Title Age-Invariant Genes: Multi-Tissue Identification and Characterization of Murine Reference Genes

Authors: John T. González[1], Kyra Thrush[2], Margarita Meer[2], Morgan E. Levine[1,2], Albert T. Higgins-Chen[1,3]

Affiliations

[1] Department of Pathology, Yale University School of Medicine, New Haven, CT, USA.

[2] Altos Labs, San Diego Institute of Sciences, San Diego, CA, USA

[3] Department of Psychiatry, Yale University School of Medicine, New Haven CT, USA.

Corresponding Author: Albert T. Higgins-Chen (a.higginschen@yale.edu)

Keywords: housekeeping genes; organ aging; RT-qPCR normalization; stable expression; mammalian transcriptome

Notes: 4 figures, 1 table, 9 supplementary figures and, 17 supplementary tables

Abstract

Studies of the aging transcriptome focus on genes that change with age. But what can we learn from age-invariant genes—those that remain unchanged throughout the aging process? These genes also have a practical application: they serve as reference genes (often called housekeeping genes) in expression studies. Reference genes have mostly been identified and validated in young organisms, and no systematic investigation has been done across the lifespan. Here, we build upon a common pipeline for identifying reference genes in RNA-seq datasets to identify age-invariant genes across seventeen C57BL/6 mouse tissues (brain, lung, bone marrow, muscle, white blood cells, heart, small intestine, kidney, liver, pancreas, skin, brown, gonadal, marrow, and subcutaneous adipose tissue) spanning 1 to 21+ months of age. We identify 9 pan-tissue age-invariant genes and many tissue-specific age-invariant genes. These genes are stable across the lifespan and are validated in independent bulk RNA-seq datasets and RT-qPCR. We find age-invariant genes have shorter transcripts on average and are enriched for CpG islands. Interestingly, pathway enrichment analysis for age-invariant genes identifies an overrepresentation of molecular functions associated with some, but not all, hallmarks of aging. Thus, though hallmarks of aging typically involve changes in cell maintenance mechanisms, select genes associated with these hallmarks resist fluctuations in expression with age. Finally, our analysis concludes no classical reference gene is appropriate for aging studies in all tissues. Instead, we provide tissue-specific and pan-tissue genes for assays utilizing reference gene normalization (i.e., RT-qPCR) that can be applied to animals across the lifespan.

Introduction

Aging, the accumulation of cellular, molecular, and physiological alterations in an organism over time, increases the risk of dysfunction, chronic disease, and mortality [1]. The advent of next-generation sequencing and other high-throughput technologies has allowed for data-driven analyses to discover age-linked gene expression changes and dysregulation. However, little effort has been directed toward identifying and understanding **age-invariant genes** – those that remain unchanged throughout the aging process. The utility of such genes would be twofold: 1) they can be used as reference genes in quantitative assays, and 2) they may share molecular features that allow them to resist changes with age.

The transcriptome has been shown to exhibit substantial remodeling during the aging process, and there is evidence that many of these changes may drive declines in cellular function. By employing bulk RNA-seq across 17 mouse tissues, Schaum et al. identified clusters of genes with similar age trajectories associated with the hallmarks of aging [2]. Gene clusters increasing in expression included immune and stress response genes, while those decreasing in expression included genes involved in the extracellular matrix, mitochondria, and protein folding [2]. Overall, a global decrease in gene expression has been reported to occur with aging, such that when comparing older animals to younger animals, differentially expressed genes tend towards downregulation [3]. For tissue-specific genes, a divergence or specialization of distinct cell types is observed during development, whereas aging has been associated with a loss of specificity in transcriptional profiles [4] and an increase in transcriptional noise (increased variance between individuals) [5–7]. Interestingly, genes subject to age-related change have been linked to specific features, including transcript length and association with CpG islands [8,9].

Studying age-invariant genes that do not change their expression and remain stable throughout the aging process may uncover complementary aging mechanisms. The notion of invariant genes has been a focus of biomedical research for over 50 years, but their study has been confined to young

organisms or cell line perturbations [10]. Due to their relative stability, invariant genes have been utilized as internal reference controls for gene expression assays. Initially coined as housekeeping genes, these invariant genes are constitutively expressed at high levels, are subject to low fluctuations, and are often essential for proper cellular function [10–12]. The changing definition of the term "housekeeping gene" led the *Minimum Information for Publication of Quantitative Real-Time PCR Experiments* (MIQE) guidelines to update the term used for normalization to **reference genes (RGs)** [13], and we will utilize this term. There is no absolute standard list of RGs; many classical RGs, including glyceraldehyde-3-phosphate dehydrogenase (GAPDH), actin β (ACTB), and β2-microglobulin (B2M), were found to be highly variable in certain contexts [11,14]. Although an ultimate RG may not exist (consistent across all possible tissues, cell types, cell cycle stages, experimental conditions, and developmental phases), identification of invariant genes in specific contexts and sample types is possible [14,15].

Little work has been done to identify and validate RGs that are stable throughout the aging process, i.e., age-invariant reference genes. These genes would be invariant across the lifespan, either within any given tissue (tissue-specific) or across all tissues (pan-tissue). Aging is known to impact classical reference gene expression: a mouse study, for example, found age, sex, and frailty explicitly alter the expression of a majority of classical RGs examined [16].[11,14] [29]. Within the aging field, studies are restricted to RGs identified in other fields rather than using a novel, aging-focused analysis. The few available studies examining RGs in aging employ targeted RT-qPCR validation of some of the aforementioned classical transcripts and recommend different RGs based on the genes and the parameters included. For example, GUSB increased with age in mouse skeletal muscle, making it a poor RG in that context, but it was the best RG candidate in human peripheral blood mononuclear cells [16–20]. Another salient example for aging is Cdkn1a/p21. Cdkn1a/p21 is often utilized as a reference gene in RT-qPCR normalization literature [20], even though it simultaneously serves as a marker of cellular

78 senescence–one of the major hallmarks of aging, which is defined by change over time [21,22]. Thus
79 there is a pressing need to identify RGs appropriate for aging studies.
80 We now have the tools and datasets to identify age-invariant RGs. The first iterations of
81 reference genes, which compose a majority of popular RGs, were not experimentally determined but
82 selected because they were detected in all tissues and assumed to have little variability [10,23]. With
83 the development of 21st-century microarray and next-generation sequencing technologies, this
84 question can finally be tackled from a data-rich perspective [23]. RNA-seq datasets have been
85 successfully used to experimentally identify RGs in healthy human tissue [10,11], mammalian animal
86 models [14,24], non-mammalian organisms [25], disease conditions [26] and even single-cell
87 populations [27]. The variables included in the datasets for these analyses determine the application
88 constraints of the resulting RGs. Novel data-rich unsupervised techniques paired with next-generation
89 sequencing data remain an untapped resource for identifying RGs for aging studies and more fully
90 understanding the dynamics of transcriptional change (or lack thereof) with aging.
91 Here, we leverage published approaches for RG identification [10] with appropriate refinements
92 **(Figure 1A)** and apply them to public bulk RNA-seq datasets with samples collected across the full
93 lifespan **(Figure 1B)** to identify age-invariant genes. We show that, unlike our age-invariant genes, no
94 classical RG is suitable for aging studies across all tissues **(Figure 1C)**; and characterize features and
95 functions of these age-invariant genes **(Figure 1D)**. Of note, we opted to focus on the subset of age-
96 invariant genes that can also serve as RGs - those that are also relatively highly expressed - due to their
97 practical applications.
98
99 **Results**
100 Identification of candidate age-invariant genes from RNA-Seq data

Bulk RNA-seq data from the Tabula Muris Senis study [2] were utilized for age-invariant RG discovery. We analyzed 17 tissues: brown adipose tissue (BAT), bone, brain, gonadal adipose tissue (GAT), heart, kidney, limb, liver, lung, marrow, mesenteric adipose tissue (MAT), pancreas, subcutaneous adipose tissue (SCAT), skin, small intestine, spleen and white blood cells (WBCs). We performed quality control and only utilized samples where we could verify the tissue label **(Supplementary Figure 1; Methods)**. The dataset contained female and male mice representing the 4 major lifespan stages: adolescent (1mo), young (3 and 6mo), middle-aged (9, 12, and 15mo), and old (21, 24, and 27mo) [28] **(Figure 1A-B)**.

Tissues were independently analyzed by sequentially applying 7 filtering criteria through each tissue's gene set **(Figure 1A)**. Here, we utilize expression counts normalized to **Transcripts Per Million (TPM)** [29], which is similar to RT-qPCR as it approximates relative molar RNA concentration, as well as **Trimmed Mean of M (TMM)** [30], which leverages inter-sample information to reduce sensitivity to gene outliers. Both normalization techniques performed similarly well at identifying RGs in a recent systematic comparison of normalization methods [26]. Our approach leverages two different normalization techniques to reduce artifacts specific to individual methods. Each criterion, or filter, was applied to each tissue individually with both normalization methods; genes were only included in the tissue-filter gene list if they satisfied the requirement in both TPM and TMM normalized datasets.

Our filtering criteria are listed below. The filtering pipeline was applied to each tissue separately, with samples spanning the lifespan stages defined in **Figure 1B**. Although some genes have been identified as age-invariant within multiple tissues, this does not suggest they are invariant to tissue type and thus should still be applied in a tissue-specific manner. Criteria 1-4 are adapted from an approach frequently used for RG identification from RNA-seq data [10,25]:

1. Continuous expression: Non-zero expression in all samples.

2. Low variance: The **standard deviation (SD)** of the log2 normalized gene (*x*) expression for all samples (*i*) is less than 1. $\forall, (_2()) < 1$

3. No exceptional expression/outliers: log2 normalized values are within two units of the gene's mean(removing genes with data points four-fold away from the gene mean). $\forall, |\ _2() - (_2())\ | \leq 2$

4. Medium to high gene expression: the gene's log2 normalized expression mean is above the mean of all the genes expressed in the particular tissue $\forall, (_2()) \geq (_2(\ ))$

To ensure age-invariant gene list quality, we added two new filters to the identification criteria:

5. Low **coefficient of variation (CV)**: The percent coefficient of variation (%CV), the ratio of the standard deviation to the mean, is lower than 20%. $\forall,\ (_2()) / (_2()) \leq 20$

6. No correlation between gene expression and age: Gene expression correlation with age is not statistically significant (no p-value under 0.05)

Finally, we performed external validation:

7. Filters 5 and 6 were applied in publicly available validation datasets with bulk tissue RNA-seq data from mice. Tissues with a validation dataset were BAT, brain, heart, kidney, muscle, liver, lung, SCAT, skin, small intestine, and WBCs (11).

The filters progressively refined the list of both tissue-specific (**Figure 2A, Supplementary Table 1**) and pan-tissue age-invariant genes (**Figure 2B, Table 1**). For reference, **Supplementary Table 2** lists information on each gene's %CV, slope with age, and correlation with age in each tissue, allowing readers to select their own cutoffs if they choose. **Supplementary Tables 3-9** contain lists of all genes that passed each consecutive filter in each tissue.

There were a few notable modifications to the original pipeline. First, we modified Criterion 4, which selects for relatively highly expressed genes and, therefore, is easily detected by RT-qPCR [10]. Because each tissue had different gene count distributions **(Supplementary Figure 2A)**, we deviated from the previous use of an arbitrary cutoff and employed an adjusted cutoff, removing genes with means below the mean of all genes expressed in a given tissue (log2 transformed) [25]. Consistent with previous publications [25], the cumulative effect of filters 2 (standard deviation cut-off) and 4 (mean cut-off) resulted in a percent coefficient of variation (%CV) of about 20% in most tissues **(Figure 2C)**. However, given the lower average normalized gene expression in some tissues (Bone, Pancreas, Spleen WBC), genes in these tissues surpassed this threshold. To ensure the genes obtained were truly low variance, we applied a hard cut-off of 20% CV (Filter 5). This approach combines Eisenberg et al.'s low variance definition of RGs and their alternative approach: mid-to-high expression [10].

Second, we added Filter 6 to ensure age invariance. We had initially hypothesized that simply analyzing samples with a wide age range using the typical RG pipeline (filters 1-4) would be sufficient to filter out genes that change with age. Indeed, adding age groups to the analysis progressively discarded genes during the filtering process **(Supplementary Figure 3A)**. This, however, could be due to the increase in samples (n) included in the analysis. To test whether the wide age range alone contributes important information, we applied the steps of the standard pipeline (filters 1-4) on samples belonging to only a particular lifespan stage and compared it to a cross-stage control with the same n. Including a wide range of ages by using cross-stage analysis discarded more genes compared to single-stage analysis for adolescent, middle-aged, and old stages **(Supplementary Figure 3B)**. Surprisingly, this was not the case for the young adult stage (3-6mo old); we found this was likely due to a subset of genes that have high expression variability in young adults but are stable in other life stages **(Supplementary Figure 3C-F)**. Regardless of lifestage analyzed, this pattern held true. Genes identified as age-invariant (with filters 1-4) only in young samples **(Supplementary Figure 3C)**; in young samples and other lifestages (analyzed

separately)**(Supplementary Figure 3D)**; in lifestages except young **(Supplementary Figure 3E)**; or with the full dataset, i.e., all lifespan stages, **(Supplementary Figure 3F)** reveal a similar pattern: some genes have higher variance (%CV) in young and old populations. This is reflected by the rightward shift in young and old samples. Young samples have an overall higher proportion of high variance (over the 20%CV dotted line) genes than old ones **(Supplementary Figure 3C-F)**. Thus, simply utilizing a wide age range in the typical pipeline does not necessarily help identify age-invariant genes. Furthermore, we found that some genes obtained through filters 1-5 still changed with age **(Figure 2 D-E)**. To address this finding, we added criterion 6, removing genes with statistically significant correlations with age for each tissue **(Figure 2D)**.

Finally, to decrease the number of false positives, we validated the gene lists using a second bulk mRNA-seq dataset for 11 out of 17 tissues (except for bone, GAT, marrow, MAT, pancreas, and spleen). The number of validated genes is displayed in **Figure 2A-B** as Step 7. Specific counts and percentages can be found in **Supplementary Table 1.** For nearly all tissues, a supermajority (>70%) of candidate age-invariant genes were validated, except in the liver (54%) and lung (62%). The fewest number of age-invariant genes was observed in WBCs, possibly due to large changes in distributions of cell types over shorter timescales [31,32]**(Figure 2A)**.

<u>RT-qPCR validation of Novel age-invariant reference genes</u>

Our analysis identified many tissue-specific age-invariant RGs (**Supplementary Table 9**), as well as 9 such genes common to all tissues (pan-tissue). Some classical RGs are not age-invariant genes **(Figure 3A-C)**. In fact, no classical RGs were age-invariant across every tissue **(Figure 3A)**. Thus, we propose a new list of 9 age-invariant genes common to all 17 tissues that can be used in studies with aged animals **(Table 1)**.

These RG sets can be utilized in the context of northern blot, RT-qPCR, and some RNA-seq normalization strategies in aging studies. Researchers have the choice of selecting from a tissue-specific gene list or from the nine pan-tissue genes. To validate this, independent samples were used to generate RT-qPCR data for three age-invariant genes identified by our computational pipeline: Atp6v1f, Srp14, and Tomm22 **(Figure 3B-D)**. Atp6v1f is an age-invariant gene shared by the two tissues assayed: the liver and heart. The other two are pan-tissue age-invariant genes. The novel samples consisted of mouse heart and liver samples in four categories: old (~19mo) female, old male, young (~8mo old) female, and young male. We compared these against three classical RGs: Cdkn1a, Tbp, and Tfrc. Classical reference genes generally had a wider cycle threshold distribution than the age-invariant genes, with Tbp being the most stable among them, followed by Tfrc and Cdkn1a **(Figure 3B-C)**. Cdkn1a codes for cyclin-dependent kinase inhibitor 1A, also known as p21. Given that Cdkn1a is widely used as a marker of cell senescence [22], it is not surprising that it has a high degree of variability despite it being widely considered an RG in RT-qPCR normalization literature [17].

To assess gene RT-qPCR stability in the context of aging, we calculated the expression stability across multiple algorithms: BestKeeper [33] **(Supplementary Figure 4A-B)**, geNorm [34] **(Supplementary Figure 4C-D),** NormFinder [35] **(Supplementary Figure 4E-F)**, and delta-CT method [36] **(Supplementary Figure 4G-H)**. These scores were utilized to calculate the summary RefFinder score **(Figure 3D, Supplementary Figure 4I)** [37]. TPM %CV for the discovery **(Supplementary Figure 4A, C, E, G, J)** and validation **(Figure 3D, Supplementary Figure 4B, D, F, H)** RNA-seq datasets strongly correlate with all stability algorithm values calculated on our in-house samples. **Figure 3D** displays the correlation between the TPM %CV values of the external validation dataset and the RefFinder score of our in-house validation samples (Pearson correlation = 0.81, p-value = 0.0027). This suggests that %CV from normalized RNA-seq samples could be used as an indicator of candidate reference genes for RT-qPCR experiments subject to the same conditions. By both metrics, the newly identified age-invariant genes

outperformed the classical RGs: these RGs are statistically different in both %CV (Welch Two Sample t-test p-value= 0.006553) and RefFinder qPCR scores (Welch Two Sample t-test p-value = 0.02401) This suggests age-invariant genes common across all tissues (Srp14 and Tomm22) or particular tissues (Atp6v1f in heart and liver) could be applied as part of normalization in age-related transcriptomic research. A combination of more than one of the age-invariant genes is recommended for RT-qPCR experiments, per the MIQE guidelines [13].

Overlapping pathways for aging stable and aging dysregulated genes

Gene enrichment analysis of the tissue-specific age-invariant genes revealed a large number of statistically significant GO biological pathway terms **(Supplementary Figure 5)**. As expected, the most enriched terms were largely involved in basic metabolic and structural processes **(Figure 4A)**. We also noted many enriched terms were related to the hallmarks of aging [38], which was surprising considering that hallmarks of aging are typically thought to involve processes that change with age. As an initial step to systematically assess the presence of stably transcribed genes in these hallmarks, we compared the enrichment scores of our tissue age-invariant gene lists with previously published enrichment terms associated with age dysregulation and disease [2,39].

We first compared our enrichment scores with the top terms associated with mouse transcriptome aging clusters, each displaying a different trajectory with aging **(Figure 4A)**. The top enrichments of these 10 clusters, obtained from the same dataset we performed our discovery on, are associated with hallmarks of aging like protein folding, inflammation, and mitochondrial function [2]. We found our tissue gene sets were significantly enriched in many, but not all, of the clusters. Of note is cluster 3, linked to mitochondrial dysfunction, where age-invariant genes are highly enriched for every term of this cluster. Age-invariant genes are also heavily represented in stress response (cluster 5), signaling (cluster 2), and protein stability (cluster 7). Interestingly, within the protein stability cluster,

age-invariant genes were enriched in terms involved in protein folding, processing, and stabilization but not in terms involved in protein localization. The clusters with the least age-invariant genes were those associated with immune response and extracellular matrix. This suggests that hallmarks themselves, or mechanisms within an aging hallmark, can be separated by the presence or absence of age-invariant genes.

Cluster 1 from Schaum et al. is defined as genes that do not change with age and, as expected, has a large overlap with our tissue age-invariant gene sets. Cluster 1 was defined by having the least amplitude (change with age) and least variability. Interestingly, throughout the 17 tissues, only ~33-40% of our age-invariant genes were in Schaum et al.'s cluster 1. The genes not shared between both methods likely reflect the difference between relatively a stable group of genes identified by hierarchical clustering and individual age-invariant genes identified due to their characteristics (as well as our RG requirement that genes be highly expressed) [2,40]. In RNA-seq, genes with low expression demonstrate significant technical noise making it difficult to assess true biological variability related to age or other factors, and are often filtered out of differential expression studies [41], so our requirement for high expression is useful for focusing on age-invariant genes.

The other ontology terms we examined came from an analysis of age-related diseases and aging hallmarks **(Figure 4B)**. Unlike Schaum et al., who used a completely unsupervised approach, Fraser et al. used genes associated with human age-related diseases in a genome-wide association study to define GO biological pathways related to both disease and at least one aging hallmark [39]. Most hallmarks have at least one GO term enriched for age-invariant genes across most of the tissues analyzed (e.g., "steroid hormone-mediated signaling pathway" in altered intercellular communication; "cellular response to insulin stimulus" and "response to nutrient levels" in deregulated nutrient sensing; "macroautophagy" and "regulation of autophagy" for loss of proteostasis; "reactive oxygen species metabolic process" in mitochondrial dysfunction; and "telomere maintenance" in telomere attrition). On

the other hand, virtually no GO term related to cellular senescence and epigenetic alterations had high proportions of stably transcribed genes. According to this alternative way of identifying gene ontology terms associated with aging hallmarks, age-invariant genes continue to be enriched in these terms.

To better understand the implications of some of these stable pathways, we used the comprehensive resource of mammalian protein complexes (CORUM) database to perform enrichment analysis **(Figure 4C)**[42]. The enriched complexes are consistent with our enrichment results in this data thus far. Complexes involved in mitochondrial function (respiratory chain complex I and cytochrome c oxidase), stress response & signaling (Regulator-AXIN/LKB1-AMPK complexes), and protein stability (COP9 signalosome, proteasome, Parvulin-associated pre-rRNP, and Chaperonin containing TCP1 Complex) are enriched in age-invariant genes.

Our analyses reveal multiple age-invariant genes within pathways that are either dysregulated with aging **(Figure 4A, C)** or associated with aging pathologies **(Figure 4B)**. Pathways related to the extracellular matrix, cellular senescence, and epigenetic alterations seem particularly devoid of stably expressed genes. These findings are not due to the high expression requirement for our age-invariant genes, as removing this requirement produced similar results **(Supplementary Figure 6)**.

Age-invariant gene features

Features of genes that change with age have long been a point of discussion in aging transcriptome research, but little is known about the genes that are able to withstand the effects of time. We tested whether our genes have the opposite features to those described in age-dysregulated transcriptome analyses. The features examined are CpG content, DNA methylation (**Supplementary Figures 7-8),** and gene length **(Supplementary Figures 7 and 9),** given that these features have been implicated in age-associated transcriptional drift [8,9].

Lee and colleagues reported that genes with CpG islands (CGI+) are less likely to change with age than genes without CpG islands (CGI-) [9]. Accordingly, we found that the proportion of genes with CpG islands located in their promoters increased as a function of our filtering process, suggesting that as we more rigorously select for age-invariant genes, the more prevalent promoter CpG islands become **(Supplementary Figure 7A)**. The transcripts themselves were not enriched for greater %CG content, suggesting there is biological specificity of the function of these islands versus an overall increase in CG content in the region **(Supplementary Figure 9D).** We next investigated whether age-invariant genes also showed greater stability in promoter methylation status during in vitro passaging or in vivo aging using reduced-representation bisulfite sequencing (RRBS) datasets. For mouse embryonic fibroblasts serially passaged into senescence, we found both age-variance (based on our skin tissue-specific notation) (**Supplementary Figure 8A**), and CGI (**Supplementary Figure 8B**) status influenced methylation variability. Regardless of age-invariant RG status, CGI+ genes are more stable than CGI- genes **(Supplementary Figure 8A)**. However, this pattern was not observed in mouse tissues, including liver, brain, heart, lung, or WBC **(Supplementary Figure 8B)**.

Stroeger et al. report that median transcript length is the factor most associated with age-related change, with longer transcripts tending to be downregulated and shorter transcripts tending to be upregulated with age [8]. Complementing these findings, we found that age-invariant genes tend to be shorter than age-variant genes when comparing minimum transcript length(**Supplementary Figure 7B**). However, the opposite is true when comparing either maximum (**Supplementary Figure 9A**) or Ensembl canonical (**Supplementary Figure 9C**) transcript length.

Discussion

Much of aging biology research has focused on changes that occur across the organismal lifespan and the contribution of these changes to age-related mortality, morbidity, and functional

decline [1,38]. Molecular signatures that are robust to aging – specifically, age-invariant genes – have received comparatively little attention. Identifying age-invariant genes allows for further study of why they do not change with age. Lessons from these age-resilient genes provide a complementary view of aging and the stability of biological systems with time. Also, from a practical perspective, because many genes change with age, it is important to identify age-invariant genes for use as **reference genes (RGs)** for gene expression normalization [13]. By adopting a pipeline for identifying RGs from RNA-seq data, we find that there are, in fact, hundreds to thousands of age-invariant genes per tissue. Strikingly, there is poor agreement between the pan-tissue age-invariant genes and commonly used classical RGs. According to our results, none of the classical RGs are suitable for use in cross-sectional aging studies across the 17 tissues studied **(Figure 3A)**, and some canonical tissue-gene pairings (e.g., GAPDH in the liver) are not age-invariant [43]. Our novel age-invariant genes are, therefore, better suited than classical RGs for performing normalization for RT-qPCR experiments in aging tissues.

We report nine pan-tissue age-invariant genes in mice (**Table 1**). Reference and housekeeping gene literature postulates that continuously and stably expressed genes serve essential cellular and organismal functions [12]. Consistent with this hypothesis, depletion of 7 out of 9 of our pan-tissue age-invariant genes have already been reported to induce cell (1110004F10Rik) or embryonic lethality when completely knocked out (Brk1, Rer1, Psmd4, Reco2, Tomm22, and Fis1) [44–49], according to the Mouse Genome Informatics database (www.informatics.jax.org) or International Mouse Phenotyping Consortium database (www.mousephenotype.org). The remaining two transcripts, Srp14 and Gemin7, have no reported knockout mouse strain or phenotypes, but we hypothesize would be lethal if absent.

Two biological processes— mitochondrial function (Fis1, Rexo2, and Tomm22) and proteostasis (Psmd4, Rer1, and Srp14)— emerge from these 9 genes. Although these biological processes are implicated in aging changes, they may also contain components that remain highly stable across the lifespan. Rexo2 (RNA exonuclease 2) was recently shown to increase mitochondrial gene transcription,

mediate RNA turnover, and enforce promoter specificity in mammalian mitochondrial transcription [48]. Rer1 returns rogue ER-resident proteins or unassembled subunits in the Golgi apparatus back to the endoplasmic reticulum [46]. Little is known about the molecular function of the small acidic protein 1110004F10Rik (also known as Smap) or its human ortholog C11orf58, but given its high stability and requirement for cell survival, this protein may merit further attention [44]. Thus, the stability of these 9 genes may have evolved as a result of these genes being critical for mitochondrial and proteostatic function, and for continued life in the face of age-related deterioration.

Simply including older mice in our study and utilizing the standard RG identification pipeline was insufficient at filtering out age-invariant genes. Rather, selecting for age-invariant genes required an additional step of explicitly removing genes that are correlated with age. We also find that the variance in expression of a given gene often changes across life stages. For instance, we identified more genes having high variance in young age than in middle or old ages (**Supplementary Figure 4**). Although perhaps surprising, this finding is consistent with reports indicating the proportion of genes decreasing in variance with age is greater than those increasing in variance with age [6,7,50]. It is possible that younger animals show greater variance related to circadian rhythms, the estrous cycles, sex differences, response to stress, or other adaptive and cyclical factors.

Some limitations and caveats constrain our study. First, some of the specific cutoffs we utilized were based on prior work, while others (e.g., exact age correlation cutoff) were based on our best judgment. We provide a complete table of filter results in **Supplementary Table 2** in case others wish to utilize different cutoffs in selecting RGs. To ensure the list of genes provided are useful reference genes in normalization strategies, including RT-qPCR and even some RNA-sequencing normalization approaches, we required high transcript expression through Filter 4. Although consistent with normalization transcript identification strategies in RNA-seq, many low-expression age-invariant genes are absent. Thus, our lists report age-stable, high-expression genes only. Our findings are influenced by

the technical limitations of RNA-seq [10,51] and the analytical limitations of high dimensional data, including subsampling of highly heterogeneous samples like aged organisms previously described in the literature [10,51,52]. However, variance in sample collection, processing, and preparation across these datasets likely compensate for any individual source's batch and degradation bias (e.g., each of the four datasets used employs a different poly-A sample preparation kit). Our final 9 pan-tissue age-invariant genes have been tested individually in 17 tissues and four datasets, totaling 1120 samples, thereby reducing the risk of, for example, a type I error (wrongly identifying a gene as age-invariant). Finally, an important assumption not usually discussed in aging transcriptome literature may influence interpretation in the context of aging: consistent RNA mass. A few studies suggest a decline in total cellular RNA mass with aging [53,54]. This is different from the reported downward trend of differentially expressed genes with age [3]. Current RNA sequencing analysis techniques use proportional estimates (counts per million, fragments per kilobase of transcript per million, transcripts per million, etc.) to normalize samples in order to compare transcript dynamics across samples. Similarly, RT-qPCR protocols typically rely on standardizing total RNA input. If total RNA mass reduction is a global feature of cellular aging, our age-invariant genes are proportionally stable but may decrease in mass with age. Similarly, a gene identified to be overexpressed in old age may maintain constant molar concentration within a cell or tissue. We recommend readers keep these considerations in mind when interpreting any gene expression study in the context of aging.

The existence and study of age-invariant genes have the potential to provide the field of aging with novel insights. It was interesting to find that age-invariant genes were enriched for some pathways associated with hallmarks or pillars of aging **(Figure 4)**, specifically nutrient sensing, proteostasis, mitochondrial function, and immune function. This is somewhat puzzling given that such hallmarks are defined by changes thought to play putatively causal roles in aging [22,55] indeed, genes that most clearly change with age are enriched in the same hallmarks [2]. It is possible enrichment in pathways

associated with hallmarks of aging may simply reflect the fact that hallmarks of aging are broad and cover much of biology. In that case, it may be necessary to more specifically delineate each hallmark of aging, e.g., perhaps only a subset of nutrient sensing processes should be considered as a hallmark. However, this broadness would not explain why some hallmarks of aging are associated while others are not. What might be the significance of genes associated with hallmarks of aging that remain stably expressed throughout aging? We note that a prior report indicated that essential genes are enriched for pro-longevity functions, as experimental overexpression of essential genes tends to increase lifespan in yeast [56]. We also find that age-invariant genes are present in pathways linked to human age-related diseases **(Figure 4A-B)**. If age-invariant genes are essential for life, then organisms may have evolved mechanisms to keep these genes stable in the face of pervasive age-related changes in the rest of the pathway or network. One potential example highlighted here is the age-invariant gene enrichment of protein complexes in the electron transport chain. NADH:ubiquinone oxidoreductase, or Mitochondrial Respiratory Complex I, is the only age-invariant gene-enriched ECT complex throughout most tissues **(Figure 4C)**. Although the downregulation of ETC genes is one of the most established transcriptional events in aging [52] and protein Complex I proteins undergo major changes in abundance with age [57], stability in some ETC components is likely required for continued life. This is consistent with Complex I being one of the ETC complexes that can be traced back to the last universal common ancestor of all living organisms [58]. Significant dysregulation of such essential components may be incompatible with life, and evolutionary forces may ensure stability throughout the lifespan. It will be interesting to determine whether further bolstering the expression or stability of such age-invariant genes may be a pro-longevity strategy or, if given their continous expression stable genes are good aging pharmacological targets. The putative aging intervention metformin, for example, may benefit from the stable expression of it's target, Complex I [59].

In contrast, age-invariant genes were not enriched in some hallmarks, including epigenetic

alterations, cellular senescence, and the extracellular matrix. Our results suggest that these three are the most vulnerable to aging as not many genes related to these hallmarks resist age-related change. In agreement with this finding, these hallmarks are key targets across many existing longevity interventions, i.e., epigenetic reprogramming, senolytics, and enhancing extracellular matrix homeostasis [60–62]. Considering that age-invariant genes tend to be essential for life, one hypothesis is that early changes in these hallmarks may not be particularly detrimental for the organism and thus lack the selective pressure to evolve stability mechanisms in aging. The cumulative long-term burden of changes, however, may contribute to pathological aging. Alternatively, these variant hallmarks may reflect adaptive processes that evolved to change dynamically with aging for the benefit of the organism.

Future analyses could focus on the processes that maintain the stability of age-invariant genes. Our initial investigations demonstrate that age-invariant genes are enriched in CpG islands, consistent with a previous report that genes with CpG islands are more resistant to age-related dysregulation than those without CpG islands, which are misexpressed during age-related heterochromatin decondensation [9]. However, further analyses are needed to determine whether the resistance to changes in the methylome of CpG-rich promoters was responsible for the stability of gene expression over time. For instance, whether increased CpG density is better able to reinforce a stable epigenetic state.

We also found that age-invariant genes tend to be shorter than others, confirming a previous study reported that the longest genes show the greatest degree of downregulation [8]. Further study is needed to better understand the relationship between expression dynamics and transcript length. Of note, classical RGs in general have been reported to exhibit shorter introns and exons, low promoter region conservation, 5’ regions with fewer repeated sequences, low nucleosome formation potential, and a higher SINE to LINE ratio [10,24]. It will be important to determine if and how these factors may contribute to the stability of age-invariant genes.

Lastly, it will be important to determine the translatability of our age-invariant transcripts, both to other organisms as well as to protein expression. In a recent study, 52% of human reference genes were matched to independently analyzed mouse reference gene orthologs [14]. Protein abundance can be inferred from transcriptomic data at the tissue and single-cell level, particularly for genes continuously and stably expressed [63,64]. These transcripts show a high correlation (~0.7) with their protein product except when variability is introduced by cellular state and microenvironment conditions. Given that age-invariant genes are assumed to be expressed in steady-state, many of these genes may also be age-invariant at the protein level.

Here, we provide the aging field with a list of 9 pan-tissue age-invariant genes for use in normalization strategies, e.g., RT-qPCT; we observe that age-invariant genes are enriched in ontology terms associated with some, but not all, hallmarks of aging; and we explore some common features of age-invariant genes (CpG island status and transcript length). Be it for understanding the basic biology of aging, establishing rigorous methodology in the field, investigating the mechanisms promoting age-invariance vs. age-variance, or finding aging therapeutic targets, age-invariant genes are an important area of study.

Methods

Data Preparation and Normalization

Four datasets were utilized in this analysis. The Discovery Dataset (GSE132040) consisted of 17 male and female tissues from mice spanning the 4 major life span stages **(Figure 1B)**. 11 of 17 tissues were validated with three datasets of bulk-RNA tissue data from male mice: GSE167665, GSE111164, and GSE141252. Count tables were obtained from GEO and normalized as described below. Sample preparation and alignment can be found in their respective publications [2,4,8]. 5 million counts/sample

were set as the count threshold for a sample to be included in normalization and further analysis. In the discovery dataset, hierarchical clustering identified a small number of samples that clustered away from their labeled tissue **(Supplementary Figure 1A)**, and examination of tissue-specific markers confirmed they may be mislabeled and, therefore, were removed from analysis **(Supplementary Figure 1B)**. The number of samples removed per tissue and lifestage can be seen in **Supplementary Table 11** and those used in the rest of the analysis in **Supplementary Table 10**. GEO accession number, tissue type, and life stage counts can be found in **Supplementary Table 12** for validation datasets. Here, intestine labels refer to samples from both the large and small intestine; and brain to those from both the cerebellum and the frontal cortex.

RNA-seq normalization is essential for proper downstream analysis of datasets. In this study, we identified our genes with two normalization approaches: TPM and TMM. The original reference gene discovery approach described by Eisenberg and Levanon in 2013 [10], utilized RPKM normalized data. Around the same time, conversations about proper data processing produced **Transcript Per Million (TPM)**, an intra-sample normalization method that approximates **relative molar RNA concentration (rmc)** [29]. TPM was only incorporated into this RG identification approach in 2019 [25]. Another major strategy for data normalization techniques involves between-sample normalization. To prevent normalization-based artifacts, and given there is no single best normalization approach, the discovery data was normalized with two different approaches: TPM and **Trimmed Mean of M (TMM)** [30]. TMM, an inter-sample normalization method, generates a normalization factor assuming most genes are not differentially expressed. Therefore, TPM is akin to RT-qPCR due to its similarity with rmc while TMM leverages inter-sample information and is less sensitive to gene outliers. Both performed similarly well at identifying RGs in a recent systematic comparison of normalization methods [26].

TPM normalized data was calculated following the formula:

$$TPM = \frac{\#\ \text{reads mapped to transcript}/\text{transcript length}}{} \times 10^6$$

Sum(#reads mapped to transcript/ transcript length)

Transcript lengths used in the above formula were obtained with EDASeq package's (version 3.13) getGeneLengthAndGCContent function. TMM was calculated using the calcNormFactors function from the edgeR package (version 3.40.1).

Gene expression plotting and validation data were performed only with TPM normalized data. Plots were generated with ggplot2(version 3.4.0), ggforce (version 0.4.1) and ggdendro (version 0.1.23).

Gene Filtering Process

Filters were applied sequentially in R (version 4.2.2) as described in Results. Most mathematical calculations used the r base and MatrixStats package (version 0.63.0). The filter criteria were applied sequentially in both TMM and TPM normalized data, separately for each tissue, thus yielding different lists for each tissue. For each filter, x is either TMM or TPM, and genes were required to pass the filter for both TMM and TPM. Requirements were defined as follows:

1. For each gene: no empty or 0 values
2. For each gene: $\forall, (_2()) < 1$
3. For each gene: $\forall, |\ _2() - (_2())\ | \leq 2$
4. For each gene: $\forall, (_2()) \geq (_2(\ )$
5. For each gene: %CV≤ 20. $\forall,\ (_2()) / (_2())\ 100 \leq 20$
6. For each gene: No correlation with age, based on Pearson's correlation p-value= 0.05/n. WGCNA package (version 1.71) function corAndPvalue was used to obtain correlation coefficients and p-values. Because each tissue had a 5% chance of finding an association by chance with a fixed 0.05 p-value, a gene present in 17 tissues would have a 58% chance of being erroneously discarded 1-$(0.095)^{17}$. We applied a fractional threshold of a 0.05 p-value, where the p-value threshold applied was 0.05/n, where n is the number of tissues in which the gene in question passed filters 1-4.

7. For each gene: %CV≤ 20 and Spearman correlation p-value= 0.05/n in a validation dataset. n= number of tissues a given gene is present in at filter criteria 6. This step was applied only to TPM normalized data

RNA isolation and cDNA synthesis

Frozen liver and heart tissues were gifts from Prof. Ron Korstanje at The Jackson Laboratories. Groups consisted of 3 samples per age (8 and 18 months) and sex (female and male), except there was only one sample for an 18-month-old female liver. RNA was isolated with RNeasy Plus Mini Kit (Qiagen #74134) with pestle and syringe homogenization. cDNA was generated using Iscript gDNA Clear cDNA Synthesis (Bio-Rad #1725035) and equivalent RNA mass per 20uL reaction (500ng of heart and 1ug of liver). RNA concentrations were determined with a Qubit 4 fluorometer (Thermo Fisher #Q33238) and RNA BR Assay Kit (Thermo Fisher Q10210).

Expression data and RG stability

RT-qPCR reactions were assembled with equivalent SsoAdvanced Universal SYBR Green Supermix (Bio-Rad #1725272), cDNA, and respective PrimePCR SYBR Green primers (Bio-Rad #10025636, AssayIDs Atp6v1f: qMmuCID0014923, Cdkn1a:qMmuCED0046265, Srp14: qMmuCID0020464, Tbp:qMmuCID0040542, Tfrc:qMmuCID0039655, Tomm22: qMmuCED0046631). RT-qPCR was performed in a CFX96 thermocycler (Bio-Rad). Stability algorithms NormFinder [35], BestKeeper [33], geNorm [34], and delta-CT method [36] were calculated and integrated into RefFinder [37]. All calculations were performed in R. geNorm and BestKeeper were calculated with the ctrlGene package (version 1.0.1) [65], Normfinder algorithm was downloaded from moma.dk, delta-CT method and RefFinder functions were recreated as originally described. Metadata for the samples used can be found in **Supplementary Table 13**, cycle threshold results in **Supplementary Table 14** for the heart, and **Supplementary Table 15** for the liver.

CpG island and methylation variability analysis

Gene CpG island (CGI) status was mapped to the annotated list from Lee et al. [9]. Gene names passing each criterion/filter for each tissue were annotated, and percent positive and negative CGI proportion was calculated. Mean and standard deviation were calculated across tissues for each criterion/filter. Counts and percentages of CGI distributions in tissue lists by filter, the odds ratio, statistical test used, and associated p-value are listed in **Supplementary Table 17**.

Composite multi-tissue murine RRBS data [66] was mapped to the mm9 gtf gencode genome. For mouse embryonic fibroblasts, data alignment was previously described [67]. For both datasets, CpG sites common to at least 10 samples and covered by more than 5 reads were analyzed. The methylation status of the promoter region was estimated by averaging the CpG beta values enclosed within 1kb of the transcription start site. Standard deviation was calculated for the methylation of each promoter.

Enrichment gene analysis

Enrichment analysis was performed using gprofiler2's (0.2.1) gost function. Electronically annotated GO terms were included in the analysis, and a common custom background of genes expressed at least once in every tissue was imputed. Bonferroni correction was used to calculate enrichment significance. Aging hallmark trajectory enrichment terms were obtained from Schaum et al. [2], while GO biological process terms associated with age-related disease and aging hallmarks were obtained from Fraser et al. 2022 [39]. A few GO terms identified by Schaum et. al. have been discontinued and are marked as obsolete. These terms were excluded from our analysis. Lastly, the top 20 age-invariant GO (biological process, cellular component, and molecular function), KEGG, and Reactome terms were determined by ranking p-values within tissues and taking the lowest 20 gene rank sums across tissues.

For the enrichment maps, all 17 sets of enrichment terms (one per tissue) were used in EnrichmentMap in Cytoscape to generate a consensus network. Different consensus parameters used were used for the CORUM [42] (P-value: 0.05, FDR Q-value: 0.05, Jaccard Overlap Combined: 0.375, test used: Jaccard Overlap Combined Index, k constant = 0.5) and GO:BP terms (P-value: 0.01, FDR Q-value: 0.01, Jaccard:

0.25, test used: Jaccard Index) networks. AutoAnnotate identified common terms for clusters of interconnected nodes. Each node is a pie chart with each slice colored by the enrichment score of each tissue [68].

Acknowledgements

The research presented here would not have been possible without the exceptional insight of our Biorad Sales Manager, Julie Brunelle. We also thank Profs. Silvia Vilarinho, Rachel Perry and Chen Liu for their insight, support, and advice. We thank Raghav Seghal, Yaroslav Markov, Jessica Kasamoto and Jenel Fraij Armstrong for comments on the manuscript.

Author Contributions

JTG conceived the study. ATCH, MEL, and JTG designed the study and interpreted the data. JTG performed all experiments and data adquisition in this paper. JTG, MM, and KT developed the R code used. JTG and ATCH wrote the article with feedback from the other authors. All authors approved of the submitted manuscript.

Conflicts of Interest

A.H.C. has received consulting fees from TruDiagnostic and FOXO Biosciences for work unrelated to this publication. All other authors report no biomedical financial interests or potential conflicts of interest.

Funding

This work was supported by the National Institute on Aging (NIA: 5R01AG065403) and by the National Institutes of Health grant to The Jackson Laboratory Nathan Shock Center of Excellence in the Basic Biology of Aging (P30AG038070)

References

1. Aging Biomarker Consortium, Bao H, Cao J, Chen M, Chen M, Chen W, Chen X, Chen Y, Chen Y, Chen Z, Chhetri JK, Ding Y, et al. Biomarkers of aging. Sci China Life Sci [Internet]. 2023; 66: 893–1066. Available from: http://dx.doi.org/10.1007/s11427-023-2305-0

2. Schaum N, Lehallier B, Hahn O, Pálovics R, Hosseinzadeh S, Lee SE, Sit R, Lee DP, Losada PM, Zardeneta ME, Fehlmann T, Webber JT, McGeever A, et al. Ageing hallmarks exhibit organ-specific temporal signatures. Nature [Internet]. 2020; 583: 596–602. Available from: http://dx.doi.org/10.1038/s41586-020-2499-y

3. Zhang MJ, Pisco AO, Darmanis S, Zou J. Mouse aging cell atlas analysis reveals global and cell type-specific aging signatures. Elife [Internet]. 2021; 10. Available from: http://dx.doi.org/10.7554/eLife.62293

4. Izgi H, Han D, Isildak U, Huang S, Kocabiyik E, Khaitovich P, Somel M, Dönertaş HM. Inter-tissue convergence of gene expression during ageing suggests age-related loss of tissue and cellular identity. Elife [Internet]. 2022; 11. Available from: http://dx.doi.org/10.7554/eLife.68048

5. Kedlian VR, Donertas HM, Thornton JM. The widespread increase in inter-individual variability of gene expression in the human brain with age. Aging [Internet]. 2019; 11: 2253–80. Available from: http://dx.doi.org/10.18632/aging.101912

6. Viñuela A, Brown AA, Buil A, Tsai P-C, Davies MN, Bell JT, Dermitzakis ET, Spector TD, Small KS. Age-dependent changes in mean and variance of gene expression across tissues in a twin cohort. Hum Mol Genet [Internet]. 2018; 27: 732–41. Available from: http://dx.doi.org/10.1093/hmg/ddx424

7. Barth E, Srivastava A, Stojiljkovic M, Frahm C, Axer H, Witte OW, Marz M. Conserved aging-related signatures of senescence and inflammation in different tissues and species. Aging [Internet]. 2019; 11: 8556–72. Available from: http://dx.doi.org/10.18632/aging.102345

8. Stoeger T, Grant RA, McQuattie-Pimentel AC. Aging is associated with a systemic length-associated transcriptome imbalance. Nature Aging [Internet]. nature.com; 2022; . Available from: https://www.nature.com/articles/s43587-022-00317-6

9. Lee J-Y, Davis I, Youth EHH, Kim J, Churchill G, Godwin J, Korstanje R, Beck S. Misexpression of genes lacking CpG islands drives degenerative changes during aging. Sci Adv [Internet]. 2021; 7: eabj9111. Available from: http://dx.doi.org/10.1126/sciadv.abj9111

10. Eisenberg E, Levanon EY. Human housekeeping genes, revisited [Internet]. Trends in Genetics. 2013. p. 569–74. Available from: http://dx.doi.org/10.1016/j.tig.2013.05.010

11. Caracausi M, Piovesan A, Antonaros F, Strippoli P, Vitale L, Pelleri MC. Systematic identification of human housekeeping genes possibly useful as references in gene expression studies. Mol Med Rep [Internet]. 2017; 16: 2397–410. Available from: http://dx.doi.org/10.3892/mmr.2017.6944

12. Zhu J, He F, Song S, Wang J, Yu J. How many human genes can be defined as housekeeping with current expression data? BMC Genomics [Internet]. 2008; 9: 172. Available from: http://dx.doi.org/10.1186/1471-2164-9-172

13. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem [Internet]. 2009; 55: 611–22. Available from: http://dx.doi.org/10.1373/clinchem.2008.112797

14. Hounkpe BW, Chenou F, de Lima F, De Paula EV. HRT Atlas v1.0 database: redefining human and

mouse housekeeping genes and candidate reference transcripts by mining massive RNA-seq datasets. Nucleic Acids Res [Internet]. 2021; 49: D947–55. Available from: http://dx.doi.org/10.1093/nar/gkaa609

15. Andersen, Jensen, Ørntoft. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied …. Cancer Res [Internet]. Available from: https://aacrjournals.org/cancerres/article-abstract/64/15/5245/511517

16. Mishra M, Kane AE, Young AP, Howlett SE. Age, sex, and frailty modify the expression of common reference genes in skeletal muscle from ageing mice. Mech Ageing Dev [Internet]. 2023; 210: 111762. Available from: http://dx.doi.org/10.1016/j.mad.2022.111762

17. González-Bermúdez L, Anglada T, Genescà A, Martín M, Terradas M. Identification of reference genes for RT-qPCR data normalisation in aging studies. Sci Rep [Internet]. 2019; 9: 13970. Available from: http://dx.doi.org/10.1038/s41598-019-50035-0

18. Touchberry CD, Wacker MJ, Richmond SR, Whitman SA, Godard MP. Age-related changes in relative expression of real-time PCR housekeeping genes in human skeletal muscle. J Biomol Tech [Internet]. 2006; 17: 157–62. Available from: https://www.ncbi.nlm.nih.gov/pubmed/16741243

19. Zampieri M, Ciccarone F, Guastafierro T, Bacalini MG, Calabrese R, Moreno-Villanueva M, Reale A, Chevanne M, Bürkle A, Caiafa P. Validation of suitable internal control genes for expression studies in aging. Mech Ageing Dev [Internet]. 2010; 131: 89–95. Available from: http://dx.doi.org/10.1016/j.mad.2009.12.005

20. Uddin MJ, Cinar MU, Tesfaye D, Looft C, Tholen E, Schellander K. Age-related changes in relative expression stability of commonly used housekeeping genes in selected porcine tissues. BMC Res Notes [Internet]. 2011; 4: 441. Available from: http://dx.doi.org/10.1186/1756-0500-4-441

21. González-Gualda E, Baker AG, Fruk L, Muñoz-Espín D. A guide to assessing cellular senescence in vitro and in vivo. FEBS J [Internet]. 2021; 288: 56–80. Available from: http://dx.doi.org/10.1111/febs.15570

22. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. Hallmarks of aging: An expanding universe. Cell [Internet]. 2023; 186: 243–78. Available from: http://dx.doi.org/10.1016/j.cell.2022.11.001

23. Wright Muelas M, Mughal F, O'Hagan S, Day PJ, Kell DB. The role and robustness of the Gini coefficient as an unbiased tool for the selection of Gini genes for normalising expression profiling data. Sci Rep [Internet]. 2019; 9: 17960. Available from: http://dx.doi.org/10.1038/s41598-019-54288-7

24. Zeng J, Liu S, Zhao Y, Tan X, Aljohi HA, Liu W, Hu S. Identification and analysis of house-keeping and tissue-specific genes based on RNA-seq data sets across 15 mouse tissues. Gene [Internet]. 2016; 576: 560–70. Available from: http://dx.doi.org/10.1016/j.gene.2015.11.003

25. Li Y, Zhang L, Li R, Zhang M, Li Y, Wang H, Wang S, Bao Z. Systematic identification and validation of the reference genes from 60 RNA-Seq libraries in the scallop Mizuhopecten yessoensis. BMC Genomics [Internet]. 2019; 20: 288. Available from: http://dx.doi.org/10.1186/s12864-019-5661-x

655 26. Wang Z, Lyu Z, Pan L, Zeng G, Randhawa P. Defining housekeeping genes suitable for RNA-seq
656 analysis of the human allograft kidney biopsy tissue. BMC Med Genomics [Internet]. 2019; 12: 86.
657 Available from: http://dx.doi.org/10.1186/s12920-019-0538-z

658 27. Lin Y, Ghazanfar S, Strbenac D, Wang A, Patrick E, Lin DM, Speed T, Yang JYH, Yang P. Evaluating
659 stably expressed genes in single cells. Gigascience [Internet]. 2019; 8. Available from:
660 http://dx.doi.org/10.1093/gigascience/giz106

661 28. Flurkey K, M. Currer J, Harrison DE. Chapter 20 - Mouse Models in Aging Research. In: Fox JG,
662 Davisson MT, Quimby FW, Barthold SW, Newcomer CE, Smith AL, editors. The Mouse in Biomedical
663 Research (Second Edition) [Internet]. Burlington: Academic Press; 2007. p. 637–72. Available from:
664 https://www.sciencedirect.com/science/article/pii/B9780123694546500741

665 29. Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM
666 measure is inconsistent among samples. Theory Biosci [Internet]. 2012; 131: 281–5. Available from:
667 http://dx.doi.org/10.1007/s12064-012-0162-3

668 30. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of
669 RNA-seq data. Genome Biol [Internet]. 2010; 11: R25. Available from:
670 http://dx.doi.org/10.1186/gb-2010-11-3-r25

671 31. He W, Holtkamp S, Hergenhan SM, Kraus K, de Juan A, Weber J, Bradfield P, Grenier JMP, Pelletier J,
672 Druzd D, Chen C-S, Ince LM, Bierschenk S, et al. Circadian Expression of Migratory Factors
673 Establishes Lineage-Specific Signatures that Guide the Homing of Leukocyte Subsets to Tissues.
674 Immunity [Internet]. 2018; 49: 1175–90.e7. Available from:
675 http://dx.doi.org/10.1016/j.immuni.2018.10.007

676 32. Born J, Lange T, Hansen K, Mölle M, Fehm HL. Effects of sleep and circadian rhythm on human
677 circulating immune cells. J Immunol [Internet]. 1997; 158: 4454–64. Available from:
678 https://www.ncbi.nlm.nih.gov/pubmed/9127011

679 33. Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP. Determination of stable housekeeping genes,
680 differentially regulated target genes and sample integrity: BestKeeper--Excel-based tool using pair-
681 wise correlations. Biotechnol Lett [Internet]. Springer Nature; 2004; 26: 509–15. Available from:
682 https://idp.springer.com/authorize/casa?redirect_uri=https://link.springer.com/article/10.1023/B:
683 BILE.0000019559.84305.47&casa_token=uGEU4e8oJrYAAAAA:RApADTvJOFk1Rx2yF3_Y3_zL623wy
684 R7WaOAJ0LConLjckuqL1o5NhxHl4zXgwZjgLyFCo1kmJCB8zoXL

685 34. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. Accurate
686 normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal
687 control genes. Genome Biol [Internet]. 2002; 3: RESEARCH0034. Available from:
688 http://dx.doi.org/10.1186/gb-2002-3-7-research0034

689 35. Andersen CL, Jensen JL, Ørntoft TF. Normalization of real-time quantitative reverse transcription-
690 PCR data: a model-based variance estimation approach to identify genes suited for normalization,
691 applied to bladder and colon cancer data sets. Cancer Res [Internet]. AACR; 2004; 64: 5245–50.
692 Available from: https://aacrjournals.org/cancerres/article-abstract/64/15/5245/511517

693 36. Silver N, Best S, Jiang J, Thein SL. Selection of housekeeping genes for gene expression studies in
694 human reticulocytes using real-time PCR. BMC Mol Biol [Internet]. 2006; 7: 33. Available from:

http://dx.doi.org/10.1186/1471-2199-7-33

37. Xie F, Wang J, Zhang B. RefFinder: a web-based tool for comprehensively analyzing and identifying reference genes. Funct Integr Genomics [Internet]. 2023; 23: 125. Available from: http://dx.doi.org/10.1007/s10142-023-01055-7

38. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. Cell [Internet]. 2013; 153: 1194–217. Available from: http://dx.doi.org/10.1016/j.cell.2013.05.039

39. Fraser HC, Kuan V, Johnen R, Zwierzyna M, Hingorani AD, Beyer A, Partridge L. Biological mechanisms of aging predict age-related disease co-occurrence in patients. Aging Cell [Internet]. 2022; 21: e13524. Available from: http://dx.doi.org/10.1111/acel.13524

40. Simpson DJ, Olova NN, Chandra T. Cellular reprogramming and epigenetic rejuvenation. Clin Epigenetics [Internet]. 2021; 13: 170. Available from: http://dx.doi.org/10.1186/s13148-021-01158-7

41. Sha Y, Phan JH, Wang MD. Effect of low-expression gene filtering on detection of differentially expressed genes in RNA-seq data. Conf Proc IEEE Eng Med Biol Soc [Internet]. 2015; 2015: 6461–4. Available from: http://dx.doi.org/10.1109/EMBC.2015.7319872

42. Tsitsiridis G, Steinkamp R, Giurgiu M, Brauner B, Fobo G, Frishman G, Montrone C, Ruepp A. CORUM: the comprehensive resource of mammalian protein complexes–2022. Nucleic Acids Res [Internet]. Oxford Academic; 2022 [cited 2023 Dec 11]; 51: D539–45. Available from: https://academic.oup.com/nar/article-abstract/51/D1/D539/6830667

43. Perry RJ, Zhang X-M, Zhang D, Kumashiro N, Camporez J-PG, Cline GW, Rothman DL, Shulman GI. Leptin reverses diabetes by suppression of the hypothalamic-pituitary-adrenal axis. Nat Med [Internet]. 2014; 20: 759–63. Available from: http://dx.doi.org/10.1038/nm.3579

44. Saran S, Tran DDH, Ewald F, Koch A, Hoffmann A, Koch M, Nashan B, Tamura T. Depletion of three combined THOC5 mRNA export protein target genes synergistically induces human hepatocellular carcinoma cell death. Oncogene [Internet]. 2016; 35: 3872–9. Available from: http://dx.doi.org/10.1038/onc.2015.433

45. Escobar B, de Cárcer G, Fernández-Miranda G, Cascón A, Bravo-Cordero JJ, Montoya MC, Robledo M, Cañamero M, Malumbres M. Brick1 is an essential regulator of actin cytoskeleton required for embryonic development and cell transformation. Cancer Res [Internet]. 2010; 70: 9349–59. Available from: http://dx.doi.org/10.1158/0008-5472.CAN-09-4491

46. Annaert W, Kaether C. Bring it back, bring it back, don't take it away from me - the sorting receptor RER1. J Cell Sci [Internet]. The Company of Biologists; 2020 [cited 2023 Jul 11]; 133: jcs231423. Available from: https://journals.biologists.com/jcs/article-abstract/133/17/jcs231423/226312

47. Hamazaki J, Sasaki K, Kawahara H, Hisanaga S-I, Tanaka K, Murata S. Rpn10-mediated degradation of ubiquitinated proteins is essential for mouse development. Mol Cell Biol [Internet]. 2007; 27: 6629–38. Available from: http://dx.doi.org/10.1128/MCB.00509-07

48. Nicholls TJ, Spåhr H, Jiang S, Siira SJ, Koolmeister C, Sharma S, Kauppila JHK, Jiang M, Kaever V, Rackham O, Chabes A, Falkenberg M, Filipovska A, et al. Dinucleotide Degradation by REXO2

733 Maintains Promoter Specificity in Mammalian Mitochondria. Mol Cell [Internet]. 2019; 76: 784–
734 96.e6. Available from: http://dx.doi.org/10.1016/j.molcel.2019.09.010

735 49. Groza T, Gomez FL, Mashhadi HH, Muñoz-Fuentes V, Gunes O, Wilson R, Cacheiro P, Frost A,
736 Keskivali-Bond P, Vardal B, McCoy A, Cheng TK, Santos L, et al. The International Mouse
737 Phenotyping Consortium: comprehensive knockout phenotyping underpinning the study of human
738 disease. Nucleic Acids Res [Internet]. 2023; 51: D1038–45. Available from:
739 http://dx.doi.org/10.1093/nar/gkac972

740 50. Barth E, Srivastava A, Wengerodt D, Stojiljkovic M, Axer H, Witte OW, Kretz A, Marz M. Age-
741 dependent expression changes of circadian system-related genes reveal a potentially conserved
742 link to aging. Aging [Internet]. 2021; 13: 25694–716. Available from:
743 http://dx.doi.org/10.18632/aging.203788

744 51. Xiong B, Yang Y, Fineis FR, Wang J-P. DegNorm: normalization of generalized transcript degradation
745 improves accuracy in RNA-seq analysis [Internet]. Genome Biology. 2019. Available from:
746 http://dx.doi.org/10.1186/s13059-019-1682-7

747 52. Cellerino A, Ori A. What have we learned on aging from omics studies? Semin Cell Dev Biol
748 [Internet]. 2017; 70: 177–89. Available from:
749 https://www.sciencedirect.com/science/article/pii/S1084952116303160

750 53. Tahoe NMA, Mokhtarzadeh A, Curtsinger JW. Age-related RNA decline in adult Drosophila
751 melanogaster. J Gerontol A Biol Sci Med Sci [Internet]. 2004; 59: B896–901. Available from:
752 http://dx.doi.org/10.1093/gerona/59.9.b896

753 54. Uemura E, Hartmann HA. RNA content and volume of nerve cell bodies in human brain. I. Prefrontal
754 cortex in aging normal and demented patients. J Neuropathol Exp Neurol [Internet]. 1978; 37: 487–
755 96. Available from: http://dx.doi.org/10.1097/00005072-197809000-00004

756 55. Kennedy BK, Berger SL, Brunet A, Campisi J, Cuervo AM, Epel ES, Franceschi C, Lithgow GJ,
757 Morimoto RI, Pessin JE, Rando TA, Richardson A, Schadt EE, et al. Geroscience: Linking aging to
758 chronic disease. Cell [Internet]. Elsevier Inc.; 2014; 159: 709–13. Available from:
759 http://dx.doi.org/10.1016/j.cell.2014.10.039

760 56. Oz N, Vayndorf EM, Tsuchiya M, McLean S, Turcios-Hernandez L, Pitt JN, Blue BW, Muir M, Kiflezghi
761 MG, Tyshkovskiy A, Mendenhall A, Kaeberlein M, Kaya A. Evidence that conserved essential genes
762 are enriched for pro-longevity factors. Geroscience [Internet]. 2022; 44: 1995–2006. Available
763 from: http://dx.doi.org/10.1007/s11357-022-00604-5

764 57. Kruse SE, Karunadharma PP, Basisty N, Johnson R, Beyer RP, MacCoss MJ, Rabinovitch PS, Marcinek
765 DJ. Age modifies respiratory complex I and protein homeostasis in a muscle type-specific manner.
766 Aging Cell [Internet]. 2016; 15: 89–99. Available from: http://dx.doi.org/10.1111/acel.12412

767 58. Goldman AD, Weber JM, LaRowe DE, Barge LM. Electron transport chains as a window into the
768 earliest stages of evolution. Proc Natl Acad Sci U S A [Internet]. 2023; 120: e2210924120. Available
769 from: http://dx.doi.org/10.1073/pnas.2210924120

770 59. McElroy GS, Chakrabarty RP, D'Alessandro KB, Hu Y-S, Vasan K, Tan J, Stoolman JS, Weinberg SE,
771 Steinert EM, Reyfman PA, Singer BD, Ladiges WC, Gao L, et al. Reduced expression of mitochondrial

772 complex I subunit Ndufs2 does not impact healthspan in mice. Sci Rep [Internet]. 2022; 12: 5196.
773 Available from: http://dx.doi.org/10.1038/s41598-022-09074-3

774 60. Statzer C, Park JYC, Ewald CY. Extracellular Matrix Dynamics as an Emerging yet Understudied
775 Hallmark of Aging and Longevity. Aging Dis [Internet]. 2023; 14: 670–93. Available from:
776 http://dx.doi.org/10.14336/AD.2022.1116

777 61. Pignolo RJ, Passos JF, Khosla S, Tchkonia T, Kirkland JL. Reducing Senescent Cell Burden in Aging and
778 Disease. Trends Mol Med [Internet]. 2020; 26: 630–8. Available from:
779 http://dx.doi.org/10.1016/j.molmed.2020.03.005

780 62. Browder KC, Reddy P, Yamamoto M, Haghani A, Guillen IG, Sahu S, Wang C, Luque Y, Prieto J, Shi L,
781 Shojima K, Hishida T, Lai Z, et al. In vivo partial reprogramming alters age-associated molecular
782 changes during physiological aging in mice. Nature Aging [Internet]. Nature Publishing Group; 2022
783 [cited 2022 Oct 19]; 2: 243–53. Available from: https://www.nature.com/articles/s43587-022-
784 00183-2

785 63. Popovic D, Koch B, Kueblbeck M, Ellenberg J, Pelkmans L. Multivariate Control of Transcript to
786 Protein Variability in Single Mammalian Cells. Cell Syst [Internet]. 2018; 7: 398–411.e6. Available
787 from: http://dx.doi.org/10.1016/j.cels.2018.09.001

788 64. Edfors F, Danielsson F, Hallström BM, Käll L, Lundberg E, Pontén F, Forsström B, Uhlén M. Gene-
789 specific correlation of RNA and protein levels in human cells and tissues. Mol Syst Biol [Internet].
790 2016; 12: 883. Available from: http://dx.doi.org/10.15252/msb.20167144

791 65. Zhong S, Zhou S, Yang S, Yu X, Xu H, Wang J, Zhang Q, Lv M, Feng J. Identification of internal control
792 genes for circular RNAs. Biotechnol Lett [Internet]. 2019; 41: 1111–9. Available from:
793 http://dx.doi.org/10.1007/s10529-019-02723-0

794 66. Meer MV, Podolskiy DI, Tyshkovskiy A, Gladyshev VN. A whole lifespan mouse multi-tissue DNA
795 methylation clock. Elife [Internet]. 2018; 7. Available from: http://dx.doi.org/10.7554/eLife.40675

796 67. Minteer C, Morselli M, Meer M, Cao J, Higgins-Chen A, Lang SM, Pellegrini M, Yan Q, Levine ME.
797 Tick tock, tick tock: Mouse culture and tissue aging captured by an epigenetic clock. Aging Cell
798 [Internet]. 2022; 21: e13553. Available from: http://dx.doi.org/10.1111/acel.13553

799 68. Thrush KL, Bennett DA, Gaiteri C, Horvath S, van Dyck CH, Higgins-Chen AT, Levine ME. Aging the
800 brain: multi-region methylation principal component based clock in the context of Alzheimer's
801 disease. Aging [Internet]. 2022; 14: 5641–68. Available from:
802 http://dx.doi.org/10.18632/aging.204196

803
804

Tables

Table 1

1110004F10Rik (MGI:1929274)	Fis1 (MGI:1913687)	Psmd4 (MGI:1201670)	Rexo2 (MGI:1888981)	Tomm22 (MGI:2450248)
Brk1 (MGI:1915406)	Gemin7 (MGI:1916981)	Rer1 (MGI:1915080)	Srp14 (MGI:107169)	

Legends

Table 1:

MGI symbol and ID for our 9 pan-tissue age-invariant genes. These genes were present across all tissues after all filtering steps and validation.

Figure 1: Visual Diagram of Article Contents

A) Bulk RNA-seq data from 17 murine tissues (GSE132040) were sequentially filtered through 7 criteria. Steps 1-4 are adapted from previous publications. We added criteria filters 5 and 6 to ensure low variation and no correlation with age. Criteria filter 7 was validation of low variation and no age correlation, performed in a second dataset for 11 of the 17 tissues. B) Sample gender, age and life stage distributions of the samples in the dataset. A full table of samples can be found in Supplementary Table 10. C) Canonical reference genes are not applicable to all tissues in an aging context but age-invariant genes introduced here are. D) Tissue aging-invariant genes are enriched to different extents for gene ontology terms associated with hallmarks of aging. Age-invariant genes have low enrichment in some (e.g. epigenetic alterations GO terms) and high enrichment in others (e.g. loss of proteostasis GO terms). Created with BioRender.com

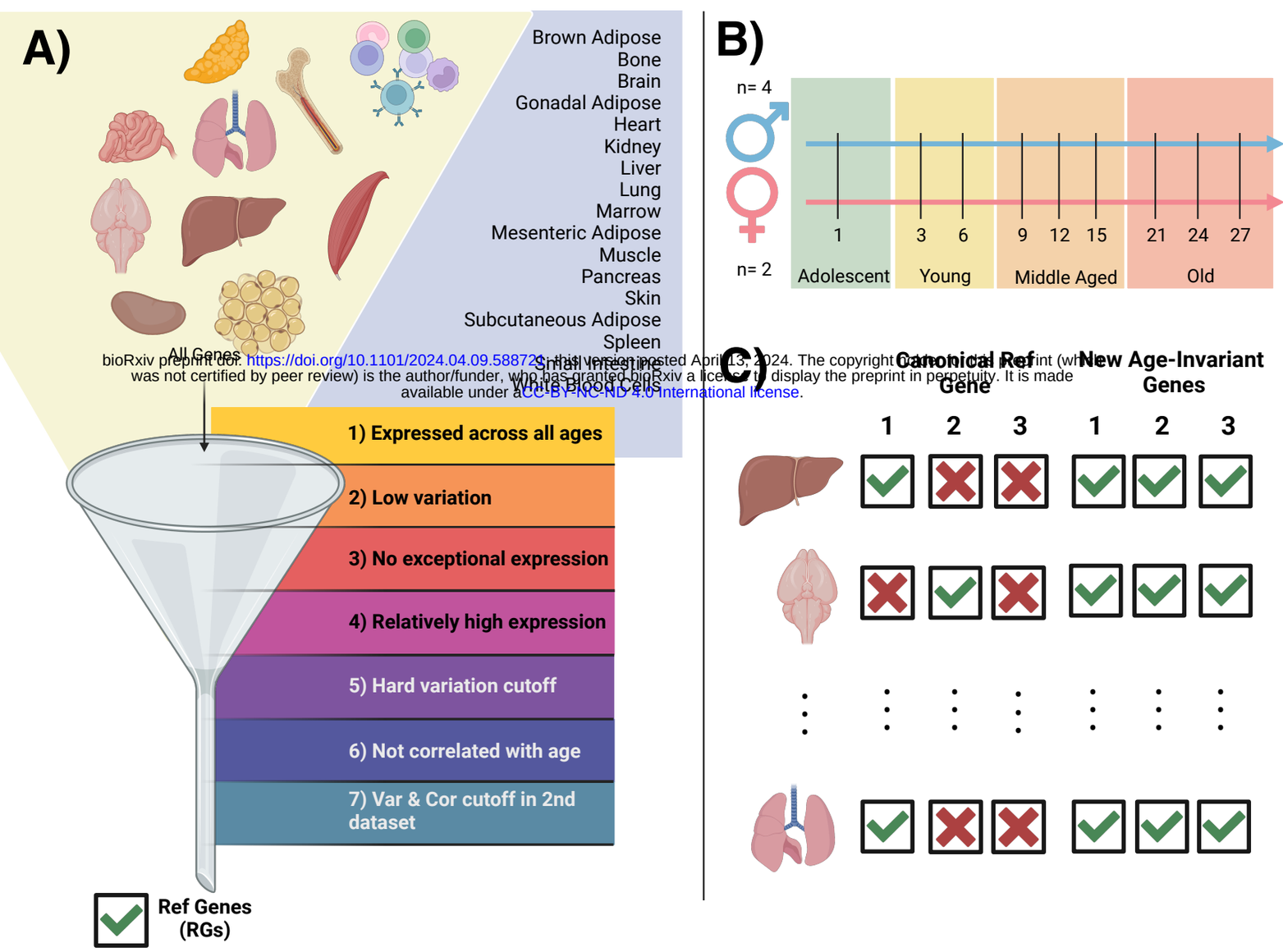
Figure 2: Gene Selection Process and Rationale

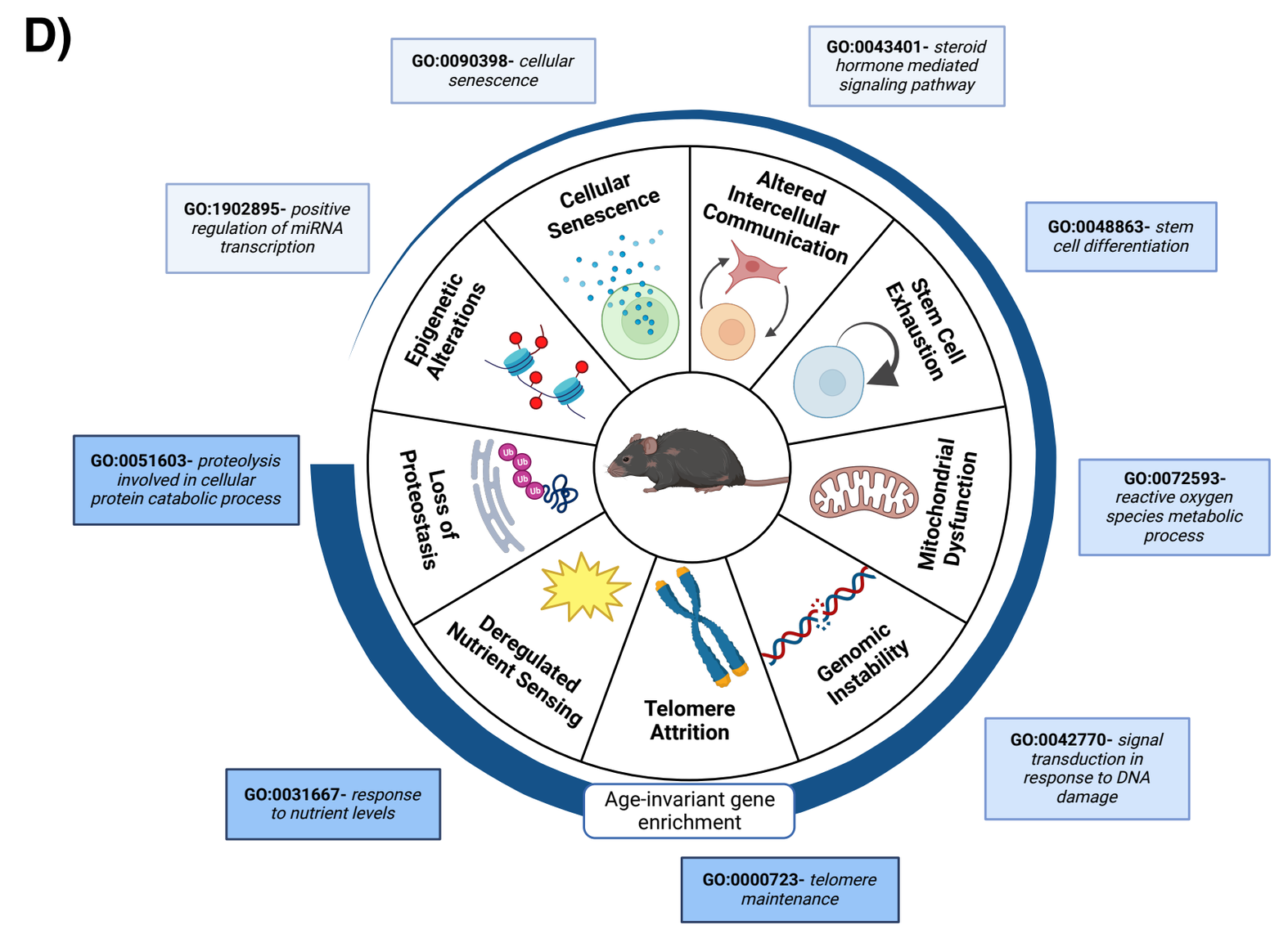
826 A) Gene count number remaining after each criteria/filter step for each tissue. B) Gene count present
827 across all tissues at each step, presented on a log2 scale. C) % Coefficient of Variance (CV) for each gene
828 calculated as SD/mean*100 distribution of log2 TPM gene expression values. Genes that satisfy every
829 subsequent filter are plotted by the last filter applied. Filters 1-3 slowly decrease %CV and the
830 cumulative effect of filters 1-4 generally results in a %CV of approximately 20%. Filter 5 imposes a strict
831 %CV < 20% requirement for all tissue-gene pairs. D) Age information must be included in exclusion
832 criteria as low variation genes can still have a high correlation with age. Filter 6 (Spearman correlation p-
833 value based removal) removes highly age-correlated genes. Dashed line corresponds to a correlation
834 coefficient (y-axis) of 0.4, which for most tissues corresponds to a significant correlation with p = 0.05.
835 Exact CV and age correlation information is found in Supplementary Table 2, in case readers wish to
836 utilize other cutoffs in selecting RGs. E) Log2 TPM (y-axis) values by life stage (color) for specific gene-
837 tissue pairs (x-axis) for genes that satisfy filters 1-5, but are eliminated by filter 6. Boxplot line represents
838 the group median while lower and upper limits of the boxplot correspond to the first (25%) and third
839 (75%) quartiles.
840
841 Figure 3: Classical & Novel RG Performance in Aging Samples
842 A) Aging RG status of classical reference gene by tissue. Genes that are age-invariant and therefore valid
843 RGs are depicted in blue while their age-variant counterparts, which were not present in the gene list
844 after filtering, appear in red. B-C) Individual gene cycle threshold (Ct) results from validation RT-qPCR
845 tissues in heart (B) and liver (C) for selected classical RGs and novel age-invariant RGs. D) RT-qPCR Gene
846 RefFinder score and mRNA-seq %CV in heart and liver. Age-invariant genes are distinct from and
847 outperform canonical RGs in %CV (Welch Two Sample t-test p-value= 0.006553) and RefFinder qPCR
848 scores(Welch Two Sample t-test p-value = 0.02401). RefFinder and %CV scores were calculated from in-
849 house and public validation datasets respectively. RefFinder score was based on BestKeeper,

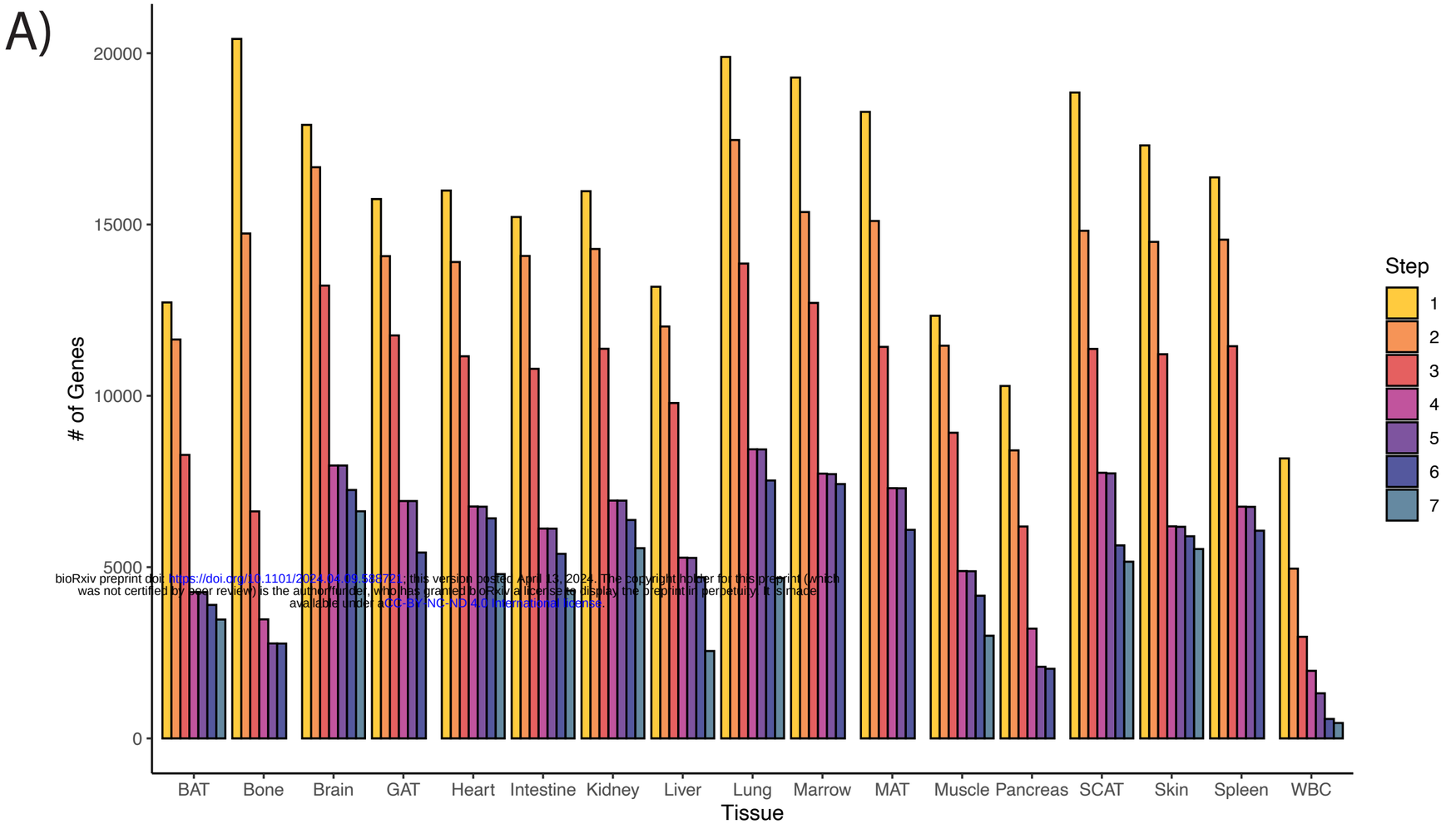
NormFinder, GeNorm and comparative delta-Ct values. Circled points indicate novel age-invariant RGs (Two pan-tissue: Tomm22 and Srp14; and one heart and liver age-invariant gene: Atp6v1f) while uncircled points specify classical RGs from Figure 3A.

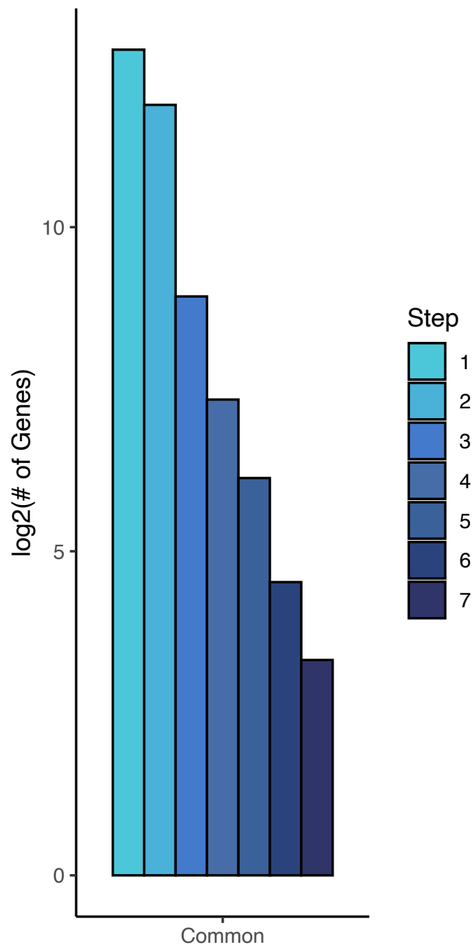
Figure 4: Age-Invariant Genes are Enriched for Dysregulated and Aging Disease Associated Gene Functions

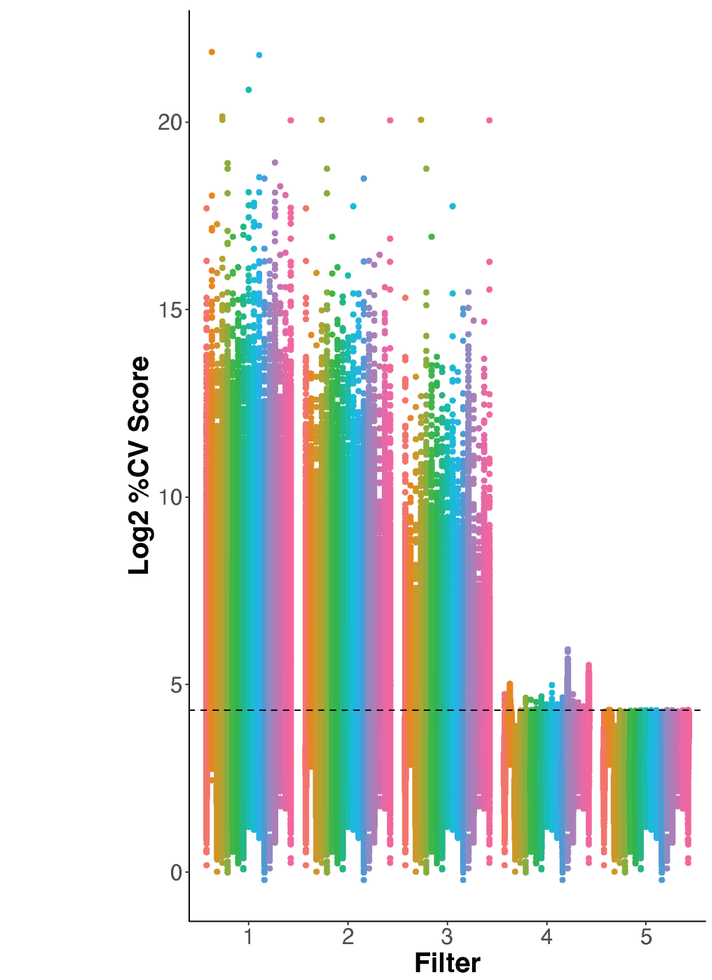
A) Tissue age-invariant genes are enriched for some GO, KEGG and REACTOME terms associated with linear and non-linear aging trajectories. Left labels correspond to enrichment terms originally classified in 9 trajectory groups. Age-invariant labels at the very bottom (pink) refer to genes identified in this paper. Heatmap columns correspond to different tissues, while rows correspond to different terms. B) Age-invariant genes are enriched for GO Biological Processes associated with age-related disease in humans. C) Tissue age-invariant genes are enriched for certain protein complexes. Gene lists are enriched for CCT complex, electron transport chain (respiratory complex I and cytochrome C), proteasome, Cop9 signalosome, PYR, Parvulin-associated pre-rRNP and Regulator-AXIN/LKB1-AMPK complexes in CORUM analysis.

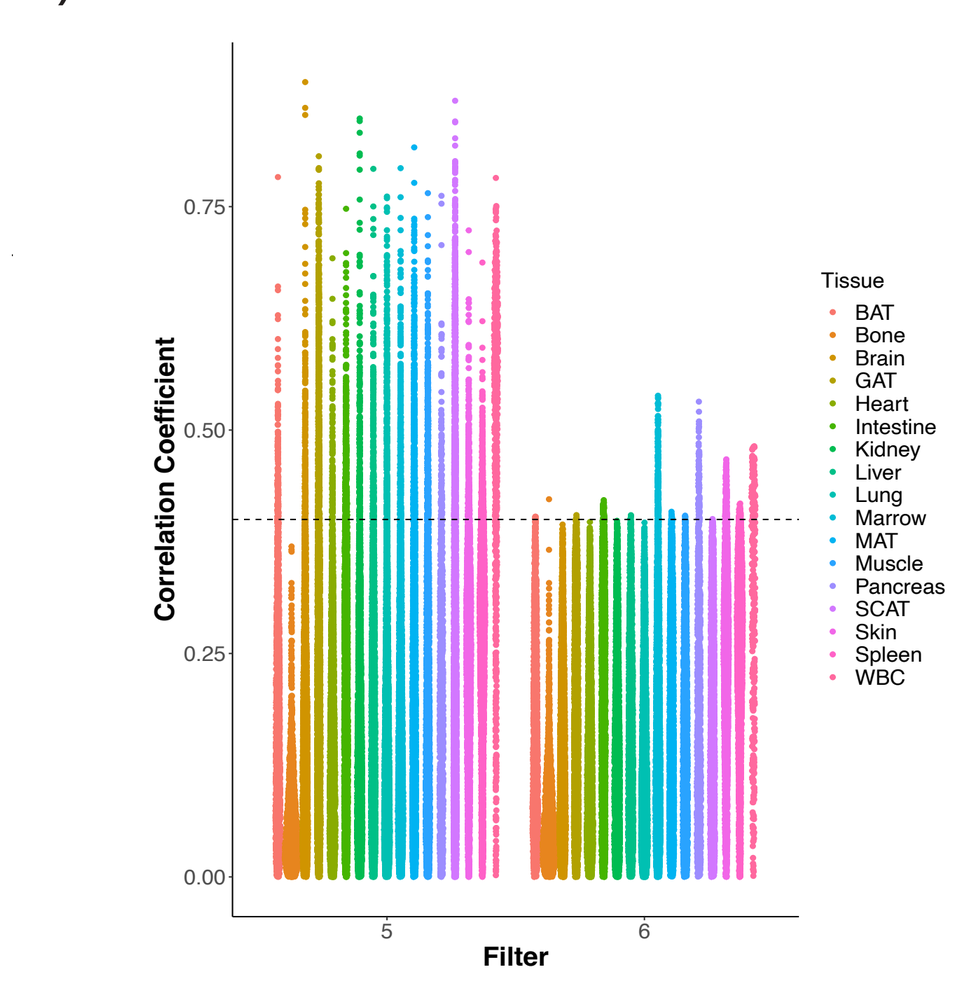


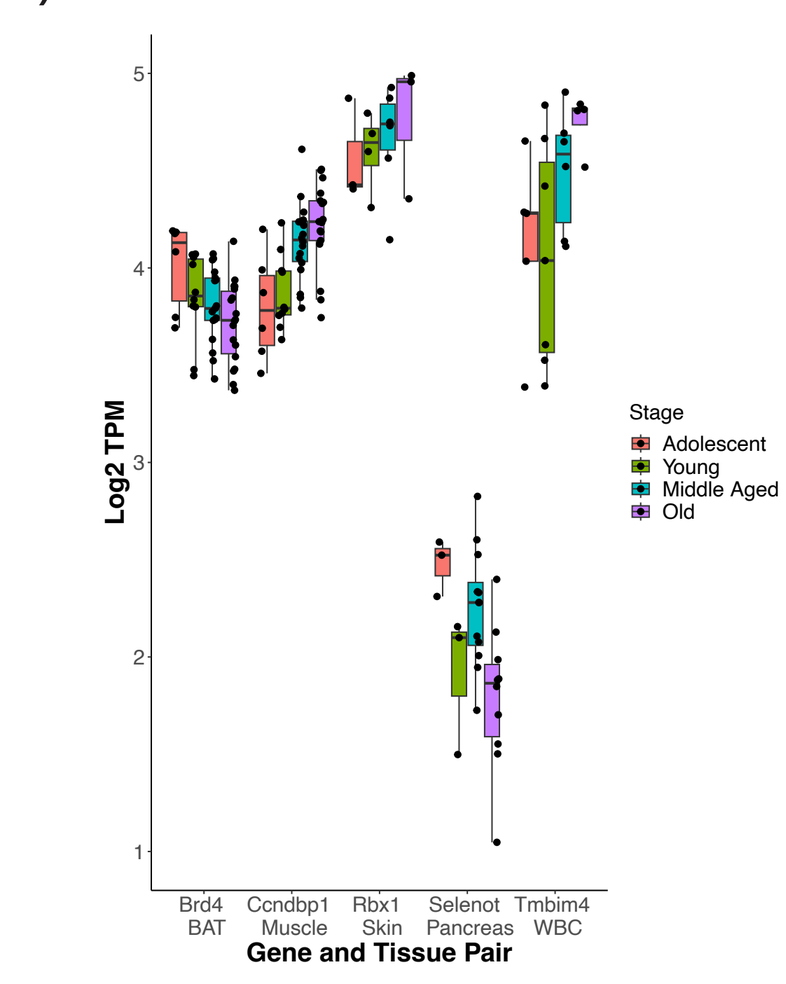


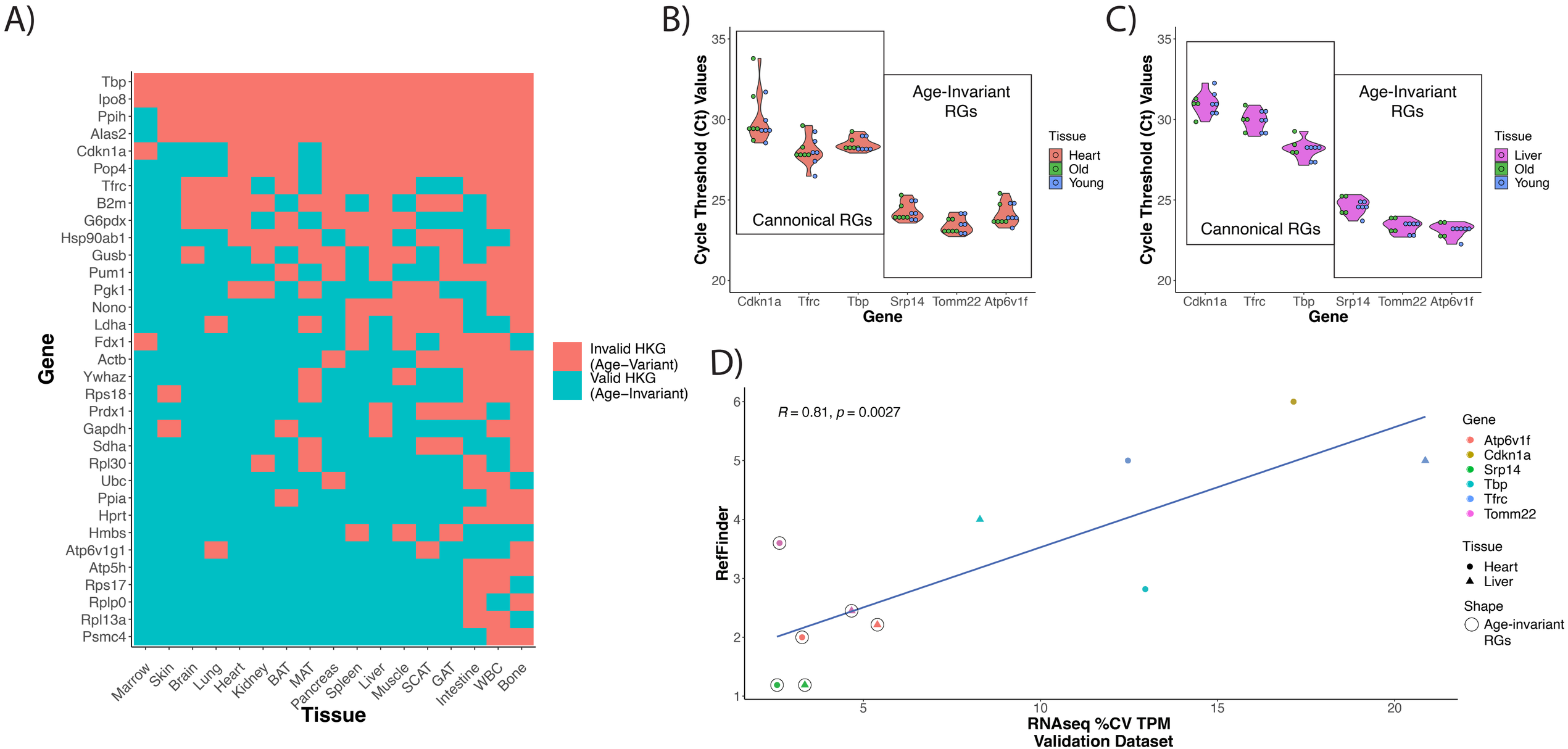












A)

Tissue

Spleen Marrow Skin GAT MAT Muscle Kidney Heart Intestine Lung SCAT Brain BAT Pancreas Liver Bone WBC

log10 p−val
−4
−8
−12
−16

Aging Trajectory Cluster Enrichment Terms

1
Asparagine N−linked glycosylation
Cell Cycle
Cell Cycle, Mitotic
cilium
ER to Golgi Anterograde Transport
Formation of a pool of free 40S subunits
GTP hydrolysis and joining of the 60S ribosomal subunit
Intra−Golgi and retrograde Golgi−to−ER traffic
L13a−mediated translational silencing of Ceruloplasmin expression
Metabolism of proteins
Nonsense Mediated Decay (NMD) enhanced by the Exon Junction Complex (EJC)
Nonsense Mediated Decay (NMD) independent of the Exon Junction Complex (EJC)
Nonsense−Mediated Decay (NMD)
peptidyl−cysteine S−nitrosylation
postsynaptic membrane
protein nitrosylation
Reduction of cytosolic Ca++ levels
SUMO is proteolytically processed
Transcriptional regulation of white adipocyte differentiation

2
anatomical structure development
anatomical structure morphogenesis
cell adhesion
cell communication
cell periphery
cellular response to stimulus
developmental process
multicellular organism development
multicellular organismal process
plasma membrane
regulation of multicellular organismal development
regulation of multicellular organismal process
response to stimulus
signal transduction
signaling
system development

3
catalytic complex
envelope
inner mitochondrial membrane protein complex
intracellular
intracellular membrane−bounded organelle
intracellular organelle
membrane−bounded organelle
mitochondrial envelope
mitochondrial inner membrane
mitochondrial membrane
mitochondrial protein complex
mitochondrion
organelle envelope
organelle inner membrane
Oxidative phosphorylation
Parkinson disease
respiratory chain

4
B cell activation
cell periphery
immune response
immune system process
leukocyte activation
lymphocyte activation
MHC class I protein complex
MHC protein complex
plasma membrane
positive regulation of immune system process
positive regulation of lymphocyte activation
regulation of glial cell migration
regulation of lymphocyte activation
regulation of response to stimulus
response to biotic stimulus
response to external biotic stimulus
response to external stimulus
response to other organism
vesicle

5
cellular response to DNA damage stimulus
defense response
DNA repair
Fanconi anemia pathway
immune response
inflammatory response
ncRNA metabolic process
ncRNA processing
phagocytosis
positive regulation of defense response
regulation of defense response
regulation of immune response
regulation of phospholipase A2 activity
response to external biotic stimulus
response to other organism
response to stress
ribonucleoprotein complex biogenesis
ribosome biogenesis
rRNA metabolic process
rRNA processing

6
Allograft rejection
antigen processing and presentation
Antigen processing and presentation
antigen processing and presentation of exogenous antigen
antigen processing and presentation of exogenous peptide antigen
antigen processing and presentation of exogenous peptide antigen via MHC class II
antigen processing and presentation of peptide antigen
antigen processing and presentation of peptide antigen via MHC class II
Autoimmune thyroid disease
Cell adhesion molecules (CAMs)
defense response
Graft−versus−host disease
immune response
immune system process
MHC class II protein complex
MHC protein complex
Phagosome
Staphylococcus aureus infection
Type I diabetes mellitus
Viral myocarditis

7
'de novo' protein folding
chaperone binding
chaperone complex
chaperone−mediated protein folding
chaperonin−containing T−complex
heat shock protein binding
positive regulation of protein localization to Cajal body
protein binding involved in protein folding
protein folding
protein localization to Cajal body
protein localization to nuclear body
protein localization to nucleoplasm
Protein processing in endoplasmic reticulum
protein refolding
protein stabilization
regulation of protein localization to Cajal body
RNA−dependent DNA biosynthetic process
telomere maintenance via telomerase
unfolded protein binding
zona pellucida receptor complex

8
Assembly of collagen fibrils and other multimeric structures
Collagen biosynthesis and modifying enzymes
Collagen chain trimerization
Collagen degradation
collagen fibril organization
Collagen formation
collagen trimer
complex of collagen trimers
Degradation of the extracellular matrix
extracellular matrix
extracellular matrix component
extracellular matrix organization
Extracellular matrix organization
extracellular matrix structural constituent
extracellular space
extracellular structure organization
Integrin cell surface interactions
Non−integrin membrane−ECM interactions
Protein digestion and absorption

9
2−oxoglutarate−dependent dioxygenase activity
catalytic activity
demethylase activity
demethylation
dioxygenase activity
heterocyclic compound binding
histone demethylase activity
histone demethylation
histone lysine demethylation
hydrolase activity, acting on acid anhydrides, in phosphorus−containing anhydrides
ion binding
nucleic acid binding
nucleoside−triphosphatase activity
organic cyclic compound binding
oxidoreductase activity
oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen
protein dealkylation
protein demethylation
pyrophosphatase activity
translational initiation

Age−Invariant
binding
cellular anatomical entity
cellular metabolic process
cellular process
cellular_component
cytoplasm
cytosol
intracellular anatomical structure
intracellular organelle
macromolecule metabolic process
membrane−bounded organelle
metabolic process
mitochondrion
molecular_function
nitrogen compound metabolic process
organic substance metabolic process
organonitrogen compound metabolic process
protein metabolic process
protein−containing complex
REACTOME root term

NA
Signaling
Mitochondrion
Immune resp.
Stress resp.
Immune resp.
Protein folding
Extra. matrix
NA
Age−Invariant

0 15 30

B)

Tissue

Marrow Brain Skin Spleen MAT BAT Kidney Heart Lung Muscle Intestine GAT SCAT Pancreas Bone Liver WBC

Aging Associated GO:BP

AI
binding
biological_process
cellular anatomical entity
cellular metabolic process
cellular process
cellular_component
cytoplasm
intracellular anatomical structure
intracellular membrane−bounded organelle
intracellular organelle
macromolecule metabolic process
membrane−bounded organelle
metabolic process
molecular_function
nitrogen compound metabolic process
organelle
organic substance metabolic process
organonitrogen compound metabolic process
primary metabolic process
protein−containing complex

AIC
cellular response to corticotropin−releasing hormone stimulus
cellular response to thyroid hormone stimulus
follicle−stimulating hormone secretion
growth hormone receptor signaling pathway
hormone biosynthetic process
inflammatory response to antigenic stimulus
negative regulation of acute inflammatory response
negative regulation of inflammatory response
negative regulation of neurotransmitter secretion
negative regulation of neurotransmitter uptake
neuromuscular synaptic transmission
peptide hormone processing
positive regulation of neurotransmitter secretion
positive regulation of peptide hormone secretion
positive regulation of transmission of nerve impulse
regulation of growth hormone receptor signaling pathway
regulation of hormone secretion
regulation of synaptic transmission, GABAergic
response to growth hormone
response to thyroid hormone
steroid hormone mediated signaling pathway
synaptic transmission, cholinergic
thyroid hormone generation
wound healing involved in inflammatory response

SC
cellular senescence
negative regulation of cellular senescence
positive regulation of cellular senescence
replicative senescence

DNS
CAMKK−AMPK signaling cascade
cellular response to insulin stimulus
glucose homeostasis
glucose mediated signaling pathway
negative regulation of insulin secretion involved in cellular response to glucose stimulus
negative regulation of lipid storage
neutral lipid metabolic process
positive regulation of glucose transmembrane transport
positive regulation of insulin secretion
regulation of glucose transmembrane transport
regulation of insulin secretion
regulation of insulin secretion involved in cellular response to glucose stimulus
regulation of lipid metabolic process
regulation of lipid storage
regulation of protein lipidation
regulation of TORC1 signaling
response to glucose
response to nutrient
response to nutrient levels

EPI
DNA methylation
genetic imprinting
miRNA mediated inhibition of translation
negative regulation of gene silencing by miRNA
negative regulation of histone acetylation
negative regulation of production of miRNAs involved in gene silencing by miRNA
positive regulation of histone acetylation
positive regulation of pri−miRNA transcription by RNA polymerase II
pri−miRNA transcription by RNA polymerase II

GI
DNA damage response, signal transduction by p53 class mediator
DNA damage response, signal transduction by p53 class mediator resulting in cell cycle arrest
DNA damage response, signal transduction by p53 class mediator resulting in transcription of p21 class mediator
double−strand break repair via nonhomologous end joining
intrinsic apoptotic signaling pathway in response to DNA damage
intrinsic apoptotic signaling pathway in response to DNA damage by p53 class mediator
mitotic G1 DNA damage checkpoint signaling
negative regulation of DNA damage response, signal transduction by p53 class mediator
positive regulation of DNA damage response, signal transduction by p53 class mediator
response to DNA damage checkpoint signaling
signal transduction in response to DNA damage

LP
endoplasmic reticulum unfolded protein response
intrinsic apoptotic signaling pathway in response to endoplasmic reticulum stress
macroautophagy
negative regulation of autophagy
negative regulation of endoplasmic reticulum stress−induced intrinsic apoptotic signaling pathway
positive regulation of autophagy of mitochondrion
positive regulation of IRE1−mediated unfolded protein response
proteolysis involved in cellular protein catabolic process
regulation of autophagy

MD
establishment of mitochondrion localization
intracellular distribution of mitochondria
membrane fission
negative regulation of protein targeting to mitochondrion
negative regulation of reactive oxygen species biosynthetic process
positive regulation of autophagy of mitochondrion
positive regulation of mitochondrial outer membrane permeabilization involved in apoptotic signaling pathway
positive regulation of reactive oxygen species metabolic process
positive regulation of release of cytochrome c from mitochondria
protein insertion into mitochondrial inner membrane
protein insertion into mitochondrial membrane
reactive oxygen species metabolic process
regulation of mitochondrial fission

SCR
hematopoietic stem cell proliferation
negative regulation of stem cell proliferation
positive regulation of stem cell population maintenance
positive regulation of stem cell proliferation
regulation of somatic stem cell population maintenance
somatic stem cell division
somatic stem cell population maintenance
stem cell differentiation
stem cell proliferation

TA
establishment of protein localization to telomere
negative regulation of telomere capping
telomere capping
telomere maintenance

Age-Invariant
Altered intercellular immunication
Cellular senescence
Deregulated nutrient sensing
Epigenetic alterations
Genomic instability
Loss of proteostasis
Mitochondrial dysfunction
Stem cell exhaustion
Telomere attrition

C)

CCT complex (chaperonin containing TCP1 complex)
CCT complex (chaperonin containing TCP1 complex)
CCT complex (chaperonin containing TCP1 complex), testis specific
CCT complex (chaperonin containing TCP1 complex)
Respiratory chain complex I, mitochondrial
Respiratory chain complex I, mitochondrial
Cytochrome bc1-complex, mitochondrial
Cytochrome c oxidase, mitochondrial
COP9 signalosome complex (Gps1, Cops1, Cops2, Cops3, Cops4, Cops5, Cops6, Cops7a, Cops8)
Parvulin-associated pre-rRNP complex
PYR complex
Immunoproteasome
Ragulator-AXIN/LKB1-AMPK complex
20S proteasome

Node Shape
Gene Set

Node Charts: Q-value (FDR) Columns
WBC Spleen Skin SCAT Pancreas Muscle MAT Marrow BAT Bone Brain GAT Heart Intestine Kidney Liver Lung

Node Chart Colors
Positive
1.00 0