Exploring molecular signatures of senescence with markeR, an R toolkit for evaluating gene sets as phenotypic markers

Rita Martins-Silva[1,2*], Alexandre Kaizeler[1,2], Nuno L. Barbosa-Morais[1,2*]
[1] Gulbenkian Institute for Molecular Medicine, 1649-028 Lisbon, Portugal
[2] Faculdade de Medicina, Universidade de Lisboa, 1649-028 Lisbon, Portugal

* Corresponding authors: Rita Martins-Silva rita.silva@gimm.pt, Nuno L. Barbosa-Morais nuno.morais@gimm.pt

ABSTRACT

Many biological processes, including cellular senescence, manifest as diverse phenotypes that vary across cell types and conditions. In the absence of single, definitive markers, researchers often rely on the expression of sets of genes to identify these complex states. However, there are multiple ways to summarise gene set expression into quantitative metrics (*i.e.,* signatures), each with its own strengths and limitations, and we know of no consensual framework to systematically evaluate their performance across datasets. We therefore developed `markeR` (https://bioconductor.org/packages/markeR), an open-source, modular `R` package that evaluates gene sets as phenotypic markers using various scoring and enrichment-based approaches. `markeR` generates interpretable metrics and intuitive visualisations that enable benchmarking of gene signatures and exploration of their associations with chosen study variables. As a case study, we applied `markeR` to 9 published senescence-related gene sets across 25 RNA-seq datasets, covering 6 human cell types and 12 senescence-inducing conditions. There was wide variability in gene set performance, as some signatures (*e.g.*, SenMayo) were robust senescence markers across contexts, while others (*e.g.*, those from MSigDB), performed poorly as such. We also used `markeR` to analyse gene expression in 49 GTEx tissues, revealing tissue- and age-related differences in senescence-associated signals. Together, these findings emphasise the difficulty of characterising molecular phenotypes and demonstrate the potential of `markeR` in facilitating the systematic evaluation of gene sets in various biological contexts.

KEYWORDS

Senescence, R package, Gene signatures, Molecular markers, Gene set enrichment analysis

1. INTRODUCTION

Originally thought to be a by-product of ageing, cellular senescence is now recognised as a tightly regulated biological programme that plays critical roles from embryonic formation to tissue repair and tumour suppression throughout the life of almost all organisms (Campisi, 2001; Demaria et al., 2014; Muñoz-Espín et al., 2013). Described as a stable and irreversible cell cycle arrest under physiological conditions, it is an intrinsic stress response of the cell induced by various factors, including telomere shortening, oncogene activation and DNA damage (Gorgoulis et al., 2019; Rodier & Campisi, 2011; Serrano et al., 1997). This cell cycle arrest is most often accompanied by profound changes in metabolic activity and secretory phenotype of the cell, the senescence-associated secretory phenotype (SASP), composed of cytokines, growth factors and proteases, all contributing to multiple roles, from immune recruitment to induction of senescence in neighbouring cells (Coppé et al., 2008).

At the molecular level, senescence is highly context-dependent and is regulated by different players such as p53, p16$^{(INK4A)}$ and p21$^{(CIP1/WAF1)}$, which in turn drive distinct gene expression programmes (Hernandez-Segura et al., 2018; Sharpless & Sherr, 2015). This diversity, coupled with the ability of different types of stress, such as DNA damage or oncogene activation, to induce senescence in a cell-type-specific manner, results in a heterogeneous and complex blend of phenotypes rather than a single, well-defined state (Gasek et al., 2021). Such complexity makes it challenging to identify a universal marker, requiring the use of gene sets, rather than individual genes, as molecular signatures of senescence (Sharpless & Sherr, 2015; Tuttle et al., 2020). This challenge is not unique to senescence, as many biological processes also present as diverse cellular states rather than discrete phenotypes.

Several gene sets have been proposed for senescence, typically derived from transcriptome analyses of senescent cells, or simple curation of genes already reported in the literature as associated with senescence (Avelar et al., 2020; Gorgoulis et al., 2019; Hernandez-Segura et al., 2017; Saul et al., 2022; Zhao et al., 2016). These gene sets often perform well as senescence markers in the specific context in which they were developed, but lack broader application otherwise. Furthermore, there is no consensual and robust strategy for combining such genes in a meaningful and reliable metric (*i.e.*, a gene signature) for the biological state. Common approaches are based on expression averages (*e.g.*, simple log-median (De Almeida et al., 2019), DAS and MSS (Lafferty-Whyte et al., 2010)), rank reflecting relative rather than absolute expression (*e.g.*, GSEA (Subramanian et al., 2005), ssGSEA (Barbie et al., 2009), CS score (X. Wang et al., 2022), SRRS (Zhang et al., 2023)), or weight-based scoring (*e.g.*, CSS (He et al., 2024)).

Some studies have proposed universal transcriptomic senescence scores (J. Wang et al., 2025), but they often lack controls, such as samples of quiescent cells, to distinguish senescence from other non-proliferative states, and extrapolate their findings to complex biological contexts, such as human tissues, where direct validation of the senescent phenotype remains challenging. While benchmarking studies have compared signature-scoring methods in specific contexts (Noureen et al., 2022), each has its own strengths and limitations, and, to our knowledge, there is no flexible

platform allowing researchers to systematically test, compare and apply multiple approaches with their gene sets of interest.

Herein, we present `markeR`, an `R` package that provides methods to quantify how well a gene set marks a phenotype by transforming its genes' expression profiles into various metrics and assessing their robustness. `markeR` can be used for (1) comparing gene sets for consistency and specificity in marking a phenotype (benchmarking), and (2) quantifying the gene set's expression and exploring its relationship with phenotypic variables (discovery). To demonstrate its applicability, we curated a compendium of publicly available gene expression datasets and senescence-related gene sets. While senescence serves as a case study due to its complexity and relevance, the framework is broadly applicable to other contexts. `markeR` is freely available in Bioconductor at https://bioconductor.org/packages/markeR.

2. MATERIALS AND METHODS

`markeR` (v1.0.0) was developed in `R` (v4.4.3) with a modular design to support easy integration of new features and functionalities (Figure 1). It was designed for users looking for both a quantitative framework with different methods to characterise a given phenotype, and a fully customisable visualisation component to suit their needs. It requires as input:

1. A list of gene sets associated with the phenotype of interest. For each gene set, information on the expected direction of regulation change (*i.e.*, up- or down-regulation) of each gene upon phenotype may or may not be included;
2. A pre-processed (*i.e.*, filtered and normalised, but not log-transformed) gene expression matrix;
3. A sample metadata table.

2.1 Modules for gene set-based phenotype characterisation

2.1.1 Score-based approaches

Log_2-median-centred score

The ***log_2-median*** method first applies a log_2 transformation to the expression values, followed by median-centring each gene's expression across all samples. Each sample's score is the mean of the median-centred expression values therein of the genes in the set (example of its use in (De Almeida et al., 2019)). Each gene's contribution is evaluated relative to its typical expression, and the resulting score reflects the coordinated expression of genes within the gene set. For bidirectional gene sets, *i.e.*, that include information on the expected direction of change of each of its genes upon phenotype, the sample score is the mean of median-centred expression for the upregulated subset of genes minus that of the downregulated subset.

Single sample Gene Set Enrichment Analysis (ssGSEA) score

Single sample Gene Set Enrichment Analysis (**ssGSEA**) was implemented using a modified version of the `GSVA` package's `gsva` function (Barbie et al., 2009), based on the original Gene Set Enrichment Analysis (GSEA) method (Subramanian et al., 2005). ssGSEA ranks all genes by their expression within each sample and computes a running-sum statistic over the ranked list. For unidirectional gene sets (*i.e.*, with no information on the expected direction of each gene's regulation upon phenotype), the ssGSEA sample score reflects overall coordinated expression changes of the genes in the set. For bidirectional gene sets, the score is calculated as the difference in ssGSEA scores between the upregulated and the downregulated subsets of genes.

Ranking score

The **ranking** method uses a non-parametric approach based on the relative expression of genes within a set. For each sample, all genes are ranked by expression. The score is the sum of the ranks of the genes in the gene set, divided by the number of genes in the set. The individual ranks of putatively downregulated genes in bidirectional gene sets are multiplied by -1.

2.1.2 Enrichment-based approach

`markeR` uses GSEA in its enrichment-based module (Subramanian et al., 2005). Linear-model-based differential gene expression analysis for the phenotypes of interest is performed with `limma` (Ritchie et al., 2015) (v3.62.2). If the phenotypic variable of interest is categorical, (*e.g.*, treatment, genotype) the design matrix is constructed with `model.matrix(~0 + variable)` (model without intercept). Pairwise comparisons between levels of this variable are specified using a contrast matrix. For continuous variables, the model (with intercept) `~1 + variable` is used instead. Alternatively, a custom design matrix modelling the user's biological questions can be provided.

Genes are then ranked according to differential expression statistics:

- for gene sets without specified direction of expression changes upon phenotype (unidirectional), the B-statistic, *i.e.*, the empirical Bayes posterior log-odds of differential expression (Smyth, 2004), irrespective of the direction of change;
- for bidirectional gene sets, the empirical Bayes moderated t-statistic (Smyth, 2004), *i.e.*, representing the ratio of estimated effect size to its standard error, multiplied by -1 for putatively downregulated genes.

GSEA is then performed using the `fgsea` package (v1.32.4), and interpreting the resulting normalised enrichment score (NES) for each gene set in each contrast (*i.e.*, phenotype)(Subramanian et al., 2005) depends on the differential expression statistic used for gene ranking:

- B-statistic: A positive NES indicates the gene set's enrichment among the most altered genes; a negative NES, mostly uninformative, indicates the gene set's depletion among the most altered genes (*i.e.*, enrichment among the least altered genes);
- t-statistic: a positive or a negative NES indicate the gene set's over- or under-activation, respectively.

Significance is determined using Benjamini-Hochberg's False Discovery Rate (FDR) correction for multiple testing of all analysed gene sets. An optional additional FDR correction can be applied to account for the number of contrasts tested per gene set, providing a more stringent control of significance (addressing the question *"Are there significantly altered gene sets in any of the contrasts?"* rather than *"Are there significantly altered gene sets in my contrast of interest?")*.

2.1.3 Individual Genes

`markeR` provides tools to assess the contribution of individual genes to their sets' score- and enrichment-based results. Visual and analytical outputs include expression heatmaps and violin plots of gene expression; correlation heatmaps to explore relationships between genes in the set; receiver operating characteristic (ROC) curves and respective areas under the curve (AUC) to evaluate gene-level classification performance; effect size (Cohen's d) of expression differences between sample groups; and principal component analysis (PCA) of gene expression profiles to assess the gene set's contribution to phenotype-associated variance.

2.2 Benchmarking and discovery modes

`markeR` provides two modes for evaluating gene sets as phenotypic markers: **benchmarking** and **discovery**.

Benchmarking assesses the robustness of gene sets' performance across multiple analytical methods. It includes violin plots for distributions of gene set scores across categorical phenotypes, while scatter plots are used when the phenotypic variable is numerical. Volcano plots display the statistical significance ($-\log_{10}$ p-values from Student's t-tests) against the effect size of differences in the distributions of sample scores (Cohen's d) between each pair of phenotyping groups of samples. Complementary heatmaps colour-code the effect size and annotate them with corresponding p-values. Additional features include ROC and AUC calculations (via the `pROC` package, v1.18.5), as well as estimation of effect size using Cohen's f for comparisons involving more than two groups or continuous variables (via the `effectsize` package, v1.0.0). Some analyses (*e.g.* ROC curves, Cohen's f heatmaps) apply only when the variable of interest is categorical. To evaluate whether a gene set's observed effect size exceeds what would be expected by chance for sets with the same characteristics, `markeR` quantifies gene set-level significance as the false positive rate (FPR), also commonly designated as empirical p-value, by comparing the observed effect sizes against null distributions of effect sizes in 100 gene sets, each generated by random selection of the same number of genes and, for bidirectional sets, expected proportion of genes up- or down-regulated upon phenotype. Results are visualised as violin plots of randomly generated Cohen's d or f values across methods, overlaid with dots for the empirical values and a 95th percentile significance threshold. For enrichment-based analyses, results from linear models are summarised using scatter or lollipop plots of NES values and their statistical significance. `plotEnrichment` (`fgsea` package) is used to plot the running enrichment score across the ranked gene list, highlighting the ranks of gene set members. Volcano plots, displaying significance against effect size, further contextualise gene set members within the broader expression landscape and help assess their robustness as markers.

Discovery assists in hypothesis generation by profiling associations of scores or enrichment values with selected phenotypic or clinical variables. Score distributions are summarised alongside effect sizes: Cohen's d for pairwise comparisons or Cohen's f otherwise. Cohen's f was calculated for each variable using a linear model of the form `score ~ variable`; for continuous variables with more than two levels, significance was assessed using the t-test for the regression slope, whereas for categorical variables, significance was obtained from an ANOVA F-test comparing group means. NES values are summarised in lollipop plots for selected contrasts, with tunable resolution. Dots are colour-coded by FDR.

Benchmarking includes the most comprehensive feature set, allowing variables of interest identified in Discovery to be seamlessly further tested. All results are visualised on plots generated by `ggplot2` (v3.5.2) and `ComplexHeatmap` (v2.22.0). Depending on the function, plots may be combined using `ggpubr` (v0.6.0), `grid` (4.4.3) or `cowplot` (v1.1.3).

2.3 Similarities with other gene sets

To get insights on the biological relevance of the gene sets, similarity to other sets can be assessed using two metrics: the Jaccard index and the odds ratio (OR). The Jaccard index (JI) is calculated as the size of the intersection divided by the union of two gene sets. The OR is derived from a 2x2 contingency table based on a user-defined gene universe, with statistical significance given by the p-value of the one-sided Fisher's exact test. Comparisons are made between user-defined and curated reference gene sets, which may include other user-defined and MsigDB collections (Subramanian et al., 2005; accessed via the `msigdbr` package, v24.1.0), and results can be filtered by minimum Jaccard index or a combination of odds ratio and p-value. Results are displayed as a heatmap (`ggplot2` v3.5.2), with colours representing either OR on a $\log_{10}$ scale or the JI. In this study, we focused on the Hallmarks collection and used a gene universe comprising all genes annotated in The Human Protein Atlas. We displayed gene sets with FDR ≤ 0.05 and odds ratio >1.

2.4 Senescence RNA-seq datasets

2.4.1 Dataset Collection

For evaluating gene sets as senescence markers, we curated human RNA-seq datasets from the Gene Expression Omnibus (GEO) (Edgar, 2002) and ArrayExpress (Parkinson et al., 2007), assisted by literature review based on PubMed searches. Curation followed a two-step strategy**.**

We first focused on fibroblast-derived datasets, given their widespread use as a tractable model for studying cellular senescence. Inclusion criteria required senescent or quiescent samples, with a minimum of 10 samples per study (specific queries in **Supplementary File 1 – GEO, ArrayExpress**). All such datasets available as of 14 February 2024 (GEO) and 29 February 2024 (ArrayExpress) were considered. Validation of senescence (*e.g.*, SA-β-gal staining, cell cycle arrest markers) was manually assessed from primary publications, and studies with insufficient or unclear validation were excluded. 16 datasets, comprising 384 samples, met all criteria (**Supplementary File 1 – Final Datasets, Metadata**).

To ensure generalisability across cell types, we then considered all non-fibroblast datasets available as of 12 July 2024, excluding cancer-specific studies (specific queries in **Supplementary File 2 – GEO, ArrayExpress**). 9 datasets, comprising 161 samples, met all criteria (**Supplementary File 2 – Final Datasets, Metadata**). These comprised five cell types: endothelial (4 datasets, 90 samples), keratinocyte (2 datasets, 36 samples), melanocyte (1 dataset, 12 samples), mesenchymal (1 dataset, 12 samples) and neuronal (1 dataset, 11 samples).

All samples (fibroblast and non-fibroblast) were categorised into three cellular states according to the respective primary study: proliferative (272 samples), quiescent (33) and senescent (240).

2.4.2 Data download and alignment

Data download and quality control were performed in a conda environment (`Python` v3.11.5). FASTQ files were obtained from GEO using `fasterq-dump` (v2.11.0; `SRA toolkit` v2.11.0; NCBI, https://github.com/ncbi/sra-tools) and manually from

ArrayExpress. Read quality was assessed using `fastqc` (v0.12.1; Andrews, 2010) and summarised using `MultiQC` (Ewels et al., 2016) (v1.14). All samples passed quality control and were retained for analysis. Reads were pseudo-aligned to the RefSeq reference human transcriptome (NCBI release 109)(O'Leary et al., 2016) using `kallisto` (Bray et al., 2016) (v0.44.0).

2.4.3 Preprocessing

Genes with low expression across conditions were removed by retaining only those with an average of > 70 reads/sample in at least one condition (quiescent, proliferative, or senescent)(**Supplementary Figure 1A**). Normalisation factors were calculated using `calcNormFactors` from the `edgeR` package (Chen et al., 2025), and read counts were subsequently normalised and $\log_2$-transformed using `voom` from the `limma` package (**Supplementary Figure 1B**).

2.4.4 Batch Effect Correction

Batch correction was performed using a modified version of the `removeBatchEffect` function from the `limma` package, following a strategy similar to Schneider et al., 2024. Dataset ID (*i.e.*, source study) was treated as a technical batch variable and regressed out of the expression matrix. In contrast, sources of variance reflecting biological differences (*e.g.*, cell type, cellular state) were modelled and kept in the analysed data (**Supplementary Figure 1C**).

2.5 GTEx Human Transcriptomic Data

Pre-processed RNA-seq data from The Genotype-Tissue Expression (GTEx) v8 were obtained from the GitHub repository associated with the Schneider et al., 2024 publication (https://github.com/DiseaseTranscriptomicsLab/voyAGEr).

2.6 Senescence Gene Sets

Senescence gene sets were selected based on resources publicly available as of 13 March 2024 (**Supplementary File 3 – Senescence Gene Sets**). We included five well-cited gene sets empirically derived from different cell types and senescence-inducing stimuli: SAUL_SEN_MAYO (Saul et al., 2022), CSGene (Zhao et al., 2016), CellAge (Chatsirisupachai et al., 2019), SeneQuest (Gorgoulis et al., 2019) and HernandezSegura (Hernandez-Segura et al., 2017). Additionally, we incorporated four gene sets from MSigDB (Subramanian et al., 2005): GOBP_CELLULAR_SENESCENCE, GOBP_NEGATIVE_REGULATION_OF_CELLULAR_SENESCENCE, GOBP_POSITIVE_REGULATION_OF_CELLULAR_SENESCENCE, and REACTOME_CELLULAR_SENESCENCE.

3. RESULTS

`markeR` is a flexible `R` package designed to quantify, compare and evaluate gene expression signatures as markers of cellular phenotypes (**Figure 1**). The package supports two main quantification approaches:

- **Score-based:** A signature score, summarising the collective expression of a gene set therein, is assigned to each sample.
- **Enrichment-based:** Gene Set Enrichment Analysis (GSEA) is run on genes ranked according to differential expression statistics, and a resulting enrichment score is assigned to each gene set for each phenotype.

3.1 Case Study: Cellular Senescence as a Complex Biological Phenotype

We use senescence as a case study for `markeR`, as it exemplifies a complex phenotype lacking a universal marker. Although several gene signatures have been proposed to identify senescent cells across diverse contexts, their systematic benchmarking is still missing. To address this, we first applied `markeR` in benchmarking mode to a curated RNA-seq dataset of senescent, proliferative and quiescent samples of human cells, thereby evaluating the robustness as markers of widely used senescence-associated gene sets across cell types and senescence-inducing stimuli. We then demonstrate `markeR`'s discovery mode using RNA-seq data of human tissues from the Genotype-Tissue Expression (GTEx) project, where senescence is not explicitly labelled but is hypothesised to increase with age (Campisi, 2005; Subramanian et al., 2005).

3.2 Gene Sets and RNA-seq Datasets

For the benchmarking analysis, we selected five senescence-associated gene sets from open-source, well-cited studies and four from MSigDB (see Methods). To focus on gene sets capturing core features of senescence, we excluded gene sets exclusively associated with the SASP, given its known heterogeneity (Hernandez-Segura et al., 2017). The selected gene sets vary in size, and some include annotations of the expected direction of expression change (*i.e.*, if each gene in the set is putatively up- or downregulated in senescence) (**Figure 2A**). Most gene sets were compiled through similar literature curation approaches, often incorporating the same well-established senescence-associated genes, likely explaining their substantial overlap. In contrast, the HernandezSegura and CellAge gene sets were derived independently from RNA-seq and microarray data analyses, respectively, and show minimal overlap with the others (**Figure 2A**). Most sets are enriched for nucleoplasm genes (**Supplementary Figure 2A**), consistent with roles in cell cycle regulation and transcription. Functional analyses (**Supplementary Figure 2B**) reveal significant enrichment in genes involved in proliferation, apoptosis, and immune pathways, aligning with known features of senescence (Coppé et al., 2008).

To evaluate the senescence-marking potential of those gene sets, we curated 23 publicly available human RNA-seq datasets (545 samples) from GEO, ArrayExpress, and manual literature review in PubMed (see Methods), prioritising diversity in cell types and senescence-inducing stressors (**Figure 2B-D**), amongst the main sources of phenotypic heterogeneity in senescence. The overall dataset for analysis includes six cell types (fibroblasts, endothelial, keratinocyte, melanocyte, mesenchymal, and neuronal), twelve distinct stressors, and three cellular states (240 senescent, 272 proliferative, and

33 quiescent samples)(see **Supplementary Files 1** and **2**). Quiescent samples were included as an essential control to distinguish senescence-specific signatures from those associated with general proliferative arrest. After pre-processing all RNA-seq data together (see Methods), cell type was, as expected, the primary driver of gene expression variation (**Figure 2B, Supplementary Figure 1D**), while cellular state (proliferative, senescent, quiescent) emerged as the main source of variance within each cell type (**Figure 2C)**.

3.3 Evaluating Gene Sets as Senescence Markers

3.3.1 Score-Based Methods Enable Sample-Level Quantitative Marking of Senescence

To quantify the performance of each gene set's expression in marking senescence phenotypes, we implemented in `markeR` several score-based methods ($\log_2$-median, ranking and ssGSEA), each with different assumptions (see Methods) but expected to produce consistent results for robust markers, as observed (**Figure 3A, Supplementary Figure 3**).

Some gene sets displayed strong cell type-specific patterns: the scores of CSGene and GOBP_CELLULAR_SENESCENCE increased from proliferative to senescent samples in some cell types but decreased in others, whereas HernandezSegura, SAUL_SEN_MAYO, and CellAge exhibited more consistent trends across cell types. Score distributions varied by cell type, suggesting cell-type-specific baseline expression levels of senescence-associated genes, particularly evident for GOBP_POSITIVE_ and GOBP_NEGATIVE_REGULATION_OF_CELLULAR_SENESCENCE, as observed when samples are stratified by senescence-inducing stimulus (**Supplementary Figure 4**). Stimulus-specific sensitivity in senescence marking was also observed: HernandezSegura and SAUL_SEN_MAYO effectively distinguished senescent from non-senescent samples across stimuli, with particularly strong discrimination in oncogene-induced senescence, while CellAge was more specific to radiation and telomere shortening (**Supplementary Figure 4**). Conversely, REACTOME_CELLULAR_SENESCENCE and GOBP_POSITIVE_REGULATION_OF_CELLULAR_SENESCENCE, showed poor discrimination across conditions. Others, such as SeneQuest, consistently discriminated senescent from non-senescent samples regardless of the inducing stimulus.

HernandezSegura, CellAge and SAUL_SEN_MAYO consistently separated senescence from quiescence, while, CSGene, GOBP_CELLULAR_SENESCENCE and GOBP_NEGATIVE_REGULATION_OF_CELLULAR_SENESCENCE, failed to do so (**Figure 3A**), importantly suggesting they reflect general non-proliferative states rather than senescence-specific programs.

Overall, HernandezSegura, CellAge, SeneQuest and SAUL_SEN_MAYO emerged as the most robust gene sets for distinguishing senescence. HernandezSegura and SAUL_SEN_MAYO showed the strongest effect sizes and classification performance (**Supplementary Figure 5**), whereas SeneQuest and CellAge were less specific, especially in differentiating senescence from quiescence. In neurons, SeneQuest and CellAge displayed inverse trends (*i.e.*, senescence scores decrease from proliferative to senescent samples, compared to other cell types), possibly reflecting the unique, post-

mitotic nature of neurons and their non-canonical, senescence-like phenotype (cell-cycle arrested but metabolically and secretorily active)(Gasek et al., 2021; Melo Dos Santos et al., 2024). This suggests that SeneQuest and CellAge may be composed of proliferation-linked genes, whereas HernandezSegura and SAUL_SEN_MAYO capture broader senescence features. Notably, while we excluded gene sets exclusively representing the SASP due to its known context-dependence and variability, SAUL_SEN_MAYO includes a substantial number of genes predicted to be secreted (**Supplementary Figure 2A**). This suggests that it may capture a core SASP component relevant across cell types, potentially explaining its strong performance even in post-mitotic contexts. In contrast, HernandezSegura includes fewer secreted genes, suggesting that SASP alone does not fully account for the robustness of these gene sets in marking senescence.

Scores incorporate gene directionality, *i.e.*, whether genes are expected to be upregulated or downregulated in senescence (**Figure 2A**), which strongly influences their interpretation. For example, ignoring directionality in CellAge yielded higher senescence scores in proliferative than senescent samples and reduced classification accuracy (**Figure 3B**).

3.3.2 Enrichment-Based Methods Enable Group-Level Interpretation

To evaluate the nine gene sets using an enrichment-based framework, we applied GSEA to gene rankings derived from linear models contrasting proliferative, quiescent, and senescent cellular states (see Methods).

Highlighting gene set members on volcano plots of differential expression (statistical significance vs. fold-changes, **Figure 4A**), summarising those models, enables visual assessment if genes within a phenotype-linked set consistently shift their expression according to expectation. CellAge, HernandezSegura, and SAUL_SEN_MAYO showed clear, consistent transcriptional differences between senescent and proliferative samples, while HernandezSegura and SAUL_SEN_MAYO also performed well in the senescent vs. quiescent contrast. SeneQuest, in line with score-based results, reflected non-proliferative features, such as a strong depletion of genes expected to be downregulated in senescent samples when compared with proliferative cells, but no clear signal in the senescence vs. quiescence contrast. Gene sets with annotated directionality and more variable differential expression statistics, such as CellAge and SeneQuest in the senescent vs. quiescent contrast, suggest a more context-dependent performance in marking the phenotype. Gene sets lacking directionality, such as GOBP_NEGATIVE_ and GOBP_POSITIVE_REGULATION_OF_CELLULAR_SENESCENCE, show weak enrichment and low fold changes, also suggesting context-dependent relevance rather than a global response of senescence.

Gene Set Enrichment Analysis confirmed significant enrichment for CellAge, SeneQuest, CSgene, SAUL_SEN_MAYO and HernandezSegura across comparisons between the three cellular states (**Figure 4B, C**). Others, such as the GOBP and REACTOME gene sets, did not reach significance in the enrichment-based analysis for the two comparisons involving senescence.

3.4 Exploring Best-Performing Senescence Gene Sets

Across score- and enrichment-based analyses, gene sets HernandezSegura, CellAge, SAUL_SEN_MAYO, and SeneQuest consistently emerged as robust indicators of senescence. These sets consistently discriminated senescent samples from both proliferative and quiescent cellular states across the considered metrics (**Figure 5A-B**). In contrast, MSigDB gene sets, despite their popularity in GSEA-based studies, performed poorly as senescence markers across most score- and enrichment-based metrics.

SeneQuest's discriminatory power improved when its up- and down-regulated components were analysed separately. The putatively downregulated genes showed similar scores in senescent and quiescent samples (**Figure 5C**), indicating downregulation in senescence relative to proliferation but little distinction from quiescence, likely reflecting non-specific cell cycle arrest (**Figure 5E**). In contrast, the putatively upregulated genes effectively discriminated senescence from quiescence (**Figure 5C**), suggesting that they dominate the set's performance in marking senescence. By comparison, both up- and down-regulated components of HernandezSegura effectively distinguish senescence from both proliferation and quiescence (**Figure 5D-E**). Conversely, as an additional example, REACTOME_CELLULAR_SENESCENCE shows poor value in discriminating between phenotypes, with its genes showing little differential expression between senescence and quiescence or proliferative samples (**Figure 5E, 3A**).

CellAge showed solid performance in marking senescence, particularly in enrichment based analyses, but in a cell-type- and stimulus-dependent manner. The scores observed in neurons may reflect a greater reliance on proliferation-associated programmes such as E2F targets, MYC signalling and mitotic checkpoint control (**Supplementary Figure 2B**), potentially limiting its applicability to non-dividing differentiated cells.

HernandezSegura emerged as the most consistent senescence-marking gene set, despite having few genes annotated as senescence-related in functional databases (**Figure 2A** and **Supplementary Figure 2B**). This suggests it captures genes with consistent expression changes upon senescence, although not necessarily the most extreme (**Figure 4A**), and therefore more likely to reflect a core transversal senescence signature, even if no functional link with the phenotype has been established yet.

SAUL_SEN_MAYO's robustness may stem from its enrichment in genes encoding secreted proteins (**Supplementary Figure 2A**), suggesting it captures core features of the senescence programme beyond cell cycle arrest, possibly driven by inflammation-related pathways such as JAK/STAT3 and NF-κB signalling (**Supplementary Figure 2B**).

When benchmarked against random gene sets for distinguishing senescence from quiescence, HernandezSegura and SAUL_SEN_MAYO showed low mean false positive rates (0.02 and 0.09). In contrast, SeneQuest and CellAge showed higher mean FPRs (0.28 and 0.29), suggesting that they are the top markers of cell cycle arrest but not specifically of senescence (**Figure 5F, Supplementary Figure 6**).

3.5 Trade-Offs

To assess the robustness of score- and enrichment-based approaches, we evaluated their performance as markers across varying sample sizes (**Figure 6A**) and gene-set sizes (**Figure 6B**). As expected, the size of gene sets is relevant for their performance, as each gene potentially brings to the signature additional information for the discrimination between phenotypes and redundancy to the collective marking ability. The CellAge signature (1259 genes) more strongly discriminates proliferative from senescent samples than Hernandez-Segura (55 genes) (**Figure 6A**), and random gene sampling also shows proportionality between gene set size and marking performance, whatever the signature and the marking approach (**Figure 6B**). However, this association appears to only hold up to a certain set size, likely the minimum number of genes comprising the full non-redundant marking information. This number depends on the signature and the marking approach. If GSEA is used, only about one third of the CellAge genes are needed for its optimal performance (better than the full signature's) in discriminating proliferative from senescent samples (**Figure 6B**, top left). With the $\log_2$-median score, just over 100 genes are needed for a performance similar to the full signature's (**Figure 6B**, right). The full Hernandez-Segura set is needed for maximum performance with GSEA (**Figure 6B**, left) but less than 20 genes are required for peak performance with a score-based approach, with this performance being superior to CellAge's for the same number of genes (**Figure 6B**, right insets). This suggests that the sparser Hernandez-Segura signature contains a few highly informative genes for the proliferative vs. senescent contrast not included in CellAge.

Our analyses show that less than 50 samples of each phenotype are enough for the signatures' maximum performance (**Figure 6A**). With GSEA, that plateauing happens with even fewer samples for gene-rich CellAge (**Figure 6A**, top left). As expected due to the approaches' nature, significance of GSEA-based results is proportional to the number of genes and saturates when that number gets into the hundreds (**Figure 6A,B**, bottom left), whereas as significance of score-based results follows the proportionality with sample size that is typical of testing differences between pairs of distributions when the effect size does not change (**Figure 6A**, bottom right).

The performance as markers of certain gene sets, profiled by enrichment methods, was influenced by the choice of senescence inducer. When limiting our analyses to proliferative, quiescent and radiation-induced senescence samples, they (*e.g.*, GOBP_NEGATIVE_REGULATION_OF_CELLULAR_SENESCENCE) marked senescence more strongly than when all senescent subtypes were included (**Figure 6C**). This suggests that some gene sets are inducer-specific signatures and underperform in other senescence contexts. The per-sample resolution of graphical depictions of distributions of scores allows to visually capture such context-specific trends (**Supplementary Figure 4**).

3.6 From Biological Insight to Translational Potential

Senescence has been linked to multiple diseases (Borghesan et al., 2020), so quantitatively marking it in large cohorts of samples could reveal associations with clinical features. Given the robustness of the HernandezSegura and SAUL_SEN_MAYO gene sets, we used the discovery mode of `markeR` to test their potential to mark senescence in human tissues GTEx data across 49 healthy tissues (Lonsdale et al.,

2013). While there is no definitive ground truth for relative senescence levels in these tissues, it is generally accepted that senescent cells accumulate with age (Campisi, 2005). Thus, if these gene sets are effective, we expect their scores or enrichments to positively correlate with age.

Each tissue exhibited a distinct median senescence score (**Figure 7A**). Consistently, from the *in vitro* senescence datasets, we observed that baseline senescence scores can differ between cell types; for instance, proliferative mesenchymal cells can score higher than senescent melanocytes (**Figure 7B**). This underscores how tissue heterogeneity can dilute senescence signatures. Therefore, score differences between tissues (**Figure 7A**) may reflect distinct cell-type compositions rather than differences in senescence levels *per se*.

For HernandezSegura, only four tissues (Aorta, Mammary Tissue, Cultured Fibroblasts, and Thyroid) showed significant age-related changes in both score and enrichment approaches (**Figure 7C, D**). Cultured Fibroblasts are of particular interest, given that fibroblasts were key in the derivation and validation of the signature. The observation of an age-related increase in the signature's strength therein, despite the large inter-individual variability, is compatible with the hypothesis of an accumulation of senescent cells with age (**Figure 7D**). The strongest signal observed for the aorta is also consistent with links between senescence and age-related vascular dysfunction (Hall & Lesniewski, 2024). Unlike more complex organs, the aorta has low cellular diversity and is mainly composed of endothelial and vascular smooth muscle cells (Ito et al., 2025). Its lifelong exposure to mechanical strain (Bäck et al., 2013) may also contribute to the observed age-related changes. However, these interpretations should be made with caution. On one hand, a similar pattern was not seen in other arterial tissues, such as the tibial. On the other, we did not detect significant age-related strengthening of the senescence signature in tissues such as the skin, lung or pancreas, despite their known links to age-related senescence (Tuttle et al., 2020).

SAUL_SEN_MAYO includes many SASP genes, including inflammatory ones, so results may partly reflect age-related inflammation (inflamm-ageing) rather than senescence (Franceschi et al., 2000). Using the same criteria as for HernandezSegura, eight tissues showed age-related alterations in scores or enrichment (**Supplementary Figure 7**), including brain regions such as the Anterior Cingulate Cortex and Hippocampus. In some tissues (*e.g.*, Sigmoid Colon, Tibial Nerve), the score decreased with age, contrary to what was expected (**Supplementary Figure 7B**). To assess the potential role of inflammation in our observations, we split the gene set into inflammation-related (N = 81) and -unrelated (N = 44) genes. Both subsets performed similarly in the *in vitro* senescence datasets, confirming their robustness in that context with a ground truth for senescence (**Supplementary Figure 8A-B**). In GTEx, the association of the inflammatory and non-inflammatory gene subsets with age varies across tissues (**Supplementary Figure 8C**). The absence of a consistent trend across tissues indicates that the observations made for SAUL_SEN_MAYO cannot be solely attributed to inflammageing, and suggest a variable combination of the signals of senescence and inflammation.

Overall, interpretation of changes in molecular signatures in tissue data must take into account that cellular composition changes with age (*e.g.,* more adipose and less epithelial cells in the breast, more granulocytes in blood)(Todhunter et al., 2018), and

GTEx samples are cross-sectional, not longitudinal (which, while ideal, are impractical). This, combined with the expected low proportion of senescent cells, as GTEx tissues were in principle functional at collection, may mask meaningful senescence-specific signals in bulk transcriptomic data.

4. DISCUSSION

Quantifying complex cellular phenotypes like senescence in transcriptomic samples often requires integrating multiple gene signatures and analytical strategies, as no single gene functions as a universal marker. To meet such need, we developed `markeR`, a software for flexibly analysing and evaluating gene sets as transcriptomic phenotypic markers.

Key strengths of `markeR` are its modular architecture, supporting the integration of new functionalities (e.g., enrichment methods beyond GSEA), and its emphasis on continuous, quantitative metrics over binary classification methods, reflecting the non-discrete nature of biological phenotypes. Rather than enforcing strict cut-offs, `markeR` encourages interpreting scores in the context of their distribution, allowing more nuanced insights, particularly for cases deemed borderline, and supporting flexible, data-informed decisions. A Shiny-based interface is planned to improve accessibility and integration with public dataset repositories (*e.g.*, GEO (Edgar, 2002)), and upcoming releases will implement standardised metrics for cross-cell-type comparisons (J. Wang et al., 2025).

Applying `markeR` to marking senescence yielded both technical and biological insights. Scoring methods offer sample-level resolution but are sensitive to the behaviour of individual genes, potentially skewing interpretability when, for instance, the goal is to evaluate an entire pathway as a phenotypic marker. Enrichment-based methods focus on coordinated gene behaviour but are limited to sample group-level comparisons. Even among scoring methods, applicability tradeoffs exist: rank-based approaches (*e.g.*, ssGSEA and rank scoring) tend to be more robust to noise, while magnitude-based methods (*e.g.*, $\log_2$-median) may better detect expression shifts in technically well-controlled settings. Ultimately, these methods should be seen as complementary and applied in an integrative manner, guided by the same biological question.

No gene set performed optimally in specifically marking senescence across all cell types analysed herein. However, HernandezSegura and SAUL_SEN_MAYO stood out for their consistency across metrics and low false positive rates, suggesting they capture core senescence molecular features. In contrast, commonly used senescence-associated MSigDB gene sets performed poorly, and others, like CSGene and SeneQuest, have their performance affected by oblivion to the direction of marker gene expression changes upon phenotype and enrichment in non-specific proliferation-related genes, respectively.

Context-dependence emerged as a major, though expected challenge. Senescence is not a discrete state but a complex phenotype influenced by inducing stimulus, cell type and microenvironment (Gasek et al., 2021). In human tissues, this complexity is reflected in the inter-individual and often age-related variability in cell type composition (Todhunter et al., 2018), and its study is hampered by the typically low abundance of naturally occurring senescent cells. Although we used quiescent samples as non-proliferative controls, other non-senescent states (*e.g.*, differentiated cells), may still affect marker specificity.

Some tissues, like cultured fibroblasts and the aorta, showed association between molecular signals of senescence and donor's age, and could be seen as positive

controls for the accumulation of senescent cells with ageing. However, defining 'positive controls' for senescence is not trivial. Tissues like skin, lung or pancreas are often cited for accumulating senescent cells with age, based on markers like p16$^{(INK4a)}$ or p21$^{(CIP1/WAF1)}$ (Idda et al., 2020; Tuttle et al., 2020). Yet these markers are not exclusive to senescent cells. For example, macrophages can express p16$^{(INK4a)}$ in non-senescent inflammatory contexts (Hall et al., 2017), and p21$^{(CIP1/WAF1)}$ is a stress-responsive gene that does not necessarily indicate irreversible growth arrest. Thus, some reported age-associated 'senescence' signals may reflect immune or stress responses, rather than senescence itself. In this context, the lack of significant age associations in some tissues, despite previous reports, likely reflect the limited specificity of senescence markers and the inability of bulk RNA-seq to resolve cell-type-specific processes. Moreover, GTEx data are cross-sectional rather than longitudinal and marked by the expectedly large natural inter-individual variability, further complicating interpretation. These findings emphasise the need to interpret senescence molecular signatures with caution when translating them from controlled *in vitro* settings to the complex cellular environments of human tissues.

Senescence is best understood as a heterogeneous collection of phenotypes shaped by the initiating stimulus, cell type and microenvironment, rather than as a single, uniform cellular state. This mirrors the current view of cancer as a spectrum of distinct yet overlapping transcriptional states, broadly defined by uncontrolled proliferation, rather than as a single, fixed phenotype. In contrast, the study of immune cell biology often emphasises unprecedent resolution at the level of specific, narrowly defined cell types. Applying such granular and rigid compartmentalisation to a systemic process like senescence risks overlooking broader molecular patterns and conserved mechanisms that transcend cell-type- and stimulus-specificity. A hybrid approach that recognises and integrates both the conserved hallmarks and the context-specific features of senescent cells is essential.

Quantifying senescence in biological samples from their bulk RNA-seq transcriptomes remains challenging, as molecular signals from rare senescent cells of specific types are diluted therein. While `markeR` is currently designed for bulk transcriptomes, it can be adapted to single-cell data. However, applying bulk-oriented scoring methods (*e.g.*, ssGSEA) to single-cell RNA sequencing (scRNA-seq) data can introduce bias, due to gene-detection variability across cells (*i.e.*, dropouts) (Noureen et al., 2022). Extending `markeR` with algorithms tailored for scRNA-seq is therefore part of our future plans. Several methods specifically designed for scRNA-seq already exist (*e.g.*, AUCell (Aibar et al., 2017)), and some scRNA-seq analysis workflows already implement score-based approaches (*e.g.,* $\log_2$-median centred in scStudio (Bica et al., 2025)), with promising performance.

Meanwhile, controlled *in vitro* datasets remain indispensable for benchmarking gene sets and workflows, and for mapping the transcriptional diversity of senescence across cell types, stimuli, and other specific microenvironmental contexts. These models will therefore continue to play a central role in refining our understanding of the molecular landscape of senescence.

Acknowledgements

We thank the present and past members from the Disease Transcriptomics Lab, for the fruitful discussions, for testing the package's functionalities during its development, and for providing valuable feedback on the manuscript. The authors used OpenAI's GPT-4 model (accessed via ChatGPT) to assist with language polishing. The model was not used for content generation or data analysis. As proud 'research parasites' (Longo & Drazen, 2016), we processed publicly available datasets; we thank the researchers whose studies provided the data used in this work, for generating and sharing it, which proved invaluable to the feasibility of our study.

This work is funded by national funds through the FCT - Fundação para a Ciência e a Tecnologia, I.P., under the project UIDP/50005/2020 and individual PhD Studentships UI/BD/153368/2022 to Rita Martins-Silva and UI/BD/153363/2022 to Alexandre Kaizeler, and CEEC Individual Researcher contract CEECIND/00436/2018 to Nuno Barbosa-Morais.

Data and code availability

No data has been generated for this manuscript. The senescence compendium was compiled using publicly available RNA-seq datasets accessed under the following accession codes: E-MTAB-5403 (Hernandez-Segura et al., 2017), E-MTAB-9714 (Mangelinck et al., 2020), GSE175533 (Chan et al., 2022), GSE60340 (Purcell et al., 2014), GSE63577 (Marthandan et al., 2016), GSE64553 (Marthandan et al., 2015), GSE214410 (Wang et al., 2025), GSE222400 (Suda et al., 2024), GSE224070 (McHugh et al., 2023), GSE230181 (Savić et al., 2023), GSE235768 (Skea et al., 2023), GSE247831 (Laurent et al., 2023), GSE74324 (Tasdemir et al., 2016), GSE75643 (Lenain et al., 2017), GSE94928 (Aarts et al., 2017), GSE130727 (Casella et al., 2019), GSE250224 (Numa et al., 2024), GSE196724 (Admasu et al., 2023), GSE206677 (Urata et al., 2022), GSE217718 (C. Wang et al., 2023), GSE110268 (Fu et al., 2019), GSE213323 (Tanke et al., 2024), E-MTAB-10969 (Bauwens et al., 2023). Data from The Genotype-Tissue Expression (GTEx) v8, already aligned, processed (filtered and normalised) and corrected for batch effects, was obtained from the GitHub repository associated with the Schneider et al., 2024 publication (https://github.com/DiseaseTranscriptomicsLab/voyAGEr). Specific donor metadata, including the exact age of each donor, was retrieved through the dbGaP accession number phs000424.v9.p2.
The complete source code for the `markeR` package (v1.0.0), used throughout this manuscript, is publicly available through Bioconductor (https://bioconductor.org/packages/release/bioc/html/markeR.html) and GitHub (https://github.com/DiseaseTranscriptomicsLab/markeR). The package can be installed in `R` using the command: `BiocManager::install("markeR")`. Additionally, the code for data preprocessing used in the senescence case study, along with the scripts to generate the figures in this paper, as well as the processed senescence datasets and gene sets used for the main figures, can be found in the `paper` branch of the `markeR` GitHub repository.

6. REFERENCES

Aarts, M., Georgilis, A., Beniazza, M., Beolchi, P., Banito, A., Carroll, T., Kulisic, M., Kaemena, D. F., Dharmalingam, G., Martin, N., Reik, W., Zuber, J., Kaji, K., Chandra, T., & Gil, J. (2017). Coupling shRNA screens with single-cell RNA-seq identifies a dual role for mTOR in reprogramming-induced senescence. *Genes & Development*, *31*(20), 2085–2098. https://doi.org/10.1101/gad.297796.117

Admasu, T. D., Kim, K., Rae, M., Avelar, R., Gonciarz, R. L., Rebbaa, A., Pedro De Magalhães, J., Renslo, A. R., Stolzing, A., & Sharma, A. (2023). Selective ablation of primary and paracrine senescent cells by targeting iron dyshomeostasis. *Cell Reports*, *42*(2), 112058. https://doi.org/10.1016/j.celrep.2023.112058

Aibar, S., González-Blas, C. B., Moerman, T., Huynh-Thu, V. A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P., Aerts, J., van den Oord, J., Atak, Z. K., Wouters, J., & Aerts, S. (2017). SCENIC: single-cell regulatory network inference and clustering. *Nature Methods*, *14*(11), 1083–1086. https://doi.org/10.1038/nmeth.4463

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data* (Version 0.12.1) [Computer software]. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Avelar, R. A., Ortega, J. G., Tacutu, R., Tyler, E. J., Bennett, D., Binetti, P., Budovsky, A., Chatsirisupachai, K., Johnson, E., Murray, A., Shields, S., Tejada-Martinez, D., Thornton, D., Fraifeld, V. E., Bishop, C. L., & de Magalhães, J. P. (2020). A multidimensional systems biology analysis of cellular senescence in aging and disease. *Genome Biology*, *21*(1), 91. https://doi.org/10.1186/s13059-020-01990-9

Bäck, M., Gasser, T. C., Michel, J.-B., & Caligiuri, G. (2013). Biomechanical factors in the biology of aortic wall and aortic valve diseases. *Cardiovascular Research*, *99*(2), 232–241. https://doi.org/10.1093/cvr/cvt040

Barbie, D. A., Tamayo, P., Boehm, J. S., Kim, S. Y., Moody, S. E., Dunn, I. F., Schinzel, A. C., Sandy, P., Meylan, E., Scholl, C., Fröhling, S., Chan, E. M., Sos, M. L., Michel, K., Mermel, C., Silver, S. J., Weir, B. A., Reiling, J. H., Sheng, Q., … Hahn, W. C. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature*, *462*(7269), 108–112. https://doi.org/10.1038/nature08460

Bauwens, E., Parée, T., Meurant, S., Bouriez, I., Hannart, C., Wéra, A.-C., Khelfi, A., Fattaccioli, A., Burteau, S., Demazy, C., Fransolet, M., De Schutter, C., Martin, N., Théry, J., Decanter, G., Penel, N., Bury, M., Pluquet, O., Garmyn, M., & Debacq-Chainiaux, F. (2023). Senescence Induced by UVB in Keratinocytes Impairs Amino Acids Balance. *Journal of Investigative Dermatology*, *143*(4), 554-565.e9. https://doi.org/10.1016/j.jid.2022.11.017

Bica, M., Serre, K., & Barbosa-Morais, N. L. (2025). scStudio: A User-Friendly Web Application Empowering Non-Computational Users with Intuitive scRNA-seq

Data Analysis. *bioRxiv*, 2025.04.17.649161. https://doi.org/10.1101/2025.04.17.649161

Borghesan, M., Hoogaars, W. M. H., Varela-Eirin, M., Talma, N., & Demaria, M. (2020). A Senescence-Centric View of Aging: Implications for Longevity and Disease. *Trends in Cell Biology*, *30*(10), 777–791. https://doi.org/10.1016/j.tcb.2020.07.002

Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, *34*(5), 525–527. https://doi.org/10.1038/nbt.3519

Campisi, J. (2001). Cellular senescence as a tumor-suppressor mechanism. *Trends in Cell Biology*, *11*(11), S27–S31. https://doi.org/10.1016/S0962-8924(01)02151-1

Campisi, J. (2005). Senescent Cells, Tumor Suppression, and Organismal Aging: Good Citizens, Bad Neighbors. *Cell*, *120*(4), 513–522. https://doi.org/10.1016/j.cell.2005.02.003

Casella, G., Munk, R., Kim, K. M., Piao, Y., De, S., Abdelmohsen, K., & Gorospe, M. (2019). Transcriptome signature of cellular senescence. *Nucleic Acids Research*, *47*(14), 7294–7305. https://doi.org/10.1093/nar/gkz555

Chan, M., Yuan, H., Soifer, I., Maile, T. M., Wang, R. Y., Ireland, A., O'Brien, J. J., Goudeau, J., Chan, L. J., Vijay, T., Freund, A., Kenyon, C., Bennett, B. D., McAllister, F. E., Kelley, D. R., Roy, M., Cohen, R. L., Levinson, A. D., Botstein, D., & Hendrickson, D. G. (2022). Novel insights from a multiomics dissection of the Hayflick limit. *eLife*, *11*, e70283. https://doi.org/10.7554/eLife.70283

Chatsirisupachai, K., Palmer, D., Ferreira, S., & De Magalhães, J. P. (2019). A human tissue-specific transcriptomic analysis reveals a complex relationship between aging, cancer, and cellular senescence. *Aging Cell*, *18*(6), e13041. https://doi.org/10.1111/acel.13041

Chen, Y., Chen, L., Lun, A. T. L., Baldoni, P. L., & Smyth, G. K. (2025). edgeR v4: Powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. *Nucleic Acids Research*, *53*(2), gkaf018. https://doi.org/10.1093/nar/gkaf018

Coppé, J.-P., Patil, C. K., Rodier, F., Sun, Y., Muñoz, D. P., Goldstein, J., Nelson, P. S., Desprez, P.-Y., & Campisi, J. (2008). Senescence-Associated Secretory Phenotypes Reveal Cell-Nonautonomous Functions of Oncogenic RAS and the p53 Tumor Suppressor. *PLOS Biology*, *6*(12), 1–1. https://doi.org/10.1371/journal.pbio.0060301

De Almeida, B. P., Vieira, A. F., Paredes, J., Bettencourt-Dias, M., & Barbosa-Morais, N. L. (2019). Pan-cancer association of a centrosome amplification gene expression signature with genomic alterations and clinical outcome. *PLOS Computational Biology*, *15*(3), e1006832. https://doi.org/10.1371/journal.pcbi.1006832

Demaria, M., Ohtani, N., Youssef, S. A., Rodier, F., Toussaint, W., Mitchell, J. R., Laberge, R.-M., Vijg, J., Steeg, H. V., Dollé, M. E. T., Hoeijmakers, J. H. J., Bruin, A. de, Hara,

E., & Campisi, J. (2014). An Essential Role for Senescent Cells in Optimal Wound Healing through Secretion of PDGF-AA. *Developmental Cell*, *31*(6), 722–733. https://doi.org/10.1016/j.devcel.2014.11.012

Edgar, R. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research*, *30*(1), 207–210. https://doi.org/10.1093/nar/30.1.207

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Franceschi, C., Bonafè, M., Valensin, S., Olivieri, F., De Luca, M., Ottaviani, E., & De Benedictis, G. (2000). Inflamm-aging: An Evolutionary Perspective on Immunosenescence. *Annals of the New York Academy of Sciences*, *908*(1), 244–254. https://doi.org/10.1111/j.1749-6632.2000.tb06651.x

Fu, L., Hu, Y., Song, M., Liu, Z., Zhang, W., Yu, F.-X., Wu, J., Wang, S., Izpisua Belmonte, J. C., Chan, P., Qu, J., Tang, F., & Liu, G.-H. (2019). Up-regulation of FOXD1 by YAP alleviates senescence and osteoarthritis. *PLoS Biology*, *17*(4), e3000201. https://doi.org/10.1371/journal.pbio.3000201

Gasek, N. S., Kuchel, G. A., Kirkland, J. L., & Xu, M. (2021). Strategies for targeting senescent cells in human disease. *Nature Aging*, *1*(10), 870–879. https://doi.org/10.1038/s43587-021-00121-8

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

Hall, B. M., Balan, V., Gleiberman, A. S., Strom, E., Krasnov, P., Virtuoso, L. P., Rydkina, E., Vujcic, S., Balan, K., Gitlin, I. I., Leonova, K. I., Consiglio, C. R., Gollnick, S. O., Chernova, O. B., & Gudkov, A. V. (2017). P16(Ink4a) and senescence-associated β-galactosidase can be induced in macrophages as part of a reversible response to physiological stimuli. *Aging*, *9*(8), 1867–1884. https://doi.org/10.18632/aging.101268

Hall, S. A., & Lesniewski, L. A. (2024). Targeting vascular senescence in cardiovascular disease with aging. *The Journal of Cardiovascular Aging*, *4*(2). https://doi.org/10.20517/jca.2023.45

He, Y., Long, K., Du, B., Liao, W., Zou, R., Su, J., Luo, J., Shi, Z., & Wang, L. (2024). The cellular senescence score (CSS) is a comprehensive biomarker to predict prognosis and assess senescence and immune characteristics in hepatocellular carcinoma (HCC). *Biochemical and Biophysical Research Communications*, *739*, 150576. https://doi.org/10.1016/j.bbrc.2024.150576

Hernandez-Segura, A., de Jong, T. V., Melov, S., Guryev, V., Campisi, J., & Demaria, M. (2017). Unmasking Transcriptional Heterogeneity in Senescent Cells. *Current Biology*, *27*(17), 2652-2660.e4. https://doi.org/10.1016/j.cub.2017.07.033

Hernandez-Segura, A., Nehme, J., & Demaria, M. (2018). Hallmarks of Cellular Senescence. *Trends in Cell Biology*, *28*(6), 436–453. https://doi.org/10.1016/j.tcb.2018.02.001

Idda, M. L., McClusky, W. G., Lodde, V., Munk, R., Abdelmohsen, K., Rossi, M., & Gorospe, M. (2020). Survey of senescent cell markers with age in human tissues. *Aging*, *12*(5), 4052–4066. https://doi.org/10.18632/aging.102903

Ito, S., Graf, D. B., Katsumata, Y., Moorleghen, J. J., Zhang, C., Li, Y., LeMaire, S. A., Shen, Y. H., Lu, H. S., Daugherty, A., & Sawada, H. (2025). Smooth muscle cells and fibroblasts in the ascending aorta exhibit minor differences between embryonic origins in angiotensin II-driven transcriptional alterations. *Scientific Reports*, *15*(1), 16617. https://doi.org/10.1038/s41598-025-99862-4

Lafferty-Whyte, K., Bilsland, A., Cairney, C. J., Hanley, L., Jamieson, N. B., Zaffaroni, N., Oien, K. A., Burns, S., Roffey, J., Boyd, S. M., & Keith, W. N. (2010). Scoring of senescence signalling in multiple human tumour gene expression datasets, identification of a correlation between senescence score and drug toxicity in the NCI60 panel and a pro-inflammatory signature correlating with survival advantage in peritoneal mesothelioma. *BMC Genomics*, *11*(1), 532. https://doi.org/10.1186/1471-2164-11-532

Laurent, M., Cordeddu, L., Zahedi, Y., & Ekwall, K. (2023). LEO1 Is Required for Efficient Entry into Quiescence, Control of H3K9 Methylation and Gene Expression in Human Fibroblasts. *Biomolecules*, *13*(11), 1662. https://doi.org/10.3390/biom13111662

Lenain, C., de Graaf, C. A., Pagie, L., Visser, N. L., de Haas, M., de Vries, S. S., Peric-Hupkes, D., van Steensel, B., & Peeper, D. S. (2017). Massive reshaping of genome-nuclear lamina interactions during oncogene-induced senescence. *Genome Research*, *27*(10), 1634–1644. https://doi.org/10.1101/gr.225763.117

Longo, D. L., & Drazen, J. M. (2016). Data Sharing. *New England Journal of Medicine*, *374*(3), 276–277. https://doi.org/10.1056/nejme1516564

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

Mangelinck, A., Coudereau, C., Courbeyrette, R., Ouararhni, K., Hamiche, A., Redon, C. E., Bonner, W. M., Van Dijk, E., Derbois, C., Olaso, R., Deleuze, J.-F., Fenaille, F., Ruebe, C. E., Thuret, J.-Y., & Mann, C. (2020). *The H2A.J histone variant contributes to Interferon-Stimulated Gene expression in senescence by its weak*

interaction with H1 and the derepression of repeated DNA sequences. Genetics. https://doi.org/10.1101/2020.10.29.361204

Marthandan, S., Baumgart, M., Priebe, S., Groth, M., Schaer, J., Kaether, C., Guthke, R., Cellerino, A., Platzer, M., Diekmann, S., & Hemmerich, P. (2016). Conserved Senescence Associated Genes and Pathways in Primary Human Fibroblasts Detected by RNA-Seq. *PLOS ONE*, *11*(5), e0154531. https://doi.org/10.1371/journal.pone.0154531

Marthandan, S., Priebe, S., Groth, M., Guthke, R., Platzer, M., Hemmerich, P., & Diekmann, S. (2015). Hormetic effect of rotenone in primary human fibroblasts. *Immunity & Ageing*, *12*(1), 11. https://doi.org/10.1186/s12979-015-0038-8

McHugh, D., Sun, B., Gutierrez-Muñoz, C., Hernández-González, F., Mellone, M., Guiho, R., Duran, I., Pombo, J., Pietrocola, F., Birch, J., Kallemeijn, W. W., Khadayate, S., Dharmalingam, G., Vernia, S., Tate, E. W., Martínez-Barbera, J. P., Withers, D. J., Thomas, G. J., Serrano, M., & Gil, J. (2023). COPI vesicle formation and N-myristoylation are targetable vulnerabilities of senescent cells. *Nature Cell Biology*, *25*(12), 1804–1820. https://doi.org/10.1038/s41556-023-01287-6

Melo Dos Santos, L., Trombetta-Lima, M., Eggen, B., & Demaria, M. (2024). Cellular senescence in brain aging and neurodegeneration. *Ageing Research Reviews*, *93*, 102141. https://doi.org/10.1016/j.arr.2023.102141

Muñoz-Espín, D., Cañamero, M., Maraver, A., Gómez-López, G., Contreras, J., Murillo-Cuesta, S., Rodríguez-Baeza, A., Varela-Nieto, I., Ruberte, J., Collado, M., & Serrano, M. (2013). Programmed Cell Senescence during Mammalian Embryonic Development. *Cell*, *155*(5), 1104–1118. https://doi.org/10.1016/j.cell.2013.10.019

Noureen, N., Ye, Z., Chen, Y., Wang, X., & Zheng, S. (2022). Signature-scoring methods developed for bulk samples are not adequate for cancer single-cell RNA sequencing data. *eLife*, *11*, e71994. https://doi.org/10.7554/eLife.71994

Numa, K., Patel, S. K., Zhang, Z. A., Burton, J. B., Matsumoto, A., Hughes, J.-W. B., Sotozono, C., Schilling, B., Desprez, P.-Y., Campisi, J., & Kitazawa, K. (2024). Senescent characteristics of human corneal endothelial cells upon ultraviolet-A exposure. *Aging*, *16*(8), 6673–6693. https://doi.org/10.18632/aging.205761

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., … Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, *44*(D1), D733–D745. https://doi.org/10.1093/nar/gkv1189

Parkinson, H., Kapushesky, M., Shojatalab, M., Abeygunawardena, N., Coulson, R., Farne, A., Holloway, E., Kolesnykov, N., Lilja, P., Lukk, M., Mani, R., Rayner, T., Sharma, A., William, E., Sarkans, U., & Brazma, A. (2007). ArrayExpress—A public

database of microarray experiments and gene expression profiles. *Nucleic Acids Research*, *35*(Database), D747–D750. https://doi.org/10.1093/nar/gkl995

Purcell, M., Kruger, A., & Tainsky, M. A. (2014). Gene expression profiling of replicative and induced senescence. *Cell Cycle*, *13*(24), 3927–3937. https://doi.org/10.4161/15384101.2014.973327

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*(7), e47–e47. https://doi.org/10.1093/nar/gkv007

Rodier, F., & Campisi, J. (2011). Four faces of cellular senescence. *Journal of Cell Biology*, *192*(4), 547–556. https://doi.org/10.1083/jcb.201009094

Saul, D., Kosinsky, R. L., Atkinson, E. J., Doolittle, M. L., Zhang, X., LeBrasseur, N. K., Pignolo, R. J., Robbins, P. D., Niedernhofer, L. J., Ikeno, Y., Jurk, D., Passos, J. F., Hickson, L. J., Xue, A., Monroe, D. G., Tchkonia, T., Kirkland, J. L., Farr, J. N., & Khosla, S. (2022). A new gene set identifies senescent cells and predicts senescence-associated pathways across tissues. *Nature Communications*, *13*(1), 4827. https://doi.org/10.1038/s41467-022-32552-1

Savić, R., Yang, J., Koplev, S., An, M. C., Patel, P. L., O'Brien, R. N., Dubose, B. N., Dodatko, T., Rogatsky, E., Sukhavasi, K., Ermel, R., Ruusalepp, A., Houten, S. M., Kovacic, J. C., Stewart, A. F., Yohn, C. B., Schadt, E. E., Laberge, R.-M., Björkegren, J. L. M., … Argmann, C. (2023). Integration of transcriptomes of senescent cell models with multi-tissue patient samples reveals reduced COL6A3 as an inducer of senescence. *Cell Reports*, *42*(11), 113371. https://doi.org/10.1016/j.celrep.2023.113371

Schneider, A. L., Martins-Silva, R., Kaizeler, A., Saraiva-Agostinho, N., & Barbosa-Morais, N. L. (2024). voyAGEr, a free web interface for the analysis of age-related gene expression alterations in human tissues. *eLife*, *12*, RP88623. https://doi.org/10.7554/eLife.88623

Serrano, M., Lin, A. W., McCurrach, M. E., Beach, D., & Lowe, S. W. (1997). Oncogenic ras Provokes Premature Cell Senescence Associated with Accumulation of p53 and p16INK4a. *Cell*, *88*(5), 593–602. https://doi.org/10.1016/S0092-8674(00)81902-9

Sharpless, N. E., & Sherr, C. J. (2015). Forging a signature of in vivo senescence. *Nature Reviews Cancer*, *15*(7), 397–408. https://doi.org/10.1038/nrc3960

Skea, D., Fotis, C., Tsolakos, N., Pliaka, V., Verrou, K.-M., & Alexopoulos, L. G. (2023). *Conserved transcriptomic signatures and protein markers in cellular senescence models*. In Review. https://doi.org/10.21203/rs.3.rs-3110821/v1

Smyth, G. K. (2004). Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. *Statistical Applications in Genetics and Molecular Biology*, *3*(1), 1–25. https://doi.org/10.2202/1544-6115.1027

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, *102*(43), 15545–15550. https://doi.org/10.1073/pnas.0506580102

Suda, K., Moriyama, Y., Razali, N., Chiu, Y., Masukagami, Y., Nishimura, K., Barbee, H., Takase, H., Sugiyama, S., Yamazaki, Y., Sato, Y., Higashiyama, T., Johmura, Y., Nakanishi, M., & Kono, K. (2024). Plasma membrane damage limits replicative lifespan in yeast and induces premature senescence in human fibroblasts. *Nature Aging*, *4*(3), 319–335. https://doi.org/10.1038/s43587-024-00575-6

Tanke, N. T., Liu, Z., Gore, M. T., Bougaran, P., Linares, M. B., Marvin, A., Sharma, A., Oatley, M., Yu, T., Quigley, K., Vest, S., Cook, J. G., & Bautch, V. L. (2024). Endothelial cell flow-mediated quiescence is temporally regulated and utilizes the cell cycle inhibitor p27. *bioRxiv: The Preprint Server for Biology*, 2023.06.09.544403. https://doi.org/10.1101/2023.06.09.544403

Tasdemir, N., Banito, A., Roe, J.-S., Alonso-Curbelo, D., Camiolo, M., Tschaharganeh, D. F., Huang, C.-H., Aksoy, O., Bolden, J. E., Chen, C.-C., Fennell, M., Thapar, V., Chicas, A., Vakoc, C. R., & Lowe, S. W. (2016). BRD4 Connects Enhancer Remodeling to Senescence Immune Surveillance. *Cancer Discovery*, *6*(6), 612–629. https://doi.org/10.1158/2159-8290.CD-16-0217

Todhunter, M. E., Sayaman, R. W., Miyano, M., & LaBarge, M. A. (2018). Tissue aging: The integration of collective and variant responses of cells to entropic forces over time. *Current Opinion in Cell Biology*, *54*, 121–129. https://doi.org/10.1016/j.ceb.2018.05.016

Tuttle, C. S. L., Waaijer, M. E. C., Slee-Valentijn, M. S., Stijnen, T., Westendorp, R., & Maier, A. B. (2020). Cellular senescence and chronological age in various human tissues: A systematic review and meta-analysis. *Aging Cell*, *19*(2), e13083. https://doi.org/10.1111/acel.13083

Urata, R., Ikeda, K., Yamazaki, E., Ueno, D., Katayama, A., Shin-Ya, M., Ohgitani, E., Mazda, O., & Matoba, S. (2022). Senescent endothelial cells are predisposed to SARS-CoV-2 infection and subsequent endothelial dysfunction. *Scientific Reports*, *12*(1), 11855. https://doi.org/10.1038/s41598-022-15976-z

Wang, C., Zhang, H., Fan, J., Li, Q., Guo, R., Pan, J., Liu, Y., Peng, J., Zhu, Q., Feng, Y., Wu, C., Luo, P., Qiu, X., Shi, J., Deng, Y., Qi, S., & Liu, Y. (2023). Inhibition of integrated stress response protects against lipid-induced senescence in hypothalamic neural stem cells in adamantinomatous craniopharyngioma. *Neuro-Oncology*, *25*(4), 720–732. https://doi.org/10.1093/neuonc/noac261

Wang, J., Zhou, X., Yu, P., Yao, J., Guo, P., Xu, Q., Zhao, Y., Wang, G., Li, Q., Zhu, X., Wei, G., Wang, W., & Ni, T. (2025). A transcriptome-based human universal senescence index (hUSI) robustly predicts cellular senescence under various conditions. *Nature Aging*. https://doi.org/10.1038/s43587-025-00886-2

Wang, X., Ma, L., Pei, X., Wang, H., Tang, X., Pei, J.-F., Ding, Y.-N., Qu, S., Wei, Z.-Y., Wang, H.-Y., Wang, X., Wei, G.-H., Liu, D.-P., & Chen, H.-Z. (2022). Comprehensive assessment of cellular senescence in the tumor microenvironment. *Briefings in Bioinformatics*, *23*(3), bbac118. https://doi.org/10.1093/bib/bbac118

Wang, Y., Eapen, V. V., Liang, Y., Kournoutis, A., Sherman, M. S., Xu, Y., Onorati, A., Li, X., Zhou, X., Corey, K. E., Du, K., Cabral Burkard, A. M., Ho, C.-K., Xie, J., Zhang, H., Maeso-Díaz, R., Ma, X., Rieprecht, U., O'Brien, T., ... Dou, Z. (2025). WSTF nuclear autophagy regulates chronic but not acute inflammation. *Nature*, *644*(8077), 780–789. https://doi.org/10.1038/s41586-025-09234-1

Zhang, X., Huang, Y., Li, Q., Zhong, Y., Zhang, Y., Hu, J., Liu, R., & Luo, X. (2023). Senescence risk score: A multifaceted prognostic tool predicting outcomes, stemness, and immune responses in colorectal cancer. *Frontiers in Immunology*, *14*, 1265911. https://doi.org/10.3389/fimmu.2023.1265911

Zhao, M., Chen, L., & Qu, H. (2016). CSGene: A literature-based database for cell senescence genes and its application to identify critical cell aging pathways and associated diseases. *Cell Death & Disease*, *7*(1), e2053–e2053. https://doi.org/10.1038/cddis.2015.414

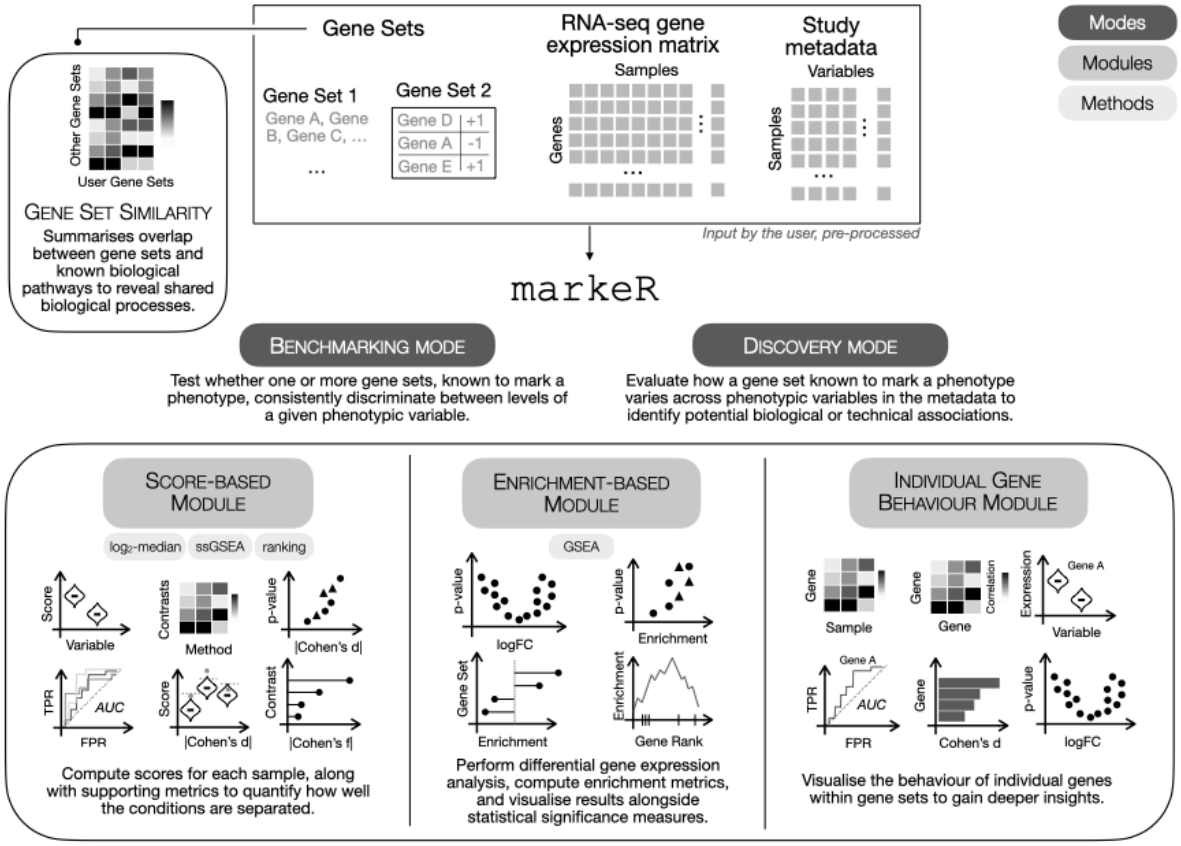


Figure 1 – **Overview of markeR**

`markeR` requires three main inputs: (1) one or more gene signatures putatively associated with a phenotype of interest, to be provided with or without directionality (e.g., signature 1 is made of genes with no information on the expected direction of regulation, therefore assumed to be up-regulated upon phenotype; signature 2 includes genes explicitly annotated as up- or down-regulated upon phenotype); (2) a pre-processed (filtered for low expression genes and normalised) RNA-seq gene expression matrix; and (3) sample-level metadata. Users can choose between two analysis modes: benchmarking, to test the robustness of signatures in discriminating between phenotypic groups of samples, and discovery, to profile associations between a given signature and selected (from metadata) phenotypic variables. Regardless of the mode, the quantitative evaluation of signatures as phenotypic markers can be done using score-based or enrichment-based methods. The contribution of individual genes to the respective signatures' performance can also be explored. `markeR` also quantitatively compares between user-defined and curated reference gene sets, to enhance biological interpretability.

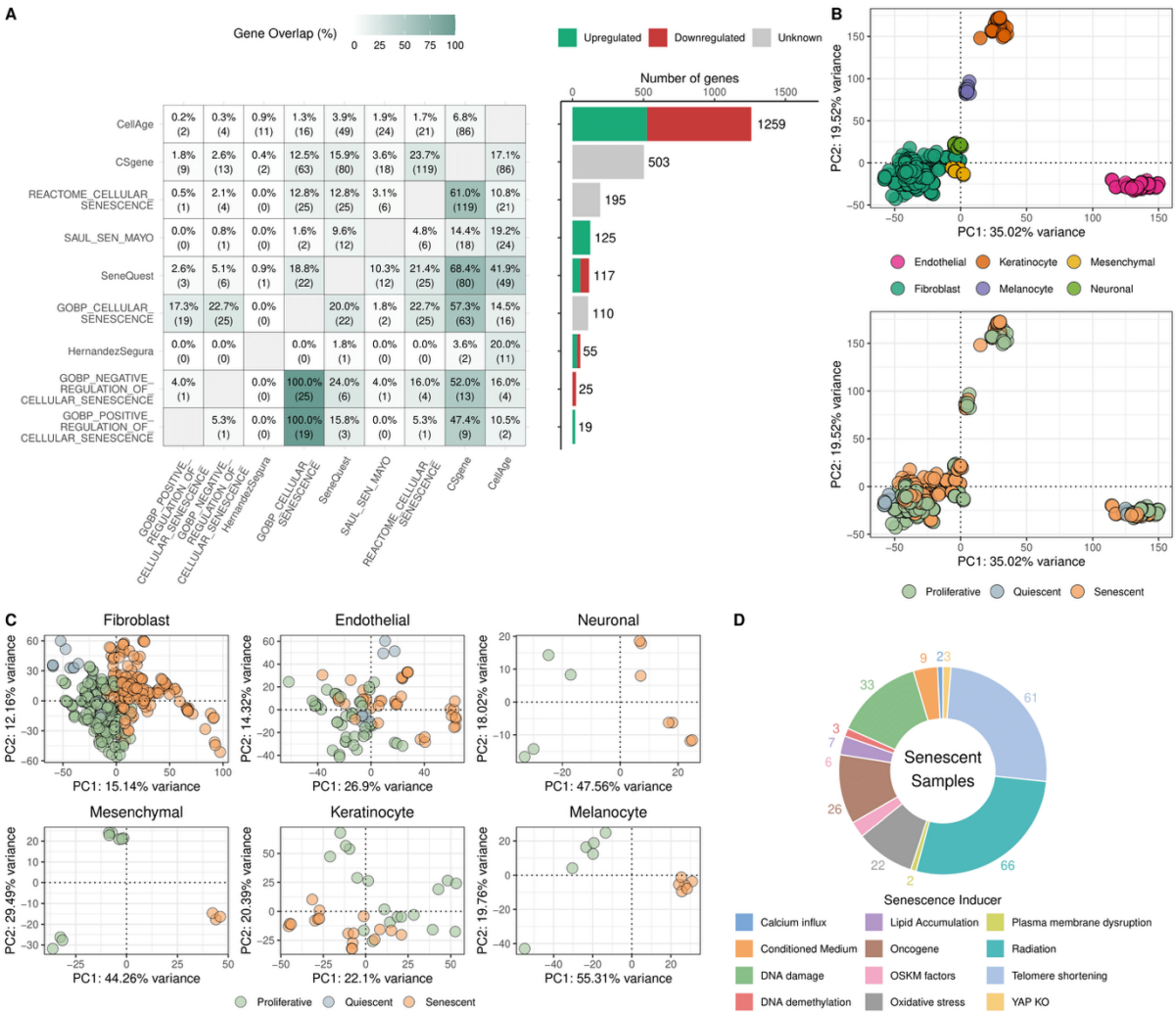


Figure 2 – **Overview of senescence gene sets and RNA-seq datasets.**
A collection of 9 publicly available senescence-associated gene sets and 25 RNA-seq datasets, covering 545 samples from six human cell types, was used in this analysis. ***(A)*** *Heatmap showing the percentage and number (in brackets) of genes of each signature in the rows in common with those of signatures in the columns (left plot), accompanied by a barplot of the total number of genes per set, coloured by the expected direction of change upon senescence (right plot; upregulated in green, downregulated in red, unknown direction in grey).* ***(B)*** *Principal component analysis (PCA) plots of gene expression showing that, after batch effect correction (see Methods), most data variance can be attributed to the cell type of origin (top plot) but not cellular state (bottom).* ***(C)*** *PCA plots of gene expression per cell type, showing that cellular state is a major source of variance within each cell type.* ***(D)*** *Donut chart of senescent samples per senescence inducer.*

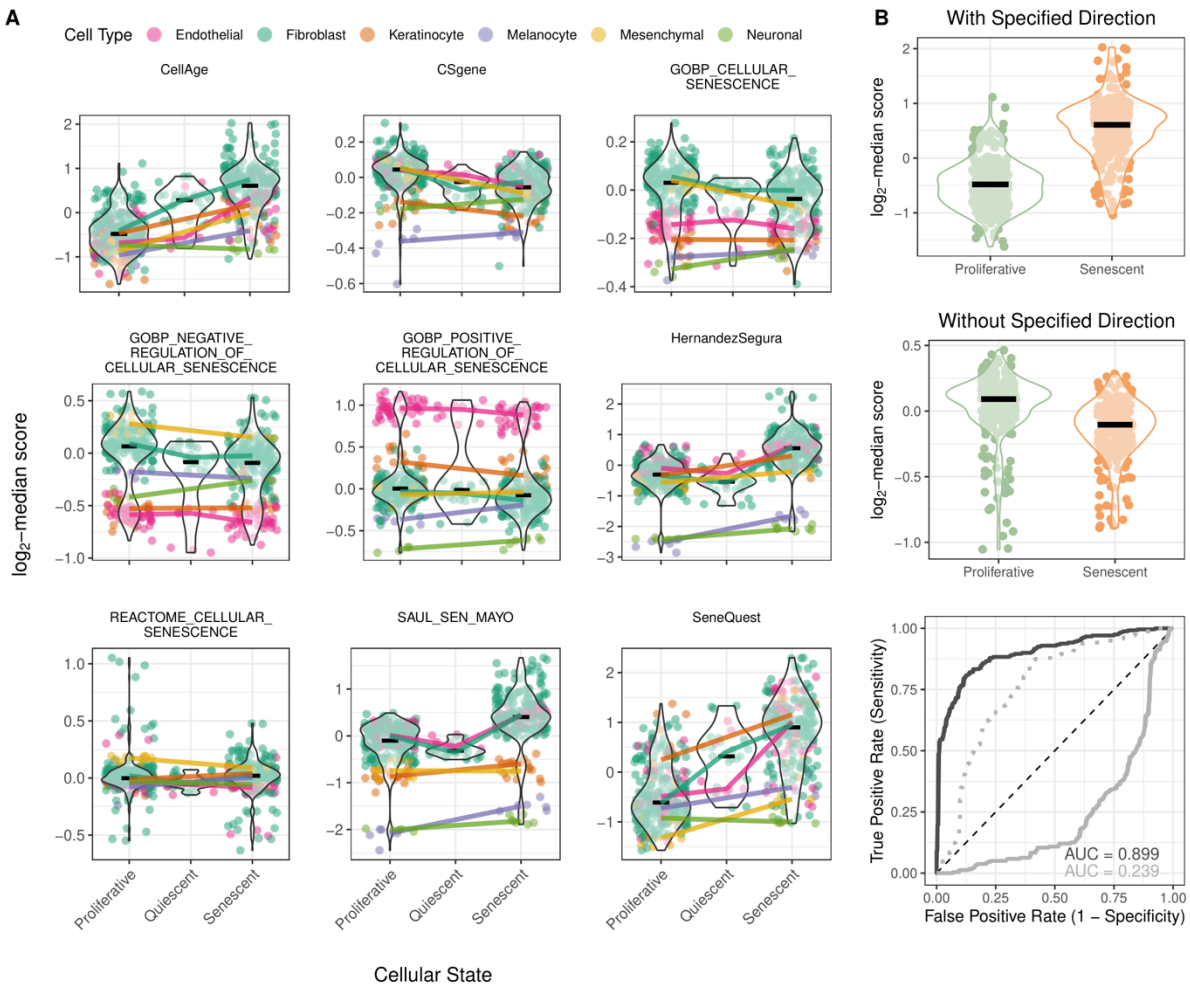


Figure 3 – **Performance of score-based phenotype quantification across gene signatures.**
(A) *Violin plots of distributions of log_2-median scores of samples (dots) for each senescence-associated gene signature per cellular state (proliferative, quiescent, senescent), coloured by cell type. Median scores per cellular state for each cell type are connected by coloured lines.* ***(B)*** *Violin plots of distributions of log_2-median scores of proliferative and senescent samples (green and orange dots, respectively) for the CellAge signature, with (top) and without (middle) the direction of gene expression change upon senescence considered. Bottom panel: Receiver Operating Characteristic (ROC) curves of the log_2-median score of the CellAge signature to mark senescent in contrast with proliferative samples, with (black) and without (grey; dotted line representing the mirrored ROC) the direction of gene expression change upon senescence considered. The respective areas-under-curve (AUC) are displayed. The loss of directional information can lead not only to a reduction in classification performance but also to misinterpretations.*

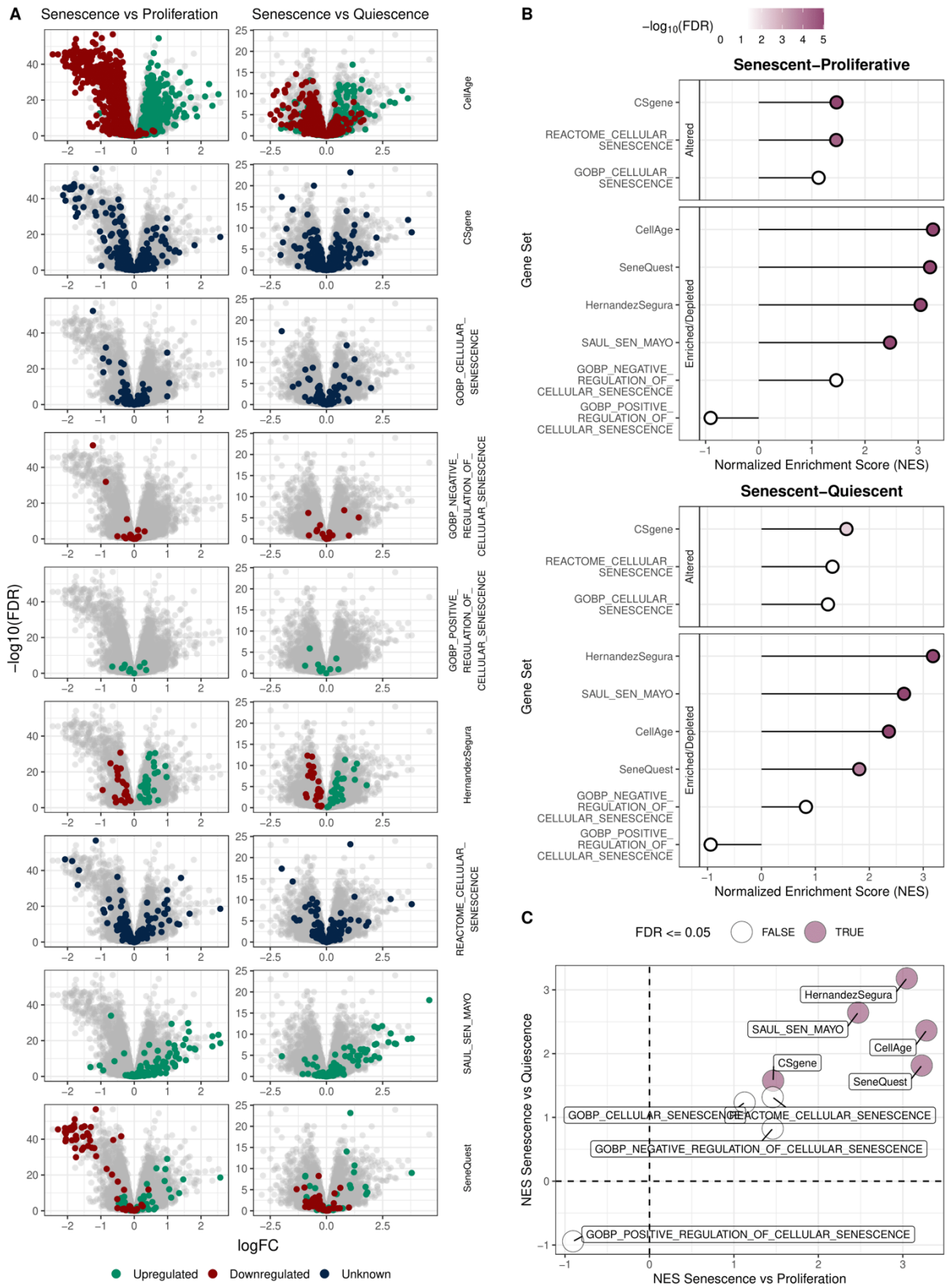


Figure 4 – **Performance of enrichment-based phenotype quantification method across gene signatures.**
(A) *Volcano plots depicting differential gene expression results for the comparisons Senescent vs Proliferative (first column) and Senescent vs Quiescent (second column), across the nine gene signatures under analysis. Each dot represents a gene, with colour indicating its expected behaviour in senescence: red for downregulated, green for upregulated, blue for genes without directional*

annotation, and grey for genes not included in the given signature. ***(B)*** *Gene Set Enrichment Analysis (GSEA) results for both comparisons involving senescence. Each plot summarises the Normalised Enrichment Scores (NES) for gene sets either lacking directional annotation ("Altered") or annotated as "Enriched/Depleted" (see Methods, section 2.1.2). NES is shown on the x-axis, with gene sets on the y-axis, with a colour gradient proportional to the FDR. An FDR of above 0.05 is represented in white.* ***(C)*** *Scatter plot showing the NES from GSEA for each gene signature in two contrasts: Senescent vs. Proliferative (x-axis) and Senescent vs. Quiescent (y-axis). Purple dots represent signatures significantly enriched in both contrasts (FDR <= 0.05), while white dots indicate non-significance in at least one of the contrasts.*

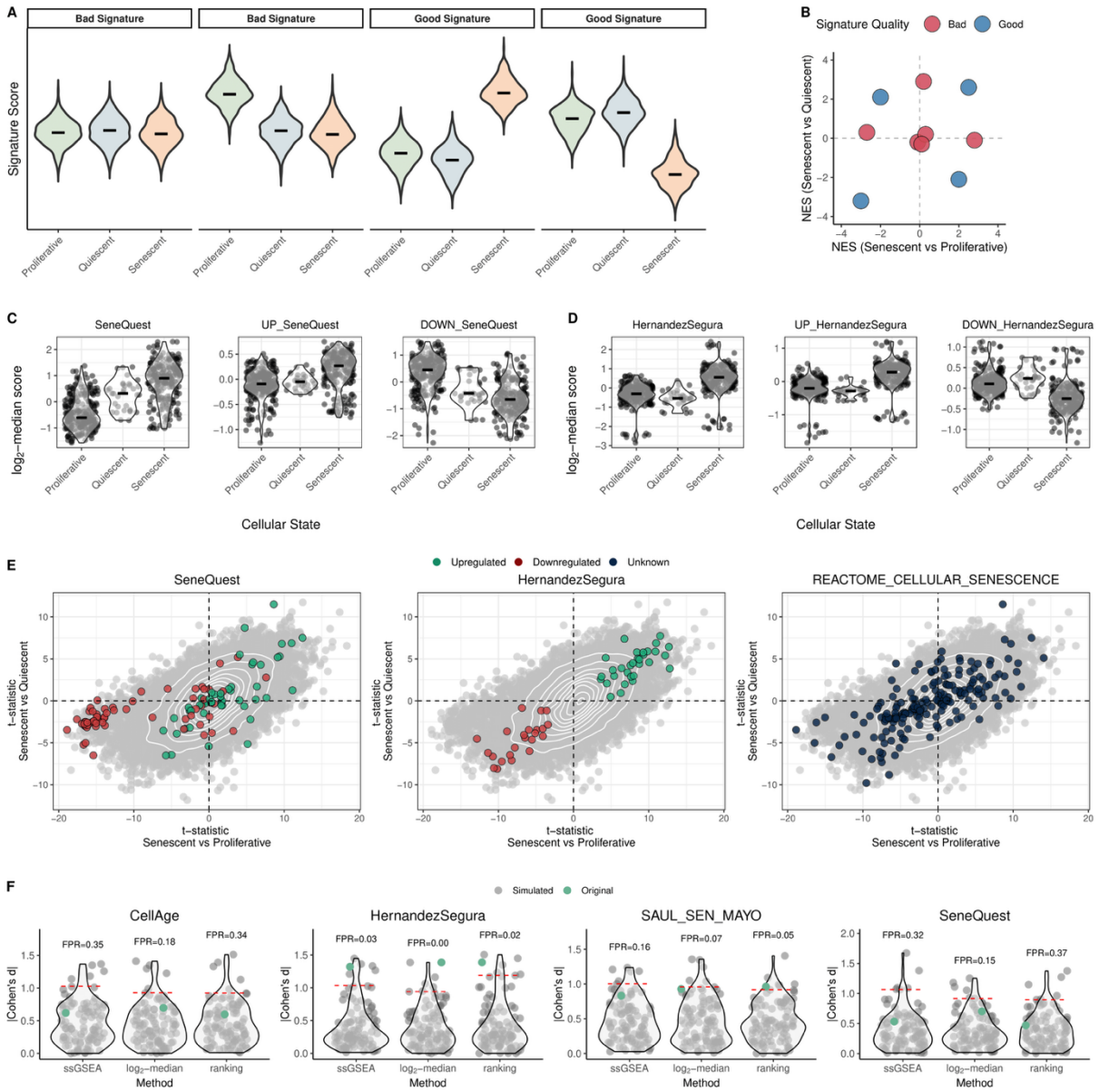


Figure 5 – **Comparative performance of senescence signatures**.
(A) *Example data illustrating the behaviour of good and bad gene signatures using score-based methods. A good signature produces a score that separates senescent from not only proliferative, but also quiescent samples. Poor signatures show overlapping distributions between senescent and other conditions.* ***(B)*** *Scatter plot of GSEA NES for mock signatures. A good signature shows strong and consistent enrichment (positive or negative NES) in both contrasts (Senescent vs. Proliferative and Senescent vs. Quiescent), falling far from the NES = 0 axes. Poor signatures show weak or inconsistent enrichment across contrasts. Red and blue colours respectively indicate Bad and Good classifications based on NES strength.* ***(C)*** *Distribution of log_2-median scores for CellAge and its UP- and DOWN-regulated subsets across proliferative, quiescent and senescent samples. The upregulated subset discriminates senescence from quiescence better, while the downregulated genes show less separation, indicating possible dominance by non-specific proliferation-associated genes. White-shaded density violins are overlaid to aid visualisation of distributions.* ***(D)*** *Distribution of log_2-median scores for HernandezSegura and its up- and downregulated subsets across proliferative, quiescent and senescent samples, with both UP and DOWN compartments showing good discriminatory power between senescent and proliferative/quiescent samples.* ***(E)*** *Scatter plots of t-statistics of differential gene expression for the Senescent vs. Proliferative (x-axis) and Senescent vs. Quiescent (y-axis) comparisons. Each grey dot represents a gene, with signature genes highlighted for SeneQuest, HernandezSegura and REACTOME_CELLULAR_SENESCENCE (left to*

right). Colour indicates the expected behaviour of the genes in the set upon senescence: red for downregulated, green for upregulated, dark blue for genes without directional annotation, and grey for genes not included in the given signature. Density contour lines in white. ***(F)*** *Violin plots showing the null distribution of absolute Cohen's d effect sizes across the three scoring methods for 100 randomly simulated signatures. Each grey dot represents a simulated signature of the same size in number of genes as the original, while green represents the original tested signature. The dashed red line indicates the 95th percentile of the null distribution. False positive rates (FPRs) are shown above each violin.*

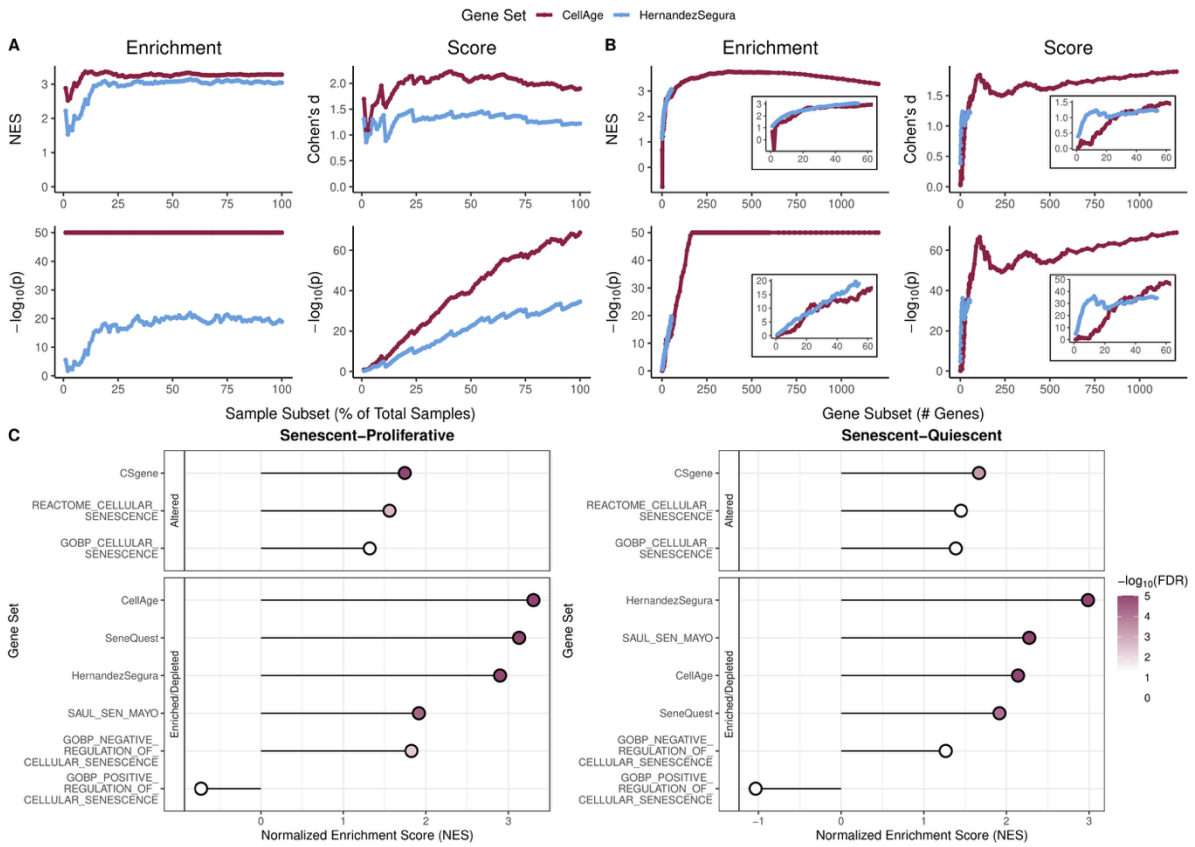


Figure 6 – **Performance of score- and enrichment-based approaches under varying sample and gene set sizes.**
(A) *Effect sizes (top row) and p-values (bottom row), for GSEA (left "Enrichment" column) and* log_2*-median score (right "Score" column), for pairwise comparisons between proliferative and senescent samples over random subsets of samples of increasing size (2-100% of 240 proliferative and 272 senescent samples, resampling for each percentage).* ***(B)*** *Effect sizes and p-values for the same methods in (A) over random gene subsets of increasing size from each representative gene set (one-gene increments up to 55, the Hernandez–Segura size, followed by 10-gene (starting at 69 genes) and then 30-gene (starting at 599 genes) steps up to the CellAge size, resampling genes for each number). Both panels show representative gene signatures: "small" (55 genes) HernandezSegura (blue) and "large" (1259 genes) CellAge (bordeaux). "Zoom-in" insets show the same results for up to 60-gene subsets.* ***(C)*** *Lollipop plots of GSEA results for all nine gene signatures using a subset of the data consisting only of proliferative, quiescent and radiation-induced senescent samples - mirroring the format of Figure 4B. In lollipop plots, NES are shown for both altered and enriched/depleted gene sets, with NES on the x-axis and gene sets on the y-axis, coloured by the FDR.*

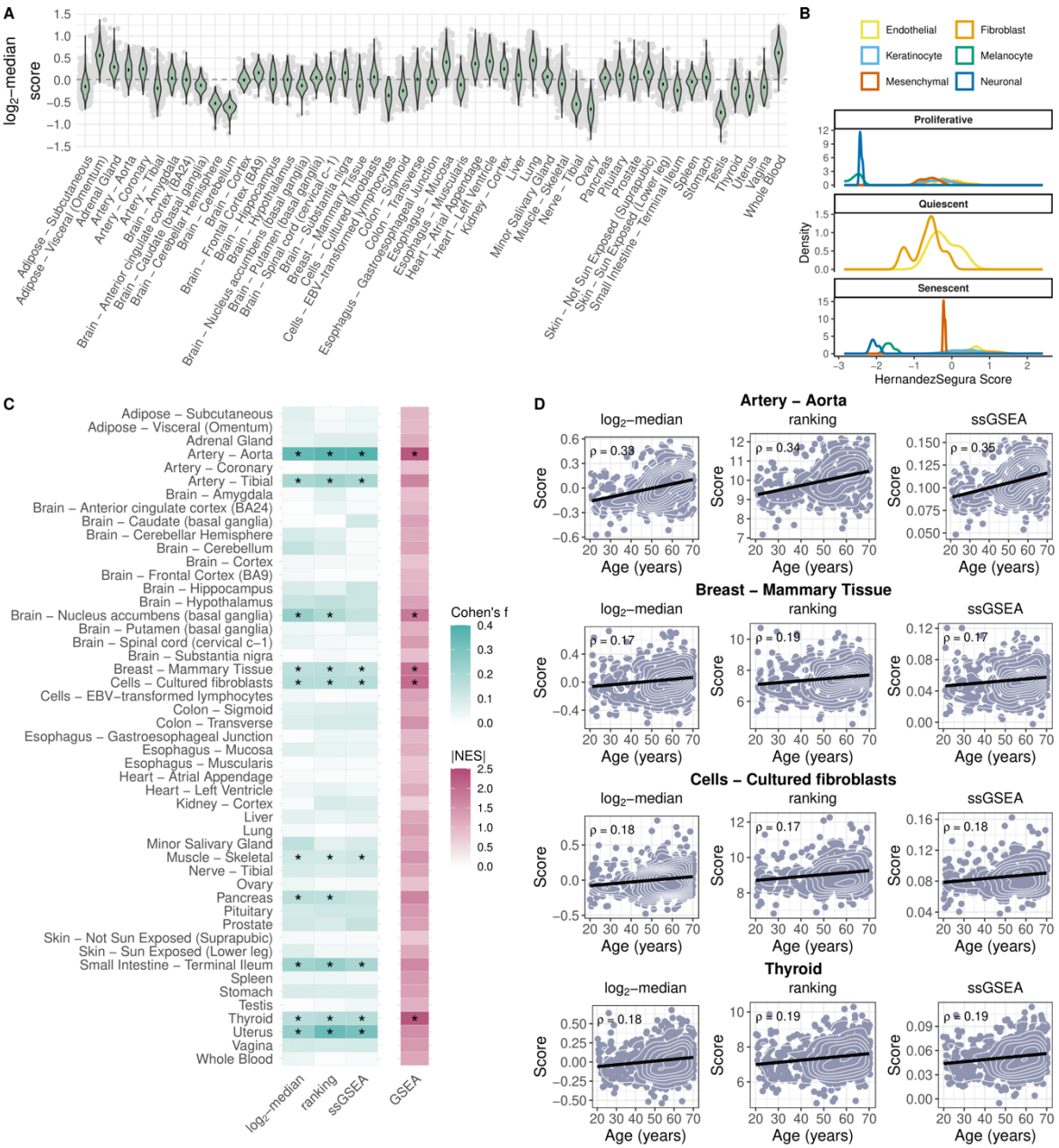
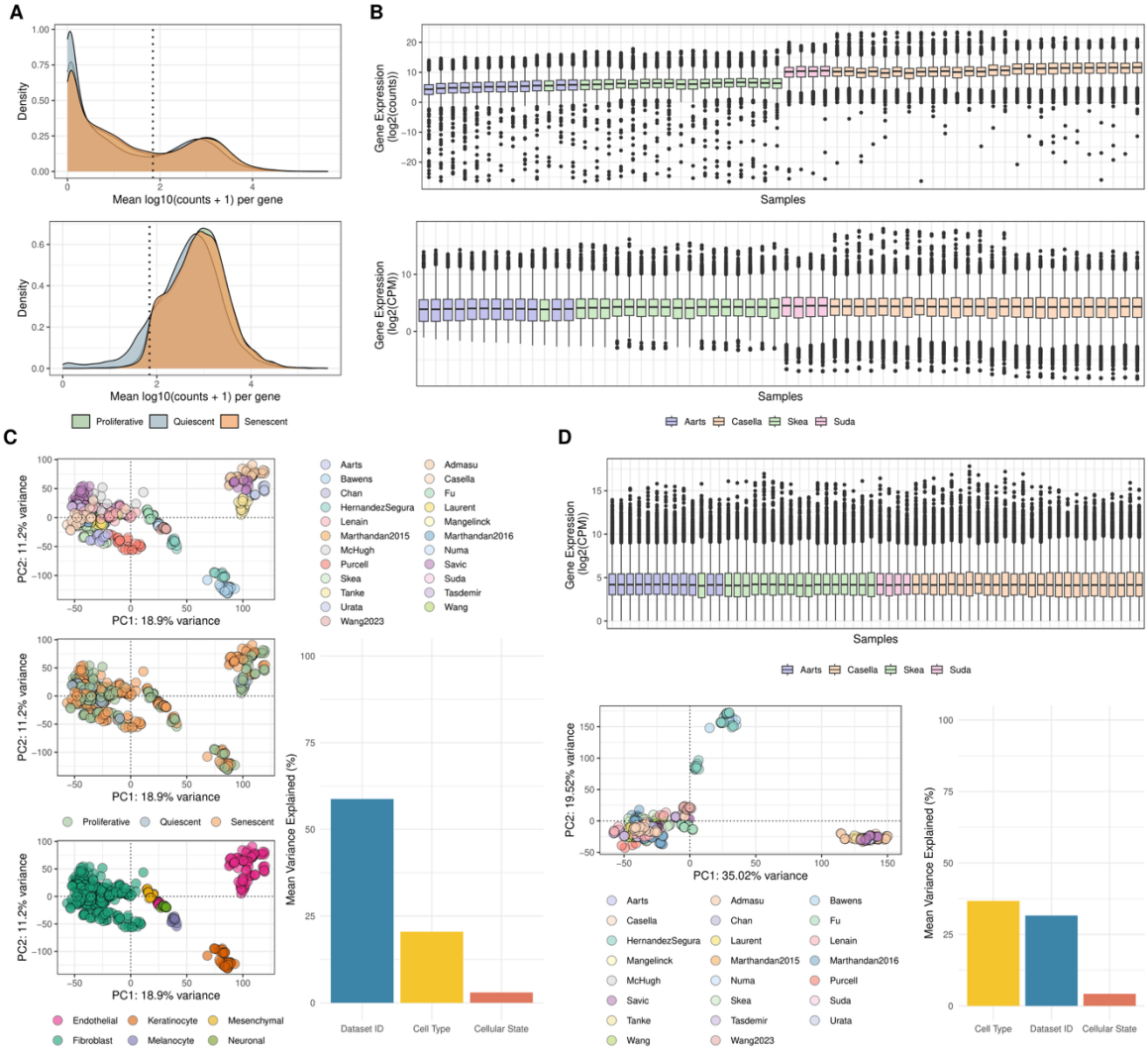
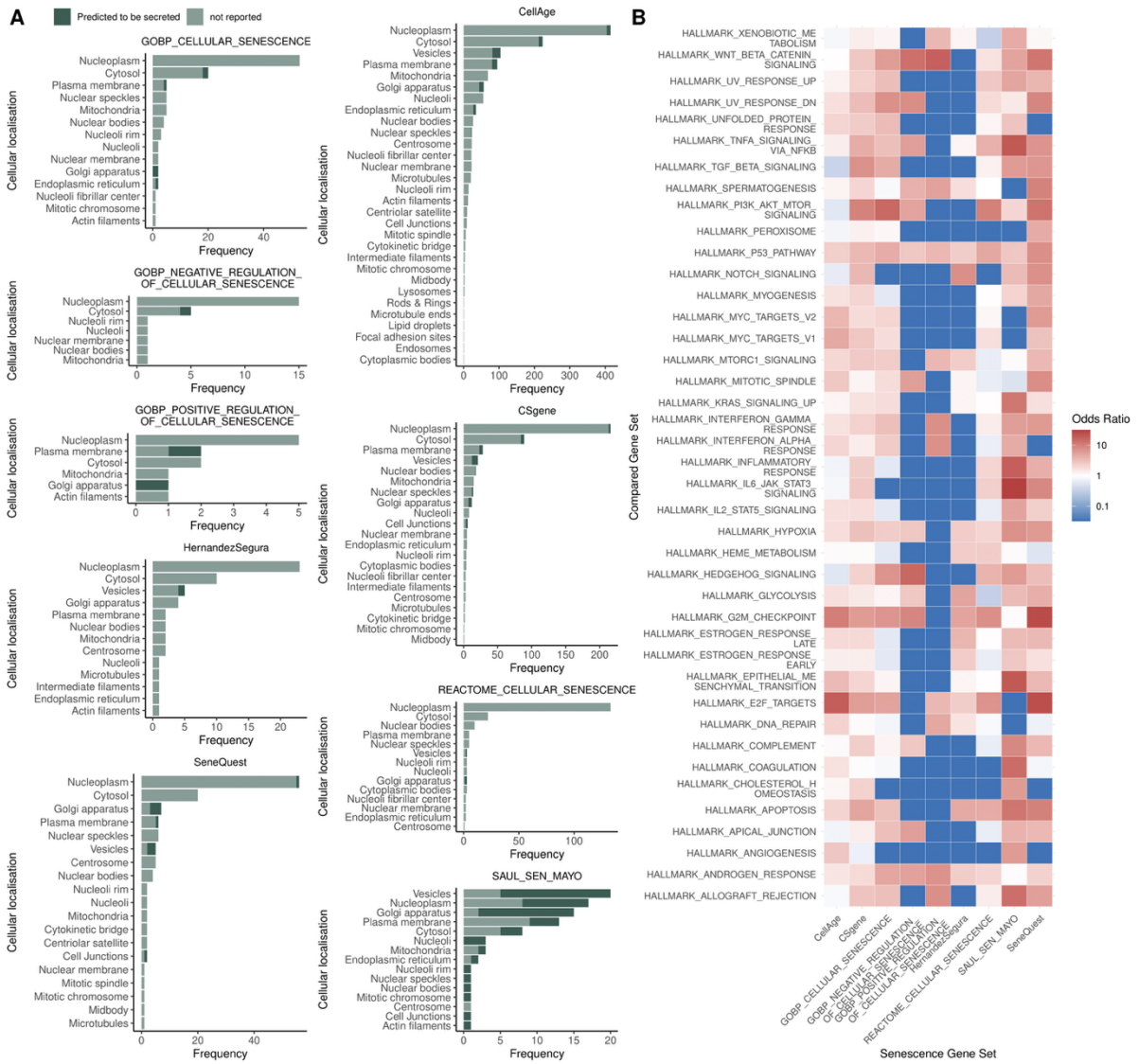


Figure 7 ***– Influence of age on HernandezSegura senescence signature scoring and enrichment across human tissues.***
(A) *Violin plots of distribution of senescence scores across 49 tissues from GTEx v8, using the log_2-median method and the HernandezSegura signature as an example. Each grey dot represents a sample, and the tissues are ordered alphabetically along the x-axis.* ***(B)*** *Distributions of senescence scores in proliferative, senescent and quiescent samples from in vitro datasets show that, for certain cell types, some proliferative samples display higher scores than senescent samples of other cell types. This highlights the strong cell-type specificity and variable baseline (i.e., in proliferative samples) levels of senescence signature scores.* ***(C)*** *Heatmap summarising the association between donors' age and senescence scores across GTEx tissues. Rows represent tissues and columns represent scoring-based approaches (log_2-median, ranking and ssGSEA) using Cohen's f as a metric of variable association (white to green), and GSEA-based enrichment analysis (NES, white to bordeaux) using age as the variable. Asterisks denote significant associations (FDR ≤ 0.05). For the score-based approaches, p-values were obtained from 1,000 random permutations of the age variable, whereas for GSEA, the p-value is derived from the method's internal enrichment testing procedure.* ***(D)*** *Scatter plots showing score vs age for the four tissues with significant associations across all four methods in (C). Each row corresponds to one tissue, with plots for the log_2-median,*

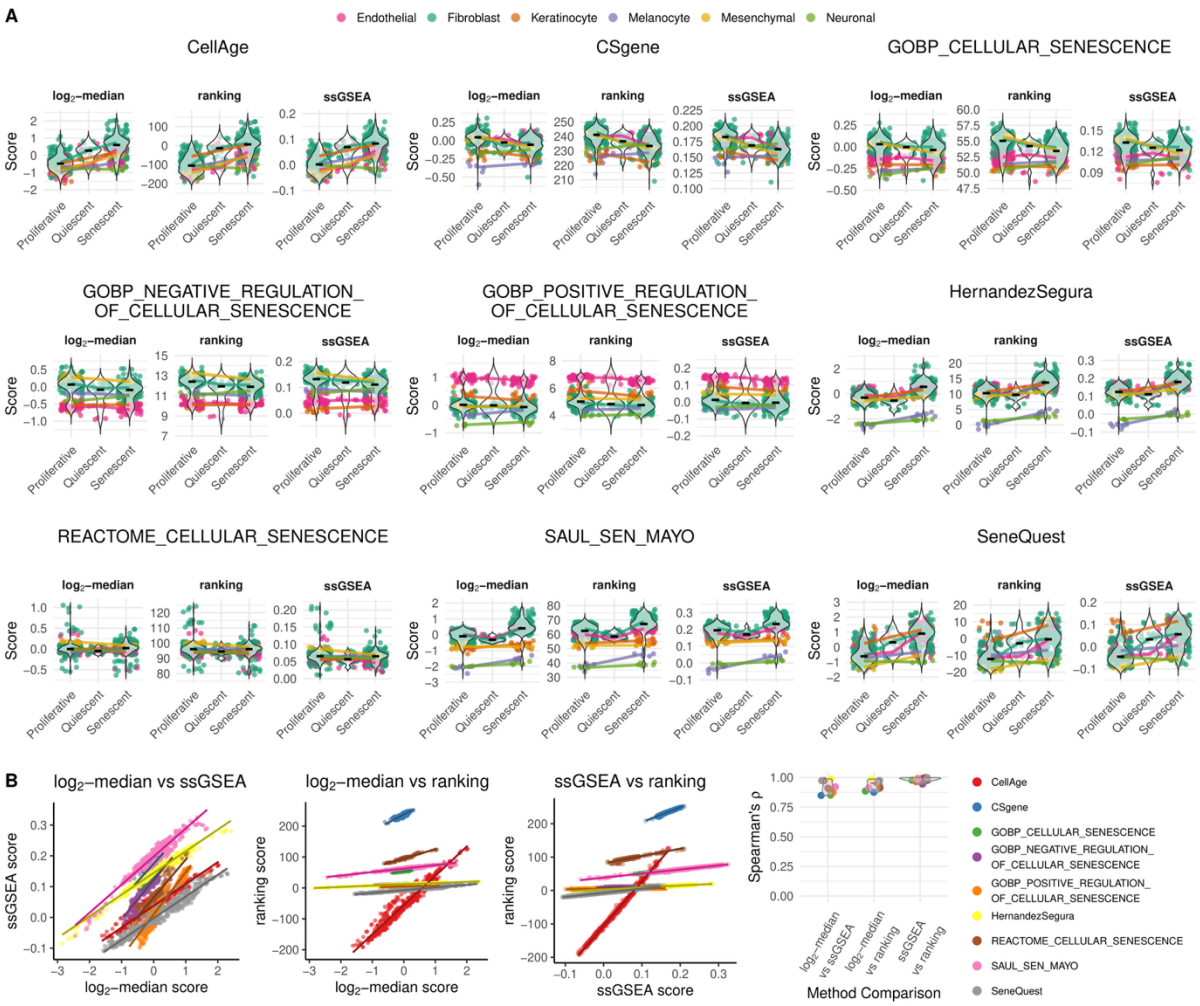
ranking and ssGSEA methods. The age in years (x-axis) is jittered for donor privacy. Black linear regression line and Spearman's correlation coefficient ρ are shown. Density contour lines in white aid visualisation of dot distributions.



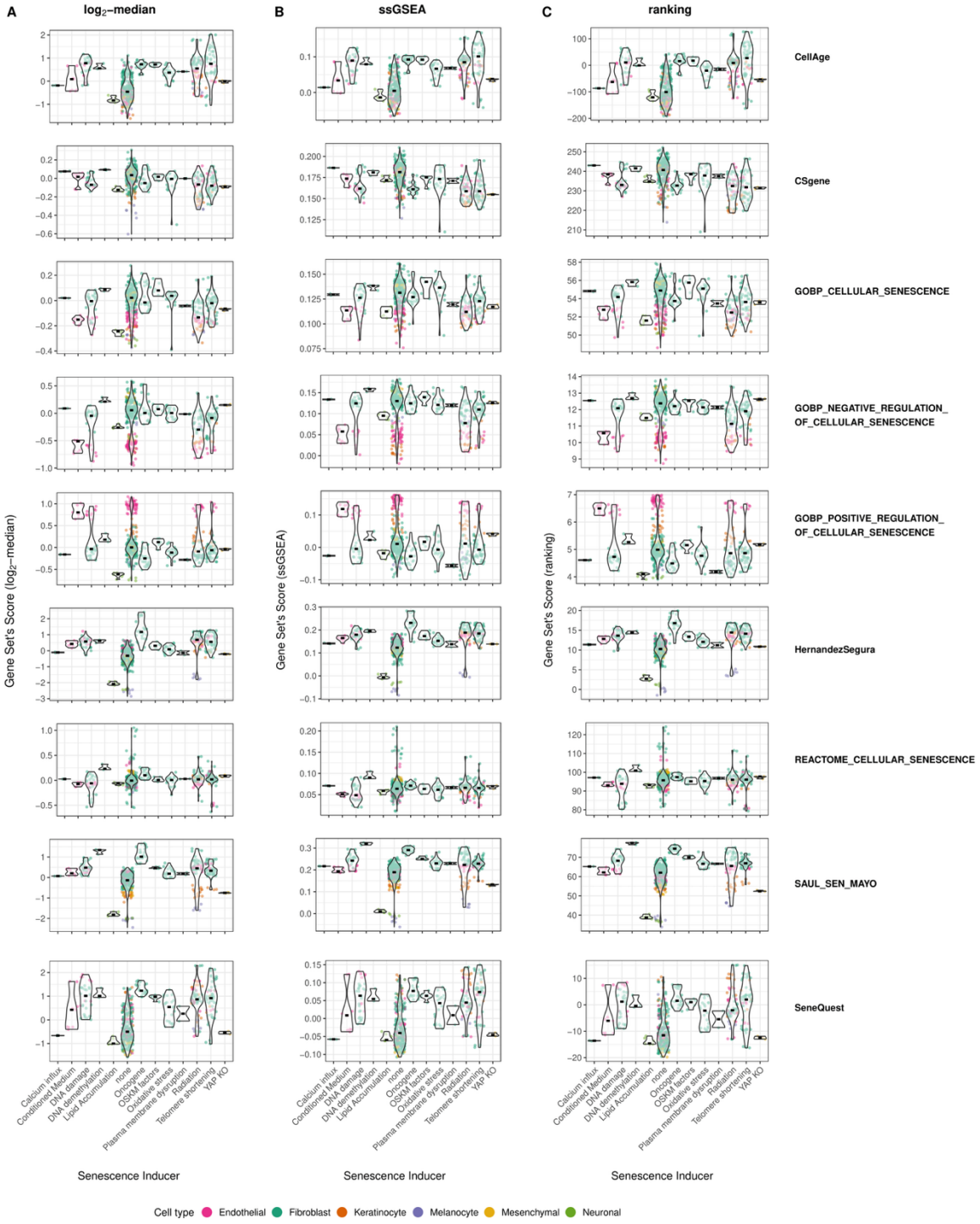
Supplementary Figure 1 – **Pre-processing of senescence datasets.**
A total of 25 RNA-seq datasets comprising 545 samples from six human cell types were combined into one dataset and processed jointly. ***(A)*** *Density plots of distributions of mean log_{10}-transformed raw gene expression, with samples grouped by cellular state, before (top) and after (bottom) filtering. Lowly expressed genes were filtered out by retaining only those with an average of >70 reads in at least one condition (dashed line).* ***(B)*** *Box plots of log_2CPM gene expression distributions across the 30 samples with the lowest and highest total read coverage. The top plot shows raw, filtered data, while the bottom shows data after normalisation, resulting in comparable median read counts across samples.* ***(C)*** *Principal Component Analysis (PCA) plot of filtered and normalised gene expression, with samples coloured by dataset ID (top left), cell type (top right), and cellular state (bottom left). (Bottom right) Bar plot of percentages of gene expression variance explained by those variables.* ***(D)*** *(Top) Box plots of log_2CPM gene expression distributions across the same 30 samples as in (B) after batch correction, with no major inter-sample differences detected. (Bottom) The PCA plot of gene expression (left) and the bar plot of percentages of gene expression variance explained by the main variables after batch correction suggest its success, with dataset ID no longer dominating variance.*



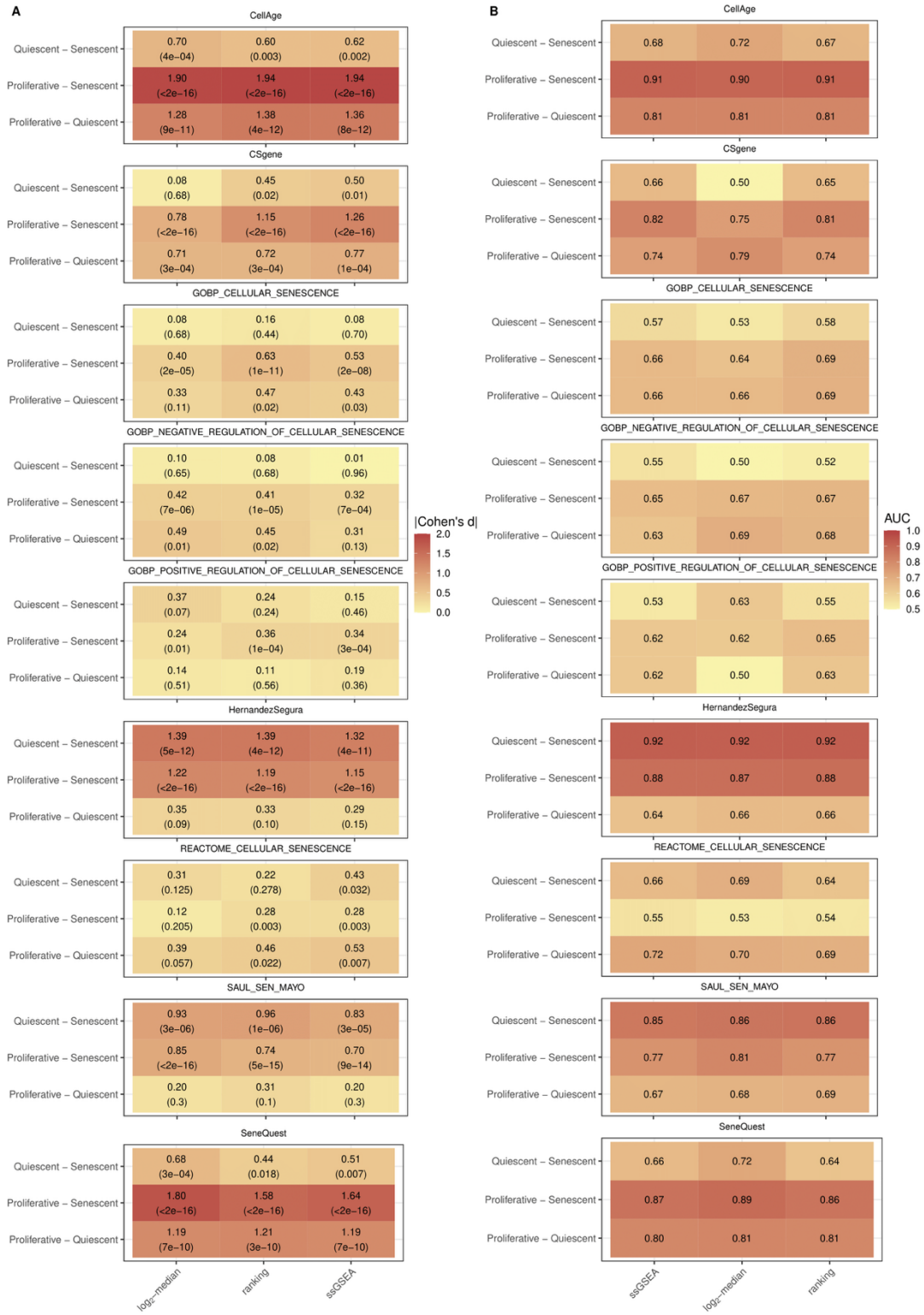
Supplementary Figure 2 – **Subcellular localisation and functional enrichment of senescence gene signatures.**
(A) *Cellular localisation of proteins encoded by genes in each set, based on annotations from the Human Protein Atlas. Bars represent the frequency of genes associated with each subcellular compartment. Genes predicted to be secreted are shown in dark green, while those with no reported extracellular localisation are shown in grey. Genes annotated to multiple compartments counted once per location.* ***(B)*** *Functional enrichment analysis (signature similarity feature of* `markeR`*) of the same gene sets using Hallmarks from MSigDB, chosen as the least redundant and most interpretable collection of biological processes therein. Heatmap of odds ratios (OR) of enrichment of genes from senescence signatures (columns) in Hallmarks (rows). The gene background consisted of all genes annotated in The Human Protein Atlas.*



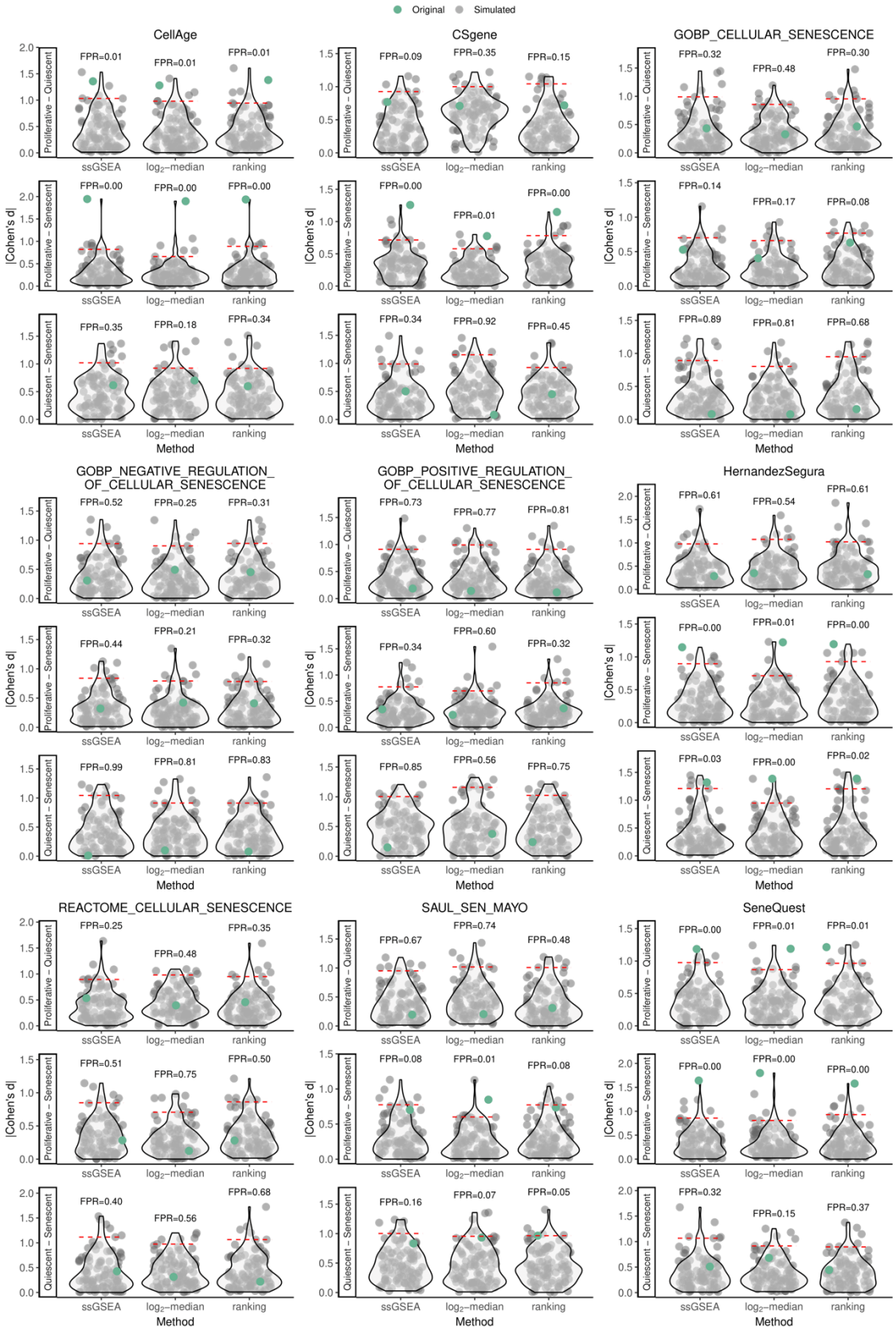
Supplementary Figure 3 – **Performance and pairwise correlation of score-based phenotype quantification methods across gene signatures.**
(A) *Violin plots showing the distribution of senescence scores for each gene signature, computed using three scoring methods: log₂-median (left), ranking (middle), and ssGSEA (right). Each plot depicts the score distribution across sample types (proliferative, quiescent, senescent), coloured by cell type. Median scores per condition for each cell type are connected by coloured lines.* ***(B)*** *Pairwise comparisons of senescence scores across methods. The first three panels show scatter plots of log₂-median vs. ssGSEA, log₂-median vs. ranking, and ssGSEA vs. ranking. Dots (i.e., samples) are coloured by the gene set scored therein, with an associated regression line. The last panel shows the distribution of Spearman's correlation coefficient ρ across all method pairs, confirming consistently strong correlations.*



Supplementary Figure 4 **– Signature scores across senescence-inducing stressors highlight context specificity.**
Violin plots showing the distribution of signature scores across different senescence-inducing stressors, calculated using three scoring methods: ***(A)*** *log$_2$-median,* ***(B)*** *ssGSEA, and* ***(C)*** *ranking. Each plot represents one gene signature, with stressor type on the x-axis and score on the y-axis. Data are coloured by cell type. These plots demonstrate that some gene signatures display stressor-specific sensitivity. For example, the HernandezSegura and SAUL_SEN_MAYO signatures are particularly effective in identifying oncogene-induced senescence, and CellAge shows stronger performance in telomere shortening and radiation-induced. In contrast, some signatures show weaker or inconsistent patterns across stressors, highlighting potential limitations in their robustness.*



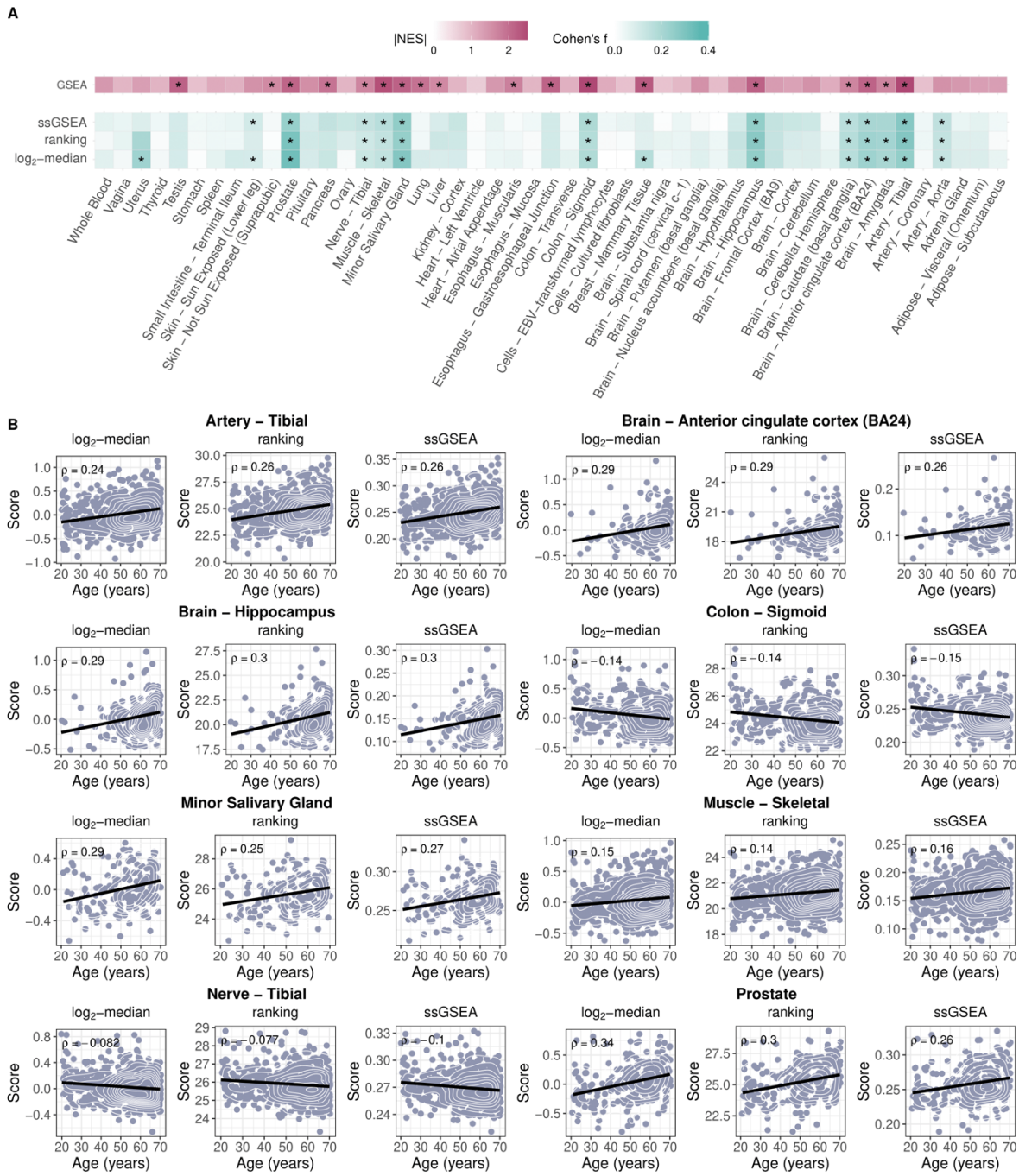
Supplementary Figure 5 - **Classification metrics of score-based phenotype quantification methods across gene signatures.**
(A) *Heatmap of absolute effect sizes (Cohen's d) for each gene signature, with rows representing pairwise contrasts between senescence states and columns representing the scoring methods. A robust signature is expected to show high and consistent effect sizes across methods, particularly for the senescence vs. proliferative and senescence vs. quiescence contrasts. In brackets, the statistical significance given by p-values of the pairwise contrasts' t-tests.* ***(B)*** *Heatmap of AUC values from ROC curve analyses for the same pairwise contrasts and methods, providing a classification-based evaluation of signature performance.*



Supplementary Figure 6 – ***False positive rates (FPRs) across senescence gene signatures using simulated null distributions***

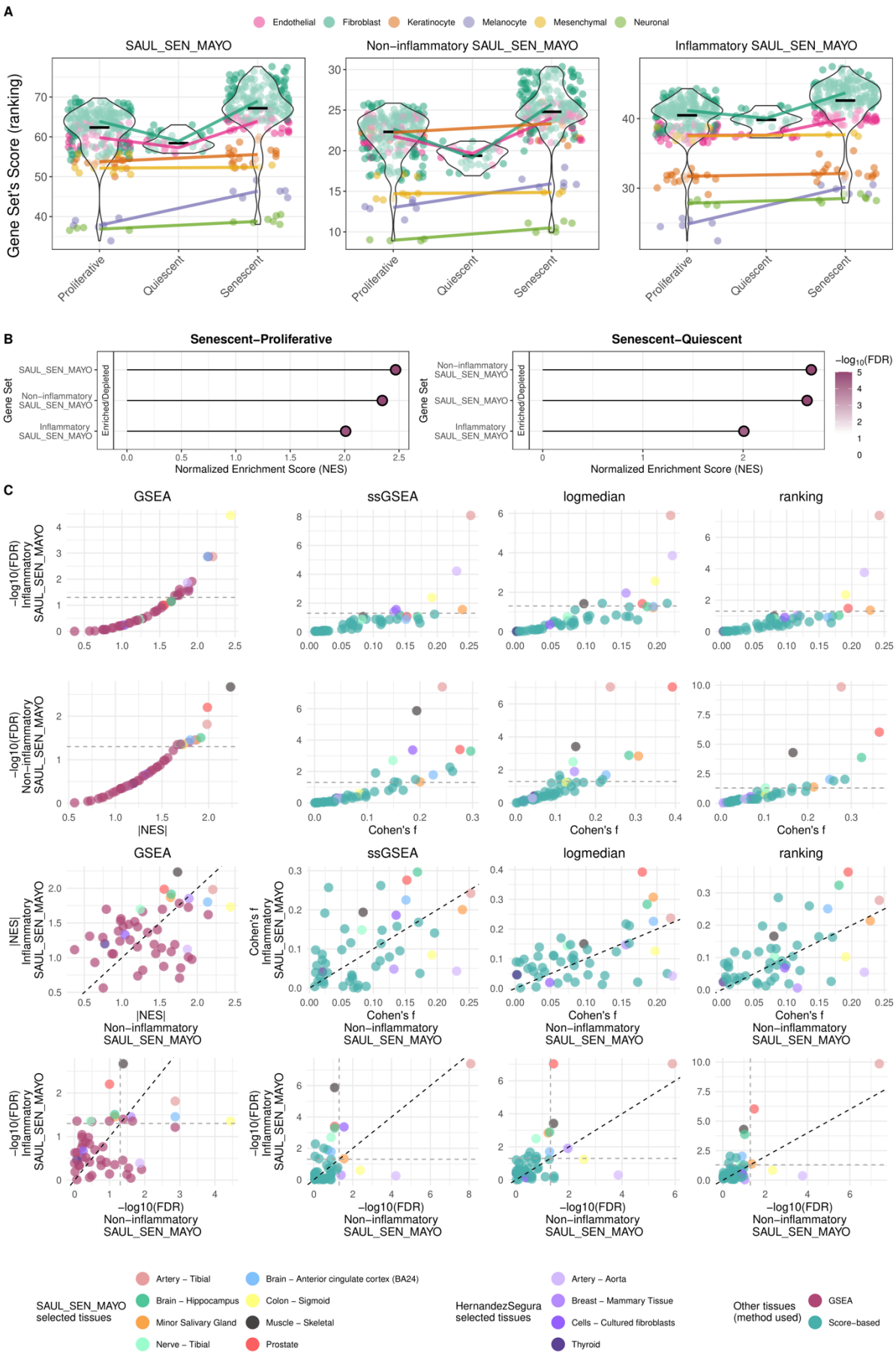
The figure is organised into nine groups, with one group representing each gene signature tested. Within each group, three violin plots illustrate the following contrasts: Senescent versus Quiescent,

Senescent versus Proliferative, and Proliferative versus Quiescent. For each contrast, the x-axis shows the various scoring methods used. The violin plots show the distribution of absolute Cohen's d effect sizes from 100 simulated signatures (grey dots) that are matched in size to the real signature, preserving the original signature's proportion of up- and down-regulated genes, if applicable. Green dots represent the effect size for the actual signature. The dashed red line marks the 95th percentile of the null distribution. False positive rates (FPRs), displayed above each plot, show the proportion of random signatures that outperformed the real signature in each case.



Supplementary Figure 7 – **Age-associated senescence scoring and enrichment across GTEx tissues using the SAUL_SEN_MAYO signature.**

(A) *Heatmap showing the association between donors' age and senescence scores across GTEx tissues, using the SAUL_SEN_MAYO signature. Columns correspond to tissues, and rows represent scoring-based approaches (log_2-median, ranking and ssGSEA) using Cohen's f as a metric of variable association (white to green), and GSEA-based enrichment analysis (NES, white to bordeaux). Asterisks denote significant associations (FDR ≤ 0.05). For the score-based approaches, p-values were obtained from 1,000 random permutations of the age variable, whereas for GSEA, the p-value is derived from the method's internal enrichment testing procedure.* ***(B)*** *Scatter plots showing score vs age for the eight tissues with significant associations across all four methods in (A). For each tissue, plots for the log_2-median, ranking and ssGSEA methods are presented. The age in years (x-axis) is jittered for donor privacy. Black linear regression line and Spearman's correlation coefficient ρ are shown. Density contour lines in white aid visualisation of dot distribution.*


A
Endothelial
Fibroblast
Keratinocyte
Melanocyte
Mesenchymal
Neuronal
SAUL_SEN_MAYO
Non−inflammatory SAUL_SEN_MAYO
Inflammatory SAUL_SEN_MAYO
Gene Set's Score (ranking)
Proliferative
Quiescent
Senescent
B
Senescent−Proliferative
Senescent−Quiescent
Gene Set
Enriched/Depleted
Normalized Enrichment Score (NES)
−log10(FDR)
C
GSEA
ssGSEA
logmedian
ranking
−log10(FDR) Inflammatory SAUL_SEN_MAYO
−log10(FDR) Non−inflammatory SAUL_SEN_MAYO
|NES|
Cohen's f
|NES| Inflammatory SAUL_SEN_MAYO
Cohen's f Inflammatory SAUL_SEN_MAYO
|NES| Non−inflammatory SAUL_SEN_MAYO
Cohen's f Non−inflammatory SAUL_SEN_MAYO
−log10(FDR) Non−inflammatory SAUL_SEN_MAYO
SAUL_SEN_MAYO selected tissues
Artery − Tibial
Brain − Hippocampus
Minor Salivary Gland
Nerve − Tibial
Brain − Anterior cingulate cortex (BA24)
Colon − Sigmoid
Muscle − Skeletal
Prostate
HernandezSegura selected tissues
Artery − Aorta
Breast − Mammary Tissue
Cells − Cultured fibroblasts
Thyroid
Other tissues (method used)
GSEA
Score−based

1202

Supplementary Figure 8 – **Comparison of inflammatory and non-inflammatory components of the SAUL_SEN_MAYO gene set.**

(A) *Violin plots showing the distribution of scores calculated using the ranking method for three gene sets: the original SAUL_SEN_MAYO signature, the non-inflammatory subset (genes not overlapping with inflammatory Hallmark pathways), and the inflammatory subset (genes overlapping with these pathways). The subsets were defined by intersecting the original gene set with Hallmark gene sets of inflammatory response, TNF-α signalling via NF-κB, interferon gamma and alpha responses, and IL6/JAK/STAT3 signalling. Each plot displays score distributions across sample types (proliferative, quiescent, senescent), coloured by cell type. Median scores per condition and cell type are connected by coloured lines.* ***(B)*** *Gene Set Enrichment Analysis results shown as lollipop plots summarising NES for the three gene sets in the in vitro senescence datasets. The colour gradient represents FDRs, with more intense colours indicating greater statistical significance.* ***(C)*** *Volcano plots (first two rows, showing statistical significance vs. effect size) and correlation plots of effect size (third row, |NES| for GSEA, Cohen's f for score-based) and statistical significance (fourth row) comparing the inflammatory SAUL_SEN_MAYO versus the non-inflammatory SAUL_SEN_MAYO gene subsets, for the four methods used in markeR (columns: GSEA, ssGSEA, log_2-median, ranking). Coloured dots indicate tissues of interest: tissues scored for SAUL_SEN_MAYO highlighted in the main text, tissues scored for Hernandez-Segura in purple shades, and remaining tissues coloured by the method used (teal for score-based, bordeaux for GSEA). For comparisons involving p-values, a dashed grey line marks the 0.05 significance threshold. Black dashed line indicates y=x.*