voyAGEr: first steps

# The web application

voyAGEr is freely available at

https://compbio.imm.medicina.ulisboa.pt/voyAGEr

voyAGEr is composed of four main sections (the tabs in the navigation bar at the top):

* **Home** (depicted by the home icon  and no literal titling): to visually explain the used method and its associated findings featured in the application.
* **Gene**: to lead a gene-centric investigation, namely to assess how the expression of a specific gene changes with age and sex in a specific tissue.
* **Tissue**: to analyse how tissue-specific transcriptomes change with age and sex.
* **Module**: to further examine sets of co-expressed genes whose expression is altered with age namely through their enrichment in specific cell types, biological pathways and association with diseases.

voyAGEr leverages RNA-seq datasets from the GTEx project (Lonsdale et al., 2013), encompassing tissue samples from hundreds of donors aged from 20 to 70 years.

# Senescence-associated genes

Cellular senescence is a stress-induced cell cycle arrest limiting proliferation of potentially oncogenic cells but progressively creating an inflammatory environment in tissues as they age and therefore an example of a process whose molecular mechanisms are of particular interest to ageing researchers (Gorgoulis et al., 2019; Van Deursen, 2014).

Senescence markers, such as *CDKN2A*, encoding cell cycle regulatory protein p16INK4A that accumulates in senescent cells (Erickson et al., 1998; Gil & Peters, 2006), can thus be studied as putative markers of ageing of certain tissues.

*CDKN2A* expression profile

To examine *CDKN2A* expression changes across age: 1- Go to the ***Gene* section**

1. Type *CDKN2A* in the ***Gene* field**

The application then features:

* 1. in the ***Profile* sub-tab***,* a heatmap of tissue-specific *CDKN2A* scaled expression (Z-scores) across age, for all tissues (**Figure 1**).
  2. in the ***Alteration* sub-tab***,* a heatmap of significance of tissue-specific *CDKN2A* expression age-related alterations due to *Age*, *Sex* or *Age&Sex* (depending on the user’s choice – ***Alterations associated with* field** on the left), for all tissues (**Figure 2**).

*This section might take a bit longer to load.*

Note that gene names in vyAGEr are HGNC (HUGO Gene Nomenclature Committee) symbols. For each gene, the respective NCBI and GeneCards webpages can be accessed by clicking on their logos next to its name on plot’s title.

*A graph of red and blue squares

Description automatically generated with medium confidence*

**Figure 1** – Heatmap of tissue-specific *CDKN2A* expression over age.

1. Go to the ***Alteration* sub-tab** to check the significance of *CDKN2A* expression alterations across tissues and age (leave the default parameters, *All tissues* and *Age,* in the ***Tissue*** and ***Alterations associated with* fields**, respectively). A heatmap like that of **Figure 2** is featured.

A screenshot of a computer screen

Description automatically generated

**Figure 2** – Heatmap of significance of tissue-specific *Age*-associated *CDKN2A* expression alterations over age.

1. Enter/select *Lung* in the ***Tissue* field** to investigate *CDKN2A* expression changes in that specific tissue.

A graph with green and black lines

Description automatically generatedPlots of *CDKN2A* expression (top panel, identical to that in the *Profile* sub-tab) and the significance of its alterations over age (bottom panel) are then featured (**Figure 3**). Significant *CDKN2A* expression increases are observed in the late forties and early sixties.

**Figure 3** – *CDKN2A* expression in the lung (top panel) and significance of its alterations (bottom panel) over age.

1. Go to the ***Profile* sub-tab**. voyAGEr can also associate *CDKN2A* expression in the lung with the donors’ sex and medical history. These clinical data are displayed in a table below the expression profile’s scatter plot.

GTEx transcriptomic data are from “healthy” tissue samples from donors that had, nonetheless, reported medical conditions (Lonsdale et al., 2013).

* 1. Click on *Sex* in the ***Coloured* by** field,leaving *All* in the ***Shaped by* field.**

*CDKN2A* lung expression progression with age appears to be influenced by the donors’ sex, particularly in the mid-thirties (**Figure 4**). This observation can be statistically tested in the ***Alteration* sub-tab** by clicking on *Sex* in the ***Alterations associated with* field**.

* 1. Back in the *Profile* sub-tab, click on *All* in the ***Coloured* by** fieldand on *Condition* in the ***Shaped by*** field**.** Enter/select *MHCOPD* in the ***Select*** field.

The *CDKN2A* lung expression profile is herein associated with medical conditions (positive if the donor suffered from the condition, negative if not and unknown if the association is uncharted). Moreover, the median gene expression values for positive and negative conditions are displayed. The significance of Kruskal-Wallis tests for the difference in gene expression medians between positive and negative donors is used to rank conditions. In this case, the condition selected (Chronic Respiratory Disease) is amongst those displaying a significant difference in median (adjusted p-value below 0.05). On the scatter plot with *CDKN2A* lung expression over age, the curves fitted independently for positive and negative conditions show that such difference in gene expression occurs mostly after the age of 50 (**Figure 5**).

**Limitations**: In the GTEx dataset, there are conditions for which very few donors are positive and others for which very few donors have their condition state annotated. The significance of the Kruskal-Wallis tests must therefore be regarded with caution and as providing limited information. In this case, for example, even though significant differences in median were found for the History of Non Metastatic Cancer and Cocaine Use in 5 years, the low number of positive samples and their concentration in limited age ranges hamper any solid conclusion.

A graph of a graph

Description automatically generated with medium confidence

**A graph of red and green dots

Description automatically generatedFigure 4** – *CDNK2A* expression in the lung, discriminated between donors sex (female in pink, male in blue) over age.

**Figure 5** – *CDNK2A* expression in the lung, discriminated between donors with (green) and without (orange) abnormal white blood count, over age.

Transcriptional changes in the Transverse colon

1. Go to the ***Tissue* section.**

A blurry image of a person

Description automatically generatedThe landscape of *Age*-, *Sex*- and *Age&Sex*-associated global gene expression alterations along age for all tissues can be profiled using the significance of proportions of differentially expressed genes. Three periods stand out with significant transcriptional changes associated with *Age* (keeping the default *All tissues* in the ***Tissue* field** and *Age* in the ***Alterations associated with* field**), after 55 years old (**Figure 6**). Moreover, most of the significant transcriptional differences between sexes appear to occur in the fifth and sixth decades of life (*All tissues* in the ***Tissue* field** and *Sex* in the ***Alterations associated with* field**) (**Figure 7**).

**Figure 6** – Heatmap of significance of tissue-specific *Age*-associated global gene expression alterations over age.

A close up of a wood

Description automatically generated

**Figure 7** – Heatmap of significance of tissue-specific *Sex*-associated global gene expression alterations over age.

1. Enter *Adipose – Subcutaneous* in the ***Tissue* field** and click on *Age* in the ***Alterations associated with* field.**

The progression of the percentage of *Age*-associated altered genes over age is now featured (**Figure 8**). The statistical significance of each proportion is also illustrated with a colour scale. Two periods of significant transcriptional changes appear to occur, at late 20’s (13.6% altered genes) and late 40’s (4.7& altered genes).

1. Click on the dot at 29.57 years old (hovering over each point in the plot will show its details).

A graph on a screen

Description automatically generatedThe list of differentially expressed genes, ordered by their significance, appears on the left (**Figure 8**).

**Figure 8** – Progression of the percentage of Age-associated altered genes over age in Adipose - Subcutaneous. For each age, the list of the most altered genes can be obtained by clicking on the respective dot.

1. Click on the *LMO3* row in the table.

Plots of *LMO3* expression and the significance of its alterations over age (like in Figure 3) appear.

1. Browse the expression alterations’ significance over age of the most altered genes by selecting them in the table.

Some (e.g., *PRELID1*, *RUNX1T1*, *FGFRL1)* have their expression significantly modified only in the aforementioned first peak at around 28 years old.

1. Click on the dot at 46.43 y.o. and similarly browse the expression alterations’ significance of the most differentially expressed genes at this age.

Some (e.g. *MT-CYB*, *MT-ND4*, *MT-ATP6*, *MT-ND2)* have their expression significantly altered only in this second peak.

Different sets of genes may drive the different age periods of major transcriptional changes, which begs assessing if they reflect the activation of distinct biological processes. For this purpose, the user can profile the biological functions of the genes underlying each peak of transcriptomic changes by assessing their enrichment in manually curated pathways from the Reactome database (Croft et al., 2014) or in user-provided gene sets.

1. Go to the ***Enrichment* sub-tab**.

A heatmap showing the normalised enrichment score (NES) of Reactome pathways (columns) along age (row) is displayed (**Figure 9**). The percentage of altered genes over age can be found on the right side of the heatmap. Reactome pathways are gathered in families of biological functions, based on shared genes, that can be found at the top of the heatmap.

Note that, for visualisation ease, only the most significantly associated pathways are featured.

The user can click on *Select:* in the ***Pathway* field** to examine results for a given Reactome pathway.

A close-up of a screen

Description automatically generated

**Figure 9** – Heatmap of significance of tissue-specific *Age*-associated global gene expression alterations over age.

1. Below the heatmap, click on the red family (family 3) in the ***Families of pathways* section**.

A word cloud provides a glimpse into the family’s biological functions.

By clicking on the ***Pathways* sub-tab** in the ***Families of pathways* section,** the user has access to the list of specific pathways from the Reactome, Gene Ontology (Gene Ontology Consortium, 2004) and KEGG (Kanehisa, 2000) databases that are associated with the family.

1. Click on *User-specified* in the ***Geneset* field** on the left.

Let’s examine the enrichment of the three peaks of transcriptional changes in senescent-associated genes.

1. Enter the 230 senescent-associated genes (retrieved from Senequest (Gorgoulis et al., 2019) whose link with senescence is supported by at least 4 sources) from this document’s appendix in the ***List of genes* field**, leave a Significance threshold p-value of 0.05 and **Run**.

Although we have two peaks, if you hover over the tip you’ll notice neither of them are significantly enriched in senescence-associated genes (**Figure 10**).

Gene symbols can be in upper or lower case but must still follow the HGNC naming. If a gene symbol is not recognised as such, the gene is not included in the analysis.

A graph with a line and a line graph

Description automatically generated with medium confidence

**Figure 10** – Enrichment of altered genes amongst senescent-associated genes over age (non significant).

# Modules of co-expressed genes

Genes with highly correlated expression are likely to be coregulated and share biological functions or associations with phenotypical or pathological traits (van Dam et al., 2017). Clusters of these genes, called modules, are identified in 4 tissues using voyAGEr and their enrichment in cell types, Reactome pathways, and disease markers can be analysed.

1. Go to the ***Results* sub-section** of the ***Module* section.**

The ***About* sub-section** graphically summarises the methods employed to obtain the modules.

Each module is made of a set of genes and characterised by an eigengene representing their average expression profile.

Modules’ eigengene expression and enrichment in Reactome pathways, cell types, and disease markers can be respectively found in the 4 sub-tabs: Expression, Cell types, Pathways, Diseases.

1. Choose *Brain - Cortex* in the ***Tissue* field**.

6 modules were identified in this tissue. Each module is named based on the colour used to depict it.

1. Go to the ***Cell types* sub-tab**.

For each tissue, cell types and their markers were retrieved from the literature and then differ from a paper to another. Regarding the Brain - Cortex analysis, we can see that the annotation of cell types from Fan (Fan et al., 2018) is more comprehensive than Descartes (Cao et al., 2020).

Select the Descartes signature. At least four of the 5 modules appear to be particularly enriched for certain cell types markers (**Figure 11**): the yellow module for Microglia, turquoise for Endothelial cells, brown for oligodendrocytes, and blue for astrocytes.

1. Change the source of cell type markers by clicking on a different row in the table on the left.

For both sources, the cell types enriched in the 4 modules match.

A screenshot of a graph

Description automatically generated**Figure 11** – Enrichment of modules of co-expressed genes identified the brain cortex in cell types markers from Cao et al. (2020)

1. Choose *MEyellow* in the ***Module* field**.

The four layers of information captured in the four Module sub-tabs are now specifically displayed for the chosen module. Besides, the module’s 121 genes are identified on the left. The expression of the module’s eigengene appears to be roughly steady throughout life (**Figure 12**).

A graph with a line and dots

Description automatically generated with medium confidence

**Figure 12** – Yellow module (associated with microglia) eigengene expression in the brain cortex over age.

1. Click on *Sex* in the ***Colored by* field**.

The module’s eigengene expression does not seem to indicate that there are substantial differences in the yellow module over age, possibly suggesting that both male and females have a comparable proportion of microglia in the brain cortex (**Figure 13**).

A graph with blue and pink dots

Description automatically generated

**Figure 13** – Yellow module eigengene expression in the brain cortex, discriminated between sexes, over age.

1. Click on *All* in the ***Colored by*** fieldand on *Condition* in the ***Shaped by*** field**.** Choose *Cerebrovascular Disease* as condition. Eigengene expression appears to be slightly higher in positive when comparing to negative conditions, suggesting a role of microglia in these conditions (**Figure 14**), in accordance with what was already described in the literature (Dong et al. 2021).

A graph with red and green dots

Description automatically generated

**Figure 14** – Yellow module eigengene expression in the brain cortex, discriminated by the donors’ “*Cerebrovascular Disease*” history, over age.

1. Explore the ***Pathways*** and the ***Diseases-DisGeNET* sub-tabs***.*

As expected, the module is associated with immune-related Reactome pathways and diseases (e.g. rheumatoid arthritis).

# References

Croft, D., Mundo, A. F., Haw, R., Milacic, M., Weiser, J., Wu, G., Caudy, M., Garapati, P., Gillespie, M., Kamdar, M. R., Jassal, B., Jupe, S., Matthews, L., May, B., Palatnik, S., Rothfels, K., Shamovsky, V., Song, H., Williams, M., … D’Eustachio, P. (2014). The Reactome pathway knowledgebase. *Nucleic Acids Research*, *42*(D1), D472–D477. https://doi.org/10.1093/nar/gkt1102

Cao, J., O’Day, D. R., Pliner, H. A., Kingsley, P. D., Deng, M., Daza, R. M., ... & Shendure, J. (2020). A human cell atlas of fetal gene expression. Science, 370(6518), eaba7721.

Dong, R., Huang, R., Wang, J., Liu, H., & Xu, Z. (2021). Effects of microglial activation and polarization on brain injury after stroke. Frontiers in neurology, 12, 620948.

Erickson, S., Sangfelt, O., Heyman, M., Castro, J., Einhorn, S., & Grandér, D. (1998).

Involvement of the Ink4 proteins p16 and p15 in T-lymphocyte senescence. *Oncogene*, *17*(5), 595–602. https://doi.org/10.1038/sj.onc.1201965

Fan, X., Dong, J., Zhong, S., Wei, Y., Wu, Q., Yan, L., ... & Tang, F. (2018). Spatial transcriptomic survey of human embryonic cerebral cortex by single-cell RNA-seq analysis. Cell research, 28(7), 730-745.

Gene Ontology Consortium. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research*, *32*(90001), 258D – 261. https://doi.org/10.1093/nar/gkh036

Gil, J., & Peters, G. (2006). Regulation of the INK4b–ARF–INK4a tumour suppressor locus: all for one or one for all. *Nature Reviews Molecular Cell Biology*, *7*(9), 667–677. https://doi.org/10.1038/nrm1987

Gorgoulis, V., Adams, P. D., Alimonti, A., Bennett, D. C., Bischof, O., Bishop, C., Campisi, J., Collado, M., Evangelou, K., Ferbeyre, G., Gil, J., Hara, E., Krizhanovsky, V., Jurk, D., Maier, A. B., Narita, M., Niedernhofer, L., Passos, J. F., Robbins, P. D., … Demaria, M. (2019). Cellular Senescence: Defining a Path Forward. *Cell*, *179*(4), 813–827. https://doi.org/10.1016/j.cell.2019.10.005

He, S., Wang, L.-H., Liu, Y., Li, Y.-Q., Chen, H.-T., Xu, J.-H., Peng, W., Lin, G.-W., Wei, P.-P., Li,

B., Xia, X., Wang, D., Bei, J.-X., He, X., & Guo, Z. (2020). Single-cell transcriptome profiling of an adult human cell atlas of 15 major organs. *Genome Biology*, *21*(1), 294. https://doi.org/10.1186/s13059-020-02210-0

Kanehisa, M. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, *28*(1), 27–30. https://doi.org/10.1093/nar/28.1.27

Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., Foster, B., Moser, M., Karasik, E., Gillard, B., Ramsey, K., Sullivan, S., Bridge, J., Magazine, H., Syron, J., … Moore, H. F. (2013). The Genotype-Tissue Expression (GTEx) project. *Nature Genetics*, *45*(6), 580–585. https://doi.org/10.1038/ng.2653

Skelly, D. A., Squiers, G. T., McLellan, M. A., Bolisetty, M. T., Robson, P., Rosenthal, N. A., & Pinto, A. R. (2018). Single-Cell Transcriptional Profiling Reveals Cellular Diversity and Intercommunication in the Mouse Heart. *Cell Reports*, *22*(3), 600–610. https://doi.org/10.1016/j.celrep.2017.12.072

van Dam, S., Võsa, U., van der Graaf, A., Franke, L., & de Magalhães, J. P. (2017). Gene co- expression analysis for functional classification and gene–disease predictions. *Briefings*

*in Bioinformatics*, bbw139. https://doi.org/10.1093/bib/bbw139

Van Deursen, J. M. (2014). The role of senescent cells in ageing. *Nature*, *509*(7501), 439–446. https://doi.org/10.1038/nature13193

# Appendix

Senescent-associated genes retrieved from Senequest (Gorgoulis et al., 2019):

*AKR1C2*

*AKT1*

*AKT1*

*ALDH1A3*

*ALOX15B*

*ANLN*

*APLP1*

*ARG2*

*ARHGAP19*

*ATF6*

*ATM*

*AURKA*

*AURKB*

*BCL2*

*BCL2L1*

*BCL2L1*

*BHLHE40*

*BIRC5*

*BLM*

*BMI1*

*BMP2*

*BMP4*

*BMP7*

*BRAF*

*BRAF*

*BRCA1*

*BTG2*

*BUB1*

*BUB1B*

*CAMK2B*

*CAV1*

*CCDC167*

*CCL2*

*CCNA2*

*CCNB1*

*CCNB2*

*CCND1*

*CCNE1*

*CD44*

*CDC20*

*CDC25C*

*CDCA2*

*CDCA3*

*CDCA5*

*CDCA8*

*CDK1*

*CDK2*

*CDK4*

*CDKN1A*

*CDKN1B*

*CDKN2A*

*CDKN2AIP*

*CDKN2B*

*CDKN3*

*CEBPA*

*CEBPB*

*CEL*

*CENPA*

*CENPN*

*CENPO*

*CENPW*

*CEP55*

*CGAS*

*CHEK2*

*CKAP2L*

*CKS1B*

*CSNK2A1*

*CTNNB1*

*CXCL8*

*CXCL8*

*CXCR2*

*CYB561A3*

*DDIAS*

*DEPDC1*

*DEPDC1B*

*DICER1*

*DKK1*

*DLGAP5*

*DNMT1*

*DPP4*

*E2F1*

*E2F1*

*EBNA1BP2*

*EDN1*

*EGFR*

*EGR1*

*ELAVL1*

*EME1*

*EP300*

*ERBB2*

*ERCC6L*

*ESPL1*

*ESR1*

*ETS2*

*EZH2*

*FAM83D*

*FANCD2*

*FGF2*

*FGF2*

*FOXM1*

*FOXO1*

*FOXO3*

*FOXO3*

*GABPA*

*GADD45A*

*GADD45B*

*GADD45G*

*GAS2L3*

*GDF15*

*GTSE1*

*HBP1*

*HDAC1*

*HIF1A*

*HJURP*

*HMGB2*

*HMMR*

*HMOX1*

*HRAS*

*HSPA1A*

*ID1*

*IFNG*

*IGF1*

*IGF1*

*IGF1R*

*IGFBP2*

*IGFBP3*

*IGFBP5*

*IGFBP7*

*IL6*

*ING1*

*ITGB4*

*JUN*

*KAT2B*

*KAT6A*

*KDM6B*

*KIF11*

*KIF20A*

*KIF23*

*KIF2C*

*KIF4A*

*KIFC1*

*KL*

*KNSTRN*

*KRAS*

*LMNA*

*LMNB1*

*MAD2L1*

*MAP3K6*

*MAPK1*

*MAPK14*

*MAPK3*

*MAPK8*

*MDM2*

*MIR22*

*MIR23A*

*MKI67*

*MTOR*

*MTOR*

*MXD4*

*MYBL2*

*MYC*

*MYC*

*NAMPT*

*NDC80*

*NEIL3*

*NEK2*

*NEK6*

*NFE2L2*

*NFKB1*

*NOS3*

*NOS3*

*NOTCH1*

*NOTCH3*

*NOX1*

*NOX4*

*NRAS*

*NUDT1*

*OGG1*

*OIP5*

*PBK*

*PIF1*

*PIK3CA*

*PIM1*

*PIMREG*

*PIN1*

*PLA2R1*

*PLK1*

*PLK4*

*PMAIP1*

*PML*

*POC1A*

*PPARGC1A*

*PPM1D*

*PRKAA1*

*PRKAA1*

*PRKCD*

*PRODH*

*PRR11*

*PSRC1*

*PTEN*

*PTEN*

*PTGS2*

*PTTG1*

*PTTG3P*

*RAC1*

*RACGAP1*

*RAD51*

*RAS*

*RB1*

*RBL2*

*RELA*

*RPS6KA6*

*RPS6KB1*

*RRM2*

*RSL1D1*

*SAT1*

*SDC1*

*SERPINA4*

*SERPINE1*

*SGO1*

*SHC1*

*SIRT1*

*SIRT2*

*SIRT3*

*SIRT6*

*SIRT7*

*SKA3*

*SKP2*

*SMAD3*

*SMARCB1*

*SMURF2*

*SOD2*

*SOD2*

*SOX9*

*SPC24*

*STAT1*

*STAT3*

*STAT5A*

*STK11*

*SUV39H1*

*TACC3*

*TBX2*

*TERF2*

*TERT*

*TGFB1*

*THBS1*

*TICRR*

*TNF*

*TNFSF13B*

*TOP2A*

*TP53*

*TP53*

*TP63*

*TP73*

*TPX2*

*TRIP13*

*TROAP*

*TTK*

*TWIST1*

*TXNIP*

*UBE2C*

*UHRF1*

*WRN*

*XRCC5*

*YAP1*

*YPEL*