



Published in final edited form as:

Mol Omics. ; 17(5): 796–808. doi:10.1039/d1mo00178g.

Transcriptome features of striated muscle aging and predictability of protein level changes

Yu Han¹, Lauren Z. Li¹, Nikhitha L. Kastury¹, Cody T. Thomas¹, Maggie P. Y. Lam^{1,2}, Edward Lau¹

¹Department of Medicine, Consortium for Fibrosis Research & Translation, University of Colorado School of Medicine, Aurora CO 80045

²Department of Biochemistry and Molecular Genetics, Consortium for Fibrosis Research & Translation, University of Colorado School of Medicine, Aurora CO 80045

Abstract

We performed total RNA sequencing and multi-omics analysis comparing skeletal muscle and cardiac muscle in young adult (4 months) vs. early aging (20 months) mice to examine the molecular mechanisms of striated muscle aging. We observed that aging cardiac and skeletal muscles both invoke transcriptomic changes in the innate immune system and mitochondria pathways but diverge in extracellular matrix processes. On an individual gene level, we identified 611 age-associated signatures in skeletal and cardiac muscles, including a number of myokine and cardiokine encoding genes. Because RNA and protein levels correlate only partially, we reason that differentially expressed transcripts that accurately reflect their protein counterparts will be more valuable proxies for proteomic changes and by extension physiological states. We applied a computational data analysis workflow to estimate which transcriptomic changes are more likely relevant to protein-level regulation using large proteogenomics data sets. We estimate about 48% of the aging-associated transcripts predict protein levels well ($r \geq 0.5$). In parallel, a comparison of the identified aging-regulated genes with public human transcriptomics data showed that only 35–45% of the identified genes show an age-dependent expression in corresponding human tissues. Thus, integrating both RNA-protein correlation and human conservation across data sources, we nominate 134 prioritized aging striated muscle signatures that are predicted to correlate strongly with protein levels and that show age-dependent expression in humans. The results here reveal new details into how aging reshapes gene expression in striated muscles at the transcript and protein levels.

Correspondence: Edward Lau, Ph.D. (edward.lau@cuanschutz.edu).

Author Contributions

Y.H., M.P.L., and E.L. conceptualized the study. Y.H., C.T.T., and M.P.L. performed experiments. L.Z.L., N.L.K., and E.L. performed data analysis. E.L. wrote analysis scripts and drafted the manuscript. Y.H., M.P.L., and E.L. revised the manuscript. M.P.L. and E.L. managed funding.

Conflicts of Interest

There are no conflicts to declare.

Introduction

Skeletal and cardiac muscles are highly specialized tissues that are associated with distinct functional declines during aging. In aged organisms, there is a progressive loss of skeletal muscle mass, function, and regenerative capacity. Sarcopenia leads to frailty, diminishes the capacity for locomotion, limits the physiological role of muscles to regulate systemic glucose metabolism, and is a strong independent predictor of mortality in the elderly (Landi et al., 2013; Moore et al., 2014). In parallel, age-associated heart diseases are a leading cause of mortality and morbidity worldwide. Old age is associated with a decline in cardiac reserve, stress tolerance, metabolic and functional capacity, and the development of myocardial fibrosis that reduces the elasticity of the cardiac muscle (Lesnefsky et al., 2016; Triposkiadis et al., 2019). Hence, existing evidence strongly points to striated muscle tissues as being key to preserving organismal function and promoting healthspan. Understanding the molecular mechanisms of aging hearts and skeletal muscles is an important component in the quest to mitigate and prevent prevalent morbidities in an aging world.

Recent reports have surveyed the transcript level changes in the aging mouse skeletal muscle (Graber et al., 2021; Lin et al., 2018; Mikovic et al., 2018) or the heart (Bartling et al., 2019; Benayoun et al., 2019; Greenig et al., 2020). Separate studies have also determined the transcript (Timmons et al., 2019; Tumasian et al., 2021) and protein abundance changes (Murgia et al., 2017; Ubaida-Mohien et al., 2019a) in aging human skeletal muscles. Despite progress however, significant knowledge gaps persist. Specifically, continued investigations are needed to establish consistent aging signatures in the heart and the skeletal muscle across multiple models, and to contrast tissue-specific signatures across the two major groups of striated muscles. Moreover, important questions remain unanswered on whether and how much of the detected transcriptome changes might be translated to the protein level.

It is now established that transcript and protein levels correlate imperfectly across tissues and biological samples, where some transcripts may even be negatively correlated with the abundance of their protein counterpart (Franks et al., 2017; Jiang et al., 2020; Krug et al., 2020). Because proteins perform the overwhelming majority of biological processes, the results from transcriptomics data might be differentially relevant to biological processes based on how well they predict protein-level changes. Recent large studies including GTEx (GTEx Consortium, 2020; Jiang et al., 2020), CPTAC (Krug et al., 2020; Mani et al., 2021), and GESTALT (Tumasian et al., 2021) have compared transcriptomics and proteomics data from matching tissues in large cohorts, and generally find moderate correlation between RNA and protein levels across samples. Very recently, the NIA GESTALT study characterized the transcriptome and proteome changes in a large cohort of human skeletal muscle biopsies from healthy donors across wide age groups (Tumasian et al., 2021), and found moderate correlation between transcripts and their protein counterparts, with 10 of the top 20 positively age-variable transcripts also showing age-dependent protein levels over the lifespan. At present however, large-scale proteomics data of comparable depths remain far less accessible and common than transcriptomics data, hence there is intense interest in comparing across omics layers and identifying the transcriptomics signatures in aging and disease models that are translatable to the protein layer.

Here we performed total transcriptomic analysis in the heart and the skeletal muscle to assess global gene expression features of striated muscle aging. We apply a computational data analysis workflow that: (i) estimates the degree to which the transcriptome changes may predict changes at the protein level; and (ii) co-analyzes public human transcriptomics data to pinpoint conserved signatures in human tissues. This approach may be useful for annotating and prioritizing transcriptomics signatures that are likely relevant to protein-level regulation.

Experimental

Animals and tissue extraction

All animal protocols were approved by the Institutional Animal Care and Use Committee at the University of Colorado School of Medicine. C57BL/6J mice were purchased from Jackson Laboratories (Bar Harbor, ME, USA) and housed in a temperature-controlled environment on a 12-h light/dark cycle and fed with normal diet and water ad libitum under National Institutes of Health (NIH) guidelines for the Care and Use of Laboratory Animals. Young adult mice (~4 months) and early aging mice (~20 months) mice (n=4, 2 male 2 female) were sacrificed, followed by measurement of body weight, heart weight and tibia length. The left cardiac ventricle and quadriceps femoris muscle were collected and stored at -80 °C.

Total RNA sequencing

To extract RNA, the tissues were cut into ~1 mm³ cubes on ice. Cold TRIzol (Invitrogen) was added at 75 µL per mg tissue and tissues were homogenized on a bead mill homogenizer with 2.8 mm ceramic beads at 10 second duration, speed 5, 5 repeats. The samples were centrifuged at 16,000 g for 15 minutes at 4 °C and the supernatant was transferred to RNase-free tubes. RNA extraction was performed using the Direct-zol RNA Miniprep Plus kit (ZYMO) following manufacturer's instructions. Total RNA sequencing was performed on the tissues (~80M reads/ 20 Gbases, 151 nt PE, Zymo Ribo-depleted library) using Illumina short-read sequencing on a NovaSeq 6000 platform. The data were mapped to the mouse genome GRCm38.p6 using STAR v.2.7.6a (Dobin and Gingeras, 2016). The mapped transcripts were assembled using Stringtie v.2.1.1 (Kovaka et al., 2019) against Gencode vM25 gff3 annotations. All sequencing data are available on GEO at GSE175854.

Liquid chromatography and mass spectrometry

To extract proteins, tissue pieces were lysed in RIPA buffer and Halt protease/phosphatase inhibitor (Thermo) on a bead mill homogenizer with 2.8 mm ceramic beads, followed by sonication and centrifugation at 14,000 g for 15 minutes at 4 °C. Protein quantity was measured using BCA assay (Thermo), after which 100 µg of proteins were digested using a filter-assisted protocol as described (Manza et al., 2005). Digested peptides were desalted using C18 spin columns (Pierce). Label-free bottom-up mass spectrometry was performed using data dependent acquisition on an Orbitrap Q-Exactive HF connected to a Easy-nLC 1200 nano-UPLC system using typical settings as described (Lau et al., 2019). Mass spectrometry raw data was converted to mzML using ThermoRawFileParser v.1.2

(Hulstaert et al., 2020), and searched against UniProt SwissProt (The UniProt Consortium, 2018) reviewed *Mus musculus* database (retrieved 04/27/2021) with appended contaminant sequences using MSFragger v.3.2 (Kong et al., 2017), followed by post-processing using PeptideProphet and ProteinProphet in the Philosopher suite v.3.4.13 (da Veiga Leprevost et al., 2020) and label-free quantification with match-between-runs using IonQuant v.1.5.5 (Yu et al., 2020). Raw data are available on jPOST (Watanabe et al., 2021) at accession JPST001197.

Retrieval and analysis of human GTEx v8 data

Human gene expression profiles are from GTEx v8 release (GTEx Consortium, 2020) data retrieved from the GTEx portal. Raw gene count data were first normalized using variance stabilizing normalization in DESeq2 v.1.30.1 (Love et al., 2014). The normalized read count data were then sequentially batch corrected with the aid of ComBat (Leek et al., 2012), first against two GTEx v8 technical metadata variables (extraction batch SMNABTCH and sequencing batch SMGEBTCH), then against tissue ischemia time SMTSISCH, and finally against donor death Hardy scale DTHHRDY. Correlations between normalized expression level and categorical age groups were calculated using ANOVA with Tukey HSD post-hoc tests. A post-hoc P value ≤ 0.01 is considered significant.

Prediction of protein level from RNA-seq data

To train a model to predict protein levels from RNA-seq data, CPTAC RNA-seq and mass spectrometry data were retrieved from five published CPTAC studies, namely the breast (Krug et al., 2020), ovarian (Hu et al., 2020b; Zhang et al., 2016), colorectal (Vasaikar et al., 2019; Zhang et al., 2014), lung adenocarcinoma (Gillette et al., 2020), and endometrial (Dou et al., 2020) cancer discovery studies. The expression data were automatically retrieved in accordance with the CPTAC data use and embargo policies by using the cptac v.0.9.1 package in Python 3.9. Statistical learning was performed using scikit-learn 0.24.2 (Lindgren et al., 2021). Transcriptomics data were first standardized using the StandardScaler method in scikit-learn, after which data were split 80/20 into train and test sets. Prediction was performed using an elastic net with the ElasticNetCV method in scikit-learn and with l1 ratios of 0.1, 0.5, 0.7, 0.9, 0.95, 0.99, and 1. An elastic net is chosen for comparability with the CPTAC DREAM baseline models (Yang et al., 2020). The test set correlation coefficients, R^2 , and normalized root mean square error (NRMSE) metrics were reported. A Python script for the retrieval and analysis of the data is provided at <https://github.com/Lau-Lab/CPTACProteinPredictions/>.

Estimation of cell type proportion

To estimate cell type proportions in the bulk RNA-seq data, we retrieved Tabula Muris FACS mouse heart and muscle single-cell RNA sequencing data sets (Tabula Muris Consortium et al., 2018). The Seurat objects were retrieved and clustered using the FindCluster function in Seurat v.4.0.1 (Butler et al., 2018). Cluster identity was assigned through manual interpretation of the Seurat top 10 marker tables. The single-cell and bulk RNA sequencing data were converted to Bioconductor ExpressionSets, following which cell type proportions in the bulk RNA sequencing data were deconvolved using MuSiC v.0.1.1 (Wang et al., 2019). Briefly, this approach model the relative abundance of a gene g in a bulk tissue in

sample j (Y_{jg}) as a function of the proportion of cell types in the tissue in a sample j in a cell type k ($p_{j,k}$), the average transcriptome size of cell type k in a sample j ($S_{j,k}$), and the relative abundance of gene g in subject j in cell type k ($\theta_{j,k}$) for all samples and all cell types. Y is given by the bulk RNA sequencing data and S and θ are estimated in the single-cell data. MuSiC uses a weighted non-negative least square regression method to estimate p (Wang et al., 2019).

Additional statistics and data analysis

Data analysis was performed in R v.4.0.5 and Bioconductor v.3.12 (Huber et al., 2015) on an x86_64-apple-darwin17.0 (64-bit) platform. Statistical tests of differential expression for RNA sequencing reads were performed using DESeq2 v.1.30.1 (Love et al., 2014) using age and sex as factors after filtering low-read genes as commonly practiced. Here we removed genes with fewer than 800 read counts across all samples within a particular tissue from consideration. Following filtering, 17,003 genes in the heart and 16,231 genes in the muscle remained and were compared. Unless specified, we report the adjusted P values (P.adj) and log2 fold change (logFC) of the DESeq2 results. Statistics for expression fold changes in label-free proteomics data were performed using limma v.3.46.0 (Ritchie et al., 2015) with age and sex as factor to report the log2 fold changes (logFC) of proteins between old and young animals. Functional enrichment was performed with the aid of fgsea v.1.16.0 (Sergushichev, 2016) against MSigDB v.7.3 annotations (Liberzon et al., 2011) or ReactomePA v.1.9.4 (Yu and He, 2016) against Reactome annotations (Fabregat et al., 2018) loaded in the package. Comparison of correlation coefficients was performed using cocor v.1.1–3 (Diedenhofen and Musch, 2015).

Results & Discussion

Global features and pathways in aging heart and muscle

We first acquired total RNA sequencing data on the total transcriptome changes in the heart and the skeletal muscle between young adult vs. early aging mice (4 vs. 20 months) (n=4 each) (Supplementary Table S1). The transcriptome profiles in each organ are distinguishable by sex, and shows more separation by ages in female than in male animals (Fig. 1A). On a global level, we considered the cellular pathways that are involved with age-associated gene expression changes using the fast gene set enrichment analysis (FGSEA) algorithm. In both tissues, aging is associated with a positive enrichment of genes involved in innate immune system and neutrophil degranulation (FGSEA P.adj 1.4e–3 heart; 1.0e–2 muscle) as well as GPCR ligand binding terms (P.adj 8.5e–2 heart; 3.1e–3 muscle); and a negative enrichment of genes involved in the respiratory chain (P.adj 3.8e–3 heart; 6.7e–3 muscle), citric acid cycle (P.adj 4.2e–3 heart; 8.1e–3 muscle), and translation (P.adj 6.9e–2 heart; 1.1e–2 muscle) (Figure 1B). Comparable processes were enriched using WikiPathways and KEGG terms as annotations (Supplementary Figure S1). A major discrepancy between the two types of muscle involves genes functioning in extracellular matrix organization and collagen degradation, which are up-regulated in aging hearts but down-regulated in aging tissues.

Closer inspection of the fold-changes of genes making up the leading edges of FGSEA-enriched terms show a concomitant down-regulation of genes in electron transport and translation pathways and up-regulation of innate immune system genes (Supplementary Figure S2A–C). Although muscle contraction genes are implicated in both tissues during aging, we find that the up-regulated genes differ, with *Sln*, *Myl7*, and *Myl4* most prominently induced in the heart as opposed to *Myh6* and *Myl2* in the skeletal muscle (Supplementary Figure S2D), whereas we also observed individual genes in extracellular matrix organization changed in opposite directions as the pathway enrichment results suggest (Supplementary Figure S2E).

Taken together, the pathway-level analysis suggests that during normal heart and skeletal muscle aging, gene expression changes are consistent with a rerouting of gene expression from mitochondrial metabolism and protein synthesis usage toward inflammatory and matricellular functional components, although the changes in extracellular matrix appeared to diverge between the two striated muscle. Our results also add to a chorus of recent findings that implicate the innate immune system in muscle aging (Graber et al., 2021; Lin et al., 2018; Ubaida-Mohien et al., 2019b) and other tissues (Benayoun et al., 2019).

Transcriptomic signatures in striated muscle aging

We next considered the signatures implicated in aging at an individual gene level. We found 358 differentially expressed protein-coding genes in old vs. adult hearts and 276 in skeletal muscles of identical animals at 10% FDR (237 and 177 at 5% FDR in the heart and the muscle, respectively) (Supplementary Data S1–S2). Among the induced genes, we found genes that are previously associated with age-associated diseases as well as genes not previously associated with aging tissues. Among the genes of interest, nuclear receptor subfamily 4 group A member 1 (*Nr4a1*) and mothers against decapentaplegic homolog 3 (*Smad3*), were both significantly decreased in aging skeletal muscle (*Nr4a1* logFC -0.73 , P.adj $3.8e-4$; *Smad3* logFC -0.44 , P.adj $2.1e-2$). A recent large-scale meta-analysis of 739 human skeletal muscle transcriptomes from endurance or resistance exercise interventions pinpointed the human *SMAD3* and *NR4A1* genes as a central hub of acute response to exercise in the skeletal muscle (Amar et al., 2021). Both *SMAD3* and *NR4A1* are acutely up-regulated after exercise, contradirectional to the observed age-associated repression here. *Smad3* encodes the Smad3 protein which interacts with STAT3 and which in skeletal muscle may function to regulate muscle mass (Chao et al., 2012), whereas *Nr4a1* may function to regulate mitochondrial biogenesis (Chao et al., 2012), hinting at potential connections of these genes to the benefits of exercise in delaying age associated muscle and mitochondrial loss.

Results on an individual-gene level were less conserved with other studies that compared young and aging mice than pathway changes. For instance, Lin et al. (Lin et al., 2018) and our data both found strong changes in immune regulation, but Lin et al. reported a strong decrease in *Fkbp5* in aging skeletal muscle which was not recapitulated in this study. Instead, we found a significant decrease in *Fkbp4* (Figure 1D) in aging skeletal muscle. Similarly, although both work noted an upregulation of muscle contraction genes in aging, the specific genes overlap only partially, with Lin et al. reported upregulation of

Myh7, *Myh3*, *Tnnt1*, among others, and the present data are represented by *Myo5a*, *Tpm2*, and *Tnnt2* (Supplementary Data S2). Lastly, among the top 20 positively age-correlated and top 20 negatively age-correlated human genes in the NIA GESTALT study (Tumasian et al., 2021), we observed some evidence for 5 being possibly recapitulated here (up-regulated: *Skap2*, *Cfap61*, *Kcnq5*; down-regulated: *Myl1*, *Casq1*) at a more relaxed 15% FDR cutoff. These across-study differences can plausibly arise from the use of arbitrary significance cutoffs as well as a combination of differences in study design, organism models, technical variations, and stochastic gene regulations. We hypothesize that both individual-gene level and pathway level changes contain complementary information into the molecular mechanisms of aging, which should be taken into account when multiple studies are compared in meta-analyses.

Interestingly, we identified a constellation of genes coding for secreted proteins, which was not the focus of prior reports in aging mice. For instance, in the heart, inactive carboxypeptidase-like protein X2 (*Cpxm2*) encodes a secreted protein that is induced in aging hearts (logFC 1.1, DESeq2 P.adj 9.1e–15). Likewise induced is EGF-containing fibulin-like extracellular matrix protein 1/fibullin 3 (*Efemp1*) (logFC 1.25, P.adj 2.7e–15), which encodes an extracellular matrix protein that may be cleaved into a secreted peptide and that binds with EGF receptor (Figure 1C). Very recently, the NIA GESTALT study has also found human EFEMP1 to be positively correlated with age in the skeletal muscle (Tumasian et al., 2021). In the normal aging muscle, osteocrin (*Ostn*) encodes a secreted hormone musclin that acts as an exercise induced myokine (Subbotina et al., 2015) but also functions in the heart where it may protect against apoptosis and inflammation (Hu et al., 2020a). In the analyzed animals, *Ostn* is induced in normal aging animals but most prominently in females (Figure 1D) (logFC 3.3, P.adj 8.0e–4). *Bdnf* encodes a myokine that is induced by exercise and regulates energy metabolism at least in female mice (Yang et al., 2019); we found dimorphic expression with higher expression in female and which is further induced in aged tissues (logFC 0.9, P.adj 5.9e–4). Other secreted factor encoding genes changed in aging include *Gdf11* (logFC 1.76, P.adj 9.5e–2) in skeletal muscle as well as *Vegfd* (logFC 0.85, P.adj 5.8e–2), *Frzb* (logFC 0.69 P.adj 9.3e–2), *Sfrp1* (logFC 0.62, P.adj 6.6e–3), *Fstl4* (logFC 0.57, P.adj 3.6e–3), and *Fgf13* (logFC –0.27, P.adj 4.8e–2) in the heart (Supplementary Data S1–2).

We identified 20 common protein-coding genes that are differentially expressed in aging in both tissues at 10% FDR (Fig. 2A). The shared genes show strong positive correlations (Spearman's correlation coefficient r 0.65, P 0.0032) with the exception of one outlier (predicted gene *Gm50364*), which is repressed in the muscle and induced in the normal aging heart (Fig. 2B). Among the common genes, formin-1 (*Fmn1*) codes for a myofibril differentiation factor that plays a role in the formation of adherens junction and is increased in aged muscles and hearts. Myeloid leukemia factor 1 (*Mlf1*) codes for a protein that may serve as a negative regulator of cell cycle exit and is suppressed in both aging hearts and muscles.

To assess whether the gene expression changes were due to changes in cell type compositions, we decomposed the bulk RNA-seq read count matrices into individual cell types derived from single-cell sequencing data using weighted non-negative least square

methods (Supplementary Figure S3). We found no evidence of substantial changes of overall cell type proportion, suggesting the observed transcriptomic changes are unlikely due to wholesale changes in cell population in young adult vs. aged hearts and muscles.

RNA-protein correlation and predictability of protein-level changes

To estimate whether the effect of aging transcriptome changes is potentially translated to proteomic changes, we first applied a statistical learning method against one of the largest proteogenomics data sets in existence to train a model to predict protein levels using their cognate transcript levels as proxy. The correlation coefficients, R^2 values, and normalized root mean square errors (NRMSE) values between the transcript-predicted protein level and empirical mass spectrometry-measured protein levels across subjects are taken as the protein predictability of a transcript for each gene. In total, we estimated the protein predictability of 11,896 transcripts (Supplementary Data S3). We found a large range of predictability where the predicted-actual Pearson's correlation coefficients ranged from -0.822 to 0.999 (interquartile range 0.18 – 0.54) (Figure 3A-B). A small portion of transcripts (4.5%) had negative correlation with protein levels. The median r of all genes is 0.374 which is comparable to previously reported RNA-protein correlation in comparable data sets (Eicher et al., 2019; Li et al., 2019; Yang et al., 2020) and in human tissues (Jiang et al., 2020). We observed no clear relationship between transcript baseline abundance and protein predictability (Supplementary Figure S4).

We found that poor predictors ($r \leq 0.3$) are enriched in pathways involving major multiprotein complexes, including Reactome Translation (P.adj $2.4e-38$), Nonsense-mediated Decay (NMD) (P.adj $6.4e-21$), and Respiratory electron transport (P.adj $1.5e-10$) terms. Good predictors ($r \geq 0.7$) are enriched in Reactome Biological oxidations (P.adj $1.8e-6$), Extracellular matrix organization (P.adj $2.0e-4$), and Metabolism of lipids (P.adj $7.8e-3$) terms (Figure 3C). This agrees with emerging themes from cancer and normal tissue studies. For instance, a large-scale GTEx survey of 32 normal human tissues has found that secreted proteins and proteins in multi-protein complexes are generally poorly predictable from transcripts (Jiang et al., 2020), presumably because of additional post-translational constraints on their steady state levels. The modeled protein predictability values (i.e., test set correlation between predicted protein level and empirical protein levels in CPTAC) agree strongly with GTEx RNA-protein Spearman's correlation coefficients (Figure 3D–E). This result corroborates that proteins exhibit a wide range of predictability by proxy transcripts, and hence different transcripts have different intrinsic value in reflecting actual protein abundance states across tissues, and moreover, that predictions of RNA-protein agreement are “transferable” across to different samples and based on basic biophysical constraints. For example, long-half-life housekeeping proteins are usually more predictable from transcripts whereas the abundance of multi-protein complex members are buffered by complex stoichiometry and assembly.

To further verify the potential impact of the transcriptome signatures on the proteome, we performed in-house exploratory proteomics analyses of identical tissues from identical animals using label-free quantitative tandem mass spectrometry (Supplementary Data S4). In total, we acquired the MS1 label-free quantity of 1,254 distinct proteins in the heart

and the skeletal muscle identified with ProteinProphet protein probability ≥ 0.95 . We identified 58 and 76 proteins with nominal limma $P \leq 0.05$ and $|\log FC| \geq 0.5$ in the heart and the skeletal muscle, respectively, although only 6 proteins reached adjusted $P \leq 0.1$, presumably due to the limited depth and breadth of the proteomics profile performed here. Nevertheless, pathway analysis using FGSEA against MSigDB revealed an enrichment of similar annotation terms to the transcriptomics data, including Reactome Extracellular matrix organization (FGSEA permutation P : $3.5e-4$) and Innate immune system (P : 0.053) terms, suggesting the mass spectrometry experiment was able to capture a representative footprint of the aging proteomes in these tissues.

Not unexpectedly, we found there is a robust correlation between RNA and protein relative abundance *across* genes *within* a tissue (Pearson's r : 0.51 heart; 0.46 muscle) (Figure 4A). There was a modest decrease in correlation of RNA and protein levels in aged samples as previously reported, but in our data this difference was not significant (Fisher's z P : 0.49 heart 0.46 muscle). The correlation between RNA and proteins weakens significantly when correlations across samples are considered (Pearson's r : 0.18 heart; 0.14 muscle) (Figure 4B). This reflects the distinction between RNA-protein correlation across genes vs. across samples, and corroborates mounting evidence that show although abundant proteins tend to have abundant transcripts, transcript changes are imperfectly correlated to proteins due to post-transcriptional activation (Franks et al., 2017). Nevertheless, when only age-differentially expressed transcripts (10% FDR) were compared, we observed a general concordance in the directionality of protein changes (Figure 4C). Notably, RNA-protein correlation is higher among transcripts with nominal changes (DESeq2 $P \leq 0.1$) that had higher estimated protein prediction ($1 \geq r \geq 0.5$) than those with lower prediction ($0 \leq r \leq 0.5$) (correlation 0.42 vs. 0.11; $\text{coco}r$ P $1.5e-3$) (Supplementary Figure S5). The analysis therefore corroborates the transferability of the learned model and offers supportive evidence for the adaption of predicted RNA-protein correlations as one optional method to help prioritize discovered transcript signatures.

Conservation of identified striated muscle aging signatures in human

Because specific genes and pathways may underlie aging processes in different organisms, we next estimated the extent to which the identified mouse aging transcriptome features are translatable to humans. To do so, we analyzed the age vs. expression relationship of GTEx v8 human transcriptomics data. In total, we retrieved 17,382 RNA sequencing samples including 432 heart left ventricle, 429 atrial appendage, and 803 skeletal muscle transcriptomes, and performed stepwise normalization for technical, tissue, and donor batches (Supplementary Figure S6A–E). We then compared whether the normalized gene expression of each gene signature is correlated with donor age groups in different tissues in humans. We found overall there are complex trends between identified age-associated signatures with human gene expression-age group relationship. Only 35% (107/305) and 44% (97/217) of the analyzed aging genes in the mouse heart and the skeletal muscle, respectively, had significant age-expression relationship in GTEx v8 (ANOVA $P \leq 0.01$), suggesting not every identified signature is potentially conserved across species (Supplementary Figure S6). On a global level, up-regulated genes in normal aging mouse hearts are significantly more likely to be positively correlated in expression with donor age

groups in human heart left ventricle (Wilcoxon P : $8.0e-5$) and atrial appendage (P : $2.8e-2$) samples but not in the other compared human tissues including the kidney cortex (P : 0.33) or liver (P : 0.27). Moreover, this correlation is not existent when only sexually dimorphic (sex-differentially expressed at 10% FDR) genes in the mouse are compared in human hearts (Wilcoxon P : 0.87 for heart - atrial appendage and 0.95 for heart - left ventricle). This global relationship is considerably subdued for genes that are differentially regulated in aging muscle, which may suggest that the aging signatures in this tissue are more specific to species or otherwise show non-linear change over the lifespan (Supplementary Figure S7).

Notably, the age-expression relationships of signature genes are often not preserved in other tissues despite the gene being expressed at appreciable levels. For example, *Efemp1* is induced in aging in the mouse data here, and is positively correlated with age group in GTEx v8 human heart and skeletal muscle tissues but not in the kidney or the liver, despite the human *EFEMP1* gene being expressed at a similar baseline level in those tissues (Figure 5A). To corroborate this observation, we acquired and analyzed RNA sequencing data from the kidney of identical animals (Supplementary Data S5). The data corroborated that there are no significant changes in *Efemp1* (logFC 0.20, P.adj 0.28). Likewise, *Fkpb4* is not significantly changed in the kidneys of aging humans in GTEx v8 or the RNA-seq data of identical animals (logFC 0.03, P.adj 0.92) (Figure 5B); *Sod3* is correlated with age in human hearts and muscles but not kidney, whereas we also found no significant changes in mouse kidney in our RNA-seq data (logFC 0.00, P.adj 1.00) (Figure 5C). Taken together, the results suggest that normal aging signatures exhibit tissue specificity, both across skeletal and cardiac muscles as well as between striated muscles and other organs, as well as potential species specificity, and a selection strategy might be employed to prioritize aging signatures that show expression trends in humans.

Integrating information from both RNA-protein correlation and human conservation, we ranked the identified aging signatures based on how well the transcript-predicted across-sample protein values reflect empirical protein levels ($r \geq 0.5$), and moreover selected signatures that show a significant correlation with age group in human GTEx v8 tissues (ANOVA $P \leq 0.01$). This combined analysis led to 134 of the prioritized transcript signatures in this study that are potentially more likely to have bonafide relevance in the biology of aging tissues at the protein levels (Figure 6A–D).

Non-coding genes in striated muscle aging

Lastly, although the focus of this study is to prioritize protein correlation of protein-coding gene signatures, not all transcripts function only through their translation products, and non-coding RNAs play important roles in virtually all aspects of biology. As we acquired total ribosomal-depleted RNA abundance data, we also explored the changes of non-coding RNAs in aging hearts and muscles including non-poly-A+ transcripts. From the data, we found 19 non-coding genes to be differentially expressed with age in the heart and 36 in the muscle at 10% FDR, with 3 overlapping (10 and 26 at 5% FDR in the heart and liver, respectively). Grouping into gene category annotations suggests that the differentially expressed non-coding genes included sense, anti-sense, and intergenic long non-coding RNAs (lncRNAs), as well as small nucleolar RNAs and processed pseudogenes (Figure 7A).

The non-coding RNAs were manually inspected for read mapping and strand specificity. We recapitulated changes in two maternally imprinted lncRNAs *Meg3* and *Riat*, which were previously found using qPCR array to be decreased in skeletal muscle over the lifespan (Mikovic et al., 2018)), although *Meg3* was also found to be increased in senescent endothelial cells (Boon et al., 2016). We also identified additional age-regulated lncRNAs. In the muscle, *Plet1os* is also decreased (logFC -1.41, P.adj 8.8e-10) (Figure 7B), whereas *Foxo6os* is decreased in aging (logFC -0.63, P.adj 0.002), and was previously found to be depressed in insulin resistant muscle (Figure 7C). In the heart, the lncRNA *Mhrt* is located on the opposite strand of mouse *Myh7* and has been associated with the regulation of *Myh6/Myh7* ratios as well as protection against pathological cardiac remodeling (Han et al., 2014); in the data, we found *Mhrt* to be drastically reduced (logFC -0.44, P.adj 2.7e-7) in aged hearts. Among the age-regulated lncRNAs, five (*Neat1*, *Plet1os*, *Foxo6os*, *Peg13*, *Mhrt*) were previously found to have potential translatability in smProt (Hao et al., 2018) or engaged in ribosomes in the mouse heart (van Heesch et al., 2019), suggesting a possibility that they may be translated. A growing number of microproteins are known to be translated in striated muscles. Future work combining transcriptomics and proteomics approaches might determine whether they are differentially regulated in aging or participate in associated pathophysiological processes.

Limitations and Future Directions

There are several important limitations that pertain to the experimental results here. The aged animals (20 months) used here are comparatively younger than the C57BL/6/J mice in some other studies (22–28 months) (Bartling et al., 2019; Graber et al., 2021; Greenig et al., 2020; Mikovic et al., 2018), which might decrease the sensitivity of identifying differentially expressed genes across age groups in the RNA sequencing data and omit signals that only appear in very elderly animals (Graber et al., 2021). Although the number of animals used (n=8) here is not atypical for large-scale discovery experiments using genetically identical animals in controlled laboratory settings, the cohort size suggests genes with smaller effect sizes would be omitted and a larger sample size is needed for further stratification, such as to detect the contribution of the age-sex interaction term that modulates sex-specific aging features.

We analyzed gene expression in the mouse quadriceps femoris skeletal muscle. In human, the quadriceps femoris contains a mixture of fast-twitch glycolytic fibers and slow-twitch oxidative fibers, which present different gene expression regulations and functions across the lifespan with more prominent fast-twitch muscle decline with age (Grosicki et al., 2021; Trappe et al., 2003). In addition, physical activities have emerged as a major modifier of muscle biological age and aging processes, but this was not considered in the current study (Cartee et al., 2016; Sanford et al., 2020). Future work might compare individual muscle fibers across normal and exercised groups to further tease out functionally relevant signatures that correlate with aging and exercise.

We performed only a label-free proteomics experiment of limited depth here, as our major goal was to compare transcriptomics and proteomics fold change. Follow-up studies might use deep quantitative proteomics comparison with common stable isotope labeling

mass spectrometry techniques to help reveal specific proteomics features of aging that are not apparent at the transcript level, or identify non-canonical translation products using proteogenomics methods. Lastly, we find that the current model remains insufficient to yield a significant improvement in predicting protein-level changes (data not shown), but the performance of transfer learning approaches is expected to continue to benefit from algorithmic advances and the accrual of large data sets in closely related tissues and species. Future work might apply a full-fledged transfer learning strategy by applying the learned model directly to a new set of transcriptomics data to predict protein level changes in a system.

Conclusion

This study examined the aging transcriptome profiles in skeletal and cardiac muscles. Our results provide further evidence to the emerging view that points to extracellular matrix, mitochondrial, and innate immunity processes as distinguishing factors in aging tissues. Notably, the data also point to several new discoveries of age-associated genes encoding for secreted signals, suggesting age-associated myokines and cardiokines present a promising avenue for further understanding the molecular mechanisms of heart and muscle aging. A number of identified signatures are specific to striated muscles aging while unchanged over age in other tissues, and hence may be particular to muscle aging processes. Comparisons to public human transcriptomes data showed that 35–45% of the aging signatures show age-dependent expression in corresponding human tissues, suggesting a subset of transcript signatures may be prioritized that are more likely conserved in humans.

Prioritizing gene lists remains an important bottleneck in deriving actionable interpretations from large-scale omics studies (Jourquin et al., 2012; Lau et al., 2018). Here we show a computational workflow that transfers the predicted correlation between transcript and protein levels trained from a large data set to a newly acquired data set. Recent reports in the literature have consistently found low to moderate RNA-protein correlation. To our knowledge, this work is the first to explicitly take into consideration differential RNA-protein correlation coefficients as a means to prioritize potential aging-associated transcript signatures from RNA-sequencing data. We found about 48% of the identified aging-associated transcript signatures are predicted to correlate well with the abundance of their protein counterparts, and used a rule-based strategy to filter for prioritized signatures that are potentially predictive of protein-level changes. We suggest that the integrated computational data analysis approach presented here may be applied to future transcriptomics studies in aging and diseases to help extract potentially high-value signatures, such as those that are likely to reflect protein-level changes or are translatable to humans.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by NIH awards F32-HL149191 to YH; R21-HL150456, R00-HL127302, and R01-HL141278 to MPL; and R00-HL144829 to EL. NLK is supported by the UC Denver MARC U-STAR training

program T34-GM096958. Data utilized in this publication include those generated by the National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC).

Data Availability

All sequencing data are available on GEO at GSE175854. Proteomics raw data are available on jPOST (Watanabe et al., 2021) at accession JPST001197.

Bibliography

- Amar D, Lindholm ME, Norrbom J, Wheeler MT, Rivas MA, and Ashley EA (2021). Time trajectories in the transcriptomic response to exercise - a meta-analysis. *Nat. Commun* 12, 3471. [PubMed: 34108459]
- Bartling B, Niemann K, Pliquett RU, Treede H, and Simm A. (2019). Altered gene expression pattern indicates the differential regulation of the immune response system as an important factor in cardiac aging. *Exp. Gerontol* 117, 13–20. [PubMed: 29738791]
- Benayoun BA, Pollina EA, Singh PP, Mahmoudi S, Harel I, Casey KM, Dulken BW, Kundaje A, and Brunet A. (2019). Remodeling of epigenome and transcriptome landscapes with aging in mice reveals widespread induction of inflammatory responses. *Genome Res* 29, 697–709. [PubMed: 30858345]
- Boon RA, Hofmann P, Michalik KM, Lozano-Vidal N, Berghäuser D, Fischer A, Knau A, Jaé N, Schürmann C, and Dimmeler S. (2016). Long noncoding RNA meg3 controls endothelial cell aging and function: implications for regenerative angiogenesis. *J. Am. Coll. Cardiol* 68, 2589–2591. [PubMed: 27931619]
- Butler A, Hoffman P, Smibert P, Papalexi E, and Satija R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol* 36, 411–420. [PubMed: 29608179]
- Cartee GD, Hepple RT, Bamman MM, and Zierath JR (2016). Exercise promotes healthy aging of skeletal muscle. *Cell Metab* 23, 1034–1047. [PubMed: 27304505]
- Chao LC, Wroblewski K, Ilkayeva OR, Stevens RD, Bain J, Meyer GA, Schenk S, Martinez L, Vergnes L, Narkar VA, et al. (2012). Skeletal muscle Nur77 expression enhances oxidative metabolism and substrate utilization. *J. Lipid Res* 53, 2610–2619. [PubMed: 23028113]
- Diedenhofen B, and Musch J. (2015). cocor: a comprehensive solution for the statistical comparison of correlations. *PLoS ONE* 10, e0121945. [PubMed: 25835001]
- Dobin A, and Gingeras TR (2016). Optimizing RNA-Seq Mapping with STAR. *Methods Mol. Biol* 1415, 245–262. [PubMed: 27115637]
- Dou Y, Kawaler EA, Cui Zhou D, Gritsenko MA, Huang C, Blumenberg L, Karpova A, Petyuk VA, Savage SR, Satpathy S, et al. (2020). Proteogenomic characterization of endometrial carcinoma. *Cell* 180, 729–748.e26. [PubMed: 32059776]
- Eicher T, Patt A, Kautto E, Machiraju R, Mathé E, and Zhang Y. (2019). Challenges in proteogenomics: a comparison of analysis methods with the case study of the DREAM proteogenomics sub-challenge. *BMC Bioinformatics* 20, 669. [PubMed: 31861998]
- Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, Haw R, Jassal B, Korninger F, May B, et al. (2018). The Reactome Pathway Knowledgebase. *Nucleic Acids Res* 46, D649–D655. [PubMed: 29145629]
- Franks A, Airoidi E, and Slavov N. (2017). Post-transcriptional regulation across human tissues. *PLoS Comput. Biol* 13, e1005535. [PubMed: 28481885]
- Gillette MA, Satpathy S, Cao S, Dhanasekaran SM, Vasaikar SV, Krug K, Petralia F, Li Y, Liang W-W, Reva B, et al. (2020). Proteogenomic characterization reveals therapeutic vulnerabilities in lung adenocarcinoma. *Cell* 182, 200–225.e35. [PubMed: 32649874]
- Graber TG, Maroto R, Thompson J, Widen S, Man Z, Pajski ML, and Rasmussen BB (2021). Skeletal muscle transcriptome alterations related to physical function decline in older mice. *BioRxiv*

- Greenig M, Melville A, Huntley D, Isalan M, and Mielcarek M. (2020). Cross-Sectional Transcriptional Analysis of the Aging Murine Heart. *Front. Mol. Biosci* 7, 565530. [PubMed: 33102519]
- Grosicki GJ, Gries KJ, Minchev K, Raue U, Chambers TL, Begue G, Finch H, Graham B, Trappe TA, and Trappe S. (2021). Single muscle fibre contractile characteristics with lifelong endurance exercise. *J Physiol (Lond)* 599, 3549–3565. [PubMed: 34036579]
- GTEX Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 369, 1318–1330. [PubMed: 32913098]
- Han P, Li W, Lin C-H, Yang J, Shang C, Nuernberg ST, Jin KK, Xu W, Lin C-Y, Lin C-J, et al. (2014). A long noncoding RNA protects the heart from pathological hypertrophy. *Nature* 514, 102–106. [PubMed: 25119045]
- Hao Y, Zhang L, Niu Y, Cai T, Luo J, He S, Zhang B, Zhang D, Qin Y, Yang F, et al. (2018). SmProt: a database of small proteins encoded by annotated coding and non-coding RNA loci. *Brief. Bioinformatics* 19, 636–643. [PubMed: 28137767]
- van Heesch S, Witte F, Schneider-Lunitz V, Schulz JF, Adami E, Faber AB, Kirchner M, Maatz H, Blachut S, Sandmann C-L, et al. (2019). The translational landscape of the human heart. *Cell* 178, 242–260.e29. [PubMed: 31155234]
- Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke T, et al. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* 12, 115–121. [PubMed: 25633503]
- Hulstaert N, Shofstahl J, Sachsenberg T, Walzer M, Barsnes H, Martens L, and Perez-Riverol Y. (2020). ThermoRawFileParser: Modular, Scalable, and Cross-Platform RAW File Conversion. *J. Proteome Res* 19, 537–542. [PubMed: 31755270]
- Hu C, Zhang X, Zhang N, Wei W-Y, Li L-L, Ma Z-G, and Tang Q-Z (2020a). Osteocrin attenuates inflammation, oxidative stress, apoptosis, and cardiac dysfunction in doxorubicin-induced cardiotoxicity. *Clin. Transl. Med* 10, e124. [PubMed: 32618439]
- Hu Y, Pan J, Shah P, Ao M, Thomas SN, Liu Y, Chen L, Schnaubelt M, Clark DJ, Rodriguez H, et al. (2020b). Integrated Proteomic and Glycoproteomic Characterization of Human High-Grade Serous Ovarian Carcinoma. *Cell Rep* 33, 108276. [PubMed: 33086064]
- Jiang L, Wang M, Lin S, Jian R, Li X, Chan J, Dong G, Fang H, Robinson AE, GTEx Consortium, et al. (2020). A quantitative proteome map of the human body. *Cell* 183, 269–283.e19. [PubMed: 32916130]
- Jourquin J, Duncan D, Shi Z, and Zhang B. (2012). GLAD4U: deriving and prioritizing gene lists from PubMed literature. *BMC Genomics* 13 Suppl 8, S20.
- Kong AT, Leprevost FV, Avtonomov DM, Mellacheruvu D, and Nesvizhskii AI (2017). MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat. Methods* 14, 513–520. [PubMed: 28394336]
- Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, and Pertea M. (2019). Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol* 20, 278. [PubMed: 31842956]
- Krug K, Jaehnig EJ, Satpathy S, Blumenberg L, Karpova A, Anurag M, Miles G, Mertins P, Geffen Y, Tang LC, et al. (2020). Proteogenomic landscape of breast cancer tumorigenesis and targeted therapy. *Cell* 183, 1436–1456.e31. [PubMed: 33212010]
- Landi F, Cruz-Jentoft AJ, Liperoti R, Russo A, Giovannini S, Tosato M, Capoluongo E, Bernabei R, and Onder G. (2013). Sarcopenia and mortality risk in frail older persons aged 80 years and older: results from ilSIRENTE study. *Age Ageing* 42, 203–209. [PubMed: 23321202]
- Lau E, Venkatraman V, Thomas CT, Wu JC, Van Eyk JE, and Lam MPY (2018). Identifying High-Priority Proteins Across the Human Diseasesome Using Semantic Similarity. *J. Proteome Res* 17, 4267–4278. [PubMed: 30256117]
- Lau E, Han Y, Williams DR, Thomas CT, Shrestha R, Wu JC, and Lam MPY (2019). Splice-Junction-Based Mapping of Alternative Isoforms in the Human Proteome. *Cell Rep* 29, 3751–3765.e5. [PubMed: 31825849]

- Leek JT, Johnson WE, Parker HS, Jaffe AE, and Storey JD (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 28, 882–883. [PubMed: 22257669]
- Lesnefsky EJ, Chen Q, and Hoppel CL (2016). Mitochondrial metabolism in aging heart. *Circ. Res* 118, 1593–1611. [PubMed: 27174952]
- Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, and Mesirov JP (2011). Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 27, 1739–1740. [PubMed: 21546393]
- Lindgren CM, Adams DW, Kimball B, Boekweg H, Tayler S, Pugh SL, and Payne SH (2021). Simplified and unified access to cancer proteogenomic data. *J. Proteome Res* 20, 1902–1910. [PubMed: 33560848]
- Lin I-H, Chang J-L, Hua K, Huang W-C, Hsu M-T, and Chen Y-F (2018). Skeletal muscle in aged mice reveals extensive transformation of muscle gene expression. *BMC Genet* 19, 55. [PubMed: 30089464]
- Li H, Siddiqui O, Zhang H, and Guan Y. (2019). Joint learning improves protein abundance prediction in cancers. *BMC Biol* 17, 107. [PubMed: 31870366]
- Love MI, Huber W, and Anders S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550. [PubMed: 25516281]
- Mani DR, Maynard M, Kothadia R, Krug K, Christianson KE, Heiman D, Clauser KR, Birger C, Getz G, and Carr SA (2021). PANOPLY: a cloud-based platform for automated and reproducible proteogenomic data analysis. *Nat. Methods* 18, 580–582. [PubMed: 34040252]
- Manza LL, Stamer SL, Ham A-JL, Codreanu SG, and Liebler DC (2005). Sample preparation and digestion for proteomic analyses using spin filters. *Proteomics* 5, 1742–1745. [PubMed: 15761957]
- Mikovic J, Sadler K, Butchart L, Voisin S, Gerlinger-Romero F, Della Gatta P, Grounds MD, and Lamon S. (2018). MicroRNA and Long Non-coding RNA Regulation in Skeletal Muscle From Growth to Old Age Shows Striking Dysregulation of the Callipyge Locus. *Front. Genet* 9, 548. [PubMed: 30505320]
- Moore AZ, Caturegli G, Metter EJ, Makrogiannis S, Resnick SM, Harris TB, and Ferrucci L. (2014). Difference in muscle quality over the adult life span and biological correlates in the Baltimore Longitudinal Study of Aging. *J. Am. Geriatr. Soc* 62, 230–236. [PubMed: 24438020]
- Murgia M, Toniolo L, Nagaraj N, Ciciliot S, Vindigni V, Schiaffino S, Reggiani C, and Mann M. (2017). Single Muscle Fiber Proteomics Reveals Fiber-Type-Specific Features of Human Muscle Aging. *Cell Rep* 19, 2396–2409. [PubMed: 28614723]
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, and Smyth GK (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43, e47. [PubMed: 25605792]
- Sanford JA, Nogiec CD, Lindholm ME, Adkins JN, Amar D, Dasari S, Drugan JK, Fernández FM, Radom-Aizik S, Schenk S, et al. (2020). Molecular transducers of physical activity consortium (motrpac): mapping the dynamic responses to exercise. *Cell* 181, 1464–1474. [PubMed: 32589957]
- Sergushichev A. (2016). An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. *BioRxiv*
- Subbotina E, Sierra A, Zhu Z, Gao Z, Koganti SRK, Reyes S, Stepniak E, Walsh SA, Acevedo MR, Perez-Terzic CM, et al. (2015). Musclin is an activity-stimulated myokine that enhances physical endurance. *Proc Natl Acad Sci USA* 112, 16042–16047. [PubMed: 26668395]
- Tabula Muris Consortium, Overall coordination, Logistical coordination, Organ collection and processing, Library preparation and sequencing, Computational data analysis, Cell type annotation, Writing group, Supplemental text writing group, and Principal investigators (2018). Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* 562, 367–372. [PubMed: 30283141]
- The UniProt Consortium (2018). UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 46, 2699. [PubMed: 29425356]

- Timmons JA, Volmar C-H, Crossland H, Phillips BE, Sood S, Janczura KJ, Törmäkangas T, Kujala UM, Kraus WE, Atherton PJ, et al. (2019). Longevity-related molecular pathways are subject to midlife “switch” in humans. *Aging Cell* 18, e12970. [PubMed: 31168962]
- Trappe S, Gallagher P, Harber M, Carrithers J, Fluckey J, and Trappe T. (2003). Single muscle fibre contractile properties in young and old men and women. *J Physiol (Lond)* 552, 47–58. [PubMed: 12837929]
- Tripodskiadis F, Xanthopoulos A, and Butler J. (2019). Cardiovascular aging and heart failure: JACC review topic of the week. *J. Am. Coll. Cardiol* 74, 804–813. [PubMed: 31395131]
- Tumasian RA, Harish A, Kundu G, Yang J-H, Ubaida-Mohien C, Gonzalez-Freire M, Kaileh M, Zukley LM, Chia CW, Lyashkov A, et al. (2021). Skeletal muscle transcriptome in healthy aging. *Nat. Commun* 12, 2014. [PubMed: 33795677]
- Ubaida-Mohien C, Lyashkov A, Gonzalez-Freire M, Tharakan R, Shardell M, Moaddel R, Semba RD, Chia CW, Gorospe M, Sen R, et al. (2019a). Discovery proteomics in aging human skeletal muscle finds change in spliceosome, immunity, proteostasis and mitochondria. *Elife* 8.
- Ubaida-Mohien C, Gonzalez-Freire M, Lyashkov A, Moaddel R, Chia CW, Simonsick EM, Sen R, and Ferrucci L. (2019b). Physical activity associated proteomics of skeletal muscle: being physically active in daily life may protect skeletal muscle from aging. *Front. Physiol* 10, 312. [PubMed: 30971946]
- Vasaikar S, Huang C, Wang X, Petyuk VA, Savage SR, Wen B, Dou Y, Zhang Y, Shi Z, Arshad OA, et al. (2019). Proteogenomic analysis of human colon cancer reveals new therapeutic opportunities. *Cell* 177, 1035–1049.e19. [PubMed: 31031003]
- da Veiga Leprevost F, Haynes SE, Avtonomov DM, Chang H-Y, Shanmugam AK, Mellacheruvu D, Kong AT, and Nesvizhskii AI (2020). Philosopher: a versatile toolkit for shotgun proteomics data analysis. *Nat. Methods* 17, 869–870. [PubMed: 32669682]
- Wang X, Park J, Susztak K, Zhang NR, and Li M. (2019). Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat. Commun* 10, 380. [PubMed: 30670690]
- Watanabe Y, Yoshizawa AC, Ishihama Y, and Okuda S. (2021). The jPOST Repository as a Public Data Repository for Shotgun Proteomics. *Methods Mol. Biol* 2259, 309–322. [PubMed: 33687724]
- Yang M, Petralia F, Li Z, Li H, Ma W, Song X, Kim S, Lee H, Yu H, Lee B, et al. (2020). Community Assessment of the Predictability of Cancer Protein and Phosphoprotein Levels from Genomics and Transcriptomics. *Cell Syst* 11, 186–195.e9. [PubMed: 32710834]
- Yang X, Brobst D, Chan WS, Tse MCL, Herlea-Pana O, Ahuja P, Bi X, Zaw AM, Kwong ZSW, Jia W-H, et al. (2019). Muscle-generated BDNF is a sexually dimorphic myokine that controls metabolic flexibility. *Sci. Signal* 12.
- Yu G, and He Q-Y (2016). ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol. Biosyst* 12, 477–479. [PubMed: 26661513]
- Yu F, Haynes SE, Teo GC, Avtonomov DM, Polasky DA, and Nesvizhskii AI (2020). Fast Quantitative Analysis of timsTOF PASEF Data with MSFragger and IonQuant. *Mol. Cell. Proteomics* 19, 1575–1585. [PubMed: 32616513]
- Zhang B, Wang J, Wang X, Zhu J, Liu Q, Shi Z, Chambers MC, Zimmerman LJ, Shaddox KF, Kim S, et al. (2014). Proteogenomic characterization of human colon and rectal cancer. *Nature* 513, 382–387. [PubMed: 25043054]
- Zhang H, Liu T, Zhang Z, Payne SH, Zhang B, McDermott JE, Zhou J-Y, Petyuk VA, Chen L, Ray D, et al. (2016). Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. *Cell* 166, 755–765. [PubMed: 27372738]

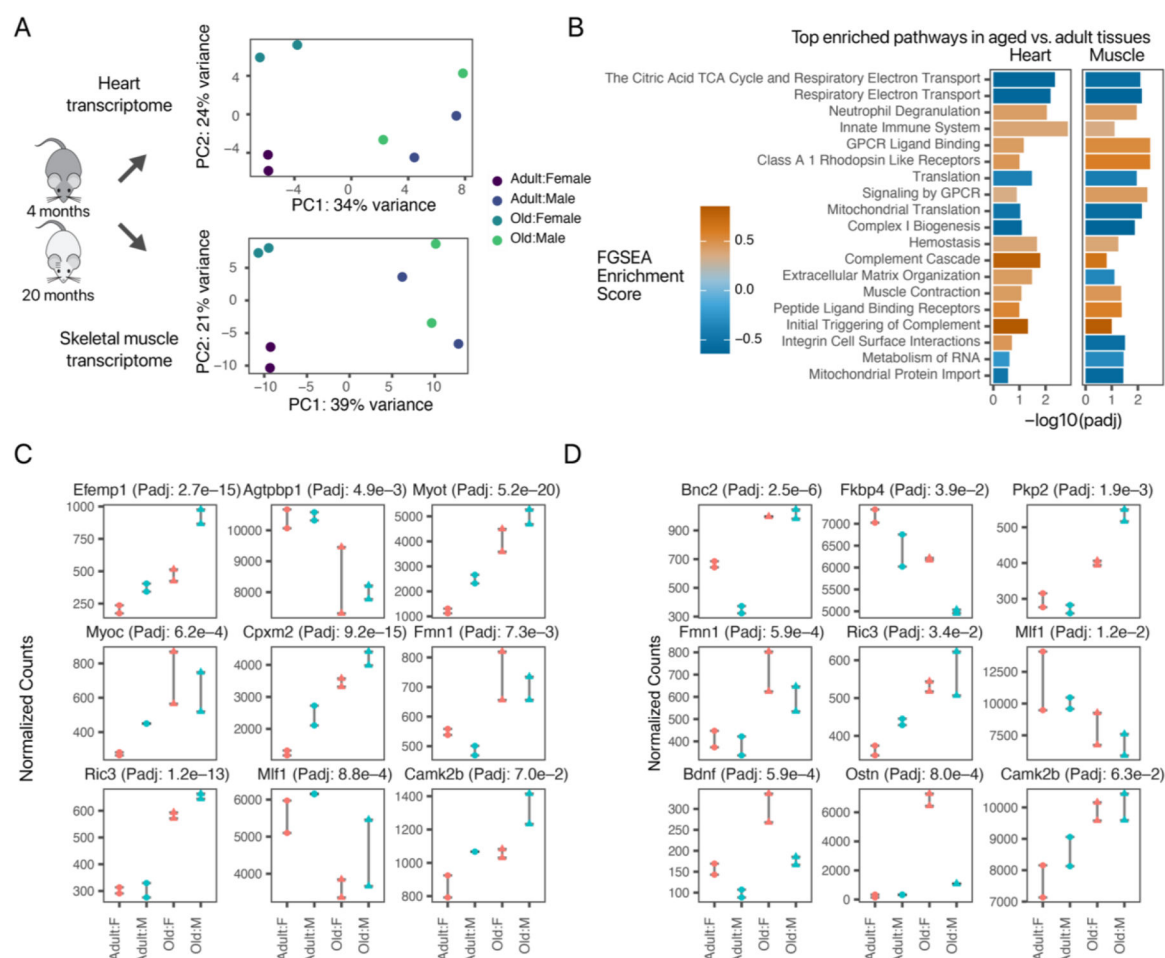


Figure 1: Transcriptome changes in young adult mice vs. early aging mice.

A. Principal component analysis of normalized gene counts in the left ventricle (top) and quadriceps femoris (bottom) show separation by sex and age. **B.** Enriched pathways in aged vs. adult gene expression in two tissues. **C-D.** Normalized read counts showing sex and age expression among selected differentially expressed aging genes in **(C)** the heart and **(D)** the skeletal muscle (FDR <5% with the exception of Camk2b where FDR <10%).

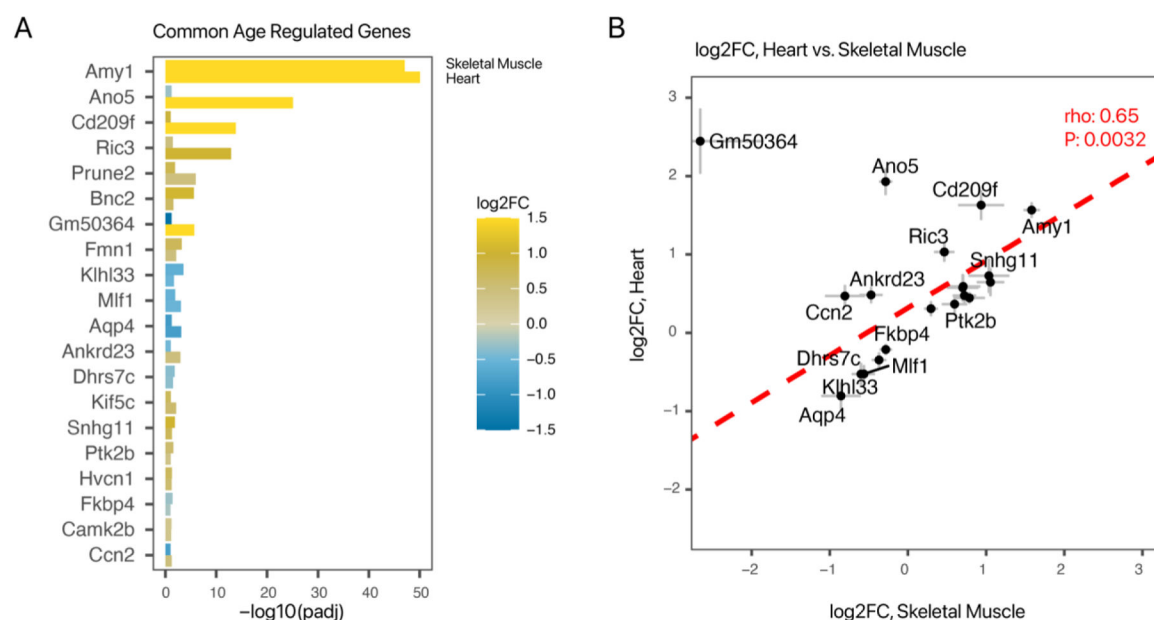


Figure 2: Shared aging signatures in aged cardiac and skeletal muscle.

A. Bar chart showing the adjusted P values (x-axis) and log₂ fold-changes (log₂FC; fill color) of the 20 aging associated genes identified in both tissues. **B.** Scatter plot showing a comparison of fold changes (aged vs. young adult) and a robust positive correlation (Spearman's correlation coefficient $\rho = 0.65$, $P = 0.0032$) between the two tissues. X-axis: log₂ fold-change in skeletal muscle; y-axis: log₂ fold-change in the heart. Error bars: standard error of log₂ fold-change; line: best-fit linear curve.

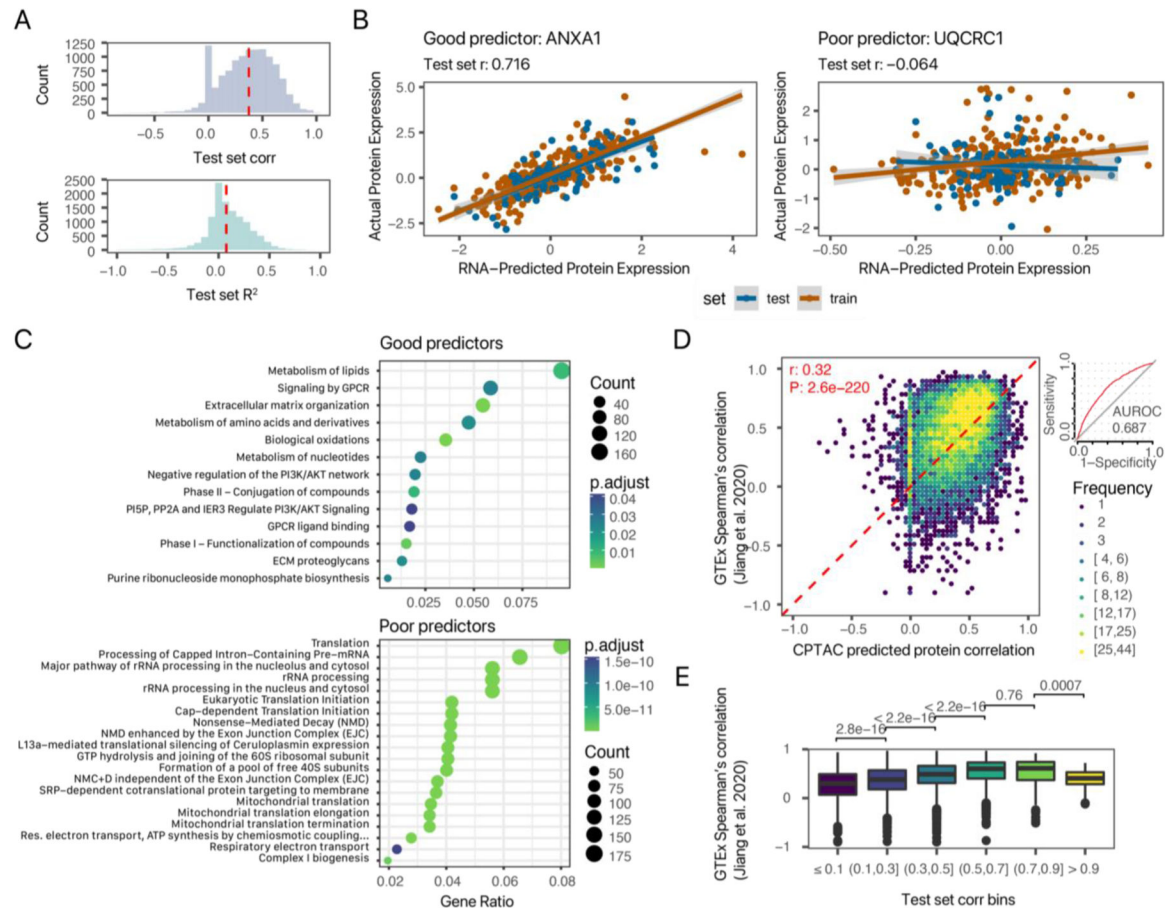


Figure 3: Predictability of protein-level from across-sample transcript variance.

To estimate whether the quantified transcript changes might translate to the proteome, we considered the predictability of protein levels from their proxy transcripts on a gene-wise basis in large proteogenomics data sets. **A**. An elastic net is applied to 717 samples with matching transcriptomics and mass spectrometry data in the CPTAC collection. The average correlation (top) and R^2 values (bottom) between predicted and actual protein levels across samples in each of 10,693 genes are shown. **B**. Examples of an aging signature whose protein abundance across samples is well predicted by its proxy transcript (*Anxa1*) in matching samples and one that is poorly predicted (*Uqcrc1*). Each data point is one CPTAC sample. Brown: train set; blue: test set. **C**. Significantly enriched Reactome terms among transcripts that predict protein well ($r \geq 0.5$) or poorly ($r < 0.5$). Size: protein count in pathway; fill color: adjusted P value; x-axis: gene annotation ratio. **D**. Scatter plot showing a robust correlation between the modeled gene-wise protein predictability here using CPTAC data with the Spearman's correlation coefficient values between protein and RNA across 32 tissues in GTEx (r : 0.32, P : $2.6e-220$). The modeled protein predictability values predict strong RNA-protein correlation ($\rho \geq 0.5$) in normal human GTEx tissues with an AUC of 0.69 (inset). **E**. Boxplot showing a breakdown of binned correlation values against GTEx correlation, the correlation plateaus at $r \geq 0.5$ which may be due to potential overfitting or cross-sample biological differences.

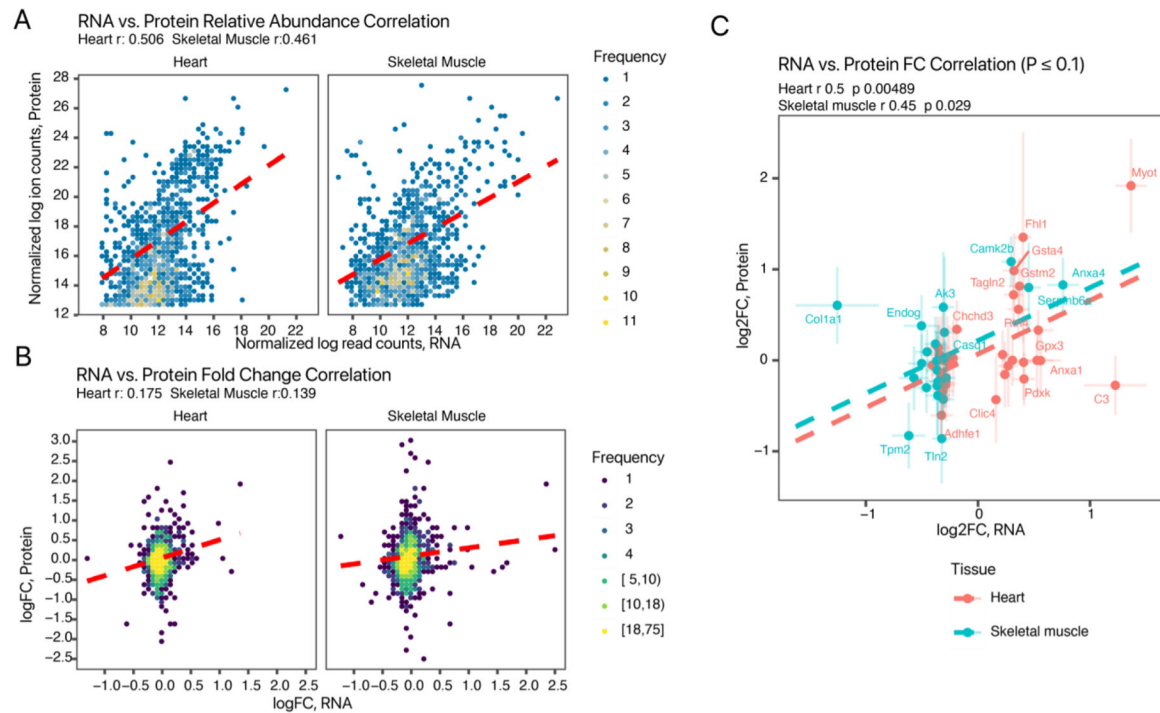


Figure 4: Correlation between RNA and protein levels from identical tissues.

A. Scatter plot showing the within-sample across-gene comparisons in the heart (left) and the skeletal muscle (right) for commonly quantified RNA and their proteins. Fill color: data frequency within bin. **B.** Scatter plot showing across-sample comparisons between RNA and proteins in the aging vs. young adult heart (left) and skeletal muscle (right). **C.** Log fold-change comparison at the RNA (x-axis) and protein (y-axis) level among commonly quantified proteins and transcripts with significant age-associated transcript level differences. Line: best-fit linear curves for the heart (red) and the skeletal muscle (blue). Error bars: standard errors of logFC.

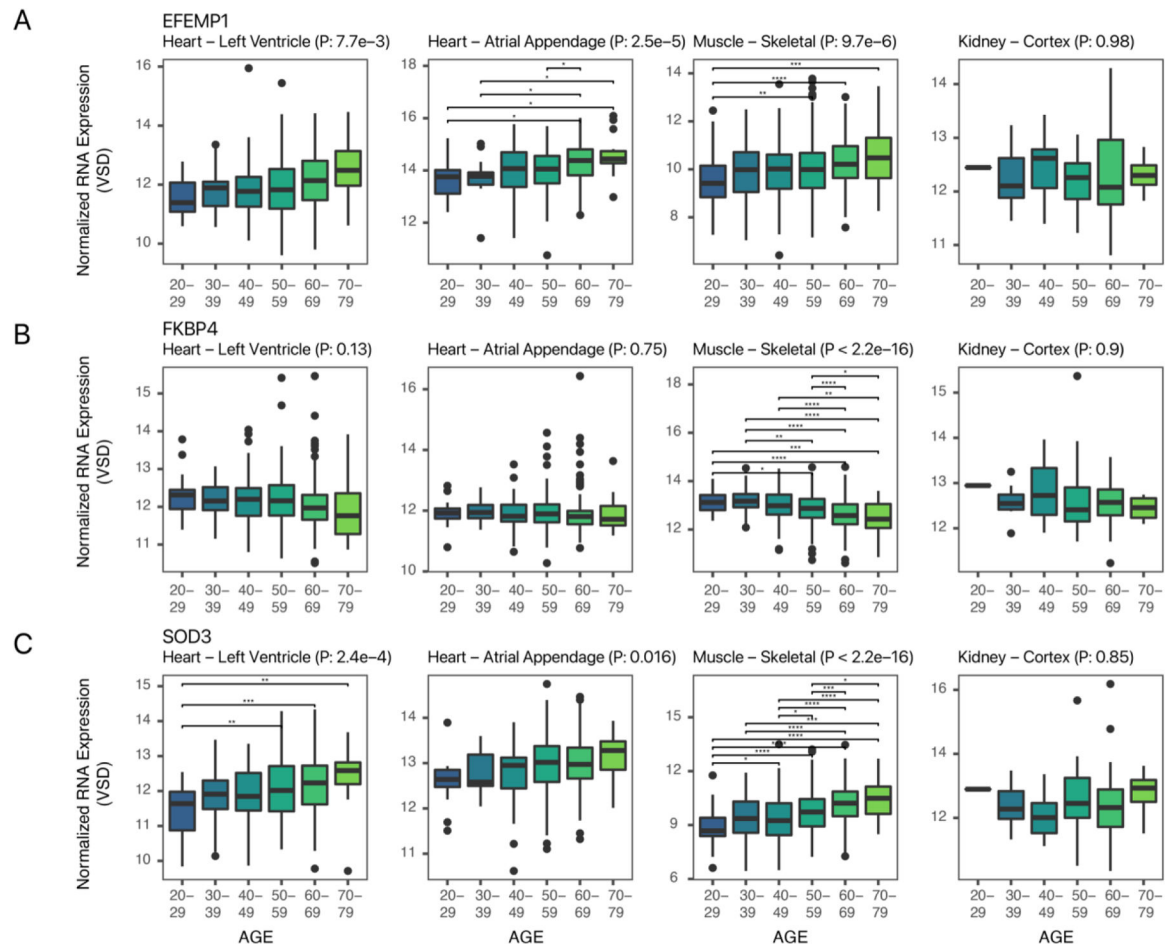


Figure 5: Conserved age-expression profiles of selected signatures in humans.

Box plots showing examples of GTEx v8 human normalized RNA expression levels across age groups in decadal brackets in four GTEx v8 human tissues (left to right) heart left ventricle, heart atrial appendage, skeletal muscle, and kidney cortex for **A. *Efemp1***, **B. *Fkbp4***, and **C. *Sod3***. P values: ANOVA. Asterisks within plots denote Tukey's post-hoc for individual group comparison. *: Tukey $P < 0.05$; **: $P \leq 0.01$; ***: $P \leq 0.001$; ****: $P \leq 0.0001$.

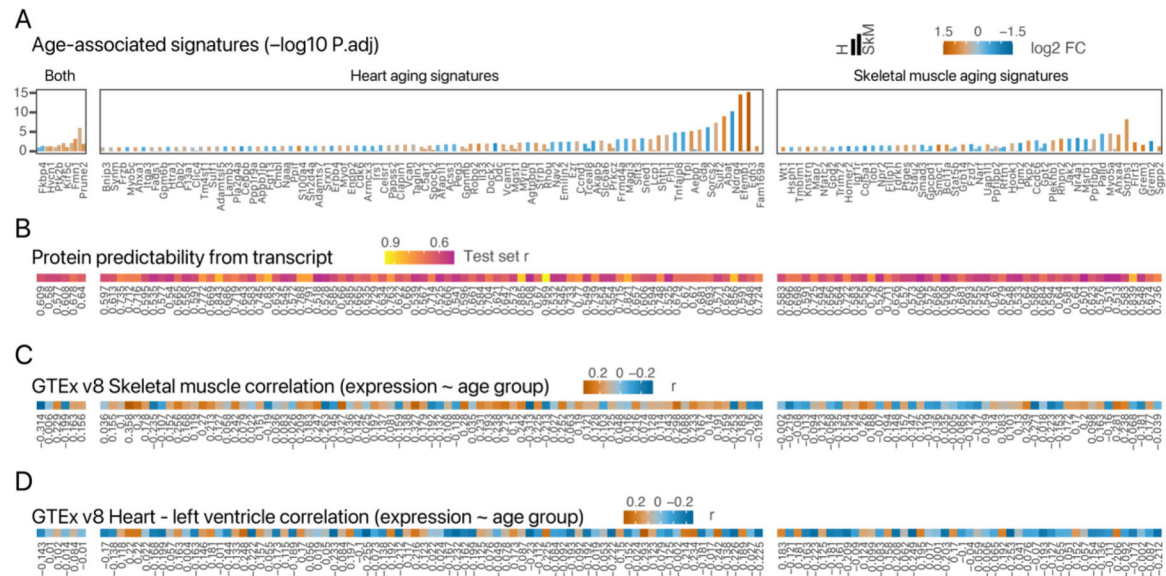


Figure 6: Prioritized age-associated signatures.

A. List of 134 prioritized aging signatures in the heart, skeletal muscle, and those common to both tissues. Top: paired bars represent $-\log_{10} P_{\text{adj}}$ in the heart and the skeletal muscles, respectively. **B-D.** The prioritized signatures had CPTAC RNA-protein correlation $r \geq 0.5$ and ANOVA $P \leq 0.01$ in GTEx v8 transcript expression against age groups in GTEx v8 heart left ventricle or skeletal muscle transcriptomes. X-axes correspond to panel A.

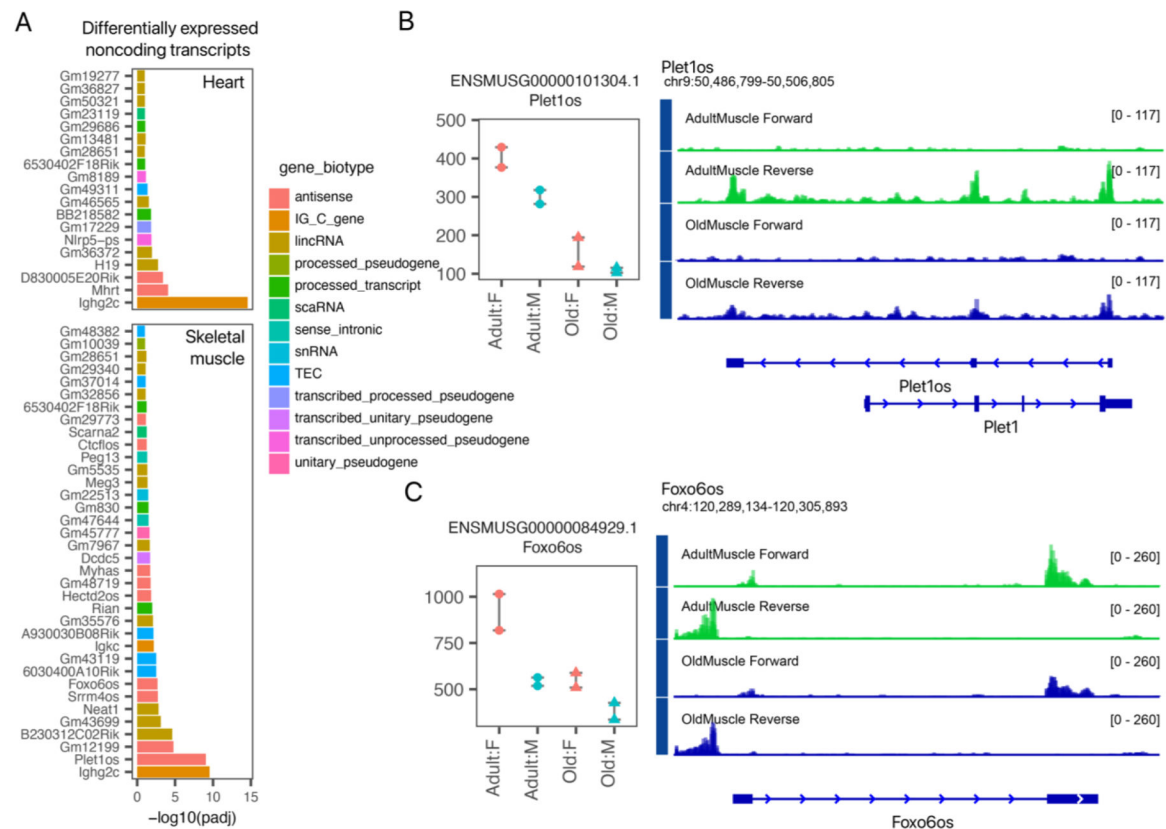


Figure 7. Non-coding RNA signatures in aging striated muscles.

A. Bar charts of differentially expressed annotated non-coding RNAs in the heart (top) and the skeletal muscle (bottom). Colors denote Gencode vM25 annotation gene biotype. **B-C.** Examples and genome tracks of two long non-coding RNAs *Plet1os* (**B**) and *Foxo6os* (**C**) that are differentially expressed in aging skeletal muscle.