SkeletAge: Transcriptomics-based Aging Clock Identifies 26 New Targets in Skeletal Muscle Aging

Muhammad Ali1, Fei Li1\*, Manpreet Katari1\*

1Department of Biology, New York University, New York, NY, 10003

\*Correspondence:

Fei Li:

* Email: [fl43@nyu.edu](mailto:fl43@nyu.edu)
* Address: 24 Waverly Place, New York, NY, 10003
* Phone: (212) 998-3746
* Fax: (212) 995-4015

Manpreet Katari:

* Email: [mkatari@nyu.edu](mailto:mkatari@nyu.edu)
* Address: 31 Washington Place, New York, NY, 10003
* Phone: (212) 998-8249

**Keywords**: Ageprint, Biological Aging, Epigenetic Clock, Longevity Drug Discovery, Skeletal Muscle.

# Abstract

Identifying the set of genes that regulate baseline healthy aging - aging that is not confounded by illness - is imperative for understating aging biology. Machine learning-based age-estimators (such as epigenetic clocks) offer a robust method for capturing biomarkers that strongly correlate with age. In principle, we can use these estimators to find novel targets for aging research, which can then be used for developing drugs that can extend the healthspan. However, methylation-based clocks do not provide direct mechanistic insight into aging, limiting their utility for drug discovery. Here, we describe a method for building tissue-specific bulk RNA-seq-based age-estimators that can be used to identify the *ageprint*. The ageprint is a set of genes that drive baseline healthy aging in a tissue-specific, developmentally-linked fashion. Using our age estimator, SkeletAge, we narrowed down the ageprint of skeletal muscles to 128 genes, of which 26 genes have never been studied in the context of aging or aging-associated phenotypes. The ageprint of skeletal muscles can be linked to known phenotypes of skeletal muscle aging and development, which further supports our hypothesis that the ageprint genes drive (healthy) aging along the growth-development-aging axis, which is separate from (biological) aging that takes place due to illness or stochastic damage. Lastly, we show that using our method, we can find druggable targets for aging research and use the ageprint to accurately assess the effect of therapeutic interventions, which can further accelerate the discovery of longevity-enhancing drugs.

# Introduction

Historically, aging has been largely accepted as a part of life. However, recent advances in geroscience have helped usher in a social paradigm shift towards accepting aging as more of a disease [1] that, at the very least, warrants investigation into its molecular architecture and, at most, needs a cure.

Over the years, researchers have uncovered lifestyle interventions that may extend lifespan, such as caloric restriction (CR) and exercise. However, CR, which shows optimistic results in model organisms, does not translate very efficiently in humans, with an average lifespan extension of up to 5%, contingent on the age at which an individual starts CR [2]. There is some evidence that CR may reduce the pace of aging by about 2 – 3% [3], which could offer promising dividends when compounded over the years. However, given the low returns, it is still far from being a readily adaptable strategy. On the other hand, regular moderate exercise has shown numerous benefits ranging from neuroprotection [4–6] to reduction in inflammation [7,8], protection from metabolic [9] and cardiovascular disorders [10] and more. However, retrospective analyses have shown that regular exercise can only extend lifespan by up to 4 years [11]; spending 300 – 600 minutes every week on voluntary moderate exercise for over 30 years only reduces the risk of all-cause mortality by up to 31% [12].

Given the meager effect of lifestyle interventions on aging, it is no surprise that even with an ideal “anti-aging” lifestyle, we would still expect an individual to age. This gives rise to the hypothesis that aging is not just the accumulation of stochastic damage over the years but is also developmentally-coded into our DNA [13]. Thus, it stands to reason that, like any other biological process, aging – at least to some degree – must also be under the control of certain genes. If we could identify these genes, we can not only begin to understand the mechanisms of aging but also start developing pharmacological interventions against aging that could potentially extend the lifespan beyond the current best lifestyle-based interventions’ capacity. Studies in model organisms, such as *C. elegans* [14,15] and mice [16–18], have demonstrated that genes can greatly influence the lifespan of an organism. Thus, it is reasonable to presume that an underlying developmental-genetic-footprint regulates how we age. We hypothesize that this *ageprint* is a tissue-specific unchanging set of co-expressed (and perhaps co-regulated) genes whose expression is the main driver of “healthy” baseline aging – the age-associated phenotypes that appear despite living a healthy lifestyle and not having any (chronic) illness. One of the main reasons we believe the *ageprint* is tissue-specific is the fact that different tissues do not always express the same genes at the same levels. Moreover, there is evidence that tissues do not age at the same rate [19–22], which further suggests that the underlying mechanisms of aging differ from tissue-to-tissue. Thus, identifying a tissue’s ageprint is pivotal to understanding and eventually preventing (biological) aging in said tissue. Ideally, the best way to identify the ageprint would be doing thousands, if not hundreds of thousands, of *in vivo* genetic screens. However, there are thousands of genes in a single organism, and a multitude of them likely regulates aging. Therefore, in a race against time, using knock-out screens and then translating them into anti-aging drugs may take decades, if not centuries. Several groups have tried to look at differentially expressed genes (DEGs) between younger and older people to identify the mechanisms of aging [23–27]. However, we believe that the ageprint may not be differentially expressed between older and younger people. Differential expression represents drastic changes in gene expression, which are more likely to be associated with the damage-accumulating model of aging, rather than the developmental model of aging. Thus, we believe that the ageprint will likely demonstrate relatively stable expression over the ages because it is driving aging that cannot be accounted for by disease, damage, or perturbations; it is the innate aging that cells are programmed to go through. Thus, one of the fastest ways to identify the ageprint is to leverage machine learning – specifically “aging clocks” – and find genes that explain aging in healthy people at both older and younger ages.

Aging clocks – or age-estimators – are essentially machine learning models that can predict chronological (or biological) age based on a set of biomarkers, such as telomere lengths [28], DNA methylation [29,30], transcriptional profiles [27,31,32], IgG glycosylation level [33], quantitative proteomic data [34], metabolite levels [35,36]. Some clocks combine several biomarkers to create composite *biological aging* scores [37–39] and can even predict all-cause mortality [40]. Fascinating as it is, these clocks can also be used to identify the most important biomarkers for aging from a huge set of biomarkers – such as all the genes in a genome – using feature selection methods. Thus, it is reasonable to presume that any set of biomarkers that strongly correlates with aging and can explain aging both at younger and older ages may offer insight into the mechanisms of (healthy) aging.

Currently, the gold standard for age-estimators is methylation-based clocks [41], commonly known as epigenetic clocks, DNAm clocks, or epigenetic age estimators. These clocks tend to have low errors and usually a high correlation between the predicted epigenetic age and chronological age. Epigenetic clocks were first popularized by the Hannum clock [29] and Horvath’s clock [30]. Since then, many different clocks have been developed, with PhenoAge [38] and GrimAge [40] being two of the most popular second-generation clocks that can predict biological age and all-cause mortality, respectively. These clocks are based off one common principle: methylation levels at certain CpGs (herein referred to as “clock CpGs”) across the genome are linearly correlated with age [42]. This phenomenon allows us to build robust (epigenetic) age estimators using methylation data to precisely predict an individual’s chronological (or biological) age and measure potential epigenetic age acceleration or deceleration [43]. Moreover, these clocks can also help us identify which CpGs are associated with aging. However, the biggest problem with epigenetic clocks is the lack of direct mechanistic relevance of CpGs [44] in aging; we do not know how CpG methylation directly translates to processes that influence aging. Different groups have tried to gain a mechanistic understanding of the importance of clock CpGs by trying to find associations with genes [45] or genomic locations [46,47], or they tried mapping the location of these CpGs to promoters and enhancers [29,30] to gain a better biological insight into the functional relevance of these clock CpGs in aging. However, these correlations only serve as a surrogate for a real mechanistic explanation.

On the other hand, biological clocks that use biomarkers whose mechanism can be more easily interpreted, such as clocks based on transcriptomics [27,31,32], proteomics [34], metabolomics [35,36], or even the gut microbiome [48], are usually not as precise as DNAm clocks [46]. That could be because people try to build pan-tissue clocks. Given that every tissue’s underlying ageprint may differ, it is easy to see why most of these clocks may not achieve great precision when trained using many different types of tissues – the number of non-CpG biomarkers that can explain aging in all tissues might be very small. The tissue-specific noise in the data may add a lot of variance that throws off the model. Thus, to readily identify the ageprint of a tissue, we need to move away from pan-tissue clocks towards robust tissue-specific clocks that are built using easily-interpretable biomarkers. This would allow us to derive mechanistic insights into a particular tissue’s aging directly from the clocks and help us identify new targets for drug development.

To explore this hypothesis, we built a transcriptomics-based clock for skeletal muscles: *SkeletAge*. We chose skeletal muscles because we wanted to capture the genes driving true baseline aging – the *ageprint* – and muscles exhibit primary sarcopenia, an age-associated decline in muscle mass that any other disease or disorder cannot explain [49]. This is an example of “healthy” aging, i.e., when a person who is free of any major diseases or conditions and follows a healthy lifestyle experiences age-related phenotypes. The existing transcriptomics-based age estimators either did not explore tissue-specificity and focused more on generalizability [27] or were trained on post-mortem samples’ RNA-seq data [32], which may not accurately reflect the baseline transcriptomic profile of the antemortem tissue [50–52]. To the best of our knowledge, there are no transcriptomics-based age estimators specifically for skeletal muscles built using RNA-seq data exclusively from healthy people. Thus, we trained SkeletAge on skeletal muscle RNA-seq data specifically from healthy people over a broad age range (19 – 90 years). We hypothesized that due to the nature of SkeletAge’s training set, we would be able to identify the genes that form the ageprint of skeletal muscles because these genes would explain healthy aging at both younger and older ages, which suggests that these genes *imprint* aging into the skeletal muscle tissue.

In line with our hypothesis, SkeletAge performed remarkably and showed a strong correlation between the predicted and chronological age (0.91). It also narrowed down the skeletal muscles’ ageprint to 128 genes that seem to be central to aging skeletal muscles; 26 of these genes have never been studied in the context of aging or age-associated phenotypes before. The ageprint identified by SkeletAge can be linked to both development and known skeletal muscle aging phenotypes. Additionally, we found that 23 out of the 128 genes were present in the druggable genome. Of the drugs associated with the druggable genes, 4 approved drugs had already shown promising results in age-related studies. Hence, we demonstrate that tissue-specific RNA-seq-based age-estimators can not only pinpoint a tissue’s ageprint but also discover novel genetic targets for aging. This opens doors to a powerful new way of using tissue-specific age estimators to identify a tissue’s ageprint, which will help in not only untangling the web of genetic mechanisms that influence tissue-specific aging but also accelerate the development and testing of anti-aging therapies by identifying druggable targets and providing a quick, intelligible readout in the form of direct transcriptomic changes, respectively.

# Results

## The right training set for *ageprint* detection

To identify the genes of interest in skeletal muscle aging, we built a transcriptomics-based skeletal muscle aging clock called *SkeletAge*. We wanted to capture the baseline tissue-specific signature of aging – the *ageprint*. Given that we hypothesized the ageprint to be a genetic signature that explains healthy aging at both older and younger ages and is selected based on a machine learning model, it was imperative that we chose the right training set. If the people in our training set were not healthy (or too healthy – such as professional athletes), they could have confounded the feature selection process of the model. Therefore, by analyzing multiple studies on healthy aging genomics [53,54], we devised a set of criteria for exclusion from the training set: BMI >20 and <30 kg/m², no regular exercise (more than 20 minutes twice a week), or history of professional athleticism or training. The participants in the training set should also be free of any major chronic illnesses or conditions such as heart disease, lung problems, metabolic disorders, neurological pathologies, or cancer. They must not have any physical or cognitive impairment and must not have undergone a serious medical procedure such as surgery. Additionally, they must not be on any medication needed to manage chronic conditions. The individuals need to be alive at the time of sample collection because of the marked differences between the antemortem transcriptome and the thanatotranscriptome of skeletal muscles [55]. Based on this selection criteria, we trained an elastic net model (λ = 19.15027 | α = 0.1) using transcriptomics data of healthy subjects from three different studies that closely met this criterion: GSE164471, GSE97084, and GSE144304. In total, we had 236 samples that we used to train our clock. The individuals comprised a broad age range (19 – 90 years old | mean age: 52 years | SD: ±24.6 years) and were reported to be healthy when their samples were taken (Figure 1).

## Performance metrics of SkeletAge

Next, we sought to test how well SkeletAge could predict someone’s chronological age. SkeletAge performed decently enough in the training set and strongly correlated with chronological age (MAE: 6.55 years | R2: 0.90 | Correlation: 0.96). We further validated SkeletAge using another 298 samples from 8 different studies (GSE196387, GSE20543, GSE157988, GSE151066, GSE58608, GSE159217, GSE200398, GSE186045). The samples in the validation set exhibited various ages, conditions, and/or treatments. For instance, there were athletes, people who were obese, or people suffering from end-stage osteoarthritis. We chose a dataset with a wide variety of ages and conditions to see whether our clock can predict the chronological age precisely, regardless of health status. Being able to do so would suggest that it predicts age based on the genes that drive baseline aging and are likely involved in the developmentally programmed aging mechanisms that are non-redundant with the stochastic damage-associated pathways of aging. Even for this widely diverse dataset, SkeletAge performed well and demonstrated a strong correlation with the chronological age (MAE: 6.62 | R2: 0.79 | Correlation: 0.91). Using a paired Mann-Whitney U test, we found that the difference between the samples' predicted and real ages was not statistically significant (V = 21723 | p-value = 0.7108). This suggests that SkeletAge is indeed capturing the ageprint of skeletal muscles that drives baseline aging regardless of the individual’s health status. The scatter plot for the validation set’s predicted ages and the best fit line for the validation set can be found in Figure 2A. Detailed information about all the conditions and groups used for the training and the testing is present in Table S1.

## Comparison of SkeletAge with RNAAgeCalc in skeletal muscles

SkeletAge is a tissue-specific age-estimator trained to predict the ageprint – the baseline signature of healthy aging – in order to identify targets for tissue-specific anti-aging therapies. To compare whether training our model using data exclusively from healthy people offered a significant advantage over a pan-tissue RNA-based age estimator trained on data with less stringent constraints for their exclusion criteria, we compared SkeletAge with RNAAgeCalc [32]. We installed the RNAAgeCalc package for R using Bioconductor and found that for our validation set, RNAAgeCalc performed worse (MAE: 16.7 years | R2: -0.16 | Correlation: 0.75; 95% CI for Correlation: 0.69 – 0.79), compared to SkeletAge (MAE: 6.62 | R2: 0.79 | Correlation: 0.91; 95% CI for Correlation: 0.89 – 0.93), as shown in Figure 2B and 2C. Since our validation set was very diverse with many conditions and/or treatments, we assumed that the phenotypic variation may be confounding RNAAgeCalc’s predictions. Therefore, we decided to test RNAAgeCalc on SkeletAge’s training set, which consisted of individuals reported to be healthy (Figure 2D). Here, RNAAgeCalc performed relatively better (MAE: 20.7 years | R2: 0.58 | Correlation: 0.79; 95% CI for Correlation: 0.74 – 0.83) but was unsurprisingly outmatched by SkeletAge (MAE: 6.56 | R2: 0.90 | Correlation: 0.96; 95% CI for Correlation: 0.95 – 0.97). This suggests that RNAAgeCalc gets skewed by confounders such as the clinical phenotypes of different individuals and thus does not use the ageprint for age prediction as SkeletAge does. This is likely due to the fundamental differences in the training sets of the two models. While we trained SkeletAge using data from people who were healthy and alive at the time of sample collection, RNAAgeCalc was built using post-mortem bulk RNA-seq data from GTEx [56,57]. Due to the post-mortem nature of tissue collection, gene expression levels may not accurately reflect the true phenotypic nature of the antemortem tissue [50–52]. The skeletal muscle thanatotranscriptome shows marked differences from the antemortem gene expression patterns [55], which may explain why RNAAgeCalc cannot identify the ageprint of a tissue. We used paired Mann-Whitney U tests to validate the differences between RNAAgeCalc and SkeletAge. RNAAgeCalc’s predicted age – RNAAge – was significantly different from the chronological ages of the samples in both the training set (V = 9253 | p-value = 6.649e-06) and the validation set (V = 3494 | p-value = 1.742e-36). On the other hand, the difference between the ages predicted by SkeletAge and the chronological ages of the samples was not significant in both the training set (V = 13902 | p-value = 0.9389) and the validation set (V = 21723 | p-value = 0.7108). Additionally, we found that the difference in the ages predicted by SkeletAge and RNAAge was also significant for both the training (V = 20466 | p-value = 6.646e-10) and the validation sets (V = 39880 | p-value = 2.913e-32). SkeletAge’s validation set error was also smaller than the error reported for the peripheral blood-based transcriptomic age-predictor built by Peters et al.: 6.6 years for SkeletAge compared to 7.8 years for the age-estimator built using peripheral blood data [27]. This further highlights the utility of moving from generalizable blood-based clocks to more precise tissue-specific age-estimators that can capture the ageprint of a specific tissue, which can then be used to understand the mechanisms of aging and discover potential targets for anti-aging therapies.

## SkeletAge predicts 26 novel targets in skeletal muscle aging

Next, we sought to understand what genes SkeletAge was using to calculate the biological age. This would allow us to identify skeletal muscles’ ageprint. To do that, we obtained the genes associated with the 129 non-zero coefficients used by the model. We manually verified that 102 of the 129 genes have previously been reported to be either involved – or studied in the context of – aging or aging-associated phenotypes; these genes are present in Table S2. Out of the remaining genes, one gene (EntrezID: 107986257) had its record withdrawn from NCBI because it did not appear in a later iteration of the genome. The remaining 26 genes – to the best of our knowledge – have never been studied in terms of aging, age-associated phenotypes, or skeletal muscle aging; these 26 novel targets are present in Table 1. We further confirmed the centrality of these genes in capturing the ageprint in skeletal muscle tissue by running a principal component analysis (PCA) using the normalized counts of these genes to separate the samples into different clusters. Unsurprisingly, the samples clustered into distinct clusters based on their age groups, regardless of whether we used our training data (Figure 3A; healthy individuals), the validation set (Figure 3C; various conditions), or the entire dataset (Figure 3F; various conditions). An important thing to note here is that the entire dataset contains individuals suffering from conditions such as type-2 diabetes, end-stage osteoarthritis, obesity, and more. Thus, SkeletAge’s genes’ ability to distinctly cluster such a diverse group demonstrates that it robustly captures the signatures of intrinsic aging – the ageprint – which are rooted in the molecular architecture of development that drives aging regardless of health status. Within those 129 genes, 64 were predicted to be decelerating the biological age (β < 0) – referred to as ProLong (promoting longevity) genes from here on – while the remaining 64 (excluding 107986257) – referred to as RedLong (reducing longevity) genes from here on – were predicted to be involved in biological age acceleration (β > 0). Both the ProLong and the RedLong genes’ expression can cluster samples very distinctively in the training set (Figure 3B) and the validation set (Figures 3D and 3E), where the age-based divide is clearly visible. This suggests that these two sets of genes capture the intrinsic aging signatures at a great resolution, even on their own. Interestingly, in the validation set, the PCA for the RedLong genes shows tighter clusters (Figure 3D) compared to the clusters produced by the ProLong genes (Figure 3E). This suggests that the expression pattern of the ProLong genes is more receptive towards phenotypic changes and health status than the RedLong genes because both the ProLong and the RedLong genes produce identical PCA plots for the training set where all samples are healthy (Figure 3B), but in the validation set, where not everyone was healthy, their plots are a little different.

## Relationships between Age, ProLong, and RedLong genes

To further elucidate how the ProLong and RedLong genes were related to each other (and aging), we calculated the correlations between the 128 annotated genes and age. 127 genes were significantly correlated with age (p-value < 0.05), and only *MIR3659* was insignificantly correlated with age (p-value: 0.078). Half of the 128 genes were positively correlated with age, while the other half were negatively correlated with age (Figure 4A). The correlations for either side of the spectrum did not exceed 0.8 or -0.8, respectively. A rather surprising finding here was that all the ProLong genes were negatively correlated with age, while all the RedLong genes were positively correlated with age (Figure 4A). This suggests that genes that accelerate aging undergo an increase in expression as we age. In contrast, the ones that decelerate aging show gradually decreased expression as we age – in principle, that makes sense. To test that theory, we compared the mean log fold changes of ProLong and RedLong genes across the nine different age groups being tested in our study (decades 2 – 10 | decade 3 is the reference: 20 – 30 years) and found that indeed, for both ProLong and RedLong genes, the mean expression incrementally reduces and increases, respectively, as we go from younger age groups towards the older age groups (Figure 4B). This difference across decades and types of genes is statistically significant (F = 32.405 | p-value = 8.900e-41). In addition to these trends, the expression of different genes from both the ProLong and RedLong groups also correlates with each other, as seen in the heatmap of the genes’ correlation (Figure 4C). All this data suggests that both ProLong and RedLong genes likely work together in a linked-regulatory fashion to orchestrate the threads of intrinsic skeletal muscle aging. This further corroborates that the ageprint is likely a set of co-expressed (and likely co-regulated) genes that drive baseline aging in a tissue.

## Characterizing the ageprint genes

To understand the possible aging mechanisms of the ageprint genes, we looked at the molecular functions of the proteins they produce. We found that the functions of the genes involved a wide variety of cellular mechanisms, including epigenetic modifications, transcription regulation, DNA repair, and development. Several genes were also found to code for growth factor receptors, kinase domain-containing proteins, and zinc-finger proteins. An overview of some interesting genes and their functions is provided in Table 2. The presence of genes associated with growth factor receptors, insulin-like growth factors, kinases, PIP phosphatases, and NAD-binding proteins suggests that the ageprint of skeletal muscles largely influences the nutrient-sensing mechanisms of aging [58,59]. Moreover, the presence of HOX genes further corroborates our theory that the ageprint is a developmentally-linked genetic signature that regulates aging in a tissue. Our findings are further supported by evidence from the literature, which shows that HOX genes are associated with aging [60], including aging in skeletal muscles [61]. Functional enrichment analyses did not yield tangible results since the number of genes was too small (128 genes).

## Novel targets are related to nucleotide and fatty acid metabolism

We performed GO-term enrichment to understand the biological processes associated with the 26 novel targets identified by SkeletAge (Figure 5). Most GO-terms can be categorized into two distinct processes: nucleotide and fatty acid metabolism. This suggests that these novel targets, constituents of skeletal muscles’ ageprint, are linked to nucleotide and fat metabolism. Out of the 26 novel targets, 16 are predicted to increase longevity (ProLong genes), while the remaining 10 are predicted to reduce longevity (RedLong genes). Most GO-terms associated with the 16 novel ProLong genes seem to be involved in nucleotide metabolism. This suggests that the genes that regulate nucleotide metabolism in skeletal muscles show downregulation with age, given how all the ProLong genes were negatively correlated with age and showed a consistent, incremental decline in their expression. In fact, sarcopenia, the characteristic decline in muscle mass, volume, and strength, is a hallmark of skeletal muscle aging [62] that may be linked to changes in cellular adenine metabolism [63]. This suggests that addressing some of the novel targets discovered by SkeletAge might help overcome declining NAD levels in aging muscle tissue [64]. On the other hand, the 10 novel targets predicted to be involved in reducing longevity (RedLong genes), are linked to GO-terms that suggest an involvement in fatty acid metabolism. Since all RedLong genes were positively correlated with age, it seems that genes that regulate fatty acid metabolism in muscles get upregulated over time. Increased fat accumulation in the muscles (myosteatosis) is a pathology that is distinct from sarcopenia [65–67]. Myosteatosis is recognized as a common occurrence in aging, which leads to reduced muscular strength [68] and altered muscle architecture [69] that ultimately adds to age-related musculoskeletal phenotypes. Therefore, targeting these novel RedLong genes might help overcome some of the myosteatosis-associated phenotypes in aging muscles.

## Finding druggable targets to discover pharmacological interventions for aging

To see if any of the genes identified by SkeletAge were druggable targets, we ran the ageprint on DGIdb [70,71] and filtered the hits for all the genes with an interaction score of > 1.0 (Figure 6). By doing so, we found that 23 of the ageprint’s constituent genes were part of the druggable genome [72,73], and 9 of those genes had approved drugs listed against their protein products (Table 3); one of these 9 genes (*PDE7B*) was among the 26 novel targets discovered by SkeletAge. Interestingly, out of the 9 genes with approved drugs, 3 genes (*PDE7B*, *NT5C2*, *TRIB3*) were associated with drugs that have previously been investigated in aging. *PDE7B* was associated with dyphylline, which was recently shown to improve fertility in naturally-aged mice by increasing primordial follicle activation – the results were also replicated in human ovarian cells [74]. Didanosine, a reverse transcriptase inhibitor linked to *NT5C2*, was shown to extend the lifespan of *C. elegans* by almost 28% [75,76]. Indapamide and perindopril were both linked to *TRIB3*. Indapamide, which can cross the blood-brain barrier, demonstrated neuroprotective and antioxidative effects against age-associated lipid peroxidation and loss of myelin and axons in middle-aged mice [77]. Perindopril was directly investigated in older people who had problems related to physical mobility in a clinical trial (International Standard Randomised Controlled Trial Register: ISRCTN67679521) and was found to improve 6-minute walking distance, lower blood pressure, and lead to an overall better self-reported quality of life. The effect of perindopril on the improvement in physical capacity was comparable to 6 months of exercise for the elderly [78]. This suggests that the ageprint predicted by SkeletAge very likely contains genes that influence and regulate the mechanisms of baseline aging in skeletal muscles. In the future, we will hopefully see more potential drugs listed in Table 3 being studied in the context of skeletal muscle aging. These findings underscore the importance of identifying a tissue’s ageprint for longevity drug discovery.

SkeletAge can be used to assess the effect of longevity therapeutics

Seeing how SkeletAge could identify druggable targets, we wanted to see if it could correctly predict the effect of potentially therapeutic compounds that target aging pathways. To test that, we retrospectively analyzed the transcriptomics data from a clinical trial (NCT03151239) using SkeletAge. NCT03151239 sought to investigate the role of NMN in skeletal muscle aging in pre-diabetic women, who were either obese or overweight, and found that while it did improve insulin response in the muscles, there was no effect on measures of physical fitness such as handgrip strength and time to recovery from fatigue, after 10 weeks of 250 mg/day NMN administration [79]. In line with their findings, SkeletAge also did not show any significant difference between the chronological ages and the predicted ages of the placebo compared to the NMN group (V = 547 | p-value = 0.5514). This suggests that SkeletAge can capture clinical phenotypes elicited by aging therapeutics and can effectively assess the effect of said therapeutics using rapidly scalable bulk RNA-seq data.

## Skeletome: a skeletal muscle database to accelerate aging research

The first step of building SkeletAge was identifying the right datasets for the training and validation set. Given the stringent requirements for our inclusion criteria, we had to gather a lot of publicly available RNA-seq data from GEO. One of the main challenges was acquiring accurate age labels for different samples because most groups had reported average ages for different groups in their papers. Thus, we reached out to several authors and corresponding principal investigators on the papers to get information about the individual samples’ ages and then put it all together manually. Over time, we curated a large database of 534 samples in total, which pulls together data from over 11 different studies, including a couple of clinical trials – we call this database *Skeletome*. The samples in the Skeletome range from 19 – 90.7 years, with an average age of 52 years and a standard deviation of ±21.3 years (Figure 7), thereby covering a large enough proportion of ages to measure genetic effects across the individuals ranging from young adults to almost centenarians. Additionally, Skeletome comprises data from studies that investigated the pre- and post-intervention effects of different “anti-aging” supplements and lifestyle patterns such as nicotinamide mononucleotide (NMN), high-intensity workout, and even sedentary lifestyle. Moreover, it also has data on various health conditions that may accelerate biological aging, such as obesity [80,81] and type-2 diabetes [82]. All the data in Skeletome comes from RNA-seq performed on biopsies of vastus lateralis tissue, thus ensuring homogeneity and reducing systematic errors and confounders arising from differences in skeletal muscle types. In addition to the metadata, Skeletome contains the raw counts for each sample, aligned to the GRCh38, thereby making it simpler (and faster) to study gene expression in skeletal muscle tissue across different ages. In compiling Skeletome, we hope this database would further accelerate research in skeletal muscle genomics by removing a key bottleneck: the time it takes to reinvent the wheel by individually reaching out to every team and getting age labels. You can access the Skeletome data through SkeletAge’s GitHub repository: <https://github.com/mali8308/SkeletAge/>.

# Discussion

Recent paradigm shifts in geroscience research have created opportunities to address aging as a disease and investigate it like one. While identifying disease-causing genes is usually one of the first steps for tackling non-infectious diseases, genes that cause aging have remained rather elusive in humans. Thus, in this paper, we describe a method to identify the *ageprint* of a tissue – the tissue-specific set of genes that regulate developmentally-linked baseline “healthy” aging, i.e., aging that occurs when there is no disease or anti-aging interventions at play – using a transcriptomics-based aging clock called *SkeletAge*. Using SkeletAge we narrowed down the skeletal muscle ageprint to 128 genes; 26 of these genes have never been studied in the context of aging or aging-associated phenotypes. Additionally, 23 ageprint genes were found to be a part of the druggable genome. Hence, we demonstrate how identifying the ageprint of a tissue can enhance our understanding of the mechanisms that drive aging and accelerate the development of pharmacological interventions against aging.

In skeletal muscles, SkeletAge performs better than a previously built tissue-specific age-estimator, RNAAgeCalc [32]. We believe that this difference in performance can be accounted for by the discrepancy in the training sets of SkeletAge and RNAAgeCalc. SkeletAge was built using data from healthy people who were alive at the time of sample collection, whereas RNAAgeCalc relies on post-mortem data from GTEx. Thus, we also highlight a new criterion for choosing the training set when building a transcriptomics-based age-estimator to identify the ageprint of a tissue.

Based on GO-term enrichment, the 26 novel targets identified by SkeletAge seem to be involved in nucleotide metabolism and myosteatosis. There is a marked decrease in NAD levels with age [64]. A study found that knocking out Nampt – an enzyme that helps produce NAD and, consequently, ATP – leads to progressive muscle weakness and a decline in physical fitness as mice age. Moreover, they also found that the weakened phenotype can be rescued through NR supplementation, a compound that bypasses the need for the Nampt enzyme. Additionally, the team overexpressed Nampt in the skeletal muscles and found that it preserves some endurance in older mice [63]. This suggests that NAD and its precursors might help in overcoming some of the age-associated skeletal muscle dysregulation, as they have in yeast [83,84]. However, as in the clinical trial NCT03151239 [79], we found that NAD did not significantly lower the biological age of skeletal muscles. This suggests that while the genes linked with NAD metabolism contribute to aging in skeletal muscles, their (dys)function likely cannot be rescued by NMN supplementation alone. Myosteatosis, the increased intramuscular accumulation of fats, is another phenotype associated with skeletal muscle aging [85], that is not linked to sarcopenia [65–67]. Perhaps studying the novel targets identified by SkeletAge might help elucidate how to regulate or reduce myosteatotic changes in muscles as we age. The ability of SkeletAge to identify a group of genes that not only precisely predict and (as a set) strongly correlate with aging but also have functions that reflect known phenotypes associated with skeletal muscle aging suggests that it is correctly identifying the ageprint. This is further corroborated by the fact that using the ageprint, SkeletAge can accurately assess the effect of longevity therapeutics on aging.

Some of the genes involved in the ageprint were also a part of the druggable genome [72,73]. 9 of these genes were associated with approved drugs, and out of those 4 drugs (dyphylline, didanosine, indapamide, perindopril) had already been tested in the context of aging. Dyphylline improved fertility in aged mice (*in vivo*) and human oocytes (*in vitro*) by activating primordial follicles by inhibiting phosphodiesterases that break down cAMP [74]. This leads to the activation of the PI3K/Akt pathways, which results in the translocation of FOXO3 to the cytoplasm and the activation of follicles [86]. Thus, dyphylline might help in rescuing fertility in older women who face ovarian insufficiency, which can lead to premature menopause [87]. Didanosine was recently shown to be able to extend the lifespan of *C. elegans* [75,76]. Didanosine is a reverse-transcriptase inhibitor that likely causes mitochondrial toxicity, which leads to mitohormesis – a process whereby mitochondrial damage leads to the upregulation of cellular defense mechanisms, including the overexpression of ATF-4, a transcription factor that helps cells respond to stress by inhibiting translation [88] and has been shown to increase lifespan in worms [89]. This suggests that the use of didanosine (or other mild mitotoxic drugs) might lead to preferential expression of stress-response genes [90], and the increased translation of protective proteins [91] through the expression of ATF-4, which might culminate in a net protective response that may extend lifespan [75]. Indapamide, a diuretic used to treat hypertension [92,93], was found to have neuroprotective and antioxidative effects against age-induced oxidative damage and myelin loss in mice [77]. With age, reactive oxygen species (ROS) production increases in microglia [94], likely through the increased expression of NADPH oxidase, which can lead to tissue damage. Indapamide, which has antioxidant properties [95] and can cross the blood-brain barrier, has been shown to reduce axon loss and demyelination in middle-aged mice, which is associated with reduced oxidative stress [77]. This suggests that indapamide and other similar antioxidants may offer neuroprotection against myelin and axon loss as we age. Perindopril was investigated in humans in a clinical trial (ISRCTN67679521) and was found to improve 6-minute walking distance in elderly people by about 30 meters – this is comparable to over 6 months of exercise. Perindopril is an angiotensin-converting-enzyme (ACE) inhibitor, a class of drugs used to treat COVID-19 with mixed results [96–98]. ACE inhibitors have previously been linked to a slower decline in muscular strength in elderly women [99]. Moreover, using perindopril also helped improve the 6-minute walking distance in elderly people who had experienced heart failure [100]. This suggests that perindopril could likely improve vascular function in the elderly and help slow aging. Thus, these findings suggest that the remaining approved drugs that were linked to the ageprint (dehydrated alcohol, selpercatinib, pralsetinib, tramadol, risperidone, flavoxate hydrochloride, bevacizumab-awwb, nelarabine, tezepelumab), might also have geroprotective mechanisms that could help alleviate some of the phenotypes associated with aging.

SkeletAge’s ability to accurately reflect the effect of longevity therapeutics on aging can be used to evaluate whether a pharmacological intervention works or not. Drugs that target aging pathways are usually evaluated using methylation-based clocks, which can be a rather hefty investment as the sample size increases. As an alternative, ageprint identification can allow researchers to use rapidly scalable bulk RNA-seq data to assess whether a therapy or drug has worked. If something looks promising, it can be investigated further or confirmed using methylation-based clocks. Hence, ageprint identification can also lower costs associated with testing drugs designed to slow down aging.

One of the main limitations of using the ageprint to study aging is its tissue-specificity, which means it cannot be generalized as a universal signature of aging. However, recent evidence has shown that different tissues age at different rates [19–22], so it might make more sense to target each tissue’s aging individually. While this approach is not perfect, we believe that by finding the ageprint for every tissue, we can find the *universal ageprint* for humans at the intersection of all the ageprints. That universal signature of healthy aging might help us develop drugs that simultaneously target aging at the molecular level in multiple tissues. Thus, in the future, we hope to dissect this universal ageprint further and get closer to understanding the basic mechanisms of aging.

Finally, our findings demonstrate that using tissue-specific age predictors to identify the ageprint is an efficient way of discovering tissue-specific novel targets for aging research, which can inform both the mechanistic and the drug-discovery side of aging research.

# Methods

## Training set inclusion criteria

To capture the ageprint of skeletal muscles, we wanted to train our elastic net model on vastus lateralis samples taken from healthy individuals over a broad age range. We hypothesized that genes that can explain healthy aging at younger and older ages would constitute the true ageprint of a tissue. We decided on 236 total samples from GSE164471, GSE97084, and GSE144304. The data in GSE164471 came from the participants of the GESTALT study [54]. The study only included participants who met the IDEAL (Insight into the Determination of Exceptional Aging and Longevity) standard developed by the National Institute on Aging [101]. The participants had no chronic illnesses (except for hypertension and cancer that has not relapsed for the last 10 years), that they took medication for (except for antihypertension medication), no prior history of professional training, had a BMI < 30 kg/m2, and were not diagnosed with any physical impairment, major disease, or cognitive damage. For the data taken from GSE97084, participants were only included if they did not work out regularly (more than 20 min, twice a week), had no cardiovascular problems, or metabolic disorders such as hypo- or hyperthyroidism, type-2 diabetes (fasting blood glucose > 110 mg/dL), and their BMI ≤ 32 kg/m2. Participants were excluded if they had metal implants, suffered from renal diseases, were pregnant, or had a history of blood clotting problems. Participants were also excluded if they used any of the following medications: corticosteroids, tricyclic antidepressants, sulfonylureas, barbiturates, peroxisome proliferator-activated receptor γ agonists, anticoagulants, β blockers, opiates, insulin, and insulin sensitizers [53]. The data from GSE144304 comes from the FITAAL study (Dutch Trial Registry identifier: NTR6124). Participants were not included in the trial if they had COPD, cancer, dementia, cardiac failure, neuromuscular disorder, anemia, or were contraindicated to receive a muscle biopsy. They were also excluded if they were currently part of another study or had recently undergone an intensive medical treatment or procedure, such as surgery. Other physiological parameters that could get participants excluded were a BMI <20 kg/m2 or >25 kg/m2, having type I or II diabetes mellitus, regular workout routine (more than 4 times a week), or pregnancy and/or nursing states [102]. The overarching trend for sample selection across the three different datasets was healthy people with a BMI >20 kg/m2 or >32 kg/m2 who were not professional athletes and did not work out regularly, were free of any major illnesses or chronic conditions that would require medication and/or treatments which can – in-itself or together with the condition – confound the gene expression patterns, and were free of physical or cognitive impairments.

## Creating the Skeletome Database

To study gene expression changes in skeletal muscle, we first created a large data set of 534 vastus lateralis samples from 11 different studies on NCBI Gene Expression Omnibus (GEO): GSE164471 [54], GSE97084 [53], GSE144304 [102], GSE196387 [103], GSE205431 [104], GSE157988 [79], GSE151066 [105], GSE58608 [106], GSE159217 [107], GSE200398 [108], GSE186045 [109] – for all of these studies, we extracted their metadata and the raw counts directly from NCBI GEO. The reads are aligned using HISAT2 [110] to GCA\_000001405.15. Runs with an alignment rate of over 50% are used to generate the counts matrix using Subread featureCounts, and the *Homo sapiens* Annotation Release 109.20190905 is used for gene annotation. Wherever necessary, we reached out to the groups that had published these results and got individual age labels for samples to map accurate ages for the penalized regression model training and testing. The raw counts and the metadata (including the GEO sample identifier, accurate age label, and condition) are present in a database called Skeletome. We hope that Skeletome on its own would help accelerate aging research in skeletal muscle by providing a substantial dataset for future transcriptomics-based clocks, anti-aging therapies, and more. Skeletome can be accessed through our GitHub repository: <https://github.com/mali8308/SkeletAge/>.

## DESeq2 for normalization and expression tracking

To process raw counts, we used DESeq2 [111]. We divided our samples into bins of 10 years, which started from decade 1 (0-10 years) and went all the way up to decade 10 (90-100 years). This made it easier to create a factor with 11 different levels of comparison instead of 116 different levels that we would have had to use if we had chosen individual sample ages as the factors. We chose the third decade (20-30 years) as the reference for DESeq2 because muscle mass tends to start declining as people enter ages beyond 30 years [112] and may reduce by over 40% as they grow past 80 years of age. Since we wanted to capture healthy aging, we used the “prime” age as the reference to compare gene expression against. Moreover, biological age measurements are often non-linear during early growth and development [43], which means that choosing a younger reference level (people who were younger than decade 3) could have confounded the skeletal muscle ageprint that we were trying to detect.

## Glmnet for training SkeletAge and feature selection

SkeletAge is based on an elastic net regression model [113], which helps select important aging genes that form a tissue’s ageprint. We used the glmnet package (4.1-8) in R to generate an elastic net regression model using leave-one-out cross-validation for hyperparameter tuning (λ = 19.15027 | α = 0.1). We trained SkeletAge on the samples in the following studies: GSE164471, GSE97084, and GSE144304. Using a penalized regression model automatically reduced the coefficients for all the genes that were not important down to 0. Thus, to extract the ageprint from SkeletAge, we selected all the β ≠ 0, except the intercept (β0). Using this strategy gave us 128 genes that seem to be central to skeletal muscle aging: the *ageprint*.

## Validating SkeletAge’s performance

To ensure that SkeletAge was truly capturing the ageprint of skeletal muscles, we created a validation set using data from 8 different studies (GSE196387, GSE205431, GSE157988, GSE15106, GSE58608, GSE159217, GSE200398, and GSE186045). The samples in these studies range from extremely healthy (such as athletes) to people who are overweight or obese, have type-2 diabetes, obesity, and more. We wanted to create a diverse set so that their condition would influence their gene expression. This would allow us to test whether the ageprint gets affected by health status by determining whether SkeletAge can still precisely map the correct ages for the samples. Theoretically, a robust ageprint would allow SkeletAge to predict precise ages despite the altered state. Our results demonstrate that even for the diverse dataset, SkeletAge could use the skeletal muscle ageprint to predict ages that were not statistically different from the samples’ chronological ages. Being able to do that using the 128 genes we have identified suggests that these genes are a universal signature that remains unchanged regardless of health status and, hence, play a pivotal role in driving healthy skeletal muscle aging.

## Comparison against RNAAgeCalc and statistical analysis

To benchmark SkeletAge’s performance, we tested it against an already-published RNA-seq-based clock that allows you to specify the tissue for which you would like to predict age: RNAAgeCalc [32]. We used the RNAAgeCalc package in R and used the predict\_age function. We passed it the raw counts of both our training and validation set, with the “tissue” and “exprtype” arguments set to “muscle” and “counts,” respectively. For statistical tests, we ran paired Mann-Whitney U tests in R at the alpha of 0.05. We measured the difference between the chronological ages of the sample and the ages predicted by SkeletAge, the chronological ages of the sample and the ages predicted by RNAAgeCalc, and the difference between the ages predicted by SkeletAge and RNAAgeCalc. We also computed the Pearson correlation coefficient between the chronological ages and the ages predicted by SkeletAge and the chronological ages and the ages predicted by RNAAgeCalc. We also calculated the coefficients of determination (R2) for both SkeletAge and RNAAgeCalc on both the training set and the validation set.

## Enrichment analysis, finding druggable targets, and generating plots

To retrieve information about genes, their descriptions, names, InterPro accession numbers, and InterPro descriptions, we used biomaRt (3.19) [114]. We used the clusterProfiler [115] package in R for enrichment analyses, including GO-term enrichment, reactome enrichment, disease enrichment, and KEGG enrichment. We also used STRINGdb [116] version 12 for protein-protein interaction analysis and networks. Due to the relatively small number of genes, we did not see any enrichment for reactome and KEGG or significant protein-protein interactions with STRINGdb. We used the DGIdb [70,71] to identify the constituent genes of the ageprint, which were also druggable targets. Lastly, we used the ggplot2 package (3.5.1) in R to create all the plots used throughout the paper.

# Abbreviations

* RNA-seq: RNA-sequencing
* CI: Confidence intervals
* MAE: Mean absolute error

# Author Contributions

M.A. and M.K. developed the SkeletAge estimator. M.A. conceived the study and performed the statistical analysis. M.A., F.L., and M.K. wrote the manuscript.

# Acknowledgement

We would like to thank the following people for sharing accurate age labels for the samples which ultimately helped us in building and validating SkeletAge:

* Center for Human Nutrition, Washington University School of Medicine: Samuel Klein, Gordon Smith.
* Division of Cardiovascular Medicine, Stanford University School of Medicine: Maléne Lindholm.
* Department of Physiology and Pharmacology, Karolinska Institutet: Carl J. Sundberg.
* Mayo Clinic: Surendra Dasari, K. Sreekumaran Nair.
* College of Health, Oregon State University: Matthew Robinson.
* Wageningen University: Evert van Schothorst, Jaap Keijer.
* Translational Research Institute, AdventHealth: Katie Whytock, Lauren Sparks, and Paul Coen.

# Conflict of interest

The authors declare no competing interests.

# Funding

This work was funded by NIH R35GM134920 (to F.L.), and NSF MCB-1934628 (to F.L.).

# Figures



**Figure 1: Creating SkeletAge and identifying the ageprint**. We created SkeletAge by taking RNA-seq data of vastus lateralis samples on NCBI GEO and training an elastic net model to predict ages using this gene expression data. This allowed us to build a tissue-specific clock that can identify the ageprint. Using the ageprint, we identified 26 novel genetic targets for skeletal muscle aging that have never been reported in the context of aging research. Creating and testing SkeletAge took about 534 samples, for which accurate age labels had to be acquired from the corresponding authors. Therefore, to accelerate future aging research, we have compiled our dataset, including the metadata with accurate age labels and the raw counts for each sample (along with its GEO sample accession), into a repository called *SkeletAge*, which is accessible through this link: <https://github.com/mali8308/SkeletAge/>.

A graph of different colored dots

Description automatically generatedA graph of different colored bars

Description automatically generated with medium confidenceA graph of blue and orange bars

Description automatically generated

**Figure 2: Comparison between SkeletAge and RNAAgeCalc**.

(A) shows the predicated ages measured by SkeletAge, whereas (B) shows the predicted ages (RNAAge) as measured by RNAAgeCalc. Each point on the scatterplot is a sample from the validation set (n = 298). The color of the points provides the health status or treatment groups. The red line is the line of best fit generated by SkeletAge, and the orange line is the line of best fit generated by RNAAgeCalc. Both models are predicting the ages for the validation set. The dotted line represents a 1:1 correlation between the chronological and the predicted ages. (C) shows the comparison of the mean absolute error (MAE) and Pearson correlation between RNAAgeCalc and SkeletAge when they are used on the validation set (n = 298). (D) is a comparison of the mean absolute error (MAE) and Pearson correlation between RNAAgeCalc and SkeletAge when they are used on the training set (n = 236).

A grid of graphs with colorful dots

Description automatically generated with medium confidence

**Figure 3: Principal component analysis (PCA) plots generated using the ageprint genes.** The plot in (A) shows PCA ran on the training set (only healthy samples). (B) PCA ran on the training set using only the RedLong or the ProLong genes; both produce identical plots, but only one is shown here. (C) The PCA plot generated for the validation set (healthy samples and people suffering from different conditions) uses just the ageprint. (D) The plot generated for the validation set using the RedLong genes, and (E) is the PCA for the validation set using the ProLong genes. (F) shows a PCA plot for the entire Skeletome dataset using the ageprint genes. Each individual point is a sample. The samples are color-coded by the decade they were binned into.

A graph of a number of dots

Description automatically generated with medium confidenceA graph of different colored rectangles

Description automatically generatedA diagram of a diagram

Description automatically generated with medium confidence

**Figure 4: Correlation of the individual genes that constitute ageprint with chronological age**. The genes in orange are RedLong genes, while the ones in purple are ProLong genes. A) shows the entire ageprint and the correlations of individual genes with chronological age. Surprisingly, all the genes negatively correlated with age are ProLong genes, while all the genes positively correlated with age are RedLong genes. B) shows the mean log fold changes in expression of the ProLong and RedLong genes across the different decades. Decade 3 (20 – 30 years) was chosen as the reference because muscle mass decreases once people are past their *prime*. C) is a heatmap of all the genes and their correlations. It shows that subsets of multiple genes’ expression correlate with each other, which further suggests that the genes are likely co-expressed and, potentially, even coregulated.

A graph with different colored bars

Description automatically generated

**Figure 5: GO-terms enriched for the 26 novel targets identified using SkeletAge**. Overall, the terms can be categorized into two major processes: nucleotide and fatty acid metabolism. Both of these processes contribute to age-associated muscle loss, which further demonstrates that identifying the ageprint can shed light on the mechanisms of baseline, healthy aging across tissues.

A graph with red and blue dots

Description automatically generated

**Figure 6: A scatterplot of all the ageprint genes that are part of the druggable genome and their associated drugs.** The size of the points reflects the interaction score. PTGES has the most approved drugs associated with it, followed by *NR2F2* and *KLF5*.

A graph of different colored shapes

Description automatically generated

**Figure 7: A violin plot showing the chronological age distribution across Skeletome**. Different datasets have different age distributions. Each dataset’s mean age is reflected in the intensity of the color. The curve's width is proportional to the number of samples for that particular age range in a dataset.

# Tables

**Table 1: The 26 novel targets predicted by SkeletAge**. These 26 genes are a part of the skeletal muscle ageprint (Table S1) identified by SkeletAge. The table contains information about the genes, including InterPro accessions and descriptions for protein-coding genes.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| EntrezID | Gene Name | Gene Description | InterPro Accession | InterPro Description |
| 8353 | H3C6 | H3 clustered histone 6 | IPR000164; IPR007125; IPR009072 | Histone H3/CENP-A; Histone H2A/H2B/H3; Histone-fold |
| 9200 | HACD1 | 3-hydroxyacyl-CoA dehydratase 1 | IPR007482 | Protein-tyrosine phosphatase-like, PTPLA; |
| 10625 | IVNS1ABP | influenza virus NS1A binding protein | IPR000210; IPR006652; IPR011333; IPR011705; IPR015915; IPR017096; IPR050457 | BTB/POZ domain; Kelch repeat type 1; SKP1/BTB/POZ domain superfamily; BTB/Kelch-associated; Kelch-type beta propeller; BTB-kelch protein; Zinc finger and BTB domain-containing |
| 23568 | ARL2BP | ADP ribosylation factor like GTPase 2 binding protein | IPR023379; IPR038849; IPR042541 | BART domain; ADP-ribosylation factor-like protein 2-binding protein; BART domain superfamily |
| 26136 | TES | testin LIM domain protein | IPR001781; IPR010442; IPR033724; IPR034958; IPR034959; IPR034960; IPR047120 | Zinc finger, LIM-type; PET domain; PET testin; Testin, LIM domain 1; Testin, LIM domain 2; Testin, LIM domain 3; Prickle/Espinas/Testin; |
| 27115 | PDE7B | phosphodiesterase 7B | IPR002073; IPR003607; IPR023088; IPR023174; IPR036971 | 3'5'-cyclic nucleotide phosphodiesterase, catalytic domain; HD/PDEase domain; 3'5'-cyclic nucleotide phosphodiesterase; 3'5'-cyclic nucleotide phosphodiesterase, conserved site; 3'5'-cyclic nucleotide phosphodiesterase, catalytic domain superfamily |
| 55344 | PLCXD1 | phosphatidylinositol specific phospholipase C X domain containing 1 | IPR000909; IPR017946; IPR042158; IPR051057 | Phosphatidylinositol-specific phospholipase C, X domain; PLC-like phosphodiesterase, TIM beta/alpha-barrel domain superfamily; PI-PLC X domain-containing protein 1/2/3; Phosphoinositide phospholipase C domain-containing protein |
| 64208 | POPDC3 | popeye domain containing 3 | IPR006916; IPR018490 | Popeye protein; Cyclic nucleotide-binding domain superfamily |
| 81553 | CYRIA | CYFIP related Rac1 interactor A | IPR009828; IPR039789 | CYRIA/CYRIB, Rac1 binding domain; CYFIP-related Rac1 interactor |
| 139221 | PWWP3B | PWWP domain containing 3B | IPR035504; IPR040263; IPR048765; IPR048795 | MUM1-like, PWWP domain; PWWP domain-containing DNA repair factor 3A/3B/4; PWWP domain-containing DNA repair factor 3A/3B/4, N-terminal; PWWP domain-containing DNA repair factor 3A/3B/4, C-terminal |
| 158135 | TTLL11 | tubulin tyrosine ligase like 11 | IPR004344; | Tubulin-tyrosine ligase/Tubulin polyglutamylase; |
| 388650 | DIPK1A | divergent protein kinase domain 1A | IPR029244; IPR022049 | FAM69, N-terminal; FAM69, protein-kinase domain |
| 390916 | NUDT19 | nudix hydrolase 19 | IPR000086; IPR015797; IPR039121 | NUDIX hydrolase domain; NUDIX hydrolase-like domain superfamily; Acyl-coenzyme A diphosphatase NUDT19 |
| 641372 | ACOT6 | acyl-CoA thioesterase 6 | IPR014940; IPR029058; IPR006862; IPR016662; IPR042490 | BAAT/Acyl-CoA thioester hydrolase C-terminal; Alpha/Beta hydrolase fold; Acyl-CoA thioester hydrolase/bile acid-CoA amino acid N-acetyltransferase; Acyl-CoA thioesterase, long chain; Acyl-CoA thioester hydrolase/BAAT, N-terminal |
| 100288798 | SLC38A4-AS1 | SLC38A4 antisense RNA 1 | - | - |
| 100500801 | MIR3659 | microRNA 3659 | - | - |
| 101929130 | LOC101929130 | uncharacterized LOC101929130 | - | - |
| 105372097 | ST8SIA5-DT | ST8SIA5 divergent transcript | - | - |
| 105373113 | LINC01703 | long intergenic non-protein coding RNA 1703 | - | - |
| 107985128 | LINC01916 | long intergenic non-protein coding RNA 1916 | - | - |
| 339290 | LINC00667 | long intergenic non-protein coding RNA 667 | - | - |
| 677771 | SCARNA4 | small Cajal body-specific RNA 4 | - | - |
| 105372798 | LOC105372798 | - | - | - |
| 105378473 | LOC105378473 | - | - | - |
| 101928316 | LINC02859 | - | - | - |
| 286495 | TTC3P1 | - | - | - |

**Table 2: Ageprint genes grouped based on gene functions**.

|  |  |  |
| --- | --- | --- |
| **Category** | **Genes** | **Notes** |
| **Methyltransferases** | METTL9, SETD7, CARNMT1 | Regulate the addition of methyl groups. |
| **Acetyltransferases** | ACOT6 | Regulates protein acetylation. |
| **Histones** | H3C6 | Chromatin protein. |
| **Transcription Factors** | ETS2, EBF3, DMRT2, BEX1, NFIL3 | Transcription regulation. |
| **DNA Repair** | PWWP3B, DDB2 | Regulating genome stability. |
| **Homeobox Genes** | HOXA7, HOXC11, HOXC10, HOXB2, HOXA-AS3, HOXB-AS1 | Linked to growth and development. |
| **Growth Factor Receptors** | EFEMP1, VLDLR, FBLN1, GREB1 | Likely control growth and nutrient sensing pathways. |
| **Insulin/IGF Genes** | INS-IGF2, IGF2 | Key regulators of metabolic processes. |
| **Phospholipases** | PLCXD1, PIP4P2 | May regulate PIP2 levels, affecting cell signaling. |
| **mTOR Complex Members** | MAPKAP1 | Could be involved in nutrient sensing and aging pathways. |
| **Kinase Domain Proteins** | TJP2, TRIB3, WEE1, RET, IP6K2, STRADB, MYLK4, DIPK1A, PFKFB2 | Regulators of phosphorylation and signaling pathways. |
| **Zinc Finger Proteins** | KLF5, TES, SH3RF3, KRBA2, NR2F2, RBM20, IVNS1ABP | Involved in DNA binding and transcriptional regulation. |
| **Immunoglobulin Domains** | EBF3, GPCPD1, CADM2, IL20RA | Likely linked to cell-to-cell signaling. |
| **Senescence Marker** | RGN | Known marker of cellular senescence. |
| **NAD-Binding Proteins** | CFAP61, HPGD | Can be involved in metabolism and aging. |
| **Non-Protein Coding Genes** | MIR3659, LINC01703, LINC00667, MIR4257, LINC02859, LINC01497, LINC01916, MIR5695 | Non-coding RNAs. |

**Table 3: The 9 genes that are a part of the ageprint and have approved drugs listed against their protein products**. 4 drugs on this list have already been investigated in the context of aging and found to have promising results.

|  |  |
| --- | --- |
| **Gene** | **Drug** |
| CADM2 | Dehydrated alcohol |
| HCN1 | Tramadol |
| HOXC10 | Bevacizumab-awwb |
| MSANTD1 | Risperidone |
| NT5C2 | Didanosine; nelarabine |
| PDE7B | Flavoxate hydrochloride; dyphylline |
| RET | Selpercatinib; pralsetinib |
| TRIB3 | Indapamide; perindopril |
| TSLP | Tezepelumab |

# Supplementary Material

**[Attach 129\_important\_genes\_for\_longevity.xlsx]**

**Table S1: Details about the datasets that were used to create and validate SkeletAge**. The table contains the NCBI GEO accession number, the general phenotype of the individuals whose data was collected, the number of samples taken from each dataset, and whether that dataset was used to train or test SkeletAge. All of these samples have had bulk RNA-seq performed on vastus lateralis biopsies.

|  |  |  |  |
| --- | --- | --- | --- |
| **GEO Accession** | **Condition** | **Set type** | **Number of Samples** |
| GSE164471 | Very healthy participants | Training set | 53 |
| GSE97084 | Women post-menopausal, otherwise the rest of the participants are healthy | Training set | 103 |
| GSE196387 | Active, Obesity, T2D | Validation set | 18 |
| GSE205431 | End-stage Osteoarthritis | Validation set | 20 |
| GSE157988 | NMN | Validation set | 92 |
| GSE151066 | Active or sedentary older adults | Validation set | 57 |
| GSE58608 | Sedentary | Validation set | 24 |
| GSE144304 | Young, or old fit or old frail males and females | Training set | 80 |
| GSE159217 | Low to moderately active old and young | Validation set | 38 |
| GSE200398 | Elite athletes | Validation set | 24 |
| GSE186045 | Healthy on bedrest or bedrest + exercise | Validation set | 26 |

**Table S2: The genes that make up the ageprint of skeletal muscles**. Out of these 102 genes have been already studied in the context of aging. The remaining 26 are present in table 1.

|  |  |  |  |
| --- | --- | --- | --- |
| **EntrezID** | **Coefficients** | **Gene Symbol** | **Gene Type** |
| 100500801 | -0.081683 | MIR3659 | ProLong |
| 55970 | -0.0027916 | GNG12 | ProLong |
| 388650 | -0.0073341 | DIPK1A | ProLong |
| 100422997 | 0.0896826 | MIR4257 | RedLong |
| 677771 | 0.13607048 | SCARNA4 | RedLong |
| 116461 | -0.0011039 | TSEN15 | ProLong |
| 10625 | -0.000618 | IVNS1ABP | ProLong |
| 5208 | 0.00089837 | PFKFB2 | RedLong |
| 84976 | -0.0112682 | DISP1 | ProLong |
| 105373113 | -0.0403977 | LINC01703 | ProLong |
| 9687 | -0.003436 | GREB1 | ProLong |
| 81553 | -0.0057108 | CYRIA | ProLong |
| 9637 | 0.00063956 | FEZ2 | RedLong |
| 2202 | 0.00260235 | EFEMP1 | RedLong |
| 10617 | -0.0021297 | STAMBP | ProLong |
| 344558 | -0.0007133 | SH3RF3 | ProLong |
| 55437 | -0.0013537 | STRADB | ProLong |
| 8452 | -0.0002669 | CUL3 | ProLong |
| 56947 | -0.0006302 | MFF | ProLong |
| 151176 | 0.00237901 | ERFE | RedLong |
| 23243 | -0.0001799 | ANKRD28 | ProLong |
| 51447 | 0.00033512 | IP6K2 | RedLong |
| 253559 | -0.0034602 | CADM2 | ProLong |
| 5028 | -9.64E-05 | P2RY1 | ProLong |
| 4291 | -0.0005327 | MLF1 | ProLong |
| 5648 | -0.0030995 | MASP1 | ProLong |
| 101929130 | -0.0131282 | LOC101929130 | ProLong |
| 344901 | 0.0078429 | OSTN | RedLong |
| 57110 | -0.0038943 | PLAAT1 | ProLong |
| 345222 | 0.01407912 | MSANTD1 | RedLong |
| 386617 | 0.00242724 | KCTD8 | RedLong |
| 80854 | -0.000223 | SETD7 | ProLong |
| 3148 | 0.00043221 | HMGB2 | RedLong |
| 3248 | 0.01113792 | HPGD | RedLong |
| 55714 | 0.00936619 | TENM3 | RedLong |
| 80014 | -0.0051949 | WWC2 | ProLong |
| 348980 | 0.00192796 | HCN1 | RedLong |
| 101928819 | 8.16850898269829e-05 | LOC101928819 | RedLong |
| 85480 | -0.0155676 | TSLP | ProLong |
| 56114 | 0.05206789 | PCDHGA1 | RedLong |
| 5098 | 0.00022187 | PCDHGC3 | RedLong |
| 340156 | -0.0019136 | MYLK4 | ProLong |
| 63933 | -1.98E-07 | MCUR1 | ProLong |
| 8353 | -0.0849072 | H3C6 | ProLong |
| 64208 | -8.98E-05 | POPDC3 | ProLong |
| 27115 | -0.0041311 | PDE7B | ProLong |
| 53832 | 0.01137221 | IL20RA | RedLong |
| 8935 | 0.01591319 | SKAP2 | RedLong |
| 100133311 | 7.41993659527668e-06 | HOXA-AS3 | RedLong |
| 3204 | 0.00014434 | HOXA7 | RedLong |
| 26136 | 0.00593984 | TES | RedLong |
| 9920 | -0.0057936 | KBTBD11 | ProLong |
| 79145 | 0.00243786 | CHCHD7 | RedLong |
| 55529 | -0.0007942 | PIP4P2 | ProLong |
| 94241 | 0.0006348 | TP53INP1 | RedLong |
| 10655 | -0.0088668 | DMRT2 | ProLong |
| 7436 | -0.0004982 | VLDLR | ProLong |
| 9414 | 0.00049092 | TJP2 | RedLong |
| 138199 | -0.0097369 | CARNMT1 | ProLong |
| 4783 | 0.0011429 | NFIL3 | RedLong |
| 158135 | -0.0006662 | TTLL11 | ProLong |
| 79109 | -0.0005887 | MAPKAP1 | ProLong |
| 9536 | 0.00019887 | PTGES | RedLong |
| 91734 | 0.00091199 | IDI2 | RedLong |
| 55853 | 0.00207521 | IDI2-AS1 | RedLong |
| 221061 | -0.0318238 | FAM171A1 | ProLong |
| 9200 | -0.0035721 | HACD1 | ProLong |
| 5979 | -0.0051137 | RET | ProLong |
| 100507027 | -0.001614 | MRLN | ProLong |
| 119385 | 0.00232788 | AGAP11 | RedLong |
| 9211 | 0.00133931 | LGI1 | RedLong |
| 22978 | 0.00197793 | NT5C2 | RedLong |
| 105378473 | 0.01111806 | LOC105378473 | RedLong |
| 282996 | -0.0003304 | RBM20 | ProLong |
| 253738 | -0.0197966 | EBF3 | ProLong |
| 3481 | 0.00037653 | IGF2 | RedLong |
| 723961 | 0.00040245 | INS-IGF2 | RedLong |
| 7465 | -0.0052627 | WEE1 | ProLong |
| 101928316 | 0.01867558 | LINC02859 | RedLong |
| 1643 | 0.00535019 | DDB2 | RedLong |
| 51642 | -0.003382 | MRPL48 | ProLong |
| 4256 | 0.00356899 | MGP | RedLong |
| 5318 | 0.00188997 | PKP2 | RedLong |
| 100288798 | 0.00179492 | SLC38A4-AS1 | RedLong |
| 3227 | -0.0832915 | HOXC11 | ProLong |
| 3226 | -5.12E-05 | HOXC10 | ProLong |
| 283349 | -0.0024406 | RASSF3 | ProLong |
| 8738 | -0.0036994 | CRADD | ProLong |
| 121441 | -0.0007618 | NEDD1 | ProLong |
| 688 | 0.01123568 | KLF5 | RedLong |
| 100874068 | 0.01785731 | GRTP1-AS1 | RedLong |
| 641372 | -0.0032804 | ACOT6 | ProLong |
| 400258 | 0.00361293 | C14orf180 | RedLong |
| 4692 | -0.0014177 | NDN | ProLong |
| 7026 | 0.00449486 | NR2F2 | RedLong |
| 57465 | 0.00158481 | TBC1D24 | RedLong |
| 51108 | -0.0007358 | METTL9 | ProLong |
| 23568 | -0.0044889 | ARL2BP | ProLong |
| 124751 | -0.0045498 | KRBA2 | ProLong |
| 3212 | -0.0080491 | HOXB2 | ProLong |
| 100874362 | -0.0447071 | HOXB-AS1 | ProLong |
| 102723487 | 0.00179037 | LINC01497 | RedLong |
| 339290 | -0.0001489 | LINC00667 | ProLong |
| 23136 | 0.00496684 | EPB41L3 | RedLong |
| 147495 | -0.0056333 | APCDD1 | ProLong |
| 105372097 | -0.0073189 | ST8SIA5-DT | ProLong |
| 107985128 | 0.03039284 | LINC01916 | RedLong |
| 100847016 | 0.03864049 | MIR5695 | RedLong |
| 83483 | 0.00093944 | PLVAP | RedLong |
| 390916 | 0.00186915 | NUDT19 | RedLong |
| 57761 | 0.01406064 | TRIB3 | RedLong |
| 56261 | -0.0052651 | GPCPD1 | ProLong |
| 26074 | 0.00386925 | CFAP61 | RedLong |
| 56911 | 0.00049606 | MAP3K7CL | RedLong |
| 54102 | -0.0002797 | CLIC6 | ProLong |
| 105372798 | -0.0101719 | LOC105372798 | ProLong |
| 2114 | 0.00082494 | ETS2 | RedLong |
| 2192 | 0.00138774 | FBLN1 | RedLong |
| 55344 | 0.00172304 | PLCXD1 | RedLong |
| 9104 | -0.0121055 | RGN | ProLong |
| 64061 | 0.00040261 | TSPYL2 | RedLong |
| 286495 | 0.00274987 | TTC3P1 | RedLong |
| 55859 | 0.02262693 | BEX1 | RedLong |
| 139221 | 0.004081 | PWWP3B | RedLong |
| 55086 | 0.01378705 | RADX | RedLong |
| 23439 | 0.00027178 | ATP1B4 | RedLong |
| 100130613 | 0.00441142 | PRR32 | RedLong |
| 90161 | -0.0051853 | HS6ST2 | ProLong |

# References

1. de Magalhães JP. The Scientific Quest for Lasting Youth: Prospects for Curing Aging. Rejuvenation Res. Mary Ann Liebert, Inc., publishers; 2014; 17: 458–67.

2. Flanagan EW, Most J, Mey JT, Redman LM. Calorie Restriction and Aging in Humans. Annu Rev Nutr. Annual Reviews; 2020; 40: 105–33.

3. Waziry R, Ryan CP, Corcoran DL, Huffman KM, Kobor MS, Kothari M, Graf GH, Kraus VB, Kraus WE, Lin DTS, Pieper CF, Ramaker ME, Bhapkar M, et al. Effect of long-term caloric restriction on DNA methylation measures of biological aging in healthy adults from the CALERIE trial. Nat Aging. Nature Publishing Group; 2023; 3: 248–57.

4. Allen MD, Dalton BH, Gilmore KJ, McNeil CJ, Doherty TJ, Rice CL, Power GA. Neuroprotective effects of exercise on the aging human neuromuscular system. Exp Gerontol. 2021; 152: 111465.

5. Bettio LEB, Thacker JS, Rodgers SP, Brocardo PS, Christie BR, Gil-Mohapel J. Interplay between hormones and exercise on hippocampal plasticity across the lifespan. Biochim Biophys Acta BBA - Mol Basis Dis. 2020; 1866: 165821.

6. Stillman CM, Esteban-Cornejo I, Brown B, Bender CM, Erickson KI. Effects of Exercise on Brain and Cognition Across Age Groups and Health States. Trends Neurosci. Elsevier; 2020; 43: 533–43.

7. Lavin KM, Perkins RK, Jemiolo B, Raue U, Trappe SW, Trappe TA. Effects of aging and lifelong aerobic exercise on basal and exercise-induced inflammation. J Appl Physiol. American Physiological Society; 2020; 128: 87–99.

8. Radak Z, Chung HY, Goto S. Systemic adaptation to oxidative challenge induced by regular exercise. Free Radic Biol Med. 2008; 44: 153–9.

9. Thyfault JP, Bergouignan A. Exercise and metabolic health: beyond skeletal muscle. Diabetologia. 2020; 63: 1464–74.

10. Li G, Li J, Gao F. Exercise and Cardiovascular Protection. In: Xiao J, editor. Physical Exercise for Human Health [Internet]. Singapore: Springer Nature; 2020 [cited 2024 Oct 23]. p. 205–16. Available from: https://doi.org/10.1007/978-981-15-1792-1\_14

11. Reimers CD, Knapp G, Reimers AK. Does physical activity increase life expectancy? A review of the literature. J Aging Res. 2012; 2012: 243958.

12. Lee DH, Rezende LFM, Joh H-K, Keum N, Ferrari G, Rey-Lopez JP, Rimm EB, Tabung FK, Giovannucci EL. Long-Term Leisure-Time Physical Activity Intensity and All-Cause and Cause-Specific Mortality: A Prospective Cohort of US Adults. Circulation. American Heart Association; 2022; 146: 523–34.

13. de Magalhães JP. Programmatic features of aging originating in development: aging mechanisms beyond molecular damage? FASEB J. 2012; 26: 4821–6.

14. Kenyon C, Chang J, Gensch E, Rudner A, Tabtiang R. A C. elegans mutant that lives twice as long as wild type. Nature. 1993; 366: 461–4.

15. Uno M, Nishida E. Lifespan-regulating genes in C. elegans. Npj Aging Mech Dis. Nature Publishing Group; 2016; 2: 1–8.

16. Dubnov S, Bennett ER, Yayon N, Yakov O, Bennett DA, Seshadri S, Mufson E, Tzur Y, Greenberg D, Kuro-o M, Paldor I, Abraham CR, Soreq H. Knockout of the longevity gene Klotho perturbs aging and Alzheimer’s disease-linked brain microRNAs and tRNA fragments. Commun Biol. Nature Publishing Group; 2024; 7: 1–14.

17. Prochownik EV, Wang H. Lessons in aging from Myc knockout mouse models. Front Cell Dev Biol [Internet]. Frontiers; 2023 [cited 2024 Oct 23]; 11. Available from: https://www.frontiersin.org/journals/cell-and-developmental-biology/articles/10.3389/fcell.2023.1244321/full

18. Ruetz TJ, Pogson AN, Kashiwagi CM, Gagnon SD, Morton B, Sun ED, Na J, Yeo RW, Leeman DS, Morgens DW, Tsui CK, Li A, Bassik MC, et al. CRISPR–Cas9 screens reveal regulators of ageing in neural stem cells. Nature. Nature Publishing Group; 2024; : 1–10.

19. Oh HS-H, Rutledge J, Nachun D, Pálovics R, Abiose O, Moran-Losada P, Channappa D, Urey DY, Kim K, Sung YJ, Wang L, Timsina J, Western D, et al. Organ aging signatures in the plasma proteome track health and disease. Nature. Nature Publishing Group; 2023; 624: 164–72.

20. Slieker RC, Relton CL, Gaunt TR, Slagboom PE, Heijmans BT. Age-related DNA methylation changes are tissue-specific with ELOVL2 promoter methylation as exception. Epigenetics Chromatin. 2018; 11: 25.

21. Tian YE, Cropley V, Maier AB, Lautenschlager NT, Breakspear M, Zalesky A. Heterogeneous aging across multiple organ systems and prediction of chronic disease and mortality. Nat Med. Nature Publishing Group; 2023; 29: 1221–31.

22. Tuttle CSL, Waaijer MEC, Slee‐Valentijn MS, Stijnen T, Westendorp R, Maier AB. Cellular senescence and chronological age in various human tissues: A systematic review and meta‐analysis. Aging Cell. 2020; 19: e13083.

23. Avelar RA, Ortega JG, Tacutu R, Tyler EJ, Bennett D, Binetti P, Budovsky A, Chatsirisupachai K, Johnson E, Murray A, Shields S, Tejada-Martinez D, Thornton D, et al. A multidimensional systems biology analysis of cellular senescence in aging and disease. Genome Biol. 2020; 21: 91.

24. Ko B, Van Raamsdonk JM. RNA Sequencing of Pooled Samples Effectively Identifies Differentially Expressed Genes. Biology. Multidisciplinary Digital Publishing Institute; 2023; 12: 812.

25. Lopes K de P, Snijders GJL, Humphrey J, Allan A, Sneeboer MAM, Navarro E, Schilder BM, Vialle RA, Parks M, Missall R, van Zuiden W, Gigase FAJ, Kübler R, et al. Genetic analysis of the human microglial transcriptome across brain regions, aging and disease pathologies. Nat Genet. Nature Publishing Group; 2022; 54: 4–17.

26. Pan J, Ma N, Yu B, Zhang W, Wan J. Transcriptomic profiling of microglia and astrocytes throughout aging. J Neuroinflammation. 2020; 17: 97.

27. Peters MJ, Joehanes R, Pilling LC, Schurmann C, Conneely KN, Powell J, Reinmaa E, Sutphin GL, Zhernakova A, Schramm K, Wilson YA, Kobes S, Tukiainen T, et al. The transcriptional landscape of age in human peripheral blood. Nat Commun. Nature Publishing Group; 2015; 6: 8570.

28. Blackburn EH, Greider CW, Szostak JW. Telomeres and telomerase: the path from maize, Tetrahymena and yeast to human cancer and aging. Nat Med. 2006; 12: 1133–8.

29. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, Klotzle B, Bibikova M, Fan J-B, Gao Y, Deconde R, Chen M, Rajapakse I, et al. Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. Mol Cell. Elsevier; 2013; 49: 359–67.

30. Horvath S. DNA methylation age of human tissues and cell types. Genome Biol. 2013; 14: 3156.

31. Holly AC, Melzer D, Pilling LC, Henley W, Hernandez DG, Singleton AB, Bandinelli S, Guralnik JM, Ferrucci L, Harries LW. Towards a gene expression biomarker set for human biological age. Aging Cell. 2013; 12: 324–6.

32. Ren X, Kuan PF. RNAAgeCalc: A multi-tissue transcriptional age calculator. PLOS ONE. Public Library of Science; 2020; 15: e0237006.

33. Krištić J, Vučković F, Menni C, Klarić L, Keser T, Beceheli I, Pučić-Baković M, Novokmet M, Mangino M, Thaqi K, Rudan P, Novokmet N, Sarac J, et al. Glycans are a novel biomarker of chronological and biological ages. J Gerontol A Biol Sci Med Sci. 2014; 69: 779–89.

34. Menni C, Kiddle SJ, Mangino M, Viñuela A, Psatha M, Steves C, Sattlecker M, Buil A, Newhouse S, Nelson S, Williams S, Voyle N, Soininen H, et al. Circulating Proteomic Signatures of Chronological Age. J Gerontol A Biol Sci Med Sci. 2015; 70: 809–16.

35. Hertel J, Friedrich N, Wittfeld K, Pietzner M, Budde K, Van der Auwera S, Lohmann T, Teumer A, Völzke H, Nauck M, Grabe HJ. Measuring Biological Age via Metabonomics: The Metabolic Age Score. J Proteome Res. American Chemical Society; 2016; 15: 400–10.

36. Menni C, Kastenmüller G, Petersen AK, Bell JT, Psatha M, Tsai P-C, Gieger C, Schulz H, Erte I, John S, Brosnan MJ, Wilson SG, Tsaprouni L, et al. Metabolomic markers reveal novel pathways of ageing and early development in human populations. Int J Epidemiol. 2013; 42: 1111–9.

37. Belsky DW, Caspi A, Houts R, Cohen HJ, Corcoran DL, Danese A, Harrington H, Israel S, Levine ME, Schaefer JD, Sugden K, Williams B, Yashin AI, et al. Quantification of biological aging in young adults. Proc Natl Acad Sci U S A. 2015; 112: E4104-4110.

38. Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, Hou L, Baccarelli AA, Stewart JD, Li Y, Whitsel EA, Wilson JG, Reiner AP, et al. An epigenetic biomarker of aging for lifespan and healthspan. Aging. 2018; 10: 573–91.

39. Sebastiani P, Thyagarajan B, Sun F, Schupf N, Newman AB, Montano M, Perls TT. Biomarker signatures of aging. Aging Cell. 2017; 16: 329–38.

40. Lu AT, Quach A, Wilson JG, Reiner AP, Aviv A, Raj K, Hou L, Baccarelli AA, Li Y, Stewart JD, Whitsel EA, Assimes TL, Ferrucci L, et al. DNA methylation GrimAge strongly predicts lifespan and healthspan. Aging. 2019; 11: 303–27.

41. Jylhävä J, Pedersen NL, Hägg S. Biological Age Predictors. eBioMedicine. Elsevier; 2017; 21: 29–36.

42. Bocklandt S, Lin W, Sehl ME, Sánchez FJ, Sinsheimer JS, Horvath S, Vilain E. Epigenetic predictor of age. PloS One. 2011; 6: e14821.

43. Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nat Rev Genet. Nature Publishing Group; 2018; 19: 371–84.

44. Bell CG, Lowe R, Adams PD, Baccarelli AA, Beck S, Bell JT, Christensen BC, Gladyshev VN, Heijmans BT, Horvath S, Ideker T, Issa J-PJ, Kelsey KT, et al. DNA methylation aging clocks: challenges and recommendations. Genome Biol. 2019; 20: 249.

45. Koch CM, Wagner W. Epigenetic-aging-signature to determine age in different tissues. Aging. 2011; 3: 1018–27.

46. Li A, Koch Z, Ideker T. Epigenetic aging: Biological age prediction and informing a mechanistic theory of aging. J Intern Med. 2022; 292: 733–44.

47. Yan Q, Paul KC, Lu AT, Kusters C, Binder AM, Horvath S, Ritz B. Epigenetic mutation load is weakly correlated with epigenetic age acceleration. Aging. 2020; 12: 17863–94.

48. Gopu V, Camacho FR, Toma R, Torres PJ, Cai Y, Krishnan S, Rajagopal S, Tily H, Vuyisich M, Banavar G. An accurate aging clock developed from large-scale gut microbiome and human gene expression data. iScience. 2024; 27: 108538.

49. Santilli V, Bernetti A, Mangone M, Paoloni M. Clinical definition of sarcopenia. Clin Cases Miner Bone Metab. 2014; 11: 177–80.

50. Schuurs TA, Morariu AM, Ottens PJ, ’t Hart NA, Popma SH, Leuvenink HGD, Ploeg RJ. Time-Dependent Changes in Donor Brain Death Related Processes. Am J Transplant. 2006; 6: 2903–11.

51. Yeh T, Wechsler AS, Graham LJ, Loesser KE, Sica DA, Wolfe L, Jakoi ER. Acute brain death alters left ventricular myocardial gene expression. J Thorac Cardiovasc Surg. 1999; 117: 365–74.

52. Franz H, Ullmann C, Becker A, Ryan M, Bahn S, Arendt T, Simon M, Pääbo S, Khaitovich P. Systematic analysis of gene expression in human brains before and after death. Genome Biol. 2005; 6: R112.

53. Robinson MM, Dasari S, Konopka AR, Johnson ML, Manjunatha S, Esponda RR, Carter RE, Lanza IR, Nair KS. Enhanced Protein Translation Underlies Improved Metabolic and Physical Adaptations to Different Exercise Training Modes in Young and Old Humans. Cell Metab. Elsevier; 2017; 25: 581–92.

54. Tumasian RA, Harish A, Kundu G, Yang J-H, Ubaida-Mohien C, Gonzalez-Freire M, Kaileh M, Zukley LM, Chia CW, Lyashkov A, Wood WH, Piao Y, Coletta C, et al. Skeletal muscle transcriptome in healthy aging. Nat Commun. 2021; 12: 2014.

55. Sanoudou D, Kang PB, Haslett JN, Han M, Kunkel LM, Beggs AH. Transcriptional profile of postmortem skeletal muscle. Physiol Genomics. American Physiological Society; 2004; 16: 222–8.

56. Carithers LJ, Ardlie K, Barcus M, Branton PA, Britton A, Buia SA, Compton CC, DeLuca DS, Peter-Demchok J, Gelfand ET, Guan P, Korzeniewski GE, Lockhart NC, et al. A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. Biopreservation Biobanking. Mary Ann Liebert, Inc., publishers; 2015; 13: 311–9.

57. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, Hasz R, Walters G, Garcia F, Young N, Foster B, Moser M, Karasik E, et al. The Genotype-Tissue Expression (GTEx) project. Nat Genet. Nature Publishing Group; 2013; 45: 580–5.

58. Micó V, Berninches L, Tapia J, Daimiel L. NutrimiRAging: Micromanaging Nutrient Sensing Pathways through Nutrition to Promote Healthy Aging. Int J Mol Sci. Multidisciplinary Digital Publishing Institute; 2017; 18: 915.

59. Zhang J, Wang S, Liu B. New Insights into the Genetics and Epigenetics of Aging Plasticity. Genes. Multidisciplinary Digital Publishing Institute; 2023; 14: 329.

60. Ko D, Mun S, Kim M, Nho Y-H, Lee D-G, Kang S, Han K, Kim M. A Glance into the Destiny of Transcriptomic Activity, Embodied by the HOX Genes, in Neonatal and Aging Dermal Cells. Adv Biol. 2024; 8: e2300325.

61. Turner DC, Gorski PP, Maasar MF, Seaborne RA, Baumert P, Brown AD, Kitchen MO, Erskine RM, Dos-Remedios I, Voisin S, Eynon N, Sultanov RI, Borisov OV, et al. DNA methylation across the genome in aged human skeletal muscle tissue and muscle-derived cells: the role of HOX genes and physical activity. Sci Rep. Nature Publishing Group; 2020; 10: 15360.

62. Yaku K, Okabe K, Nakagawa T. NAD metabolism: Implications in aging and longevity. Ageing Res Rev. 2018; 47: 1–17.

63. Frederick DW, Loro E, Liu L, Davila A, Chellappa K, Silverman IM, Quinn WJ, Gosai SJ, Tichy ED, Davis JG, Mourkioti F, Gregory BD, Dellinger RW, et al. Loss of NAD Homeostasis Leads to Progressive and Reversible Degeneration of Skeletal Muscle. Cell Metab. 2016; 24: 269–82.

64. Chubanava S, Treebak JT. Regular exercise effectively protects against the aging-associated decline in skeletal muscle NAD content. Exp Gerontol. 2023; 173: 112109.

65. Ahn H, Kim DW, Ko Y, Ha J, Shin YB, Lee J, Sung YS, Kim KW. Updated systematic review and meta-analysis on diagnostic issues and the prognostic impact of myosteatosis: A new paradigm beyond sarcopenia. Ageing Res Rev. 2021; 70: 101398.

66. Miljkovic I, Zmuda JM. Epidemiology of myosteatosis. Curr Opin Clin Nutr Metab Care. 2010; 13: 260–4.

67. Wang L, Zhou Y, Wang Y, Shan T. Integrative cross-species analysis reveals conserved and unique signatures in fatty skeletal muscles. Sci Data. Nature Publishing Group; 2024; 11: 290.

68. Biltz NK, Collins KH, Shen KC, Schwartz K, Harris CA, Meyer GA. Infiltration of intramuscular adipose tissue impairs skeletal muscle contraction. J Physiol. 2020; 598: 2669–83.

69. Jiang Z, Marriott K, Maly MR. Impact of Inter- and Intramuscular Fat on Muscle Architecture and Capacity. Crit Rev Biomed Eng [Internet]. Begel House Inc.; 2019 [cited 2024 Oct 14]; 47. Available from: https://www.dl.begellhouse.com/journals/4b27cbfc562e21b8,6e3159d928ffbfcc,186e9a8444b5d348.html

70. Cannon M, Stevenson J, Stahl K, Basu R, Coffman A, Kiwala S, McMichael JF, Kuzma K, Morrissey D, Cotto K, Mardis ER, Griffith OL, Griffith M, et al. DGIdb 5.0: rebuilding the drug–gene interaction database for precision medicine and drug discovery platforms. Nucleic Acids Res. 2024; 52: D1227–35.

71. Griffith M, Griffith OL, Coffman AC, Weible JV, McMichael JF, Spies NC, Koval J, Das I, Callaway MB, Eldred JM, Miller CA, Subramanian J, Govindan R, et al. DGIdb: mining the druggable genome. Nat Methods. 2013; 10: 1209–10.

72. Hopkins AL, Groom CR. The druggable genome. Nat Rev Drug Discov. Nature Publishing Group; 2002; 1: 727–30.

73. Russ AP, Lampel S. The druggable genome: an update. Drug Discov Today. 2005; 10: 1607–10.

74. Zhang W, Gao L, Zhang X, Weng Y, Du Y, Sun Y-L, Wei H, Hao T, Chen Y, Liang X, Zhang M. Theophylline derivatives promote primordial follicle activation via cAMP-PI3K/Akt pathway and ameliorate fertility deficits in naturally aged mice. Int J Biol Sci. 2024; 20: 5312.

75. Brochard T, McIntyre RL, Houtkooper RH, Seluanov A, Gorbunova V, Janssens GE. Repurposing nucleoside reverse transcriptase inhibitors (NRTIs) to slow aging. Ageing Res Rev. 2023; 92: 102132.

76. McIntyre RL, Molenaars M, Schomakers BV, Gao AW, Kamble R, Jongejan A, Weeghel M van, Kuilenburg ABP van, Possemato R, Houtkooper RH, Janssens GE. Anti-retroviral treatment with zidovudine alters pyrimidine metabolism, reduces translation, and extends healthy longevity via ATF-4. Cell Rep [Internet]. Elsevier; 2023 [cited 2024 Oct 25]; 42. Available from: https://www.cell.com/cell-reports/abstract/S2211-1247(22)01829-0

77. Michaels NJ, Lemmon K, Plemel JR, Jensen SK, Mishra MK, Brown D, Rawji KS, Koch M, Yong VW. Aging-Exacerbated Acute Axon and Myelin Injury Is Associated with Microglia-Derived Reactive Oxygen Species and Is Alleviated by the Generic Medication Indapamide. J Neurosci. Society for Neuroscience; 2020; 40: 8587–600.

78. Sumukadas D, Witham MD, Struthers AD, McMurdo MET. Effect of perindopril on physical function in elderly people with functional impairment: a randomized controlled trial. CMAJ. CMAJ; 2007; 177: 867–74.

79. Yoshino M, Yoshino J, Kayser BD, Patti GJ, Franczyk MP, Mills KF, Sindelar M, Pietka T, Patterson BW, Imai S-I, Klein S. Nicotinamide mononucleotide increases muscle insulin sensitivity in prediabetic women. Science. 2021; 372: 1224–9.

80. Foster CA, Barker-Kamps M, Goering M, Patki A, Tiwari HK, Mrug S. Epigenetic age acceleration correlates with BMI in young adults. Aging. 2023; 15: 513–23.

81. Lin W-Y, Wang Y-C, Teng I-H, Liu C, Lou X-Y. Associations of five obesity metrics with epigenetic age acceleration: Evidence from 2,474 Taiwan Biobank participants. Obesity. 2021; 29: 1731–8.

82. Fraszczyk E, Thio CHL, Wackers P, Dollé MET, Bloks VW, Hodemaekers H, Picavet HS, Stynenbosch M, Verschuren WMM, Snieder H, Spijkerman AMW, Luijten M. DNA methylation trajectories and accelerated epigenetic aging in incident type 2 diabetes. GeroScience. 2022; 44: 2671–84.

83. Odoh CK, Guo X, Arnone JT, Wang X, Zhao ZK. The role of NAD and NAD precursors on longevity and lifespan modulation in the budding yeast, Saccharomyces cerevisiae. Biogerontology. 2022; 23: 169–99.

84. Wilson N, Kataura T, Korsgen ME, Sun C, Sarkar S, Korolchuk VI. The autophagy–NAD axis in longevity and disease. Trends Cell Biol. Elsevier; 2023; 33: 788–802.

85. Imamura K, Ashida H, Ishikawa T, Fujii M. Human major psoas muscle and sacrospinalis muscle in relation to age: a study by computed tomography. J Gerontol. 1983; 38: 678–81.

86. Ting AY, Zelinski MB. Characterization of FOXO1, 3 and 4 transcription factors in ovaries of fetal, prepubertal and adult rhesus macaques. Biol Reprod. 2017; 96: 1052.

87. Cox L, Liu JH. Primary ovarian insufficiency: an update. Int J Womens Health. Dove Medical Press; 2014; 6: 235–43.

88. Costa-Mattioli M, Walter P. The integrated stress response: From mechanism to disease. Science. American Association for the Advancement of Science; 2020; 368: eaat5314.

89. Statzer C, Meng J, Venz R, Bland M, Robida-Stubbs S, Patel K, Petrovic D, Emsley R, Liu P, Morantte I, Haynes C, Mair WB, Longchamp A, et al. ATF-4 and hydrogen sulfide signalling mediate longevity in response to inhibition of translation or mTORC1. Nat Commun. Nature Publishing Group; 2022; 13: 967.

90. Wang J, Robida-Stubbs S, Tullet JMA, Rual J-F, Vidal M, Blackwell TK. RNAi Screening Implicates a SKN-1–Dependent Transcriptional Response in Stress Resistance and Longevity Deriving from Translation Inhibition. PLOS Genet. Public Library of Science; 2010; 6: e1001048.

91. Rogers AN, Chen D, McColl G, Czerwieniec G, Felkey K, Gibson BW, Hubbard A, Melov S, Lithgow GJ, Kapahi P. Life Span Extension via eIF4G Inhibition Is Mediated by Posttranscriptional Remodeling of Stress Response Gene Expression in C. elegans. Cell Metab. Elsevier; 2011; 14: 55–66.

92. Fiddes R, Blumenthal J, Dawson JE, Dyckman E, St John Hammond PG, Harris S, Lasseter KC, Levine BS, Montoro R, Niederman AL, Ratner PH, Rosenblatt S, Toth PD, et al. Evaluation of indapamide 1.25 mg once daily in elderly patients with mild to moderate hypertension. J Hum Hypertens. 1997; 11: 239–44.

93. Syed YY. Perindopril/Indapamide/Amlodipine in Hypertension: A Profile of Its Use. Am J Cardiovasc Drugs. 2022; 22: 219–30.

94. O’Neil SM, Witcher KG, McKim DB, Godbout JP. Forced turnover of aged microglia induces an intermediate phenotype but does not rebalance CNS environmental cues driving priming to immune challenge. Acta Neuropathol Commun. 2018; 6: 129.

95. Tamura A, Sato T, Fujii T. Antioxidant activity of indapamide and its metabolite. Chem Pharm Bull (Tokyo). 1990; 38: 255–7.

96. Hippisley-Cox J, Young D, Coupland C, Channon KM, Tan PS, Harrison DA, Rowan K, Aveyard P, Pavord ID, Watkinson PJ. Risk of severe COVID-19 disease with ACE inhibitors and angiotensin receptor blockers: cohort study including 8.3 million people. Heart. BMJ Publishing Group Ltd and British Cardiovascular Society; 2020; 106: 1503–11.

97. Najafi N, Davoudi A, Izadyar H, Alishahi A, Mokhtariani A, Soleimanpourian B, Tabarrayi M, Moosazadeh M, Daftarian Z, Ahangarkani F. The effect of ACE inhibitors and ARBs on outcomes in hospitalized patients with COVID-19. Ir J Med Sci. 2022; 192: 1517.

98. Writing Committee for the REMAP-CAP Investigators. Effect of Angiotensin-Converting Enzyme Inhibitor and Angiotensin Receptor Blocker Initiation on Organ Support–Free Days in Patients Hospitalized With COVID-19: A Randomized Clinical Trial. JAMA. 2023; 329: 1183–96.

99. Onder G, Penninx BWJH, Balkrishnan R, Fried LP, Chaves PHM, Williamson J, Carter C, Di Bari M, Guralnik JM, Pahor M. Relation between use of angiotensin-converting enzyme inhibitors and muscle strength and physical function in older women: an observational study. Lancet Lond Engl. 2002; 359: 926–30.

100. Hutcheon SD, Gillespie ND, Crombie IK, Struthers AD, McMurdo MET. Perindopril improves six minute walking distance in older patients with left ventricular systolic dysfunction: a randomised double blind placebo controlled trial. Heart Br Card Soc. 2002; 88: 373–7.

101. Schrack JA, Knuth ND, Simonsick EM, Ferrucci L. “IDEAL” Aging is Associated with Lower Resting Metabolic Rate: The Baltimore Longitudinal Study of Aging. J Am Geriatr Soc. 2014; 62: 667–72.

102. de Jong JCBC, Verschuren L, Caspers MPM, van der Hoek MD, van der Leij FR, Kleemann R, van den Hoek AM, Nieuwenhuizen AG, Keijer J. Evidence for sex-specific intramuscular changes associated to physical weakness in adults older than 75 years. Biol Sex Differ. 2023; 14: 45.

103. Whytock KL, Pino MF, Sun Y, Yu G, De Carvalho FG, Yeo RX, Vega RB, Parmar G, Divoux A, Kapoor N, Yi F, Cornnell H, Patten DA, et al. Comprehensive interrogation of human skeletal muscle reveals a dissociation between insulin resistance and mitochondrial capacity. Am J Physiol-Endocrinol Metab. American Physiological Society; 2023; 325: E291–302.

104. Drummer DJ, Lavin KM, Graham ZA, O’Bryan SM, McAdam JS, Lixandrão ME, Seay R, Aban I, Siegel HJ, Ghanem E, Singh JA, Bonfitto A, Antone J, et al. Muscle transcriptomic circuits linked to periarticular physiology in end-stage osteoarthritis. Physiol Genomics. 2022; 54: 501–13.

105. Rubenstein AB, Hinkley JM, Nair VD, Nudelman G, Standley RA, Yi F, Yu G, Trappe TA, Bamman MM, Trappe SW, Sparks LM, Goodpaster BH, Vega RB, et al. Skeletal muscle transcriptome response to a bout of endurance exercise in physically active and sedentary older adults. Am J Physiol-Endocrinol Metab. American Physiological Society; 2022; 322: E260–77.

106. Lindholm ME, Huss M, Solnestam BW, Kjellqvist S, Lundeberg J, Sundberg CJ. The human skeletal muscle transcriptome: sex differences, alternative splicing, and tissue homogeneity assessed with RNA sequencing. FASEB J. 2014; 28: 4571–81.

107. Lagerwaard B, Nieuwenhuizen AG, Bunschoten A, de Boer VCJ, Keijer J. Matrisome, innervation and oxidative metabolism affected in older compared with younger males with similar physical activity. J Cachexia Sarcopenia Muscle. 2021; 12: 1214–31.

108. Semenova EA, Zempo H, Miyamoto-Mikami E, Kumagai H, Larin AK, Sultanov RI, Babalyan KA, Zhelankin AV, Tobina T, Shiose K, Kakigi R, Tsuzuki T, Ichinoseki-Sekine N, et al. Genome-Wide Association Study Identifies CDKN1A as a Novel Locus Associated with Muscle Fiber Composition. Cells. Multidisciplinary Digital Publishing Institute; 2022; 11: 3910.

109. McFarland AJ, Ray PR, Bhai S, Levine BD, Price TJ. RNA sequencing on muscle biopsy from a 5‐week bed rest study reveals the effect of exercise and potential interactions with dorsal root ganglion neurons. Physiol Rep [Internet]. Wiley; 2022 [cited 2024 Oct 14]; 10. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8823189/

110. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. Nature Publishing Group; 2019; 37: 907–15.

111. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15: 550.

112. Lexell J, Taylor CC, Sjöström M. What is the cause of the ageing atrophy?: Total number, size and proportion of different fiber types studied in whole vastus lateralis muscle from 15- to 83-year-old men. J Neurol Sci. Elsevier; 1988; 84: 275–94.

113. Zou H, Hastie T. Regularization and Variable Selection Via the Elastic Net. J R Stat Soc Ser B Stat Methodol. 2005; 67: 301–20.

114. Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, Huber W. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics. 2005; 21: 3439–40.

115. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS J Integr Biol. Mary Ann Liebert, Inc., publishers; 2012; 16: 284–7.

116. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, Gable AL, Fang T, Doncheva NT, Pyysalo S, Bork P, Jensen LJ, von Mering C. The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. 2022; 51: D638.