Six drivers of aging identified among genes differentially expressed with age

Ariella Coler-Reilly[1,2,6], Zachary Pincus[3,4,] Erica L. Scheller[1,2,3,5], and Roberto Civitelli[1,2,5]

[1]Division of Bone and Mineral Diseases, Musculoskeletal Research Center

Departments of [2]Medicine, [3]Developmental Biology, [4]Genetics, [5]Cell Biology and Physiology;

[6]Medical Scientist Training Program

Washington University School of Medicine, St. Louis, MO, USA

Running title: Genetic Drivers of Aging

Address all correspondence and reprint requests to:

Ariella Coler-Reilly
Division of Bone and Mineral Diseases
Washington University in St. Louis
660 South Euclid – Campus Box 8301
St. Louis, MO 63110
USA
E-mail: acoler-reilly@wustl.edu

Conflict of Interest Statement: The authors have declared that no conflicts of interest exist.

Abstract

Many studies have compared gene expression in young and old samples to gain insights on aging, the primary risk factor for most major chronic diseases. However, these studies only describe associations, failing to distinguish drivers of aging from compensatory geroprotective responses and incidental downstream effects. Here, we introduce a workflow to characterize the causal effects of differentially expressed genes on lifespan. First, we performed a meta-analysis of 25 gene expression datasets comprising samples of various tissues from healthy, untreated adult mammals (humans, dogs, and rodents) at two distinct ages. We ranked each gene according to the number of distinct datasets in which the gene was differentially expressed with age in a consistent direction. The top age-upregulated genes were TMEM176A, EFEMP1, CP, and HLA-A; the top age-downregulated genes were CA4, SIAH, SPARC, and UQCR10. Second, the effects of the top ranked genes on lifespan were measured by applying post-developmental RNA interference of the corresponding ortholog in the nematode C. elegans (two trials, with roughly 100 animals per genotype per trial). Out of 10 age-upregulated and 9 age-downregulated genes that were tested, two age-upregulated genes (*csp-3*/CASP1 and *spch-2*/RSRC1) and four age-downregulated genes (*C42C1.8*/DIRC2, *ost-1*/SPARC, *fzy-1*/CDC20, and *cah-3*/CA4) produced significant and reproducible lifespan extension. Notably, the data do not suggest that the direction of differential expression with age is predictive of the effect on lifespan. Our study provides novel insight into the relationship between differential gene expression and aging phenotypes, pilots an unbiased workflow that can be easily repeated and expanded, and pinpoints six genes with evolutionarily conserved, causal roles in the aging process for further study.

Graphical Abstract

CDC20, SPARC, RSRC1, DIRC2, CASP1, CA4
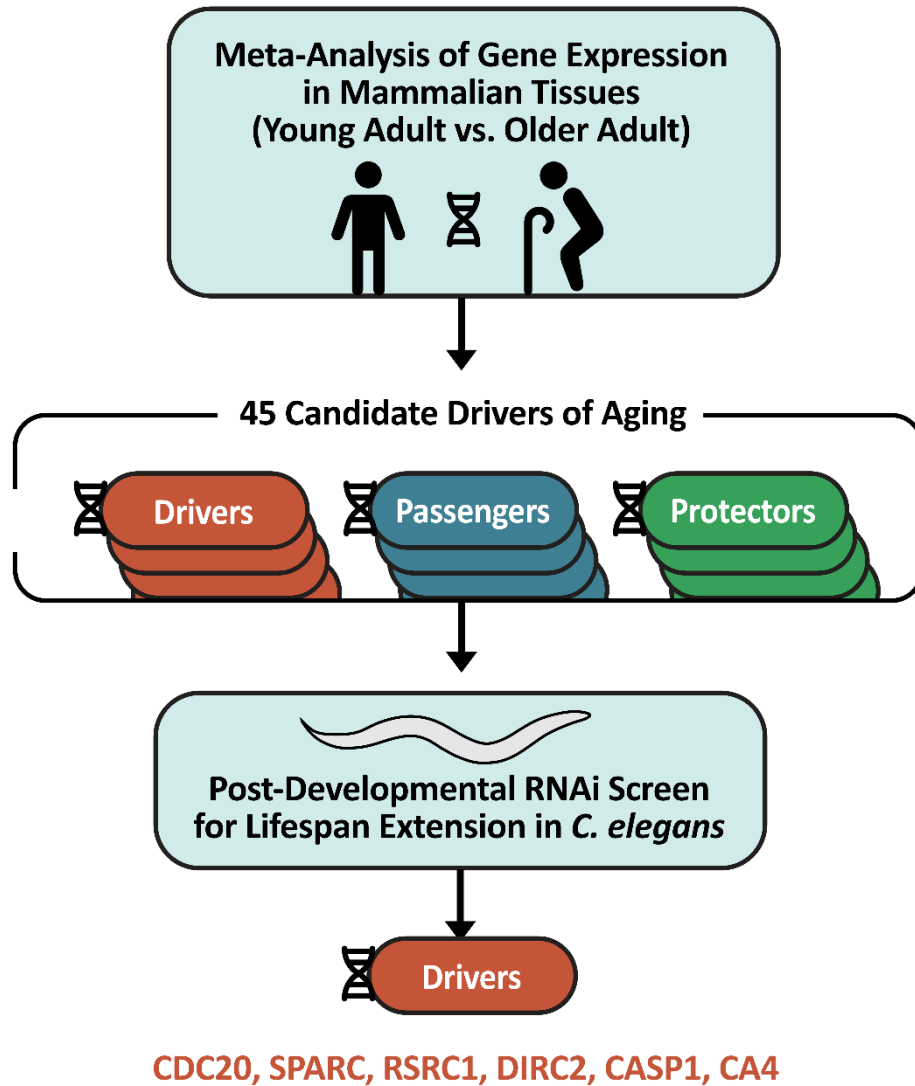
Introduction

Advanced age is the primary risk factor for most chronic diseases, and as our population ages, the social and economic burden of chronic disease continues to grow year by year [1]. The field of geroscience has emerged to study the mechanisms underlying aging itself and develop strategies to combat age-related decline, or senescence, at the source. According to the widely accepted evolutionary theory of aging, senescence is pervasive because there is negligible selection pressure during the post-reproductive period, a phenomenon known as the "selection shadow" [2, 3]. It is therefore crucial to study the role of genetic variants and gene expression changes in aging in order to uncover potentially advantageous adjustments that have been masked by the selection shadow.

Specialized approaches are needed to detect age-related gene expression signals, which are often subtle and widespread rather than striking and targeted. Indeed, as noted in the *Handbook of the Biology of Aging*, differentially expressed genes (DEGs) with the largest fold changes are frequently found to be downstream targets rather than upstream regulators [4]. Moreover, age-related phenomena such as transcriptional drift create reproducible expression patterns that are nonetheless stochastic and unregulated, further obscuring meaningful signals [5–7]. As stochastic signals are unlikely to replicate across species and tissues, and since frequency is more meaningful than fold change, drivers of aging may ideally be identified using a multi-species, multi-tissue meta-analysis using the value-counting method. This strategy, first pioneered by Magalhães et al. in 2009 [8] and further developed by Palmer et al. in 2021 [9], has been used to catalog numerous individual DEGs as well as broader patterns in functional enrichment and pathway analysis. However, translation of such findings into actionable therapeutic strategies is challenging. Any upregulated gene presumed to be a driver of aging could just as easily be a

4

compensatory geroprotective response or an unimportant downstream effect, often called a "passenger" to contrast with the aforementioned "driver" [5, 10]. In other words, as the age-old adage warns, correlation does not necessarily equal causation.

Functionally evaluating genes related to aging also presents special challenges. Stable cell lines cannot be used to study aging *in vitro* because of the immortal nature of such lines. *In vivo* models are more useful, but require time and resources to age the animals and monitor them until their natural death. In mammals, this can entail years of labor, and this is surely one reason why the short-lived nematode *C. elegans* has been such a popular model organism in geroscience for decades [11]. While nematodes are only distant relatives of humans, they share remarkably similar features of post-reproductive senescence, including sarcopenia and reduced motility, deteriorated learning and memory, and weakened immunity [12–14]. In contrast to humans or any mammal, these age-dependent changes occur on a compressed time-scale of days rather than years, with an average lifespan of only a few weeks. On a genetic level, orthologs of roughly half of all human genes have been identified, and tools have been developed to rapidly, easily, and inexpensively manipulate those genes in *C. elegans* worms, making them an ideal choice for genetic screens [15]. However, it is difficult to substantiate findings in *C. elegans* as relevant to human physiology without any means of contextualizing the results in mammalian systems.

Here, we introduce a workflow to unify two separate approaches, analysis of mammalian DEGs and genetic screening in *C. elegans*, leveraging the strengths of each to mitigate their respective weaknesses. We first performed a meta-analysis comparing gene expression in young adults vs older adults using publicly available datasets comprising samples of various tissues from healthy, untreated mammals (humans, dogs, and rodents). DEGs were ranked by the consistency of differential expression with age across the largest number of datasets. The highest ranking DEGs

with known orthologs in *C. elegans* were then tested using post-developmental RNAi lifespan assays. Ultimately, we identified six genes with evolutionarily conserved, causal effects on aging that may be prioritized for future mechanistic studies. In addition, we have established a proof of principle for a unified approach for studying evolutionarily conserved mechanisms of aging.

Material and Methods

Meta-Analysis Dataset Selection

This meta-analysis was designed as a simple and scalable approach that is nonetheless highly capable of identifying a collection of genes consistently associated with mammalian aging. The intention was to extricate subtle but meaningful age-related signals from a background of transcriptional drift and stochastic changes that are unlikely to replicate across species, tissues, and experimental platforms. Thus, as the inclusion criteria and exclusion criteria detailed in Table 1 show, any datasets comprising samples from mammals representative of typical individuals at both young adult and older adult time-points were eligible for inclusion.

Gene expression data were obtained from publicly available datasets hosted on the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) repository [16]. The filters "*organism: mammal*" and "*subset variable type: age*" were used to identify roughly 200 curated datasets as candidates for the present study as of March 2021. These datasets were then manually inspected using the inclusion and exclusion criteria (Table 1) to identify 25 suitable datasets, which are listed in Supplementary Table 1.

Identification of Differentially Expressed Genes (DEGs)

The analysis of gene expression data was conducted using the R software environment version 3.2.3 [17] and a series of packages from the Bioconductor project, including *GEOquery* version 2.40.0 [18], *limma* version 3.26.8 [19], and *BioBase* version 2.30.0 [20]. Briefly, by adapting scripts from GEO's own GEO2R tool [21], the data was retrieved and translated to R-compatible formats via *GEOquery* and analyzed for DEGs via *limma*. DEGs were calculated by comparing samples from young vs. old tissues for each individual dataset; individual samples were included

or excluded from the analysis using the aforementioned inclusion and exclusion criteria (Table 1). To cast a wide net with high sensitivity to detect DEGs even in datasets with small sample sizes, candidate DEGs were identified in individual datasets using the permissive threshold of adjusted p-value < 0.25, controlling for false discovery rates (FDR) using the Benjamini-Hochberg method. Finally, to facilitate analysis across datasets, all DEGs from non-human datasets were converted to their human homologs using the *homologene* package version 1.4.68.19.3.27 [22].

Value-Counting Method for Ranking DEGs

The genes were then scored using a variation of the value-counting method first established in the cancer field [23] and later applied to age-dependent gene expression [8]. This approach enables the integration of gene expression data from diverse species, tissues, platforms, and experimental designs while remaining highly scalable and reproducible. In brief, genes are ranked by the number of datasets in which they are identified as a DEG according to a chosen threshold. Thus, consistency of differential expression across a variety of datasets is prioritized, while individual effect sizes are discarded.

Here, a new variation of the value-counting method was introduced to further prioritize consistency: ranking was determined based on the absolute value of the difference between upregulation and downregulation scores, where scores were determined by the number of datasets in which the DEG was significantly upregulated and downregulated with age, respectively. This is written formulaically below along with an example.

8

Let scores $S_{i_{Up}}$ and $S_{i_{Down}}$ represent the number of datasets in which gene i is significantly upregulated and downregulated with age, respectively. The total score S_i and rank R_i for each gene i is calculated as follows:

$$S_i = S_{i_{Up}} - S_{i_{Down}}$$

$$R_i = |S_i| = \left| S_{i_{Up}} - S_{i_{Down}} \right|$$

For example, if a gene was significantly (adj. p < 0.25) upregulated in 2 datasets and downregulated in 8 datasets out of the total 25 datasets analyzed, this gene would have a rank of $|2 - 8| = 6$. It is important to highlight that a gene with a total score and rank of 0 does not necessarily indicate that the gene was differentially expressed in none of the datasets, as it could also be upregulated and downregulated in an equal number of datasets.

Based on the average number of DEGs identified per dataset in the previous step being 1,816 genes out of an average over 30,000 probes per dataset, a binomial distribution with a success rate of 6% and 25 trials can be applied to estimate the final p value for high-ranking genes. For DEGs of rank 6 or above, the cumulative probability $P(X \geq 6)$ yields a final p value of 0.003:

$$P(X \geq 6) = \sum_{j=6}^{n} \binom{25}{j} 0.06^j (1 - 0.06)^{25-j}$$

DEGs with a rank of at least 7 ($R_i \geq 7$) were further analyzed for gene expression patterns across tissues by normalizing to the number of datasets from each tissue type that were analyzed.

This value-counting analysis was conducted using the python software environment version 3.11.5 [24], and the data were visualized utilizing the *pandas*, *matplotlib*, and *seaborn* packages [25–27].

Pathway Analysis

DEGs with a rank of at least 6 ($R_i \geq 6$) were analyzed using pathway analysis in the R software environment to explore their known roles in key biological processes. The Bioconductor package *clusterProfiler* version 4.8.0 [28] was used to perform gene ontology (GO) enrichment analysis. DEGs were mapped to GO biological processes, cellular components, and molecular functions using standard settings (Benjamini-Hochberg adjusted p < 0.05).

Identifying Worm Orthologs of DEGs

C. elegans orthologs of DEGs with a rank of at least 7 ($R_i \geq 7$) were identified using OrthoList 2, which is a compendium of worm genes with human orthologs compiled by a meta-analysis of several orthology prediction methods [29]. Where multiple orthologs were available for a given DEG, the highest confidence ortholog was chosen, as indicated by the number of orthology prediction methods supporting orthology. Where multiple orthologs and/or clones were available for a given gene without any discernible way to prioritize one over another, the first item listed in the results was chosen. The final list of orthologs along with the availability of corresponding RNAi clones is shown in Supplementary Tables 2 and 3. In some cases, when a clone could not be cultured or verified by sequencing (as outlined in the next section below), experiments were conducted using the next clone on the list.

Worm Culture and Post-Developmental RNAi

Wild-type (N2) *C. elegans* worms were maintained on plates of solid nematode growth media (NGM) seeded with *Escherichia coli* OP50 bacteria at 20°C using standard protocols[30]. *Escherichia coli* HT115 bacteria clones carrying RNAi constructs of interest were obtained from the Ahringer RNAi library [31] and seeded onto solid NGM plates containing Isopropyl β-D-1-

10

thiogalactopyranoside (IPTG) and ampicillin according to standard protocols for the RNAi feeding method [12, 32]. Briefly, for each gene of interest, an individual colony of RNAi bacteria was cultured in liquid LB medium overnight and then seeded onto plates the following day. In parallel, to confirm the identity of the clones, DNA was isolated from the same culture using the QIAprep Spin Miniprep Kit (QIAGEN, Hilden, Germany), and the inserts were sequenced with an M13-forward primer using standard Sanger sequencing services by Azenta Life Sciences (South Plainfield, NJ, USA). The seeded plates were incubated at room temperature for 2-3 days, during which time 2' fluro-5' deoxyuridine (FUDR) was added to the plates 24-48 hours before transferring worms. Worms were age-synchronized using the bleaching method with L1 synchronization and allowed to develop to the late L4 stage on standard OP50 plates before being transferred to the plates seeded with the RNAi feeding bacteria, as described in previous post-developmental RNAi screens [33, 34].

Lifespan Extension Screen

Lifespan assays were conducted using standard protocols [30]. Briefly, worms were scored as alive or dead every two to three days by visual observation: apparently motionless worms were gently prodded with a platinum wire pick, and worms that failed to react were scored as dead and removed from the plate. Worms that left the plate surface and dried on the plate wall were censored, but worms that displayed abnormalities such as internal hatching or vulva rupture were included in all analyses. For the initial screening, the 19 candidate clones were tested across several batches of experiments, with a GFP RNAi negative control group present in every batch, and the well-known *daf-2* RNAi positive control [35] in some selected batches. For every clone tested, the initial screening included roughly 80-100 worms spread across multiple plates (biologic replicates), with approximately 20 worms per plate. For the subsequent validation of

11

the clones that significantly extended lifespan in the initial screening, each group included roughly 100-120 worms, with approximately 25 worms per plate. These were tested in just two batches back-to-back under identical conditions with the same stocks of plates and reagents to minimize batch effects and maximize comparability of results. A GFP RNAi negative control group was included in both batches; the *daf-2* RNAi positive control was included in one batch, while an additional negative control empty L4440 vector RNAi was included in the other batch. To further ensure reproducibility, the N2 worms used during the initial screening and validation experiments were obtained from colonies maintained by separate, independent laboratories (see Acknowledgments for collaborators).

Lifespan Extension Analysis

Lifespan was defined as the number of days until death, starting from the first day of adulthood (3 days after L1 synchronization). The Online Application for Survival Analysis 2 (OASIS 2) tool was used to calculate mean, median, and maximum lifespans for each group as well as to compare test groups using the log-rank test [36]. An RNAi clone was considered to have extended lifespan if the log-rank test comparing that clone to the GFP RNAi negative control within the same batch was significant ($p < 0.05$ with Bonferroni multiple test correction) in both the initial screen and the subsequent validation screen. Survival data was then exported and plotted as survival curves using GraphPad Prism version 10.2.3.403 for Windows.

Results

Meta-analysis datasets were derived from a variety of mammalian tissues

Twenty-five publicly available gene expression datasets were selected from the NCBI GEO repository according to the inclusion and exclusion criteria outlined in Table 1, and their traits and NCBI identification numbers are listed in Supplementary Table 1. The predominant species represented in this analysis was mouse, constituting roughly half the datasets (13), followed by human (6), then rat (5), then dog (1). Most datasets were derived from muscle (7) and brain (5) tissues, but also well represented were adipose tissues (3) as well as immune cells and their precursors (3), with smaller contributions from the heart, liver, trachea, cochlea, and reproductive tissues (Supp. Fig. 1A). The number of DEGs extracted from each dataset varied widely, ranging from six genes to 3,631 genes (median, 1,509; interquartile range 466 - 3,159). If each instance of a DEG being extracted is considered a data point, then the sum total of data points contributed by most tissues ranged from roughly 4,000 to 11,000; however, cochlea, trachea, and reproductive tissues contributed strikingly fewer, with less than 1,000 data points each (Supp. Fig. 1B). These results reflect the wide variety of studies contributing to this analysis, with varying experimental methods as well as unequal availability of samples from different tissues, particularly from human subjects.

High-ranking genes were consistently differentially expressed with age across diverse tissues

Using the value-counting method, every gene was assigned upregulation and downregulation scores corresponding to the number of datasets in which that gene was significantly upregulated and downregulated with age, respectively (FDR-adjusted $p < 0.25$). In general, more genes were commonly upregulated than downregulated with age. Out of a highest possible score of 25 (total

13

number of datasets), the highest downregulation score was 9 (Fig. 1A), and the highest upregulation score was 11 (Fig. 1B). Similarly, only 31 genes achieved a downregulation score of 7 or more, whereas 74 genes reached an upregulation score of 7 or more.

To narrow the list of DEGs to those with the most consistent trends, genes were ranked according to the absolute value of the difference between their upregulation and downregulation scores. Thus, genes exhibiting opposing trends in different species or tissues were not ranked highly. While there were 105 genes with a downregulation or upregulation score of at least 7, there were only 45 genes that ranked 7 or above after the opposing score was subtracted. The highest-ranking age-upregulated genes were EFEMP1 (Rank 11), TMEM176A (11), CP (9), and HLA-A (9); the highest-ranking age-downregulated genes were CA4 (8), SIAH2 (8), SPARC (8), and UQCR10 (8). Ranks were used to select DEGs for further analyses and experiments: rank 6 was used as the cut-off to select 130 DEGs for pathway analysis, and rank 7 was used as the cutoff to select 45 DEGs for *in vivo* testing in *C. elegans*.

The 45 highest-ranking DEGs, comprising 16 age-downregulated and 29 age-upregulated genes, are listed in Figure 1C with a heatmap displaying the tissues that contributed to each gene's rank. For example, EFEMP1, which was tied for the highest-ranking gene, was significantly upregulated in datasets from mouse liver and hematopoietic stem cells, rat heart and adipose tissues, and both human and mouse brain and muscle tissues; EFEMP1 was not significantly downregulated in any of the 25 datasets analyzed. As illustrated in the heatmap, no gene was able to achieve this high rank without being consistently differentially expressed in datasets from at least three distinct tissue types, and often more. The genes CA4 and CP were notable for being consistently differentially expressed across all six major tissue types studied as well as being among the top four highest-ranking downregulated and upregulated genes, respectively.

14

Excluding tissue types with only one or two datasets, the only gene differentially expressed in 100% of datasets from a given tissue type was NPC2, which was age-upregulated in all five datasets from the brain, as well as handful of datasets from heart, muscle, and immune tissues. Collectively these findings illustrate how the meta-analysis ranking system was able to reveal genes with striking age-related expression patterns.

Gene ontology patterns were consistent with previous literature

Gene ontology (GO) enrichment analysis was performed to assess how high-ranking DEGs could be categorized into recognizable functional groups and pathways. For this analysis, the cutoff was relaxed to include DEGs of rank 6 and above, yielding a pool of 40 age-downregulated and 90 age-upregulated genes. The GO term matching the largest number of genes from the downregulated pool was the mitochondrial inner membrane, and several additional terms related to mitochondria were enriched as well (Fig. 2A-B). Also strongly represented were both cellular components and molecular functions related to extracellular matrix proteins, particularly collagen. Interestingly, collagen-containing extracellular matrix was also the top GO term for the pool of age-upregulated genes (Fig. 2C-D). However, the overwhelming majority of the GO terms enriched among age-upregulated genes were biological processes related to immune activity, especially adaptive immunity. The results of this pathway analysis largely aligned with expectations and patterns observed in previous studies, reinforcing the validity of the meta-analysis design and execution.

Knocking down orthologs of several mammalian DEGs extended lifespan in C. elegans

To prepare for *in vivo* experiments in the model organism *C. elegans*, the database OrthoList 2 was searched for worm orthologs corresponding to the highest-ranking mammalian DEGs. Out

15

of the 45 DEGs queried, 27 (60%) were matched to at least one worm ortholog, including 11 age-downregulated genes (Supp. Table 2) and 16 age-upregulated genes (Supp. Table 3). After excluding orthologs with no corresponding verifiable RNAi clone, a total of 19 genes were available for *in vivo* testing in *C. elegans*, including roughly equal quantities of age-downregulated (9) and age-upregulated (10) genes.

To probe the role of each ortholog in senescence while excluding effects on embryonic and juvenile development, bacteria carrying the RNAi clones were fed to the worms post-developmentally. A two-tiered screen was performed, with an initial screening (n ≈ 80 – 100 worms per group) for all 19 orthologs followed by independent validation experiments (n ≈ 100 – 150 worms per group) for RNAi clones that exhibited significant lifespan extension during the screening. Knockdown of six out of 19 (or 32%) of the orthologs of mammalian DEGs significantly extended lifespan in *C. elegans* during both tiers of the screen relative to within-batch GFP controls (log-rank test, Bonferroni-adjusted p < 0.05). In order from largest to smallest mean lifespan extension, these knockdowns targeted: *fzy-1* (ortholog of CDC20), *ost-1* (SPARC), *spch-2* (RSRC1), *C42C1.8* (DIRC2/SLC49A4), *csp-3* (CASP1), and *cah-3* (CA4), which were predominantly (66%) age-downregulated DEGs (Fig. 3A-F). The only two age-upregulated targets were *spch-2* (RSRC1) and *csp-3* (CASP1), while the rest were age-downregulated. Mean lifespan extension ranged from 9% to 15%, median extension from 6% to 19%, and maximum lifespan 4% to 15% relative to within-batch GFP controls, which averaged a lifespan mean of 23.35, median 22.14, and maximum 27.29 days (Table 2). The variability in the validation experiments was very low, as illustrated by the overlapping survival curves for the three independent negative control groups (Fig. 3G). The degree of lifespan extension was comparable to the positive control RNAi against *daf-2* (Fig. 3H), which extended mean lifespan

16

by 9%, median by 9%, and maximum by 7% (n = 103, log-rank test, Bonferroni-adjusted p < 0.01, Table 2).

Lastly, the expression patterns of the six genes that extended lifespan in worms were re-examined in our mammalian datasets (Supp. Fig. 2). CASP1 and RSCR1 were age-upregulated in seven datasets each with no age-downregulation. CA4, DIRC2, and CDC20 were age-downregulated in seven or eight datasets each with no age-upregulation. SPARC was age-downregulated in nine datasets but also age-upregulated in one dataset derived from mouse liver. All six were differentially expressed with age in multiple mouse tissues; all except CDC20 in human tissues; and all except CA4 in rat tissues. All six genes were differentially expressed in at least one of the seven datasets from muscle, which is unsurprising, but also one of only two datasets from liver, which is a much higher rate. Expression patterns in adipose tissue were also prominent: CASP1, RSRC1, and SPARC were differentially expressed in two of three fat datasets, and CA4 and DIRC2 in one dataset each. Finally, RSRC1, CA4, and CDC20 were differentially expressed in certain brain tissues including human frontal cortex and mouse neocortex and striatum. In summary, the most striking pattern was that knockdown of age-upregulated DEGs were not more likely than age-downregulated DEGs to extend lifespan, and the prominent contributions from adipose and liver tissue were also notable.

17

Discussion

This study has established a geroscience-specific workflow to channel large quantities of gene expression data into a streamlined list of actionable targets using accessible, scalable tools in computational biology and *C. elegans* research. The goal of this approach is to maximize the value of existing research by harnessing these readily available datasets and methods effectively to produce novel, valuable discoveries.

Over the past few decades there have been copious studies comparing gene expression in tissues from older versus younger subjects in a variety of species [37, 38], and these generally culminate in conclusions based on functional enrichment analysis. In general, advanced age has been associated with upregulation of immune and inflammatory pathways but downregulation of the electron transport chain and other mitochondrial activities as well as collagen and other extracellular matrix proteins [8, 37, 39], and our results were consistent with these established trends. However, therapeutic directions cannot be extrapolated from purely observational gene expression data, where drivers of aging cannot be distinguished from compensatory protective responses and irrelevant downstream effects. Moreover, there is no guarantee that functional groups reflect concerted biological activities, and they are biased in favor of well-defined gene-sets, which are both important limitations to consider.

Two of our highest-ranking individual genes, EFEMP1 (Rank 11) and CP (Rank 9), which were consistently age-upregulated over all six major tissue types, have known associations with age-related pathologies, and have also been classified as age-associated in previous similar meta-analyses [8, 9]. EFEMP1, also known as fibulin-3, is an extracellular matrix glycoprotein strongly associated with aging pathologies: overexpression contributes to age-related macular degeneration, high plasma levels are associated with signs of brain aging and higher risk of

18

dementia, and upregulation of this gene is associated with Werner syndrome, a premature aging condition [40, 41]. On the other hand, CP, or ceruloplasmin, is a copper-binding glycoprotein involved in iron metabolism and defense against oxidative stress; decreased CP activity is associated with advanced age and age-related diseases, such as Parkinson's and Alzheimer's disease [42–44]. From this context, we may infer that although both genes exhibit similar expression profiles, EFEMP1 likely plays a role driving age-related pathology, whereas CP may be upregulated with age as a compensatory response to amplify its protective effects. However, even for well documented genes like these, such inferences still involve speculation, and there are many other DEGs that are much less clearly characterized without supplemental information.

We focused on funneling our DEGs into a *C. elegans* RNAi lifespan screen to gain insights on the role of each gene in aging and longevity. Two of the ten tested age-upregulated genes extended lifespan when knocked down in *C. elegans*: *csp-3*,an ortholog of CASP1, and *spch-2 ,* an ortholog of RSRC1. Caspases are proteases involved in apoptosis and inflammation [45], and CASP-1 is particularly well known a major component of the NLRP3-CASP1 inflammasome and a promising therapeutic target for Hutchinson-Gilford progeria, another premature aging syndrome, and Alzheimer's disease [46, 47]. In fact, pharmacological CASP-1 inhibitors have demonstrated to be protective against cognitive decline in mouse models of Alzheimer dementia [48, 49]. Interestingly, CASP1 was not differentially expressed in any of the brain datasets we examined, which included samples from human frontal cortex. However, there is evidence that CASP1 is overexpressed in the frontal cortex and hippocampus of patients with Alzheimer's disease [47].The novel discovery that RNAi inhibition of an orthologous caspase extends lifespan in worms may suggest an evolutionarily conserved role for caspases in driving inflammaging and age-related neurodegeneration.

RSRC1, on the other hand, is much less well studied, but appears to have a role in brain development: RSRC1 polymorphism is associated with schizophrenia [50], and patients homozygous for loss-of-function RSRC1 mutations exhibit developmental delay and intellectual disability [51, 52]. In our meta-analysis, RSRC1 was age-upregulated in three of the five datasets from brain tissue, and post-developmental knockdown produced lifespan extension. Collectively, these results suggest that RSRC1 is crucial for early development but functions aberrantly late in life as a driver of aging.

Four of the nine tested age-underlinedownregulated genes we tested extended lifespan when knocked down in *C. elegans*, including orthologs of two of the four highest-ranking (Rank 8) age-downregulated DEGs: *ost-1*, ortholog of SPARC; and *cah-3*, an ortholog of CA4. SPARC, also known as osteonectin, is a highly conserved extracellular matrix glycoprotein; it is expressed ubiquitously, though primarily in adipocytes [53, 54]. It also plays a role in bone development and turnover and wound healing, especially in corneal tissue [55–57]. In adipocytes, SPARC has been linked to adipose fibrosis, age-related inflammation and metabolic dysfunction, as well as diabetes and its complications (nephropathy and retinopathy), and obesity [54, 58, 59]. In our meta-analysis, SPARC was downregulated with age in all major tissues studied except the brain and liver; the most pronounced pattern was in adipose tissue, where expression decreased with age in two of the three datasets. Interestingly, SPARC is also associated with liver fibrosis [60], and it was in fact upregulated in our dataset from mouse liver. The large amount of available literature underscores SPARC's pleiotropic functions and open questions about its role in aging. Our work revealed tissue-specific changes in SPARC expression with age and the discovery that post-developmental SPARC knockdown extends lifespan in *C. elegans*.

20

Carbonic anhydrases are crucial for such fundamental biological processes as regulating pH and transporting carbon dioxide, and they are ubiquitous from microbes to mammals [61]. We found one member of this enzyme family, CA4, to be consistently downregulated with age across all six major tissue types studied, including human and mouse brain datasets. In humans, CA4 mutations contribute to retinitis pigmentosa [62], but in rodents, CA4 has also been studied as a mechanism for extracellular buffering in the brain [63, 64]. Carbonic anhydrase inhibitors have been pursued as potential treatments for brain edema, glaucoma, epilepsy, cancer, and mountain sickness [61]. Here we showed, for the first time, that downregulating the CA4 ortholog, *cah-3*, post-developmentally extended lifespan in *C. elegans*, suggesting that less CA4 activity may be needed in late life.

The largest lifespan extension achieved in our study was via RNAi knockdown of the CDC20 ortholog *fzy-1*, a result consistent with a previous study (Xue et a. 2007). Cell division cycle 20 (CDC20) is an evolutionarily conserved, positive regulator of cell division essential for life in both worms and mammals [65, 66]. The concentration and activity of CDC20 must be tightly regulated, as hyperactivity is associated with aneuploidy and oncogenesis [67–69], and downregulation is associated with premature cellular senescence [70]. Interestingly, although adult *C. elegans* are post-mitotic creatures, genes associated with cell proliferation and differentiation modulate worm lifespan through mechanisms that are thought to be evolutionarily conserved [71]. We found that CDC20 was most downregulated in mouse dendritic cells, hematopoietic stem cells, and rodent liver tissues, which may reflect a decline in proliferation of those tissues.

Lastly, we found that post-developmental knockdown of the DIRC2 ortholog *C42C1.8* extends lifespan in *C. elegans*. DIRC2, or disrupted in renal carcinoma 2, is also known as solute carrier

family 49 member 4, or SLC49A4. Aside from the eponymous roles in renal carcinoma and solute transport (specifically, lysosomal export of vitamin B6), very little is known about this gene [72], or the worm ortholog *C42C1.8,* which has no common name. Our work demonstrates that DIRC2 is actually one of the most consistently age-downregulated genes in mammals, with expression declining in both human and mouse muscle tissues and rodent fat, heart, liver, and trachea. The example of DIRC2 demonstrates the power of our approach to identify promising, understudied targets for further investigation.

C. elegans has long been used to investigate the mechanisms of aging using well developed functional genomics tools. There were two genome-wide RNAi longevity screens, by the Ruvkun [73] and Kenyon [74] groups, each boasting 70-80% coverage of all open-reading frames. Due to very high false negative rates, they identified a combined total of 120 longevity genes, with only four genes in common [75]. The Ruvkun group performed a follow-up screen using post-developmental instead of embryonic RNAi on 2,700 genes essential for development [33], and similar smaller studies have been published by others since [34, 76]. The yield of lifespan extending gene activations out of total genes tested was less than 1% for genome-wide screens, and 2.4% for the post-developmental screen of essential genes, reflecting the importance of antagonistic pleiotropy in aging [3, 77]. A more recent study achieved a yield of 44% when testing orthologs of genes differentially expressed with age in human blood, and this study also reported a background rate of 7% yield for randomly chosen genes [78]. Yield is highly dependent on experimental methods such as number of animals and time-points as well as environmental factors like temperature; in this latter case, the authors also tested several of genes under two different conditions (pre- and post-developmentally), raising the yield. Here we reported a yield of 32%, suggesting that roughly one-third of the candidates identified in our

meta-analysis are drivers of aging; the remaining two-thirds may have negligible or protective effects, or they may also be drivers of aging but under conditions not tested in this experiment.

There are important limitations to our study, many of which pertain to the nature of RNAi screens and the challenges of modeling human physiology in worms. First, only gene inactivation, not overexpression, could be tested, so we could identify drivers of aging but not geroprotective genes. Secondly, only some of the high-ranking DEGs corresponded to verifiable worm orthologs and were able to be tested, and even those orthologs were selected with varying levels of confidence and specificity. Thirdly, the evolutionary distance between humans and worms limits our interpretation of the functional roles of each gene product, as molecules with very similar structures can play very different biological roles in such distinct species.

It should also be noted that the datasets included in our meta-analysis were derived from neither a complete nor an even distribution of tissue types; for example, there were no datasets derived from the kidneys or intestines, whereas muscle was highly represented. As more datasets are made available, we expect this approach to provide increasing contributions to the geroscience literature. Consistent with this goal, the methods described herein are intended to be accessible and flexible enough for others to reproduce and expand our workflow in future studies. Only basic coding skills in R and python are required to reproduce the meta-analysis, and the GEO2R toolset at the core of our scripts has recently been updated to accept both microarray and RNAseq datasets.

In conclusion, the overall trends we observed in our meta-analysis were consistent with previous literature, but our novel workflow identified six genes with evolutionarily conserved, causal roles in the aging process. Of note, knocking down age-upregulated genes was not more likely to produce life extension than interfering with age-downregulated genes. Thus, our results do not

support the commonly held assumption that reversing any changes in age-related gene

expression is beneficial, and future studies should further investigate this trend.

Author Contributions

Conceptualization and Methodology: A. Coler-Reilly, Z. Pincus; Formal analysis and investigation: A. Coler-Reilly, E. Scheller, R. Civitelli; Writing - original draft preparation: A. Coler-Reilly; Writing - review and editing: Z. Pincus, E. Scheller, R. Civitelli; Funding acquisition: A. Coler-Reilly, R. Civitelli; Resources: Z. Pincus, R. Civitelli; Supervision: E. Scheller, R. Civitelli.

Acknowledgments

Data Availability

All raw scripts are permanently stored and available for review at the following repository:
https://github.com/AriellaStudies/Aging-DEGs

References

[1] B. K. Kennedy *et al.,* "Geroscience: linking aging to chronic disease," *Cell*, vol. 159, no. 4, pp. 709–713, 2014, doi: 10.1016/j.cell.2014.10.039.

[2] T. B. Kirkwood and S. N. Austad, "Why do we age?," *Nature*, vol. 408, no. 6809, pp. 233–238, 2000, doi: 10.1038/35041682.

[3] S. N. Austad and T. Kirkwood, "Evolutionary Theory in Aging Research," in *Cold Spring Harbor monograph*, vol. 51, *Molecular biology of aging*, L. Guarente, L. Partridge, and D. C. Wallace, Eds., New York, Bloxham: Cold Spring Harbor Laboratory Press, 2008, pp. 95–111.

[4] L. Hou, D. Wang, H. Cheng, B. Xian, and J.-D. J. Han, "Systems Approaches to Understanding Aging," in *The handbooks of aging, Handbook of the biology of aging*, M. Kaeberlein and G. M. Martin, Eds., Amsterdam: Academic Press, 2016, pp. 241–261.

[5] A. Perez-Gomez, J. N. Buxbaum, and M. Petrascheck, "The aging transcriptome: read between the lines," *Current opinion in neurobiology*, vol. 63, pp. 170–175, 2020, doi: 10.1016/j.conb.2020.05.001.

[6] R. Bahar *et al.,* "Increased cell-to-cell variation in gene expression in ageing mouse heart," *Nature*, vol. 441, no. 7096, pp. 1011–1014, 2006, doi: 10.1038/nature04844.

[7] S. Rangaraju *et al.,* "Suppression of transcriptional drift extends C. elegans lifespan by postponing the onset of mortality," *eLife*, vol. 4, e08833, 2015, doi: 10.7554/eLife.08833.

[8] J. P. de Magalhães, J. Curado, and G. M. Church, "Meta-analysis of age-related gene expression profiles identifies common signatures of aging," *Bioinformatics (Oxford, England)*, vol. 25, no. 7, pp. 875–881, 2009, doi: 10.1093/bioinformatics/btp073.

[9] D. Palmer, F. Fabris, A. Doherty, A. A. Freitas, and J. P. de Magalhães, "Ageing transcriptome meta-analysis reveals similarities and differences between key mammalian tissues," *Aging*, vol. 13, no. 3, pp. 3313–3341, 2021, doi: 10.18632/aging.202648.

[10] J. P. de Magalhães and O. Toussaint, "How bioinformatics can help reverse engineer human aging," *Ageing research reviews*, vol. 3, no. 2, pp. 125–141, 2004, doi: 10.1016/j.arr.2003.08.006.

[11] T. E. Johnson, "Caenorhabditis elegans 2007: the premier model for the study of aging," *Experimental gerontology*, vol. 43, no. 1, pp. 1–4, 2008, doi: 10.1016/j.exger.2007.09.008.

[12] D. S. Wilkinson, R. C. Taylor, and A. Dillin, "Analysis of aging in Caenorhabditis elegans," *Methods in cell biology*, vol. 107, pp. 353–381, 2012, doi: 10.1016/B978-0-12-394620-1.00012-6.

[13] H. G. Son, O. Altintas, E. J. E. Kim, S. Kwon, and S.-J. V. Lee, "Age-dependent changes and biomarkers of aging in Caenorhabditis elegans," *Aging cell*, vol. 18, no. 2, e12853, 2019, doi: 10.1111/acel.12853.

[14] C.-H. Chen, Y.-C. Chen, H.-C. Jiang, C.-K. Chen, and C.-L. Pan, "Neuronal aging: learning from C. elegans," *Journal of Molecular Signaling*, vol. 8, no. 1, p. 14, 2013, doi: 10.1186/1750-2187-8-14.

[15] G. L. Sutphin and R. Korstanje, "Longevity as a Complex Genetic Trait," in *The handbooks of aging, Handbook of the biology of aging*, M. Kaeberlein and G. M. Martin, Eds., Amsterdam: Academic Press, 2016, pp. 4–54.

[16] R. Edgar, M. Domrachev, and A. E. Lash, "Gene Expression Omnibus: NCBI gene expression and hybridization array data repository," *Nucleic Acids Research*, vol. 30, no. 1, pp. 207–210, 2002, doi: 10.1093/nar/30.1.207.

[17] R Core Team, *R: A Language and Environment for Statistical Computing*. Vienna, Austria. [Online]. Available: https://www.r-project.org/

[18] S. Davis and P. S. Meltzer, "GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor," *Bioinformatics (Oxford, England)*, vol. 23, no. 14, pp. 1846–1847, 2007, doi: 10.1093/bioinformatics/btm254.

[19] M. E. Ritchie *et al.,* "limma powers differential expression analyses for RNA-sequencing and microarray studies," *Nucleic Acids Research*, vol. 43, no. 7, e47, 2015, doi: 10.1093/nar/gkv007.

[20] W. Huber *et al.,* "Orchestrating high-throughput genomic analysis with Bioconductor," *Nat Methods*, vol. 12, no. 2, pp. 115–121, 2015, doi: 10.1038/nmeth.3252.

[21] T. Barrett *et al.,* "NCBI GEO: archive for functional genomics data sets--update," *Nucleic Acids Research*, vol. 41, Database issue, D991-5, 2013, doi: 10.1093/nar/gks1193.

[22] Ogan Mancarci, *homologene: Quick Access to Homologene and Gene Annotation Updates.* [Online]. Available: https://cran.r-project.org/package=homologene

[23] D. R. Rhodes *et al.,* "Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 25, pp. 9309–9314, 2004, doi: 10.1073/pnas.0401994101.

[24] G. van Rossum and F. L. Drake Jr, *Python tutorial*: Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands, 1995.

[25] W. McKinney, "Data Structures for Statistical Computing in Python," in *Proceedings of the 9th Python in Science Conference*, Austin, Texas, 2010, pp. 56–61.

[26] J. D. Hunter, "Matplotlib: A 2D Graphics Environment," *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 90–95, 2007, doi: 10.1109/MCSE.2007.55.

[27] M. Waskom, "seaborn: statistical data visualization," *JOSS*, vol. 6, no. 60, p. 3021, 2021, doi: 10.21105/joss.03021.

[28] T. Wu *et al.,* "clusterProfiler 4.0: A universal enrichment tool for interpreting omics data," *The Innovation*, vol. 2, no. 3, p. 100141, 2021, doi: 10.1016/j.xinn.2021.100141.

[29] W. Kim, R. S. Underwood, I. Greenwald, and D. D. Shaye, "OrthoList 2: A New Comparative Genomic Analysis of Human and Caenorhabditis elegans Genes," *Genetics*, vol. 210, no. 2, pp. 445–461, 2018, doi: 10.1534/genetics.118.301307.

[30] G. L. Sutphin and M. Kaeberlein, "Measuring Caenorhabditis elegans life span on solid media," *Journal of visualized experiments : JoVE*, no. 27, 2009, doi: 10.3791/1152.

[31] R. S. Kamath and J. Ahringer, "Genome-wide RNAi screening in Caenorhabditis elegans," *Methods*, vol. 30, no. 4, pp. 313–321, 2003, doi: 10.1016/s1046-2023(03)00050-1.

[32] L. Timmons and A. Fire, "Specific interference by ingested dsRNA," *Nature*, vol. 395, no. 6705, p. 854, 1998, doi: 10.1038/27579.

[33] S. P. Curran and G. Ruvkun, "Lifespan regulation by evolutionarily conserved genes essential for viability," *PLoS genetics*, vol. 3, no. 4, e56, 2007, doi: 10.1371/journal.pgen.0030056.

[34] Di Chen, K. Z. Pan, J. E. Palter, and P. Kapahi, "Longevity determined by developmental arrest genes in Caenorhabditis elegans," *Aging cell*, vol. 6, no. 4, pp. 525–533, 2007, doi: 10.1111/j.1474-9726.2007.00305.x.

[35] A. Dillin, D. K. Crawford, and C. Kenyon, "Timing requirements for insulin/IGF-1 signaling in C. elegans," *Science*, vol. 298, no. 5594, pp. 830–834, 2002, doi: 10.1126/science.1074240.

[36] S. K. Han *et al.,* "OASIS 2: online application for survival analysis 2 with features for the analysis of maximal lifespan and healthspan in aging research," *Oncotarget*, vol. 7, no. 35, pp. 56147–56152, 2016, doi: 10.18632/oncotarget.11269.

[37] J. M. Zahn *et al.,* "AGEMAP: a gene expression database for aging in mice," *Plos Genet*, vol. 3, no. 11, e201, 2007.

[38] J. P. de Magalhães *et al.,* "Human Ageing Genomic Resources: updates on key databases in ageing research," *Nucleic Acids Research*, vol. 52, D1, D900-D908, 2024, doi: 10.1093/nar/gkad927.

[39] M. J. Peters *et al.,* "The transcriptional landscape of age in human peripheral blood," *Nature communications*, vol. 6, p. 8570, 2015, doi: 10.1038/ncomms9570.

[40] L. Cheng *et al.,* "EFEMP1 Overexpression Contributes to Neovascularization in Age-Related Macular Degeneration," *Frontiers in pharmacology*, vol. 11, p. 547436, 2020, doi: 10.3389/fphar.2020.547436.

[41] E. R. McGrath *et al.,* "Plasma EFEMP1 Is Associated with Brain Aging and Dementia: The Framingham Heart Study," *Journal of Alzheimer's disease : JAD*, vol. 85, no. 4, pp. 1657–1666, 2022, doi: 10.3233/JAD-215053.

[42] I. Semsei, F. Jeney, and T. Fülöp, "Effect of age on the activity of ceruloplasmin of human blood," *Archives of Gerontology and Geriatrics*, vol. 17, no. 2, pp. 123–130, 1993, doi: 10.1016/0167-4943(93)90044-I.

[43] J. Kristinsson, J. Snaedal, G. Tórsdóttir, and T. Jóhannesson, "Ceruloplasmin and iron in Alzheimer's disease and Parkinson's disease: a synopsis of recent studies," *Neuropsychiatric Disease and Treatment*, vol. 8, pp. 515–521, 2012, doi: 10.2147/NDT.S34729.

[44] B. Wang and X.-P. Wang, "Does Ceruloplasmin Defend Against Neurodegenerative Diseases?," *Current neuropharmacology*, vol. 17, no. 6, pp. 539–549, 2019, doi: 10.2174/1570159X16666180508113025.

[45] M. D. Molla, Y. Akalu, Z. Geto, B. Dagnew, B. Ayelign, and T. Shibabaw, "Role of Caspase-1 in the Pathogenesis of Inflammatory-Associated Chronic Noncommunicable Diseases," *Journal of Inflammation Research*, vol. 13, pp. 749–764, 2020, doi: 10.2147/JIR.S277457.

[46] A. González-Dominguez *et al.,* "Inhibition of the NLRP3 inflammasome improves lifespan in animal murine model of Hutchinson-Gilford Progeria," *EMBO molecular medicine*, vol. 13, no. 10, e14012, 2021, doi: 10.15252/emmm.202114012.

[47] M. T. Heneka *et al.,* "NLRP3 is activated in Alzheimer's disease and contributes to pathology in APP/PS1 mice," *Nature*, vol. 493, no. 7434, pp. 674–678, 2013, doi: 10.1038/nature11729.

[48] J. Flores, A. Noël, B. Foveau, O. Beauchet, and A. C. LeBlanc, "Pre-symptomatic Caspase-1 inhibitor delays cognitive decline in a mouse model of Alzheimer disease and aging," *Nat Commun*, vol. 11, no. 1, p. 4571, 2020, doi: 10.1038/s41467-020-18405-9.

[49] J. Flores, M.-L. Fillion, and A. C. LeBlanc, "Caspase-1 inhibition improves cognition without significantly altering amyloid and inflammation in aged Alzheimer disease mice," *Cell Death Dis*, vol. 13, no. 10, p. 864, 2022, doi: 10.1038/s41419-022-05290-x.

[50] S. G. Potkin *et al.,* "Gene discovery through imaging genetics: identification of two novel genes associated with schizophrenia," *Mol Psychiatry*, vol. 14, no. 4, pp. 416–428, 2009, doi: 10.1038/mp.2008.127.

[51] Y. Perez *et al.,* "RSRC1 mutation affects intellect and behaviour through aberrant splicing and transcription, downregulating IGFBP3," *Brain : a journal of neurology*, vol. 141, no. 4, pp. 961–970, 2018, doi: 10.1093/brain/awy045.

[52] M. Scala *et al.,* "RSRC1 loss-of-function variants cause mild to moderate autosomal recessive intellectual disability," *Brain : a journal of neurology*, vol. 143, no. 4, e31, 2020, doi: 10.1093/brain/awaa070.

[53] M. C. Fitzgerald and J. E. Schwarzbauer, "Importance of the basement membrane protein SPARC for viability and fertility in Caenorhabditis elegans," *Current Biology*, vol. 8, no. 23, pp. 1285–1288, 1998, doi: 10.1016/s0960-9822(07)00540-4.

[54] K. Kos and J. P. H. Wilding, "SPARC: a key player in the pathologies associated with obesity and diabetes," *Nat Rev Endocrinol*, vol. 6, no. 4, pp. 225–235, 2010, doi: 10.1038/nrendo.2010.18.

[55] E. M. Rosset and A. D. Bradshaw, "SPARC/osteonectin in mineralized tissue," *Matrix biology : journal of the International Society for Matrix Biology*, 52-54, pp. 78–87, 2016, doi: 10.1016/j.matbio.2016.02.001.

[56] A. Basu, L. H. Kligman, S. J. Samulewicz, and C. C. Howe, "Impaired wound healing in mice deficient in a matricellular protein SPARC (osteonectin, BM-40)," *BMC Cell Biol*, vol. 2, no. 1, p. 15, 2001, doi: 10.1186/1471-2121-2-15.

[57] J. B. Lin *et al.,* "Dry eye disease in mice activates adaptive corneal epithelial regeneration distinct from constitutive renewal in homeostasis," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 120, no. 2, e2204134120, 2023, doi: 10.1073/pnas.2204134120.

[58] L. Xu *et al.,* "Elevated plasma SPARC levels are associated with insulin resistance, dyslipidemia, and inflammation in gestational diabetes mellitus," *PLoS ONE*, vol. 8, no. 12, e81615, 2013, doi: 10.1371/journal.pone.0081615.

[59] S. Ryu *et al.,* "The matricellular protein SPARC induces inflammatory interferon-response in macrophages during aging," *Immunity*, vol. 55, no. 9, 1609-1626.e7, 2022, doi: 10.1016/j.immuni.2022.07.007.

[60] C. Atorrasagasti *et al.,* "Lack of the matricellular protein SPARC (secreted protein, acidic and rich in cysteine) attenuates liver fibrogenesis in mice," *PLoS ONE*, vol. 8, no. 2, e54962, 2013, doi: 10.1371/journal.pone.0054962.

[61] A. Aspatwar *et al.,* "Carbonic anhydrases in metazoan model organisms: molecules, mechanisms, and physiology," *Physiological reviews*, vol. 102, no. 3, pp. 1327–1383, 2022, doi: 10.1152/physrev.00018.2021.

[62] Z. Yang *et al.,* "Mutant carbonic anhydrase 4 impairs pH regulation and causes retinal photoreceptor degeneration," *Human molecular genetics*, vol. 14, no. 2, pp. 255–265, 2005, doi: 10.1093/hmg/ddi023.

[63] M. S. Ghandour, O. K. Langley, X. L. Zhu, A. Waheed, and W. S. Sly, "Carbonic anhydrase IV on brain capillary endothelial cells: a marker associated with the blood-brain barrier," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, no. 15, pp. 6823–6827, 1992, doi: 10.1073/pnas.89.15.6823.

[64] G. N. Shah *et al.,* "Carbonic anhydrase IV and XIV knockout mice: roles of the respective carbonic anhydrases in buffering the extracellular space in brain," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 46, pp. 16771–16776, 2005, doi: 10.1073/pnas.0508449102.

[65] R. S. Kamath *et al.,* "Systematic functional analysis of the Caenorhabditis elegans genome using RNAi," *Nature*, vol. 421, no. 6920, pp. 231–237, 2003, doi: 10.1038/nature01278.

[66] M. Li, J. P. York, and P. Zhang, "Loss of Cdc20 causes a securin-dependent metaphase arrest in two-cell mouse embryos," *Molecular and cellular biology*, vol. 27, no. 9, pp. 3481–3488, 2007, doi: 10.1128/MCB.02088-06.

[67] D. J. Clarke *et al.,* "S-phase checkpoint controls mitosis via an APC-independent Cdc20p function," *Nature cell biology*, vol. 5, no. 10, pp. 928–935, 2003, doi: 10.1038/ncb1046.

[68] H. Fujita *et al.,* "Premature aging syndrome showing random chromosome number instabilities with CDC20 mutation," *Aging cell*, vol. 19, no. 11, e13251, 2020, doi: 10.1111/acel.13251.

[69] L. Wang, J. Zhang, L. Wan, X. Zhou, Z. Wang, and W. Wei, "Targeting Cdc20 as a novel cancer therapeutic strategy," *Pharmacology & Therapeutics*, vol. 151, pp. 141–151, 2015, doi: 10.1016/j.pharmthera.2015.04.002.

[70] D. Volonte, M. Sedorovitz, and F. Galbiati, "Impaired Cdc20 signaling promotes senescence in normal cells and apoptosis in non-small cell lung cancer cells," *The Journal of biological chemistry*, vol. 298, no. 10, p. 102405, 2022, doi: 10.1016/j.jbc.2022.102405.

[71] H. Xue *et al.,* "A modular network model of aging," *Molecular Systems Biology*, vol. 3, p. 147, 2007, doi: 10.1038/msb4100189.

[72] S. Akino, T. Yasujima, T. Yamashiro, and H. Yuasa, "Disrupted in renal carcinoma 2 (DIRC2/SLC49A4) is an H+-driven lysosomal pyridoxine exporter," *Life science alliance*, vol. 6, no. 2, 2023, doi: 10.26508/lsa.202201629.

[73] B. Hamilton *et al.,* "A systematic RNAi screen for longevity genes in C. elegans," *Genes & development*, vol. 19, no. 13, pp. 1544–1555, 2005, doi: 10.1101/gad.1308205.

[74] M. Hansen, A.-L. Hsu, A. Dillin, and C. Kenyon, "New genes tied to endocrine, metabolic, and dietary regulation of lifespan from a Caenorhabditis elegans genomic RNAi screen," *Plos Genet*, vol. 1, no. 1, pp. 119–128, 2005, doi: 10.1371/journal.pgen.0010017.

[75] M. Petrascheck and D. L. Miller, "Computational Analysis of Lifespan Experiment Reproducibility," *Frontiers in Genetics*, vol. 8, p. 92, 2017, doi: 10.3389/fgene.2017.00092.

[76] R. Tacutu *et al.,* "Prediction of C. elegans longevity genes by human and worm longevity networks," *PLoS ONE*, vol. 7, no. 10, e48282, 2012, doi: 10.1371/journal.pone.0048282.

[77] G. C. Williams, "Pleiotropy, Natural Selection, and the Evolution of Senescence," *Evolution*, vol. 11, no. 4, pp. 398–411, 1957.

[78] G. L. Sutphin *et al.,* "Caenorhabditis elegans orthologs of human genes differentially expressed with age are enriched for determinants of longevity," *Aging cell*, vol. 16, no. 4, pp. 672–682, 2017, doi: 10.1111/acel.12595.

Table 1. Inclusion and exclusion criteria for the meta-analysis of genes differentially expressed during mammalian aging.

Criteria	Included	Excluded
Data Availability	Stored in NCBI GEO database, Includes "age" as subset variable	All other datasets
Species	Mammal (human, mouse, rat, dog)	Non-Mammal (e.g., drosophila, C. elegans…)
Tissue Type	All (heart, muscle, brain, liver, fat, immune cells…)	None
Age	Young adult and older adult in distinct groups	Early development, including embryonic stages as well as juveniles
Condition	Healthy, untreated	All diseases (e.g. lupus, tumor samples), All interventions including drugs as well as lifestyle interventions like diet and exercise
Genotype	Wild-type, No genetic condition specified	Mutant, transgenic animals, Humans with specified genetic conditions
Sample Size	Any (no specific minimum)	Datasets with low sample size became naturally excluded when they yielded no DEGs

NCBI GEO = National Center for Biotechnology Information, Gene Expression Omnibus; **DEGs** = Differentially Expressed Genes

Table 2. Lifespan extension in *C. elegans* via post-developmental RNAi of orthologs of mammalian DEGs.

Human DEG	Worm Ortholog (Clone)	Number Subjects	Mean Lifespan +/- Standard Error (%Extension)	Median Lifespan (%Extension)	Maximum Lifespan (%Extension)	Corrected p value
CDC20 ↓	*fzy-1* (II-4O16)	103	26.21 ± 0.56 (15%)	26.41 (19%)	29.96 (12%)	$p < 0.001$
SPARC ↓	*ost-1* (IV-9H03)	143	26.38 ± 0.46 (11%)	24.91 (12%)	31.89 (15%)	$p < 0.001$
DIRC2 ↓	*C42C1.8* (IV-6D23)	151	26.09 ± 0.4 (10%)	25.41 (15%)	30.47 (10%)	$p < 0.001$
CA4 ↓	*cah-3* (X-7M16)	98	24.66 ± 0.44 (8%)	23.48 (6%)	28.45 (6%)	$p < 0.001$
RSRC1 ↑	*spch-2* (I-3C03)	99	25.31 ± 0.38 (11%)	24.31 (10%)	27.87 (4%)	$p < 0.001$
CASP1 ↑	*csp-3* (I-5P02)	101	25.93 ± 0.49 (9%)	24.7 (12%)	31.06 (12%)	$p < 0.01$
Positive Control	*daf-2* (III-7614)	103	24.95 ± 0.42 (9%)	24.17 (9%)	28.77 (7%)	$p < 0.01$

DEG = Differentially Expressed Gene. DEGs are written with HUGO (Human Genome Organization) Gene Nomenclature Committee (HGNC) symbols and an arrow indicating whether the gene was age-downregulated or age-upregulated. Worm orthologs are written using the common name if available, otherwise the sequence name, followed in parentheses by the specific Ahringer RNAi clone utilized. Percent lifespan extension was calculated against internal GFP controls using the log-rank test with Bonferroni adjusted p values. Average GFP control lifespan was mean = 23.35, median = 22.14, maximum = 27.29 days.

Figure Legends

Figure 1. Genes most consistently differentially expressed with age in mammalian tissues. Every gene had a downregulation score S_{Down} and an upregulation score S_{Up} representing the number of datasets in which the gene was significantly downregulated and upregulated with age, respectively (Benjamini-Hochberg adjusted p < 0.25). The rank R of each gene was calculated as the absolute value of the difference between these two scores, the total score S: $R = |S| = |S_{Up} - S_{Down}|$. (**A**) There were 31 genes with $S_{Down} \geq 7$, and the most consistently downregulated genes had a score of $S_{Down} = 9$. (**B**) There were 74 genes with $S_{Up} \geq 7$, and the most consistently upregulated genes had a score of $S_{Up} = 11$. (**C**) The heatmap shows the gene symbol, rank R, and normalized tissue-specific expression trends for all 45 genes of rank $R \geq 7$, comprising 29 age-upregulated genes ($S > 0$) and 16 age-downregulated genes ($S < 0$). Heatmap values range from 100% down to 100% up, representing as a percentage the fraction $|S_{tissue}/n_{tissue}|$, or the total score S divided by the number of datasets n derived from only the specified tissue type.

Figure 2. Gene ontology (GO) enrichment analysis of the 130 genes ranked highly for consistent differential expression with age, including 40 age-downregulated genes and 90 age-upregulated genes. For each group, bar charts display the top ten GO terms, colored according to three major GO term categories shown in the key. (**A**) Among age-downregulated genes, the most well-represented GO category was Cellular Components (CC, green). (**B**) The CC category is displayed in more detail in the accompanying network diagram, highlighting the downregulation of collagen-related and mitochondrial membrane-related genes. (**C**) Among age-upregulated genes, the major category was Biological Processes (BP, red). (**D**) The BP category is expanded to show the variety of immune response-related genes upregulated with age. The data were analyzed and visualized in R using the *clusterProfiler* package standard settings including Benjamini-Hochberg adjusted p < 0.05.

Figure 3. *C. elegans* survival curves for lifespan-extending post-developmental RNAi targeting orthologs of mammalian DEGs, including age-downregulated (blue) and age-upregulated (red) genes. Each RNAi clone significantly extended lifespan (**A-F**, log-rank test, Bonferroni-adjusted p < 0.05) in two independent experiments, and the results of the second experiment are shown here (n ≈ 100 − 150 worms per group). This experiment was performed in two back-to-back batches with an internal GFP control group in each batch (GFP1 for batch 1, GFP2 for batch 2). (**G**) Negative control GFP groups performed very similarly to each other (solid lines) and to the alternative empty L4440 vector control (dotted lines). (**H**) The RNAi clone Ahringer III-7614 against *daf-2* was used a positive control (green). For detailed quantification, see Table 2.

31

Figure 1

A. Frequency of Age-Downregulated Genes by Score
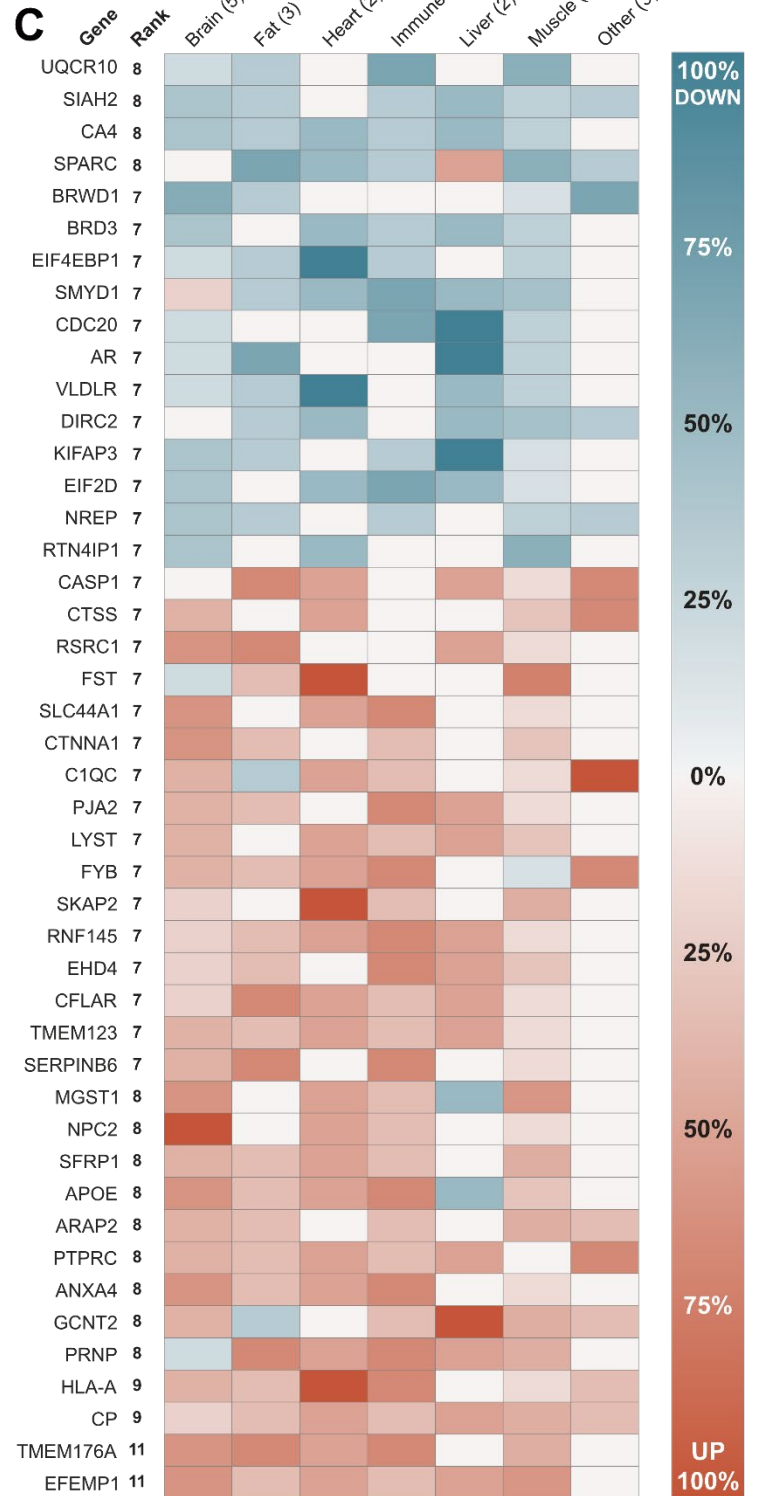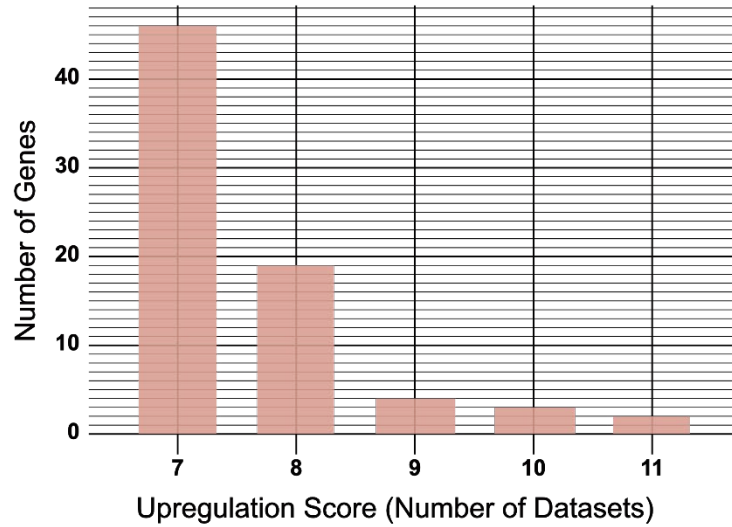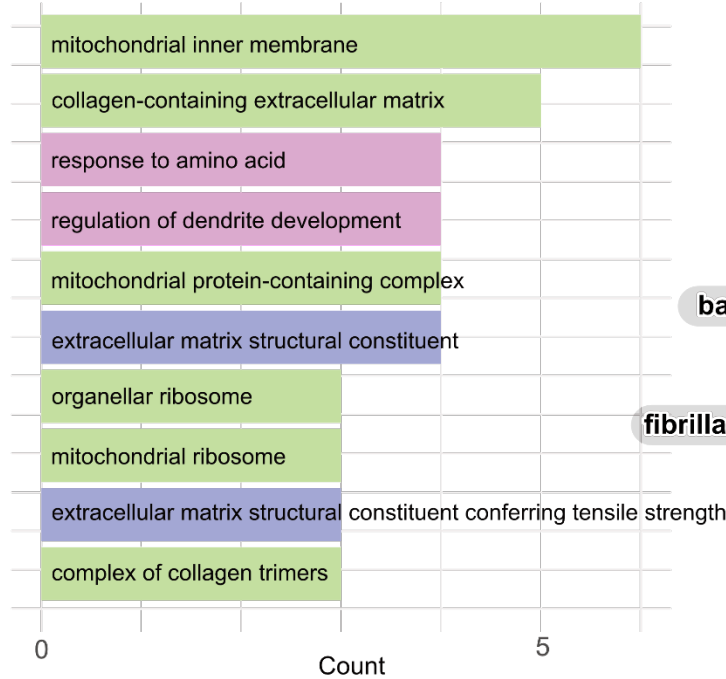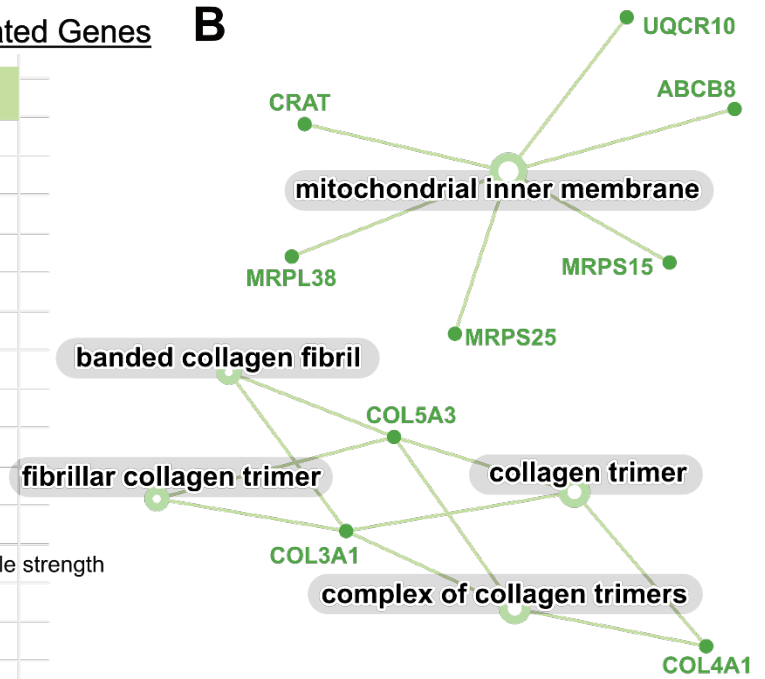
B. Frequency of Age-Upregulated Genes by Score
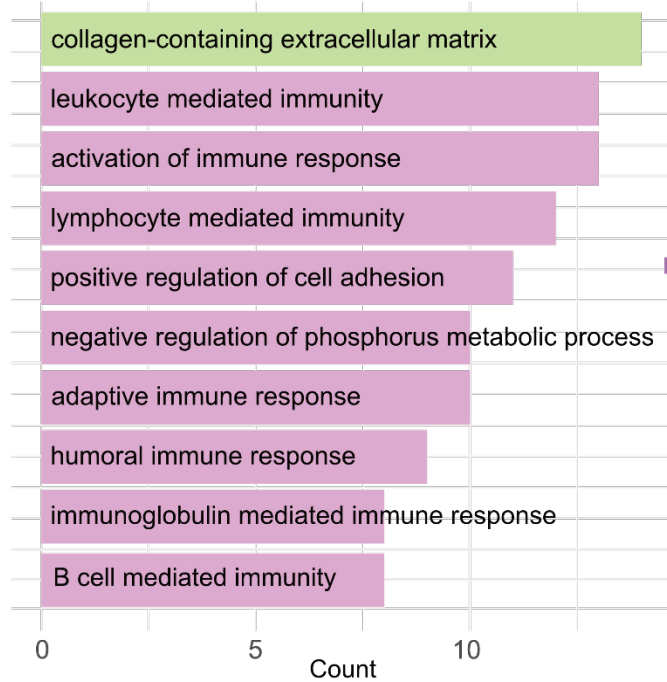
C.

Figure 2

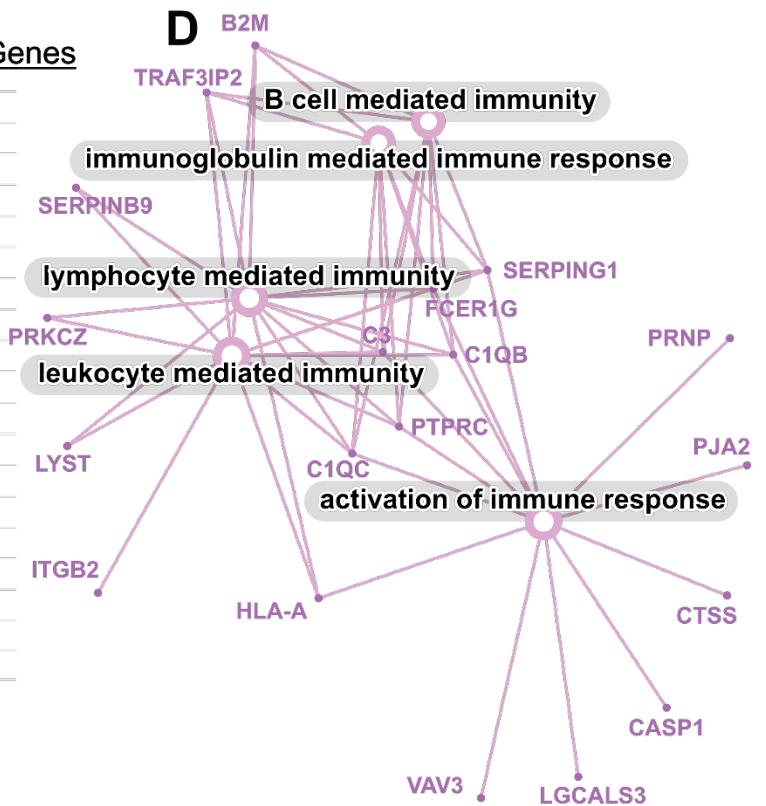A. Enriched GO Terms for Age-Downregulated Genes
B.
C. Enriched GO Terms for Age-Upregulated Genes
D.

Biological Processes   Cellular Components   Molecular Function

Figure 3