basal ganglia) | 14516 |
| Brain - Spinal cord (cervical c-1) | 14602 |
| Brain - Substantia nigra | 14566 |
| Breast - Mammary Tissue | 14738 |
| Cells - EBV-transformed lymphocytes | 12360 |
| Cells - Transformed fibroblasts | 12787 |
| Cervix - Ectocervix | 14281 |
| Cervix - Endocervix | 15289 |
| Colon - Sigmoid | 14190 |
| Colon - Transverse | 14826 |
| Esophagus - Gastroesophageal Junction | 14030 |
| Esophagus - Mucosa | 14209 |
| Esophagus - Muscularis | 13982 |
| Fallopian Tube | 15052 |
| Heart - Atrial Appendage | 13852 |
| Heart - Left Ventricle | 13118 |
| Kidney - Cortex | 14775 |
| Liver | 13298 |
| Lung | 14953 |
| Minor Salivary Gland | 14751 |
| Muscle - Skeletal | 12473 |
| Nerve - Tibial | 14627 |
| Ovary | 14143 |
| Pancreas | 13754 |
| Pituitary | 15342 |
| Prostate | 15127 |
| Skin - Not Sun Exposed (Suprapubic) | 14635 |
| Skin - Sun Exposed (Lower leg) | 14637 |
| Small Intestine - Terminal Ileum | 15065 |
| Spleen | 14588 |
| Stomach | 14558 |
| Testis | 17965 |
| Thyroid | 14801 |
| Uterus | 14236 |
| Vagina | 14882 |
| Whole Blood | 12043 |

**Supplementary Table 1**: *Number of studied genes per tissue.*

For each tissue, we kept the genes with at least 1 CPM of expression in at least 40% of the samples.

## References


1. Campisi, J. Aging, Cellular Senescence, and Cancer. *Annu. Rev. Physiol.* **75**, 685–705 (2013).
2. Niccoli, T. & Partridge, L. Ageing as a Risk Factor for Disease. *Curr. Biol.* **22**, R741–R752 (2012).
3. Wyss-Coray, T. Ageing, neurodegeneration and brain rejuvenation. *Nature* **539**, 180–186 (2016).
4. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The Hallmarks of Aging. *Cell* **153**, 1194–1217 (2013).
5. Stegeman, R. & Weake, V. M. Transcriptional Signatures of Aging. *J. Mol. Biol.* **429**, 2427–2437 (2017).
6. Tower, J. Sex-Specific Gene Expression and Life Span Regulation. *Trends Endocrinol. Metab.* **28**, 735–747 (2017).
7. Austad, S. N. & Fischer, K. E. Sex Differences in Lifespan. *Cell Metab.* **23**, 1022–1033 (2016).
8. Mele, M. *et al.* The human transcriptome across tissues and individuals. *Science (80-. ).* **348**, 660–665 (2015).
9. Gershoni, M. & Pietrokovski, S. The landscape of sex-differential transcriptome and its consequent selection in human adults. *BMC Biol.* **15**, 7 (2017).
10. Kassam, I., Wu, Y., Yang, J., Visscher, P. M. & McRae, A. F. Tissue-specific sex differences in human gene expression. *Hum. Mol. Genet.* **28**, 2976–2986 (2019).
11. Mayne, B. T. *et al.* Large Scale Gene Expression Meta-Analysis Reveals Tissue-Specific, Sex-Biased Gene Expression in Humans. *Front. Genet.* **7**, (2016).
12. Zahn, J. M. *et al.* AGEMAP: A Gene Expression Database for Aging in Mice. *PLoS Genet.* **3**, e201 (2007).
13. Kimmel, J. C. *et al.* Murine single-cell RNA-seq reveals cell-identity- and tissue-specific trajectories of aging. *Genome Res.* **29**, 2088–2103 (2019).
14. Benayoun, B. A. *et al.* Remodeling of epigenome and transcriptome landscapes with aging in mice reveals widespread induction of inflammatory responses. *Genome Res.* **29**, 697–709 (2019).

15. Lui, J. C., Chen, W., Barnes, K. M. & Baron, J. Changes in gene expression associated with aging commonly originate during juvenile growth. *Mech. Ageing Dev.* **131**, 641–649 (2010).
16. Shavlakadze, T. *et al.* Age-Related Gene Expression Signature in Rats Demonstrate Early, Late, and Linear Transcriptional Changes from Multiple Tissues. *Cell Rep.* **28**, 3263-3273.e3 (2019).
17. The Tabula Muris Consortium. A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature* **583**, 590–595 (2020).
18. Schaum, N. *et al.* Ageing hallmarks exhibit organ-specific temporal signatures. *Nature* **583**, 596–602 (2020).
19. Zahn, J. M. *et al.* Transcriptional profiling of aging in human muscle reveals a common aging signature. *PLoS Genet.* **preprint**, e115 (2005).
20. Welle, S., Brooks, A. I., Delehanty, J. M., Needler, N. & Thornton, C. A. Gene expression profile of aging in human muscle. *Physiol. Genomics* **14**, 149–159 (2003).
21. Jozsi, A. C. *et al.* Aged human muscle demonstrates an altered gene expression profile consistent with an impaired response to exercise. *Mech. Ageing Dev.* **120**, 45–56 (2000).
22. Rodwell, G. E. J. *et al.* A Transcriptional Profile of Aging in the Human Kidney. *PLoS Biol.* **2**, e427 (2004).
23. Galatro, T. F. *et al.* Transcriptomic analysis of purified human cortical microglia reveals age-associated changes. *Nat. Neurosci.* **20**, 1162–1171 (2017).
24. Olah, M. *et al.* A transcriptomic atlas of aged human microglia. *Nat. Commun.* **9**, 539 (2018).
25. Berchtold, N. C. *et al.* Gene expression changes in the course of normal brain aging are sexually dimorphic. *Proc. Natl. Acad. Sci.* **105**, 15605–15610 (2008).
26. Lu, T. *et al.* Gene regulation and DNA damage in the ageing human brain. *Nature* **429**, 883–891 (2004).
27. Işıldak, U., Somel, M., Thornton, J. M. & Dönertaş, H. M. Temporal changes in the gene expression heterogeneity during brain development and aging. *Sci. Rep.* **10**, 4080 (2020).
28. Haustead, D. J. *et al.* Transcriptome analysis of human ageing in male skin shows mid-life period of variability and central role of NF-κB. *Sci. Rep.* **6**, 26846 (2016).

29. Holzscheck, N. *et al.* Multi-omics network analysis reveals distinct stages in the human aging progression in epidermal tissue. *Aging (Albany. NY).* **12**, 12393–12409 (2020).
30. Nakamura, S. *et al.* Identification of blood biomarkers of aging by transcript profiling of whole blood. *Biochem. Biophys. Res. Commun.* **418**, 313–318 (2012).
31. Harries, L. W. *et al.* Human aging is characterized by focused changes in gene expression and deregulation of alternative splicing. *Aging Cell* **10**, 868–878 (2011).
32. Thomas, R. Age-Associated Changes in Gene Expression Patterns in the Liver,. *J. Gastrointest. Surg.* **6**, 445–454 (2002).
33. Yoshida, S., Yashar, B. M., Hiriyanna, S. & Swaroop, A. Microarray analysis of gene expression in the aging human retina. *Invest. Ophthalmol. Vis. Sci.* **43**, 2554–60 (2002).
34. Gheorghe, M. *et al.* Major aging-associated RNA expressions change at two distinct age-positions. *BMC Genomics* **15**, 132 (2014).
35. Yang, J. *et al.* Synchronized age-related gene expression changes across multiple tissues in human and the link to complex diseases. *Sci. Rep.* **5**, 15145 (2015).
36. Aramillo Irizar, P. *et al.* Transcriptomic alterations during ageing reflect the shift from cancer to degenerative diseases in the elderly. *Nat. Commun.* **9**, 327 (2018).
37. Glass, D. *et al.* Gene expression changes with age in skin, adipose tissue, blood and brain. *Genome Biol.* **14**, R75 (2013).
38. Yu, Y. *et al.* A rat RNA-Seq transcriptomic BodyMap across 11 organs and 4 developmental stages. *Nat. Commun.* **5**, 3230 (2014).
39. Rando, T. A. & Wyss-Coray, T. Asynchronous, contagious and digital aging. *Nat. Aging* **1**, 29–35 (2021).
40. Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
41. Van Deursen, J. M. The role of senescent cells in ageing. *Nature* **509**, 439–446 (2014).
42. Gorgoulis, V. *et al.* Cellular Senescence: Defining a Path Forward. *Cell* **179**, 813–827 (2019).
43. Gil, J. & Peters, G. Regulation of the INK4b–ARF–INK4a tumour suppressor locus: all for one or one for all. *Nat. Rev. Mol. Cell Biol.* **7**, 667–677 (2006).
44. Erickson, S. *et al.* Involvement of the Ink4 proteins p16 and p15 in T-lymphocyte senescence. *Oncogene* **17**, 595–602 (1998).
45. Narita, M. *et al.* Rb-Mediated Heterochromatin Formation and Silencing of E2F Target

Genes during Cellular Senescence. *Cell* **113**, 703–716 (2003).

46. Sun, X. & Kaufman, P. D. Ki-67: more than a proliferation marker. *Chromosoma* **127**, 175–186 (2018).
47. Croft, D. *et al.* The Reactome pathway knowledgebase. *Nucleic Acids Res.* **42**, D472–D477 (2014).
48. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–50 (2005).
49. Kanehisa, M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
50. Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* **32**, 258D – 261 (2004).
51. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
52. van Dam, S., Võsa, U., van der Graaf, A., Franke, L. & de Magalhães, J. P. Gene co-expression analysis for functional classification and gene–disease predictions. *Brief. Bioinform.* bbw139 (2017) doi:10.1093/bib/bbw139.
53. Kelley KW, Nakao-Inoue H, Molofsky AV, O. M. Variation among intact tissue samples reveals the core transcriptional features of human CNS cell classes. *Nat. Neurosci.* **21**, 1171–1184 (2018).
54. Horn, M. A. & Trafford, A. W. Aging and the cardiac collagen matrix: Novel mediators of fibrotic remodelling. *J. Mol. Cell. Cardiol.* **93**, 175–185 (2016).
55. Piñero, J. *et al.* The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.* (2019) doi:10.1093/nar/gkz1021.
56. Piñero, J. *et al.* DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.* **45**, D833–D839 (2017).
57. Craig, T. *et al.* The Digital Ageing Atlas: integrating the diversity of age-related changes into a unified resource. *Nucleic Acids Res.* **43**, D873–D878 (2015).
58. Lehallier, B. *et al.* Undulating changes in human plasma proteome profiles across the lifespan. *Nat. Med.* **25**, 1843–1850 (2019).
59. Khramtsova, E. A., Davis, L. K. & Stranger, B. E. The role of sex in the genomics of

human complex traits. *Nat. Rev. Genet.* **20**, 173–190 (2019).

60. Subramanian, A. *et al.* A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **171**, 1437-1452.e17 (2017).
61. de Almeida BP, Saraiva-Agostinho N, B.-M. N. cTRAP: Identification of candidate causal perturbations from differential gene expression data. (2020).
62. de Magalhães, J. P., Curado, J. & Church, G. M. Meta-analysis of age-related gene expression profiles identifies common signatures of aging. *Bioinformatics* **25**, 875–881 (2009).
63. Tacutu, R. *et al.* Human Ageing Genomic Resources: new and updated databases. *Nucleic Acids Res.* **46**, D1083–D1090 (2018).
64. Dönertaş, H. M., Fuentealba Valenzuela, M., Partridge, L. & Thornton, J. M. Gene expression-based drug repurposing to target aging. *Aging Cell* **17**, e12819 (2018).
65. Janssens, G. E. *et al.* Transcriptomics-Based Screening Identifies Pharmacological Inhibition of Hsp90 as a Means to Defer Aging. *Cell Rep.* **27**, 467-480.e6 (2019).
66. https://shiny.rstudio.com/gallery/.
67. Kunst, J. highcharter: A Wrapper for the 'Highcharts' Library.
68. Yihui, X., Joe, C. & Xianying, T. DT: A Wrapper of the JavaScript Library 'DataTables'.
69. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
70. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).
71. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
72. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
73. Korotkevich, G. *et al.* Fast gene set enrichment analysis. *bioRxiv* 060012 (2021) doi:10.1101/060012.
74. Chen, Y.-A. *et al.* Integrated Pathway Clusters with Coherent Biological Themes for Target Prioritisation. *PLoS One* **9**, e99030 (2014).
75. Dowle, M. & Srinivasan, A. data.table: Extension of 'data.frame'.
76. Lein, E. S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain.

*Nature* **445**, 168–176 (2007).

77. Butovsky, O. *et al.* Identification of a unique TGF-β–dependent molecular and functional signature in microglia. *Nat. Neurosci.* **17**, 131–143 (2014).
78. Cahoy, J. D. *et al.* A Transcriptome Database for Astrocytes, Neurons, and Oligodendrocytes: A New Resource for Understanding Brain Development and Function. *J. Neurosci.* **28**, 264–278 (2008).
79. Hickman, S. E. *et al.* The microglial sensome revealed by direct RNA sequencing. *Nat. Neurosci.* **16**, 1896–1905 (2013).
80. Ma, S. *et al.* Caloric Restriction Reprograms the Single-Cell Transcriptional Landscape of Rattus Norvegicus Aging. *Cell* **180**, 984-1001.e22 (2020).
81. Rubenstein, A. B. *et al.* Single-cell transcriptional profiles in human skeletal muscle. *Sci. Rep.* **10**, 229 (2020).
82. He, S. *et al.* Single-cell transcriptome profiling of an adult human cell atlas of 15 major organs. *Genome Biol.* **21**, 294 (2020).
83. Cui, Y. *et al.* Single-Cell Transcriptome Analysis Maps the Developmental Track of the Human Heart. *Cell Rep.* **26**, 1934-1950.e5 (2019).
84. Skelly, D. A. *et al.* Single-Cell Transcriptional Profiling Reveals Cellular Diversity and Intercommunication in the Mouse Heart. *Cell Rep.* **22**, 600–610 (2018).
85. Palmer, C., Diehn, M., Alizadeh, A. A. & Brown, P. O. Cell-type specific gene expression profiles of leukocytes in human peripheral blood. *BMC Genomics* **7**, 115 (2006).
86. Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).