Aging Skeletal Muscle Proteomics Finds Changes in Spliceosome, Immune factors, Proteostasis and Mitochondria

Ceereena Ubaida-Mohien[$], Alexey Lyashkov[$], Marta Gonzalez-Freire[$], Ravi Tharakan[$], Michelle Shardell[$], Ruin Moaddel[$], Richard D. Semba[#], Chee W. Chia[$], Myriam Gorospe[$], Ranjan Sen[$] & Luigi Ferrucci[$]*

Author Affiliations

[$]*Intramural Research Program, National Institute on Aging, National Institutes of Health, Baltimore, MD 21224, USA*

[#]*Johns Hopkins Medical Institute, Baltimore, MSD 21205, USA*

Contact Info

*Correspondence**: FerrucciLu@grc.nia.nih.gov

Abstract

A progressive decline of skeletal muscle strength with aging is a primary cause of mobility loss and frailty in older persons, but the molecular mechanisms of such decline are not fully understood. Here, using quantitative discovery proteomic data from skeletal muscle specimens collected from 58 healthy persons aged 20 to 87 years show that ribosomal proteins and proteins related to energetic metabolism, including those related to the TCA cycle, mitochondria respiration, and glycolysis were underrepresented in older persons. Proteins with important roles in innate and adaptive immunity, involved in proteostasis and regulation of alternative splicing were all overrepresented in muscle from older persons. Changes with aging of alternative splicing were confirmed by RNA-seq. Overall, older muscle has a profound deficit of energetic metabolism, a pro-inflammatory environment and increased proteostasis. Upregulation of the splicing machinery maybe an attempt to compensate for these changes and this could be tested in future studies.

Introduction

One of the most striking phenotypes of aging is the decline of skeletal muscle strength, which occurs in all aging individuals and contributes to the impairment of lower extremity performance and loss of mobility [1-3]. The magnitude of decline in strength is higher than that expected from the loss of muscle mass, suggesting that the contractile capacity of each unit of muscle mass is progressively lower with aging. The reasons for such a decline of contractile capacity are unclear, and several hypotheses have been proposed [4]. Studies conducted in humans by ^{31}P magnetic resonance spectroscopy as well as "ex vivo" respirometry have shown that skeletal muscle oxidative capacity declines with aging and such decline affects mobility performance [5-8]. Ample evidence from animal models, and more limited evidence from human studies also suggest that aging causes progressive muscle denervation, with enlargement of the motor units and degeneration of the neuromuscular junction, but whether these changes account for the change of contractile performance of human muscle with aging has not been studied [9-13].

Currently, no treatment is available to prevent or delay the decline of muscle strength and function with aging. Thus, understanding the mechanisms driving the decline in muscle contractile capacity with aging is essential to identify new targets of intervention. Previous studies attempted to address this question by performing cross-sectional untargeted proteomic analysis in skeletal muscle biopsy specimens from young and old individuals. However, these studies were limited in size, focused on cancer cachexia, analyzed single fibers, did not account for levels of physical activity or did not explore the effect of aging over its continuous range and, therefore, could not distinguish changes due to aging from those due to disease or sedentary state [14-18]. To overcome these earlier limitations, we have performed a quantitative mass spectrometry-based proteome analysis (tandem mass tag, TMT) of skeletal muscle biopsies obtained from individuals distributed over a wide age range, who were extremely healthy based on strict objective clinical criteria. We characterized proteins that were overrepresented and underrepresented in older individuals

and using these data we made inferences about molecular pathways affected by aging in skeletal muscle.

Results and Discussion

Quantitative Skeletal Muscle Proteome Analysis of Healthy Aging

Skeletal muscle biopsies were collected from 60 healthy participants of the Genetic and Epigenetic Study of Aging Laboratory Testing (GESTALT) aged 20 to 87 years who were defined as 'healthy' based on very strict evaluation criteria at the National Institute on Aging Clinical Unit in Baltimore [19]. Exclusion criteria included any diseases that required chronic treatment with the exclusion of mild hypertension fully controlled with one drug only, any physical or cognitive impairment, and any abnormal values in pre-defined list of blood clinical tests (see methods for details). Participants who consented for a muscle biopsy were homogeneously distributed across the age strata 20-34 (n=13), 35-49 (n=11), 50-64 (n=12), 65-79 (n=12) and 80+ (n=10), and biopsies were analyzed by tandem shotgun mass spectrometry-based quantitative proteomics method (Figure 1A, Table S1). Using multiplexed isobaric labelling tags (TMT) and a customized analytical strategy [20] [21] we identified 400,000 tryptic peptides from 6.7 million spectra (396 multiplexed MS runs from 12 TMT 6plex experiments), which allowed the quantification of 5,891 proteins (Table S2).

To control for batch variability and avoid bias, we included a reference sample in all 12 TMT sets. A loading normalization was implemented that assumed that the sums of all intensities from all the proteins across the samples in a single TMT experiment were equal and that the sample loading effects, peptide bias effects and the residual error were normally distributed across a constant variance across samples (Figure S1A). To test the effectiveness of these approaches, we examined TMT batch effects in several analyses, allowing for experiment-specific random effects. We then averaged the expression values from each TMT across the sample groups and found that the ranks between TMTs were highly correlated (Figure 1B-C,

Figure S1B). Together, these findings indicate that the protein quantification across the 12 TMT experiments was robust.

Of the initial 5,891 proteins detected, we excluded from the analysis 1,511 proteins that were not quantifiable in at least 3 participants per age strata (at least 15 participants total) and performed the analysis in the remaining 4,380 proteins detected in more than 15 donors (3 per age strata), which were quantified from 46,834 unique peptides and 2.7 million spectra (Figure 1D). We used Partial Least Squares (PLS) analysis to explore the overall clustering of the 4,380 proteins across age groups (Figure 1E). The age groups (color-coded) were well separated along at least one axis in the three-dimensional clustering classification (Figure S1C). As expected, most of the proteins identified were classified as "muscle proteins", and the top 10 most abundant muscle proteins accounted for 45% of the total spectral abundance. Low-abundant mitochondrial proteins, such as cardiac phospholamban (PLN), were also quantified.

Focus on the Aging Biological Mechanisms

The relationship of age with the skeletal muscle proteins was estimated by linear mixed regression models that included sex, race, level of physical activity, type I/type II muscle fiber ratio, body mass index (BMI) and TMT batch effect as covariates (method section). Of note, the age beta-coefficients (aging effect size) are small because they express the difference in protein "per year" of age. For example, the difference in protein between two individuals that differ by 20 years would be 20 times the size of the beta coefficient. We adjusted for physical activity because it both tends to decline with age and strongly affects biological processes in muscle cells [22-24]. Previous studies demonstrated that gender and race strongly affect body composition and muscle mass [25]. Skeletal muscle tissue includes different myofiber types: type I fibers (slow-twitch), type IIa fibers (fast-oxidative), and IIb fibers (fast glycolytic muscle fibers containing four different myosin isoforms), each supported by different energetic metabolism and with different protein composition. An

4

analysis for a proxy measure of "muscle fiber ratio" [26] was estimated by calculating the ratio of myosin 7 (MYH7), the slow-twitch fiber isoform, and the sum of fast-twitch fiber isoforms (MYH1, MYH2 and MYH4) (Figure S2A1-A4); as expected, the fiber ratio of slow/fast was higher with older age (Figure S2A5). The slight change of slow/fast fiber ratio was significant and outweighed the wide variability among individuals (p=0.005); BMI was adjusted because obese persons tend to have muscle fat infiltration and lower muscle quality and muscle-fat interaction may affect muscle composition and function [4,27]. Gender may also have an impact on protein expression in skeletal muscle, as males and females are known to have differences in muscle mass; however, because of the limited sample size, we did not stratify the analysis by gender. This analysis should be done in future larger studies.

Proteins were then deemed significantly underrepresented or overrepresented in older age based on p-values for age-coefficients in the regression equation, calculated from Satterthwaite's t-tests (Figure 2A; Figure S2B). There were 1265 proteins significantly associated with age (p <0.05, with BH correction <0.1, 917 proteins), suggesting that approximately 29% of the skeletal muscle proteome changed with aging after 20 years of age (Table S3). Of these, 29% (361) were significantly underrepresented and 71% (904) were overrepresented with older age. The age-associated analysis across the experimental dataset and across multiple comparisons was highly robust (Figure S2C).

Notably, the proteins most strongly associated with older age (Figure 2A, right) were LSM14 homolog A (LSm14A, β=0.023, p=0.0109), tissue metalloproteinase inhibitor 3 (TIMP-3, β=0.0219, p=0.00026) and serum amyloid P-component (APCS, β=0.0164, p=1.26E-11). Protein LSm14A is implicated in processing the assembly of processing bodies, involved in mRNA turnover, and can also bind to viral nucleic acids and initiate IFN-β production, contributing to innate immunity [28]. TIMP3 regulates the adipogenic differentiation of fibro/adipogenic progenitors (FAPs) in skeletal muscle, and its overrepresentation may explain the tendency for fat infiltration in aging muscle [29]. Consistent with this hypothesis, Perilipin 1 (PLIN1, β=0.014, p=0.0003), a lipid droplet-coating protein, and adipogenesis

5

regulatory factor (ADIRF, β=0.01173, p=1.78E-05), a protein that is only expressed in adipose tissue, were among the most overrepresented proteins in old muscle. APCS is indicative of systemic amyloid, and its overrepresentation in aging muscle has never been previously described.

The most underrepresented proteins (age β<-0.01 and p<0.05) with old age (Figure 2A, left) were HLA class II histocompatibility antigen (HLA-DRB1, p=0.024), dedicator of cytokinesis protein 4 (DOCK4, p=0.025), myosin-binding protein H (MYBPH, p=0.0005), and Microfibril-associated glycoprotein 4 (MFAP4, p=0.000002). Although HLA-DRB1 is the most altered protein, it is present only in 53% of the donors. MYBPH maintains the structural integrity of the muscle and its decreased expression has been associated with muscle weakness in age-related disorders [30].

To explore differences of protein expression profiles across the lifespan, we generated a heatmap of the 1,265 age-associated proteins and looked for clusters of proteins showing parallel changes with age (Figure 2B). Hierarchical clustering of protein expression suggested that the strongest difference was between young (20-34) and old (80+). There were small differences before the age of 50, but afterwards there was on average three log fold protein expression differences, and even more substantial differences after the age of 64. The separation of protein expression between three age groups (20-49, 50-64, and 65+) was confirmed by PLS analysis (Figure 2C). These findings are consistent with data showing that the age-associated decline of muscle strength is already detectable after the fourth decade of life and substantially accelerates after the age of 70 [31].

Next, we grouped all quantified proteins according to biological mechanisms associated with aging in skeletal muscle (Figure 2D.1). We also included a category for all contractile and architectural muscle proteins (named hereafter "muscle proteins"). Though the highest abundance proteins detected correspond to muscle proteins, the largest category were mitochondria proteins (15%). Each of the other categories represented <9% of total proteins.

Protein classes differed between those underrepresented and overrepresented with older age are summarized in (Figure 2D.2 and 2D.3) and are described in detail in subsequent sections. Specifically, proteins implicated in muscle contraction, muscle architecture, mitochondria metabolism, as well as ribosome function decreased with older age. By contrast, proteins related to genomic maintenance, transcriptional regulators, splicing, neuromuscular junction, proteostasis, senescence and immune function increased with age. Other smaller subcategories of proteins were also differentially abundant (Figure 2D.3).

Contractile, Architectural and Neuromuscular Junction Proteins (NMJ)

Since many proteins decreasing with age were contractile proteins, we classified these further by function. The top 95 proteins in this class are involved in the architectural and functional stabilization of the sarcomere, including sarcospan (SSPN, β=0.002, *p*=0.016) (Figure 2E), a dystrophin-associated protein complex important for muscle regeneration, actin-binding LIM domain and actin-binding protein 1 (LIMA1, β=0.003, *p*=0.009), a cytoskeleton-binding protein that stabilizes actin filaments, and plectin (PLEC, β=0.0007, *p*=0.036), a large cytoskeleton protein that preserves interactions within the acto-myosin complex. Increases in delta sarcoglycan (SGCD*,* β=0.0019, *p*=0.00004, and gamma sarcoglycan (SGCG*,* β=0.0016, *p*=0.0062) were consistent with mouse studies showing that dystrophin, sarcoglycan subcomplex γ- and δ-sarcoglycan were overexpressed with aging, perhaps a compensatory mechanism to avoid damage in the sarcomere during contraction or as biomarkers of continuous repair [32]. Interestingly, MAPT (tau, mostly expressed in neurons and involved in the assembly and stabilization of microtubules), was also significantly underrepresented in older muscle (Figure 2E). A crucial component of muscle function is the neuromuscular junction (NMJ), and since the abundance of all NMJ related proteins increased with age we examined the agrin signaling pathway of NMJ. Agrin (AGRN) and acetylcholine esterase (ACHE) increased with age but not significantly (Figure S2D). By contrast, the levels of Syne-1 which anchors both synaptic and non-synaptic myonuclei for proper neuron innervation and respiration increased with age (SYNE1 β=0.002, *p*=0.005) as

did beta-2-syntrophin, which is believed to be involved in acetylcholine receptor clustering (SNTB2, β=0.0029, *p*=0.0003).

Decline of Mitochondrial Proteins with Age

Because of the striking difference in abundance of mitochondrial and energy metabolism proteins with age, we studied these proteins by protein annotations using Uniprot keywords, GO ontology terms and extensive manual curation based on the most recent literature. The coverage of mitochondrial proteins quantified by our analysis compared to those described in the literature ranged from 92% for TCA proteins to 52% for proteins located in outer mitochondrial membrane [possibly due to incomplete tissue disruption [33,34]] (Figure 3A). The coverage of the bioenergetics and mitochondrial proteome in our dataset is similar to that reported by other authors [15,33]. Of the mitochondrial proteins identified, the abundance of 25% of them (173 proteins) changed with age, mostly (70%) declining with age. Notably, however, outer membrane proteins were more abundant (Figure 3B); for example, NADH-cytochrome b5 reductase 3 (CYB5R3), an NADH-dehydrogenase located in the outer membrane of ER and mitochondria, whose overexpression is known to mimic many effects of caloric restriction, was significantly overrepresented in older age (Figure S3A) [35,36]. The permanence of mitochondrial protein debris in aging muscle has been previously reported - attributed to defective autophagy, and through to cause activation of the inflammasome and a proinflammatory state [37].

Of the enzymatic mitochondrial proteins, 99 were respiratory chain proteins (Complex I-V and assembly complex proteins), and most of them declined with aging (28 proteins *p*<0.05; Figure 3C). Surprisingly, succinate dehydrogenase complex assembly factor 2 (SDHAF2), required for covalent FAD insertion into complex II, the electron transport chain, and the TCA cycle, were significantly overrepresented with older age (Figure 3C inset). The reason for this exception is unclear and if replicated in other analysis requires further work.

We then analyzed proteins from complex I to V and found that 16 proteins were significantly lower at older age (Figure 3C, Figure S3B). Among 41 proteins involved in energy production, most were underrepresented at older ages. Of 22 proteins quantified for TCA cycle, only malate dehydrogenase (MDH1), isocitrate dehydrogenase (IDH1), fumarate hydratase (FH) and succinate--CoA ligase (SUCLG1) (Figure S3C) were significantly lower at older ages. The decreasing levels of IDH-1 with age is unsurprising, as previous studies have shown a decrease in abundance of IDH-1 in older *C.elegans* [38]. IDH1 converts isocitrate to α-ketoglutarate by reducing NADP to NADPH in the process. In addition, to IDH1, NADP+ is also reduced to NADPH via the mitochondrial NAD(P)-malic enzyme (ME2) [39] and predominantly through NNT (NAD(P) transhydrogenase) and the pentose phosphate pathway. In our study, NNT (β=-0.003, *p*=0.001) significantly decreased with aging. Interestingly, the decrease in expression levels of both NNT and IDH1 with age, suggests a decreased capacity of the mitochondria to maintain proton gradients and results in oxidative damage. Further, NADK (NAD+ Kinase), which is highly regulated by the redox state of the cell and regulates NADP synthesis in vivo decreased with age (NADK2, β=-0.001, *p*=0.052). The changes in the NADP/NADPH levels influence cellular metabolism, calcium signaling and anti-inflammatory processes and regeneration of glutathione [40].

NAD$^+$ declines with age in several tissues and its metabolism has been implicated in the aging process and age-related pathologies including loss of skeletal muscle mass [41,42]. NAD$^+$ is synthesized *in vivo* predominantly via the salvage pathway and the *de novo* and Preiss-Handler pathways [43,44]. We specifically examined age differences in abundance of proteins from these pathways. We found that NAM-N-methyl transferase (NNMT, β=0.007, *p*=0.016), nicotinamide ribose kinases (NMRK1, β=-0.003, *p*=0.002), poly-ADP-ribose polymerases (PARP1, β=0.002, *p*=0.003) and CD73 (NT5E, β=0.004, *p*=0.056) were significantly increased with at older ages, while only NMRK1 decreased with age (Figure 3D). NAMPT, which converts NAM to NMN, was not significantly different with age while NMRK1, which converts NR to NMN, was significantly lower in the muscle of older participants. These

findings may explain the mechanism by which NMN tends to be lower in tissue from older compared to younger persons. Two additional mechanisms that may exacerbate the decline in NMN and NAD$^+$ with aging, namely the increased expression levels of CD73 that converts NMN to NR and the increase in expression levels of PARP1, which converts NAD$^+$ into NAM and ADP-ribose.

Changes of Proteins Involved in Genomic Maintenance and Cellular Senescence

Most genomic maintenance proteins increased in abundance with age, especially those involved in DNA damage recognition and repair, such as double-strand break repair protein (MRE11), X-ray repair cross-complementing protein 5 (XRCC5), and structural maintenance of chromosomes protein 1A (SMC1A) (Figure S4A). Prelamin-A/C (LMNA), Lamin-B1 (LMNB1) and Lamin-B2 (LMNB2), members of the LMN family of protein components of nuclear lamina that help maintain nuclear and genome architecture, were all overrepresented with older age (Figure S4B). Sirtuin 2 (SIRT2, β=-0.0013, p=0.032), implicated in genomic stability, metabolism and aging, was also found to be lower in older skeletal muscle (Figure S4C).

Forty proteins that in the literature have been implicated in cellular senescence were significantly overrepresented with age. These included extracellular superoxide dismutase (SOD3, β=0.005, p=0.000009) and Transgelin-2 (TAGLN2, β=0.005, p=0.0002), a potential oncogenic factor and senescence-associated protein, The proteins decreased with age were GOT1, MAP2K3 (β=-0.003 and -0.0021, respectively), and casein kinase II subunit alpha (CSNK2A1, β=-0.0016, p=0.014). Interestingly, in addition to regulating cell cycle, CSNK2A1 plays a central role in many other biological mechanisms, including apoptosis, which is suppressed in senescent cells (Figure S4D). These observations suggest that senescent cells from different possible origins (e.g., myocytes, adipocytes or fibroblasts) may accumulate in old muscle.

Implications of Proteins that Modulate Transcription and Splicing

Of all the 69 age-associated transcription regulatory proteins quantified, only 8 were underrepresented with older age, including kelch-like protein 31 (KLHL31, β=-0.0017, *p*=0.003), which is implicated in the maintenance of skeletal muscle structure, increases with muscle growth and prevents congenital myopathy in mice [45]. By contrast, myocyte-specific enhancer factor 2D (MEF2D, β=0.003, *p*=0.018), essential for myogenesis and muscle regeneration and regulator of KLHL31 production, increased with age [46]. Contrary to earlier reports, CTCF (β=0.009, *p*=0.026), a transcriptional activator and repressor protein that fine-tunes chromatin architecture, also increased with age (Figure 4A).

A major unexpected finding of our analysis was the strong increase in major spliceosome complex proteins with aging (Figure S5A). The spliceosome comprises five small nuclear RNAs (snRNAs), *U1*, *U2*, *U4*, *U5*, and *U6*, that form functional complexes with proteins to regulate alternative splicing, a process by which different exons of one pre-mRNA are variably combined to generate different proteins [47]. We found differential expression of many proteins widely distributed across the five spliceosome complexes and other spliceosome-associated protein factors essential for mRNA maturation and gene expression (Figure 4B). In particular, of the ~300 proteins and spliceosome-associated factors described [48] [49], we quantified 99 and 57 of them, respectively, were overrepresented in older muscle (Figure 4C). Overall, spliceosomal proteins increased by ~15% between the ages of 20 and 87 years (Figure 4D). Spliceosome components are actively rearranged during assembly, catalysis, disassembly and recycling, each step involving recruitment and recycling of several proteins [50]. To understand whether aging affects preferentially one of these biological steps, we categorized the spliceosomal complexes and snRNPs into E complex, A complex, and B complex (assembly complex, 37 proteins), Bact complex and C complex (catalysis complex, 7 proteins) and snRNPs (recycling, SART1 protein) (Figure 4E), but we found no evidence of proteins from a specific complex being more overrepresented with aging than proteins from other complexes (Figure S5B). LSm RNA-binding protein (LSM14A)

was the most overrepresented assembly protein, displaying a 20-fold increase with age. The overrepresentation of spliceosomal proteins, such as the pre-mRNA-processing-splicing factor 8 (PRPF8) (Figure 4E, inset), the key catalytic core and the largest and most conserved protein in the spliceosome, suggests that pre-RNA processing may be upregulated in older skeletal muscle.

A systematic change in the splicing machinery with older age was previously suggested by transcriptomic analyses skeletal muscle biopsies [51] and human peripheral blood leukocytes [52] of young and old individuals. In both studies, processing of mRNAs was the feature that best discriminated younger and older persons, suggesting that modulation of alternative splicing is one of the signatures of aging [53]. Although the mechanisms and consequences of the rise in splicing factors with aging are unknown, they may indicate either a dysregulation of the splicing apparatus or a shift toward increased splicing and/or altered splice isoform diversity with aging [54].

Age-Associated Alternative Splicing and Splicing Events

The marked rise in overrepresentation of splicing machinery proteins with aging prompted questions about its functional consequences. Emerging literature suggest that change in expression of splicing factors is a major determinant for selection of specific splicing variants and changes in splicing variants contributes to some aging phenotypes, including age-related diseases [55] [56]. We analyzed potential differences in mRNA splicing with age (see methods) using RNA-seq data that were available for most of the same specimens used for the proteomic study (n=53). Specifically, we studied a set of variations of the exon-intron structure, known as transcriptional events, namely Alternative First exon (AF), Skipped Exon (SE), Alternative Last exon (AL), Alternative 3' splice-site (A3), Alternative 5' splice-site (A5), Retained Intron (RI) and Mutually Exclusive Exons (MX) [57]. Donor-specific splicing index (PSI, which measures each isoform as a % of total isoforms) was calculated for each AS event in each sample and a linear mixed regression model was used to identify age-

associated PSIs for each splicing event. Analysis of 144,830 transcripts from RNA-seq datasets showed that around 3.7% of the skeletal muscle transcripts (5,459 transcripts, corresponding to 6,255 events) showed relative abundance changes with aging (Figure 5A; Table S4). Next, we calculated the frequency and distribution of splicing events with aging as well as the directionality of such changes and found that 2,714 events were significantly less frequent at older ages and 3,545 events significantly more frequent at older ages (Figure 5B; Table S5). The overall number of events increased slightly with older age but AS events, at least for the 6,255 AS events quantified, increased significantly with age (r2=0.33, p=6.001e-06) (Figure 5C).

We then investigated whether any specific class of skeletal muscle AS events was enriched in our age-association analysis compared to the list of splicing events described for human skeletal muscle in the Ensembl human transcriptome (Figure 5D). The rates of observed skeletal muscle events are very similar to those reported in the Ensembl transcriptome (Figure 5D) except for ME, A3, SE and AF. The largest difference was in the skipped exon (SE) class of events, where a higher percentage of transcripts were exon-skipped compared to Ensembl events, with 27% of all the skeletal muscle AS events of the exon skipping type. A previous study reported 35% of the erythroid genes show evidence of AF exons, indicating that alternative promoters and AF are widespread in the human genome and play a major role in regulating expression of select isoforms in a tissue-specific manner [58]. This finding is in line with our result of 36% AF in our skeletal muscle data.

We next examined whether AS events occurred in proteins connected with pathways that are known to be dysregulated with aging; interestingly, among the top fold enriched (FE) gene ontology (GO) biological processes associated with age, splicing changes were more frequent on those that negatively regulated IκB kinase/NF-κB signaling (FE=2.86, p=3.18E-04), and those that regulated mitophagy (autophagy of mitochondria; FE=3.71, p=2.23E-04) and fatty acid beta oxidation (FE=3.21, p=1.72E-04). The GO biological process with positive age-associated splicing events were mitochondrial morphogenesis (FE=5.15, p=8.98E-03),

13

response to mitochondrial depolarization (FE=4.93, *p*=2.46E-04), and endoplasmic reticulum calcium ion homeostasis (FE=4.48, *p*=2.31E-04). These data suggest that the upregulation of alternative splicing in skeletal muscle with aging may be reactive to change that occur with aging either by rising an inflammatory response or by activating damage-response mechanisms at a time when energy becomes scarce.

Among the 5,459 transcripts (from 3,791 genes) that were alternatively spliced with age, 4967 transcripts were protein-coding. We compared these genes with the age-associated proteins and found that 8.9% of the age-associated alternatively spliced transcripts (385) were reflected in protein changes (Figure 5E). This comparison of age-associated proteins and alternatively splicing mRNAs suggests that 30% (385) of the age-associated proteins undergo alternative splicing. Among this group, 64 proteins are involved in cellular organization or biogenesis (GO:007180), and proteins like tubulin (TUBB2B, TUBB), profilin 2 (PFN2) and actin-related protein 2/3 complex subunit 4 (ARPC4) are involved in the cytoskeletal regulation by Rho GTPase pathway. A further PANTHER database classification of these proteins shows an enrichment in categories like RNA/DNA binding, cytoskeletal, translational and ribosomal proteins (Figure 5E protein categories). Overall, these findings suggest that a large percentage of proteins that change with aging also undergo splicing variations, and this is especially true for mitochondrial proteins. The physiological reasons for these changes remain unknown.

Depletion of Ribosomal Proteins with Age

Similarly, to previous studies, we found that a large number of ribosomal proteins were differentially expressed with older age (Figure 2, Figure S6A) [59,60]. In particular, all the 60S and 40S ribosomal proteins were globally reduced in older muscle; exceptions included 60S ribosomal proteins L12 and L3 (RPL12, β=0.0008, *p*=0.024, RPL3, β=0.003, *p*=0.016), as well as H/ACA ribonucleoprotein complex subunit 4 (DKC1, β=0.002, *p*=0.034) and nucleolar protein 58 (NOP58, β=0.003, *p*=0.00007), which were overrepresented in old muscle.

14

RPL12, RPL3, and DKC1 play a role in viral mRNA translation, while NOP58 is important for ribosomal biogenesis (Figure S6A-C). Changes in ribosome proteins may signal a decline in protein synthesis with aging, which may lead to slow recycling and progressive damage accumulation in contractile proteins.

Differential Regulation of Proteins Related to Proteostasis in Aging

Cells rely on a complex proteostatic machinery to handle protein quality control, assembly, folding and elimination. These activities are essential for the recycling of damaged proteins or entire organelles and provide critical protection against damage during conditions of metabolic or oxidative stress. Extensive literature supports the decline of proteostasis with aging in animal models and in humans [61] [62]. Of the 239 detected proteins that has been related to proteostasis in the literature, 31% were altered with age ($p<0.05$, 24 underrepresented and 50 overrepresented with older age) (Figure 6A). Most proteins underrepresented with age were chaperones, including DnaJ homolog subfamily A (DNAJA1), also named heat shock protein (Hsp) 40 (β=-0.0021, p=0.003), Hsp27 (β=-0.004, p=0.0001), Hsp70 protein 8 (HSPA8, β=-0.002, p=2.34E-07) (Figure S7) as well as Hsp27 protein 1 (HSPB1) and protein 7 (HSPB7), Hsp10 protein 1 (HSPE1) and Hsp60 protein 1 (HSPD1). Excluding HSPD1, the decline of these proteins with aging was previously described [62-64]. Other differentially expressed proteostasis-related proteins, including PDIA6, NPM1, ANP32E, and DNAJC2, are also regulatory chaperones (Figure 6B).

The loss of chaperone function during aging may be compensated by an increase in autophagic activity (Figure 6C), as misfolded proteins must be removed and degraded through an alternative mechanism. Indeed, most proteostasis proteins overrepresented with aging were related to autophagy except HSPA8 and Eukaryotic translation initiation factor 4 gamma 1 (EIF4G1). For example, TDP-43, a DNA/RNA-binding protein that tends to form aggregates in tissues such as skeletal muscle and brain and is both removed by autophagy and involved in autophagy maintenance, increased significantly with aging (TARDBP,

15

β=0.002, *p*=0.0002) (Figure 6D). On the contrary, calreticulin (CALR), a quality control chaperone induced under ER stress that stimulated autophagy, was significantly higher in muscle of older participants (β=0.001, *p*=0.022) (Figure 6D) [65]. Of note, calreticulin is used by macrophages to tag cells to be removed by programmed cell phagocytosis [66]. Consistent with this finding, CALR is considered a main biomarker of age-related diseases and frailty [67].

Pro-inflammatory and Anti-Inflammatory Immune Proteins of Aging Muscle

Of the 32 immune-related age-associated proteins that were quantified (Figure 7A), three broad themes emerged from the aging muscle immune proteome (Figure 7B-D). First, many proteins previously linked to macrophage function (such as CD14, LGALS3, CAPG, INPPL1 and MAST2) were dysregulated in aging muscle, with skewing towards a pro-inflammatory phenotype. For example, the overrepresentation with aging of proteins such as Monocyte differentiation antigen CD14 (CD14, β=0.003, *p*=0.009), Interferon-induced, double-stranded RNA-activated protein kinase (E2AK2, β=0.0008 *p*=0.046) and ASC (PYCARD) (β=0.006, *p*=0.025) can be viewed as being pro-inflammatory via their proposed role in lipid sensing and NF-κB activation (Figure 7B.1) [68,69]. Interestingly, we also identified proteins that were concurrently downregulated, such as Microtubule-associated serine/threonine-protein kinase 2 (MAST2, β=-0.002, *p*=0.023) and Phosphatidylinositol 3,4,5-trisphosphate 5-phosphatase 2 (INPPL1, β=-0.0009, *p*=0.036), that could accentuate the inflammatory phenotype by attenuating the negative regulation of NF-κB (Figure 7B.2) [70,71]. Thus, increased expression of NF-κB activators and decreased expression of NF-κB attenuators may synergize to elevate chronic inflammation in aging muscle. We also noted increased expression of High mobility group protein B2 (HMGB2, β=0.004, *p*=0.001), a well-known 'alarmin' [72] that is released from dying cells or within neutrophil extracellular traps (NETs), that may further exacerbate the inflammatory milieu. Cumulatively, our observations are consistent with the emergence and/or enrichment of pro-inflammatory macrophages in aging human muscle.

Second, we found evidence of an anti-inflammatory activity that could potentially offset the pro-inflammatory milieu of aging muscle (Figure 7C). This was most evident in strong age-associated up-regulation of annexin A1 (ANXA1, β=0.008, p=0.00001), a protein that has been linked to resolution of inflammation [73]. Elevated levels of adiponectin (ADIPOQ, β=0.002, p=0.008), a chemokine produced exclusively by adipocytes, likely reflected increased adipogenic activity in aging muscle. However, it is interesting to note that ADIPOQ has also been proposed to inhibit endothelial NF-κB activation [74,75] and may, thereby, have context-dependent anti-inflammatory functions. Finally, erbin (ERBIN, β=0.002, p=0.019), a nuclear lamina-associated protein that was overrepresented with age in our studies, has been implicated in reducing NF-κB activation by some stimuli [76], with associated reduction in pro-inflammatory gene expression.

Third, coordinate up-regulation of several members of the alternate complement pathway, such as CFAH (β=0.003, p=0.028, and CFAD (β=0.003, p=0.039) and modulators of complement activity such as CD antigen CD55 (DAF and CD55, β=0.006, p=0.00007) indicate ongoing innate immune activity in aging muscle (Figure 7D). Whether this trend reflects increased presence of dying cells and cell debris or below-threshold autoimmune activity remains to be determined. The latter could be mediated, for example, by Immunoglobulin heavy constant gamma 4 (IGHG4, β=0.008, p=0.019) which we found to be increased with age. This antibody isotype has been implicated in the generation of autoantibodies against muscle-specific kinases that are prevalent in certain forms of myasthenia gravis [77]. The possible connection between the aging muscle and chronic neurodegenerative disorders in which destruction of self-tissue by complement has been ascribed a causative role [78] is an intriguing area for future investigation.

Conclusions

The biological mechanisms that mediate the deleterious effect of aging on skeletal muscle are still controversial, as some evidence suggested that the decline of mitochondrial content,

volume and energetic efficiency plays a primary role, while other evidence showed show no significant change for the same parameters with aging, especially if the level of physical activity was considered [79]. To investigate systematically the changes in expressed proteins that might drive the decline in skeletal muscle function, we conducted an in-depth quantitative measurement of age-related changes in protein abundance in human skeletal muscle. While we did not use model systems or *in vivo* experiments, because of the careful design of the study, the selection of an extraordinary healthy population, the depth of protein detection and rigorous analysis made it possible to produce a descriptive quantitative dataset to show aging-associated molecular changes. We used a MS-based isobaric relative quantitative approach for proteome analysis that provides broad coverage of the proteins of human skeletal muscle in very healthy individuals over a wide age range and we adjusted our analysis for potential confounders. The biological function of most of the protein reported in this study was gathered by an extensive review of the literature and instead of relying only on annotation of Uniprot or GO database, we manually curated the functional classification used in the analysis. We present evidence that our approach is robust and sensitive to true biological variability. We confirmed the altered expression of proteins implicated in pathways differentially active in human skeletal muscle with aging, including more highly abundant mitochondrial proteins and less abundant inflammatory proteins. We also identified subsets of proteins increasing with age that were not previously described, namely proteins implicated in alternative splicing and autophagy. Our work provides a rich resource to study the effect of aging on the human skeletal muscle proteome and sets the stage for future research on the mechanisms driving the age-associated decline in muscle function.

Methods

Study design and participants

Muscle biopsies analyzed in this study were collected from participants from the Genetic and Epigenetic Study of Aging and Laboratory Testing (GESTALT). Participants were enrolled in

GESTALT if they were free of major diseases, except for controlled hypertension or a history of cancer that had been clinically silent for at least 10 years, were not chronically on medications (except one antihypertensive drug), had no physical or cognitive impairments, had a BMI less than 30 kg/m^2, and did not train professionally. Inclusion criteria were gathered from information on medical history, physical exams, and blood test interpreted by a trained nurse practitioner [80]. Participants were evaluated at the Clinical Research Unit of the National Institute on Aging Intramural Research Program. Data and muscle specimens from 60 participants were available for this study. However, two participants were excluded because the muscle specimen provided was too small to obtain reliable proteomic data. Therefore, data from 58 participants dispersed over a wide age-range (20-34 y, n=13; 35-49 y, n=11; 50-64 y, n=12; 65-79 y, n=12; 80+ y, n=10) were used for this study. Anthropometric parameters were objectively assessed. The level of physical activity was determined using an interview-administered standardized questionnaire originally developed for the Health, Aging and Body Composition Study [81] and modeled after the Leisure-Time Physical Activity questionnaire [82]. Total participation time in moderate to vigorous physical activity per week was calculated by multiplying frequency by amount of time performed for each activity, summing all of the activities, then dividing by two to derive minutes of moderate to vigorous physical activity per week, the following categories were used: <30 minutes per week of high intensity physical activity was considered "not active" and coded as 0; high-intensity physical activity ≥30 and <75 minutes was considered "moderately active" and coded as 1, high-intensity physical activity ≥75 and <150 minutes was considered "active" and coded as 2, and high-intensity physical activity ≥150 minutes was considered to "highly active" and coded as 3. An ordinal variable from 0 to 3 was used in the analysis. The GESTALT protocol is approved by the Intramural Research Program of the US National Institute on Aging and the Institutional Review Board of the National Institute of Environmental Health Sciences. All participants provided written, informed consent at every visit.

Muscle biopsies

The depth of the subcutaneous fat (uncompressed and compressed) was determined using MRI images of the middle thigh performed on the previous day. A region above the vastus lateralis muscle was marked at the mid-point of a line drawn between the great trochanter and the mid-patella upper margin. The skin was prepped with povidone–iodine (Betadine®) and ethyl alcohol, and the outside areas covered with sterile drapes. The biopsy site was anesthetized intradermally using a 27-gauge needle and then subcutaneously using a 23-gauge x 1 1/2 -inch needle, follow by an 18-gauge spinal needle, with ~15 mL of 1% lidocaine with sodium bicarbonate. The operator was careful that the anesthetic was infiltrated in the subcutaneous tissue and above the muscle fascia but not the muscle fibers not to distort the tissue structure and induce a gene expression response. A 6-mm Bergstrom biopsy needle was inserted through the skin and fascia incision into the muscle, and muscle tissue samples were obtained using a standard method. Biopsy specimens cut into small sections were snap frozen in liquid nitrogen and subsequently stored at −80 °C until used for analyses.

Sample preparation and protein extraction

On average 8 mg of muscle tissue was pulverized in liquid nitrogen and mixed with the lysis buffer containing protease inhibitor cocktail (8 M Urea, 2M Thiourea, 4% CHAPS, 1% Triton X-100, 50 mM Tris, pH 8.5 (Sigma)). Protein concentration was determined using commercially available 2-D quant kit (GE Healthcare Life Sciences). Sample quality was confirmed using NuPAGE® protein gels stained with fluorescent SyproRuby protein stain (Thermo Fisher).

In order to remove detergents and lipids 300 μg of muscle tissue lysate were precipitated using standard methanol/chloroform extraction protocol [83]. Proteins were resuspended in concentrated urea buffer (8 M Urea, 2 M Thiourea, 150 mM NaCl (Sigma)), reduced with 50 mM DTT for 1 hour at 36°C and alkylated with 100 mM iodoacetamide for 1 hour at 36°C in

20

the dark. The concentrated urea was diluted 12 times with 50 mM ammonium bicarbonate buffer and proteins were digested for 18 hours at 36°C using trypsin/LysC mixture (Promega) in 1:50 (w/w) enzyme to protein ratio. Protein digests were desalted on 10 x 4.0 mm C18 cartridge (Restek, cat# 917450210) using Agilent 1260 Bio-inert HPLC system with the fraction collector. Purified peptides were speed vacuum-dried and stored at -80°C until further processing.

Tandem Mass Tags (TMT) labeling was used to perform quantitative proteomics. Each TMT labeling reaction contains 6 labels to be multiplexed in a single MS run. Donor IDs were blinded, and samples were randomized to prevent TMT bias. Each TMT 6-plex set included one donor from each of the 5 age groups and one reference sample. 5 muscle samples 100 µg each corresponding to 5 different age groups and one separately prepared master reference sample were labeled with 6-plex tandem mass spectrometry tags using a standard TMT labeling protocol (Thermo Fisher). 200 femtomole of bacterial beta-galactosidase digest (SCIEX) was spiked into each sample prior to TMT labeling to control for labeling efficiency and overall instrument performance. Labeled peptides from 6 different TMT channels were combined into one experiment and fractionated.

High-pH RPLC fractionation and concatenation strategy

High-pH RPLC fractionation was performed on Agilent 1260 bio-inert HPLC system using 3.9 mm X 5 mm XBridge BEH Shield RP18 XP VanGuard cartridge and 4.6 mm X 250 mm XBridge Peptide BEH C18 column (Waters). Solvent composition was as follows: 10 mM ammonium formate (pH 10) as mobile phase (A) and 10 mM ammonium formate and 90% ACN (pH 10) as mobile-phase B [84].

TMT labeled peptides prepared from the skeletal muscle tissues were separated using a linear organic gradient that went from 5% to 50% B in 100 min. Initially, 99 fractions were collected during each LC run at 1 min interval each. Three individual high-pH fractions were concatenated into 33 combined fractions with the 33 min interval between each fraction

(fractions 1, 34, 67=combined fraction 1, fractions 2, 35, 68=combined fraction 2 and so on). Combined fractions were speed vacuum dried, desalted and stored at -80°C until final LC-MS/MS analysis.

LC-MS/MS analyses

Purified peptide fractions from skeletal muscle tissues were analyzed using UltiMate 3000 Nano LC Systems coupled to the Q Exactive HF mass spectrometer (Thermo Scientific, San Jose, CA). Each fraction was separated on a 35 cm capillary column (3 um C18 silica, Hamilton, HxSil cat# 79139) with 150 um ID on a linear organic gradient using 650 nl/min flow rate. Gradient went from 5 to 35% B in 205 min. Mobile phases A and B consisted of 0.1% formic acid in water and 0.1% formic acid in acetonitrile, respectively. Tandem mass spectra were obtained using Q Exactive HF mass spectrometer with the heated capillary temperature +280°C and spray voltage set to 2.5 kV. Full MS1 spectra were acquired from 300 to 1500 m/z at 120000 resolution and 50 ms maximum accumulation time with automatic gain control [AGC] set to $3x10^6$. Dd-MS2 spectra were acquired using dynamic m/z range with fixed first mass of 100 m/z. MS/MS spectra were resolved to 30000 with 155 ms of maximum accumulation time and AGC target set to 2x105. Twelve most abundant ions were selected for fragmentation using 30% normalized high collision energy. A dynamic exclusion time of 40 s was used to discriminate against the previously analyzed ions.

Proteomics informatics

The mgf files generated (using MSConvert, ProteoWizard 3.0.6002) from the raw data from each sample fraction was searched with Mascot 2.4.1 and X!Tandem CYCLONE (2010.12.01.1) using the SwissProt Human sequences from Uniprot (Version Year 2015, 20,200 sequences, appended with 115 contaminants) database. The search engine was set with the following search parameters: TMT6plex lysine and n-terminus as fixed modifications and variable modifications of carbamidomethyl cysteine, deamidation of asparagine and glutamate, carbamylation of lysine and n-terminus and oxidized methionine. A peptide mass

tolerance of 20 ppm and 0.08 Da, respectively, and two missed cleavages were allowed for precursor and fragment ions in agreement with the known mass accuracy of the instrument. Mascot and X!Tandem search engine results were analyzed in Scaffold Q+ 4.4.6 (Proteome Software, Inc). The TMT channels' isotopic purity was corrected according to the TMT kit.peptide and protein probability was calculated by PeptideProphet [85] and ProteinProphet probability model [86]. The PeptideProphet model fits the peptide-spectrum matches into two distributions, one an extreme value distribution for the incorrect matches, and the other a normal distribution for correct matches. The protein was filtered at thresholds of 0.01% peptide FDR, 1% protein FDR and requiring a minimum of 1 unique peptide for protein identification. We allow single peptide hits for two reasons: first, any peptide that is quantifiable is detected across all samples (n=58); second, we identify proteins with more than one search engine, so the protein identification is confirmed at least twice, even for single-peptide hits. For these reasons the even single peptides are unlikely to be random hits. As for single peptide quantification, the spectrum-to-spectrum variability is no higher between spectra from the same peptide than between spectra from different peptides from the same protein. Therefore, it is unlikely that there is any differential 'bias' in reporter ions from peptide to peptide. More importantly, TMT is taken as relative, not absolute, quantification. So even if there were such a bias, it would be the same across samples, so the relative quantification would not be affected. Reporter ion quantitative values were extracted from Scaffold and decoy spectra, contaminant spectra and peptide spectra shared between more than one protein were removed. Typically, spectra are shared between proteins if the two proteins share most of their sequence, usually for protein isoforms. Reporter ions were retained for further analyses if they were exclusive to only one protein, and they were identified in all 6 channels across each TMT set. Since we have multiple age group across each TMT experiment, we analyzed the proteins for missing reporter ion intensity.

For this analysis the protein with missing reporter ion in some of the channels (not more than 2 channels) for each TMT experiment was identified and missing value imputation was performed using multiple imputation with chained equations (MICE) R library by predictive mean matching. Mean imputation was performed <0.01% in one or two TMT channels in most of the TMT experiments, except TMTset1 (the missing reporter ion for channel 5 is 0.03%). The reporter ion intensity from the proteins derived from the imputation method (on an average <10 proteins) were concatenated with reporter ion intensity identified in all 6 channels and further analysis performed using adjudicated values. The log2 transformed reporter ion abundance was normalized by median subtraction from all reporter ion intensity spectra belonging to a protein across all channels [20,21]. Relative protein abundance was estimated by median of all peptides for a protein combined. Protein sample loading effects from sample preparations were corrected by median polishing, i.e., subtracting the channel median from the relative abundance estimate across all channels to have a median zero as described elsewhere [20,21]. Quantified proteins were clustered together if they shared common peptides and corresponding gene names were assigned to each protein for simplicity and data representation. Annotation of the proteins were performed by manual curation and combining information from Uniprot, GO and PANTHER database. Further bioinformatics analysis was performed using R programming language (3.4.0) and the free libraries available on Bioconductor.

Linear mixed effect model and statistical analyses

Linear mixed regression model was implemented to examine age effects and the data was adjusted for physical activity, gender, race, bmi, type I and type II myosin fiber ratio and TMT mass spectrometry experiments. Protein significance from the regression model was determined with *p*-values derived from lmerTest. Partial Least Square analysis (PLS) was used to derive models with classification that maximized the variance between age groups. PLS loadings were derived from linear model adjusted protein results. The regression model was performed using R 3.3.4 [87] with lme4 v1.1. library. Heat maps and hierarchical cluster

analysis were performed using the non-linear minimization package in R [88]. GraphPad

PRISM 6.07 and R packages were used for statistical analysis and generation of figures.

STRING analysis (10.5 version) was used for obtaining protein-protein interaction network.

Enrichment analysis was performed by GeneSet Enrichment Analysis (GSEA) and

PANTHER, the pathways were mapped and visualized by Cytoscape 3.0 [89]. One-way

ANOVA, nonparametric, and chi-square tests (continuous and categorical variables) were

used to test for sample differences between age groups.

RNA extraction and purification

Total RNA was prepared by lysing cell pellets (2x106) in 700 μl Qiazol and extracted using

Qiagen miRNeasy mini kit according to the manufacturer's recommendation (Qiagen Inc,

CA, USA) from the same samples (n=54). Small ribosomal RNA was further depleted using

Qiagen GeneRead rRNA Depletion Nano Kit. Total RNA quality and quantity was checked

using RNA-6000 nano kits on the Agilent 2100-Bioanalyzer. 375 ng of high-quality RNA was

used for first-strand and second-strand cDNA synthesis followed by single primer isothermal

amplification (SPIA) using NuGEN Ovation RNA‑Seq System V2 kits according to

manufacturer's protocol. This kit amplified both polyA-tailed and non-polyA tailed RNA and

removed ribosomal RNA. The amplified cDNA was sheared using Bioruptor (Diagenode) to

an average size of 150-250 bases. The sequencing library was prepared using Illumina

ChIP-Seq kits according to the manufacturer's protocol (Illumina, San Diego, CA). In short,

the ends of the fragments were repaired using T4 DNA polymerase, E. coli DNA Pol I large

fragment (Klenow polymerase), and T4 polynucleotide kinase (PNK) and an A-overhang was

added to the 3' end. Adapters were ligated to the DNA fragments and size-selected (250-350

bases) on a 4.5% agarose gel. An 18-cycle PCR amplification was performed followed by a

second 4.5% agarose gel size selection before cluster generation in cbot2 and sequencing

with Illumina Hiseq2500 sequencer using V4 reagents. Single-read sequencing was

performed for 138 cycles and Real-Time Analysis (RTA) v1.18.66.3 generated the base-call

files (BCL files). BCL files were de-multiplexed and converted to standard FASTQ files using bcl2fastq program (v2.17.1.14)

RNA-Seq quantification and splicing analysis

The quality of the bases was checked using FASTQC program (v11.2) before and after adapter removal and last base trimming by cutadapt program (v1.9). The cleaned FASTQ files were aligned, quantified and annotated to the human hg19 genome using Salmon [90] with the concept of quasi-mapping with two phase inference procedure for gene model annotations. The GC bias corrected, quantified transcript isoform abundance values (TPM) were used for further computation of relative abundance of the events or transcripts isoforms known as percent spliced-in (PSI) by SUPPA [57]. Since the variability of low-expressed genes between biological replicates were reported, the transcript data were filtered for the transcripts which were expressed in at least three donors per each age group. Thus, we excluded ~23% of the transcripts from total transcript quantification for further splicing analysis. Events coordinates are extracted from the Ensembl annotation (GRCh37.75) and alternative splicing events were generated. PSI values of alternative splicing events for each transcript from each sample (n=53) were estimated and the PSI values showing a good agreement with the RNA seq data were kept for further analysis. The magnitude of the PSI change (differential splicing) across the age were calculated with a linear mixed model analysis performed on the PSI to estimate the age-related splicing changes of the transcript isoform. The PSI regression model was adjusted with the aging confounders as same as described above for protein regression model except fiber ratio. For transcript data we used RNA experiment batches as a random effect. The age beta coefficient for each alternative splicing event transcript PSI was reported as the magnitude to the splicing event-specific PSI change with age. Significance of the alternative splicing events was calculated by lmerTest and was reported if the observation had a *p*-value <0.05 at transcript level for age beta coefficient.

Age-association of proteins and transcripts

Proteins or transcripts either significantly upregulated or down regulated with age, present in 50% of the samples or at least in three samples for each age group, and significant ($p<0.05$) were considered as age-associated. Age-association was measured by linear mixed model adjusted for confounders of aging phenotype either in protein analysis or in RNAseq analysis and were further filtered for significance calculation. Age beta coefficient for each protein or transcript were calculated from log2 normalized data on which a mixed linear regression model was applied. Thus, the age beta coefficient represents the mean log2 fold expression per year of age. LmerTest was used for calculating p-values from t-tests via Satterthwaite's degrees of freedom method. Any protein or transcript was represented as age-associated if the p-value for the protein or transcript was <0.05. P-values for multiple comparisons were adjusted using Benjamini-Hochberg method in R and adjusted p-values were reported on supplemental tables. Age-associated proteins and age-associated alternatively spliced transcripts were further analyzed into two categories, either age-association beta coefficient (<0) was under represented with age--indicating a decrease in the abundance of the protein with a year of age or age-association beta coefficient (>0) was over represented with age--indicating the abundance of the protein was increased with a year of age. For simplicity of reporting, we calculated the enrichment of these proteins/transcripts over the total age-associated protein/transcripts and reported as underrepresented and overrepresented with age.

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD011967. RNASeq data is deposited in GEO (GSE129643).

Supplementary Information

Document S1. Supplemental Experimental Procedures, Table S1 and Figure S1- S7

27

Table S1. Characteristics of Participants (Related to Figure1)

Table S2. Complete Dataset of Skeletal Muscle Proteome Quantified by TMT6plex.

Related to Figure1. Sheet1: Raw data of all the proteins quantified. Sheet2. Description of the column headers for the sheet1.

Table S3 to Figure2. Dysregulated Proteins with Age. Sheet1. Age-associated proteins. Proteins which were significantly (p<0.05) dysregulated with age. Sheet2. Description of the column headers for the sheet1

Table S4 to Figure 5. Sheet1. Age-associated splicing events (6255 events). Sheet2. Description of the column headers for the sheet1

Table S5 to Figure 5. Sheet1. Age-associated negative splicing events. Sheet 2. Age associated positive splicing events. Sheet3. Description of the column headers for the sheet1 and sheet2

Author Contributions

Conceptualization, CU, RS and LF; Methodology, CU, CWC, RM, AL, MS and LF; Formal Analysis; CU, RT, MS, RS and LF; Investigation, CU, RM, CWC, AL, MGF, RT, RDS, RS

and LF; Writing – Original Draft, CU, AL, RM, MGF, RT, MS, CWC, RDS, MG, RS and LF;

Resources, CWC and LF; Supervision, LF.

Competing Interests

The authors declare that the research was conducted in the absence of any commercial or

financial relationships that could be construed as a potential conflict of interest.

References

1   Cruz-Jentoft, A. J., Landi, F., Topinkova, E. & Michel, J. P. Understanding sarcopenia as a geriatric syndrome. *Curr Opin Clin Nutr Metab Care* **13**, 1-7, doi:10.1097/MCO.0b013e328333c1c1 (2010).

2   Studenski, S. A. *et al.* The FNIH sarcopenia project: rationale, study description, conference recommendations, and final estimates. *J Gerontol A Biol Sci Med Sci* **69**, 547-558, doi:10.1093/gerona/glu010 (2014).

3   Cesari, M. *et al.* Sarcopenia-related parameters and incident disability in older persons: results from the "invecchiare in Chianti" study. *J Gerontol A Biol Sci Med Sci* **70**, 457-463, doi:10.1093/gerona/glu181 (2015).

4   Moore, A. Z. *et al.* 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, doi:10.1111/jgs.12653 (2014).

5   Gianni, P., Jan, K. J., Douglas, M. J., Stuart, P. M. & Tarnopolsky, M. A. Oxidative stress and the mitochondrial theory of aging in human skeletal muscle. *Exp Gerontol* **39**, 1391-1400, doi:10.1016/j.exger.2004.06.002 (2004).

6   Hepple, R. T. Impact of aging on mitochondrial function in cardiac and skeletal muscle. *Free Radic Biol Med* **98**, 177-186, doi:10.1016/j.freeradbiomed.2016.03.017 (2016).

7   Zane, A. C. *et al.* Muscle strength mediates the relationship between mitochondrial energetics and walking performance. *Aging Cell* **16**, 461-468, doi:10.1111/acel.12568 (2017).

8   Gonzalez-Freire, M. *et al.* Skeletal muscle ex vivo mitochondrial respiration parallels decline in vivo oxidative capacity, cardiorespiratory fitness, and muscle strength: The Baltimore Longitudinal Study of Aging. *Aging Cell* **17**, doi:10.1111/acel.12725 (2018).

9   Wang, Z. M., Zheng, Z., Messi, M. L. & Delbono, O. Extension and magnitude of denervation in skeletal muscle from ageing mice. *J Physiol* **565**, 757-764, doi:10.1113/jphysiol.2005.087601 (2005).

10   Messi, M. L. *et al.* Resistance Training Enhances Skeletal Muscle Innervation Without Modifying the Number of Satellite Cells or their Myofiber Association in Obese Older Adults. *J Gerontol A Biol Sci Med Sci* **71**, 1273-1280, doi:10.1093/gerona/glv176 (2016).

11   Delbono, O. Neural control of aging skeletal muscle. *Aging Cell* **2**, 21-29, doi:10.1046/j.1474-9728.2003.00011.x (2003).

12   Spendiff, S. *et al.* Denervation drives mitochondrial dysfunction in skeletal muscle of octogenarians. *J Physiol* **594**, 7361-7379, doi:10.1113/JP272487 (2016).

13   Gonzalez-Freire, M., de Cabo, R., Studenski, S. A. & Ferrucci, L. The Neuromuscular Junction: Aging at the crossroad between nerves and muscle. *Front Aging Neurosci* **6**, 208, doi:10.3389/fnagi.2014.00208 (2014).

14   Doran, P., Donoghue, P., O'Connell, K., Gannon, J. & Ohlendieck, K. Proteomics of skeletal muscle aging. *Proteomics* **9**, 989-1003, doi:10.1002/pmic.200800365 (2009).

15   Murgia, M. *et al.* Single muscle fiber proteomics reveals fiber-type-specific features of human muscle aging. *Cell Rep* **19**, 2396-2409, doi:10.1016/j.celrep.2017.05.054 (2017).

16   Waldera-Lupa, D. M. *et al.* Proteome-wide analysis reveals an age-associated cellular phenotype of in situ aged human fibroblasts. *Aging (Albany NY)* **6**, 856-878, doi:10.18632/aging.100698 (2014).

17   Brocca, L. *et al.* Structure and function of human muscle fibres and muscle proteome in physically active older men. *J Physiol* **595**, 4823-4844, doi:10.1113/JP274148 (2017).

18   Ebhardt, H. A. *et al.* Comprehensive proteome analysis of human skeletal muscle in cachexia and sarcopenia: a pilot study. *J Cachexia Sarcopenia Muscle* **8**, 567-582, doi:10.1002/jcsm.12188 (2017).

19   Tanaka, T. *et al.* Plasma proteomic signature of age in healthy humans. *Aging Cell*, e12799, doi:10.1111/acel.12799 (2018).

20   Kammers, K., Cole, R. N., Tiengwe, C. & Ruczinski, I. Detecting significant changes in protein abundance. *EuPA Open Proteom* **7**, 11-19, doi:10.1016/j.euprot.2015.02.002 (2015).

21   Herbrich, S. M. *et al.* Statistical inference from multiple iTRAQ experiments without using common reference standards. *J Proteome Res* **12**, 594-604, doi:10.1021/pr300624g (2013).

22   Egan, B. & Zierath, J. R. Exercise metabolism and the molecular regulation of skeletal muscle adaptation. *Cell Metab* **17**, 162-184, doi:10.1016/j.cmet.2012.12.012 (2013).

23   McGee, S. L. & Walder, K. R. Exercise and the Skeletal Muscle Epigenome. *Cold Spring Harb Perspect Med* **7**, doi:10.1101/cshperspect.a029876 (2017).

24   Bauman, A., Merom, D., Bull, F. C., Buchner, D. M. & Fiatarone Singh, M. A. Updating the Evidence for Physical Activity: Summative Reviews of the Epidemiological Evidence, Prevalence, and Interventions to Promote "Active Aging". *Gerontologist* **56 Suppl 2**, S268-280, doi:10.1093/geront/gnw031 (2016).

25   Gallagher, D. *et al.* Appendicular skeletal muscle mass: effects of age, gender, and ethnicity. *J Appl Physiol (1985)* **83**, 229-239, doi:10.1152/jappl.1997.83.1.229 (1997).

26   Schiaffino, S. & Reggiani, C. Fiber types in mammalian skeletal muscles. *Physiol Rev* **91**, 1447-1531, doi:10.1152/physrev.00031.2010 (2011).

27   Silva, D. A. S. & Martins, P. C. Impact of physical growth, body adiposity and lifestyle on muscular strength and cardiorespiratory fitness of adolescents. *J Bodyw Mov Ther* **21**, 896-901, doi:10.1016/j.jbmt.2017.01.007 (2017).

28   Li, Y. *et al.* LSm14A is a processing body-associated sensor of viral nucleic acids that initiates cellular antiviral response in the early phase of viral infection. *Proc Natl Acad Sci U S A* **109**, 11770-11775, doi:10.1073/pnas.1203405109 (2012).

29   Kopinke, D., Roberson, E. C. & Reiter, J. F. Ciliary Hedgehog Signaling Restricts Injury-Induced Adipogenesis. *Cell* **170**, 340-351 e312, doi:10.1016/j.cell.2017.06.035 (2017).

30   Hundley, A. F., Yuan, L. & Visco, A. G. Skeletal muscle heavy-chain polypeptide 3 and myosin binding protein H in the pubococcygeus muscle in patients with and without pelvic organ prolapse. *Am J Obstet Gynecol* **194**, 1404-1410, doi:10.1016/j.ajog.2006.01.049 (2006).

31   Dodds, R. M. *et al.* Grip strength across the life course: normative data from twelve British studies. *PLoS One* **9**, e113637, doi:10.1371/journal.pone.0113637 (2014).

32   Hughes, D. C., Wallace, M. A. & Baar, K. Effects of aging, exercise, and disease on force transfer in skeletal muscle. *Am J Physiol Endocrinol Metab* **309**, E1-E10, doi:10.1152/ajpendo.00095.2015 (2015).

33   Morgenstern, M. *et al.* Definition of a High-Confidence Mitochondrial Proteome at Quantitative Scale. *Cell Rep* **19**, 2836-2852, doi:10.1016/j.celrep.2017.06.014 (2017).

34   Zhao, Q. *et al.* 1-Dodecyl-3-methylimidazolium chloride-assisted sample preparation method for efficient integral membrane proteome analysis. *Anal Chem* **86**, 7544-7550, doi:10.1021/ac5013267 (2014).

35   Martin-Montalvo, A. *et al.* Cytochrome b5 reductase and the control of lipid metabolism and healthspan. *NPJ Aging Mech Dis* **2**, 16006, doi:10.1038/npjamd.2016.6 (2016).

36   Diaz-Ruiz, A. *et al.* Overexpression of CYB5R3 and NQO1, two NAD(+) -producing enzymes, mimics aspects of caloric restriction. *Aging Cell*, e12767, doi:10.1111/acel.12767 (2018).

37   Ferrucci, L. & Fabbri, E. Inflammageing: chronic inflammation in ageing, cardiovascular disease, and frailty. *Nat Rev Cardiol* **15**, 505-522, doi:10.1038/s41569-018-0064-2 (2018).

38   Copes, N. *et al.* Metabolome and proteome changes with aging in Caenorhabditis elegans. *Exp Gerontol* **72**, 67-84, doi:10.1016/j.exger.2015.09.013 (2015).

39   Sauer, U., Canonaco, F., Heri, S., Perrenoud, A. & Fischer, E. The soluble and membrane-bound transhydrogenases UdhA and PntAB have divergent functions in NADPH metabolism of Escherichia coli. *J Biol Chem* **279**, 6613-6619, doi:10.1074/jbc.M311657200 (2004).

40   Sheeran, F. L., Rydstrom, J., Shakhparonov, M. I., Pestov, N. B. & Pepe, S. Diminished NADPH transhydrogenase activity and mitochondrial redox regulation in human failing myocardium. *Biochim Biophys Acta* **1797**, 1138-1148, doi:10.1016/j.bbabio.2010.04.002 (2010).

41   Fang, E. F. *et al.* NAD(+) Replenishment Improves Lifespan and Healthspan in Ataxia Telangiectasia Models via Mitophagy and DNA Repair. *Cell Metab* **24**, 566-581, doi:10.1016/j.cmet.2016.09.004 (2016).

42   Goody, M. F. & Henry, C. A. A need for NAD+ in muscle development, homeostasis, and aging. *Skelet Muscle* **8**, 9, doi:10.1186/s13395-018-0154-1 (2018).

43   Verdin, E. NAD(+) in aging, metabolism, and neurodegeneration. *Science* **350**, 1208-1213, doi:10.1126/science.aac4854 (2015).

44   Bogan, K. L. & Brenner, C. Nicotinic acid, nicotinamide, and nicotinamide riboside: a molecular evaluation of NAD+ precursor vitamins in human nutrition. *Annu Rev Nutr* **28**, 115-130, doi:10.1146/annurev.nutr.28.061807.155443 (2008).

45   Papizan, J. B. *et al.* Deficiency in Kelch protein Klhl31 causes congenital myopathy in mice. *J Clin Invest* **127**, 3730-3740, doi:10.1172/JCI93445 (2017).

46   Schiaffino, S., Dyar, K. A. & Calabria, E. Skeletal muscle mass is controlled by the MRF4-MEF2 axis. *Curr Opin Clin Nutr Metab Care* **21**, 164-167, doi:10.1097/MCO.0000000000000456 (2018).

47   Papasaikas, P. & Valcarcel, J. The Spliceosome: The Ultimate RNA Chaperone and Sculptor. *Trends Biochem Sci* **41**, 33-45, doi:10.1016/j.tibs.2015.11.003 (2016).

48   Zhou, Z., Licklider, L. J., Gygi, S. P. & Reed, R. Comprehensive proteomic analysis of the human spliceosome. *Nature* **419**, 182-185, doi:10.1038/nature01031 (2002).

49   Rappsilber, J., Ryder, U., Lamond, A. I. & Mann, M. Large-scale proteomic analysis of the human spliceosome. *Genome Res* **12**, 1231-1245, doi:10.1101/gr.473902 (2002).

50   Wahl, M. C., Will, C. L. & Luhrmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**, 701-718, doi:10.1016/j.cell.2009.02.009 (2009).

51   Giresi, P. G. *et al.* Identification of a molecular signature of sarcopenia. *Physiol Genomics* **21**, 253-263, doi:10.1152/physiolgenomics.00249.2004 (2005).

52   Harries, L. W. *et al.* Human aging is characterized by focused changes in gene expression and deregulation of alternative splicing. *Aging Cell* **10**, 868-878, doi:10.1111/j.1474-9726.2011.00726.x (2011).

53   Latorre, E. & Harries, L. W. Splicing regulatory factors, ageing and age-related disease. *Ageing Res Rev* **36**, 165-170, doi:10.1016/j.arr.2017.04.004 (2017).

54   Welle, S., Brooks, A. I., Delehanty, J. M., Needler, N. & Thornton, C. A. Gene expression profile of aging in human muscle. *Physiol Genomics* **14**, 149-159, doi:10.1152/physiolgenomics.00049.2003 (2003).

55   Deschenes, M. & Chabot, B. The emerging role of alternative splicing in senescence and aging. *Aging Cell* **16**, 918-933, doi:10.1111/acel.12646 (2017).

56   Mazin, P. *et al.* Widespread splicing changes in human brain development and aging. *Mol Syst Biol* **9**, 633, doi:10.1038/msb.2012.67 (2013).

57   Alamancos, G. P., Pages, A., Trincado, J. L., Bellora, N. & Eyras, E. Leveraging transcript quantification for fast computation of alternative splicing profiles. *RNA* **21**, 1521-1531, doi:10.1261/rna.051557.115 (2015).

58   Tan, J. S., Mohandas, N. & Conboy, J. G. High frequency of alternative first exons in erythroid genes suggests a critical role in regulating gene function. *Blood* **107**, 2557-2561, doi:10.1182/blood-2005-07-2957 (2006).

59   Steffen, K. K. & Dillin, A. A Ribosomal Perspective on Proteostasis and Aging. *Cell Metab* **23**, 1004-1012, doi:10.1016/j.cmet.2016.05.013 (2016).

60   Kirby, T. J. *et al.* Blunted hypertrophic response in aged skeletal muscle is associated with decreased ribosome biogenesis. *J Appl Physiol (1985)* **119**, 321-327, doi:10.1152/japplphysiol.00296.2015 (2015).

61   Kaushik, S. & Cuervo, A. M. Proteostasis and aging. *Nat Med* **21**, 1406-1415, doi:10.1038/nm.4001 (2015).

62   Charmpilas, N., Kyriakakis, E. & Tavernarakis, N. Small heat shock proteins in ageing and age-related diseases. *Cell Stress Chaperones* **22**, 481-492, doi:10.1007/s12192-016-0761-x (2017).

63   Senf, S. M. Skeletal muscle heat shock protein 70: diverse functions and therapeutic potential for wasting disorders. *Front Physiol* **4**, 330, doi:10.3389/fphys.2013.00330 (2013).

64   Matsumoto, T. *et al.* Small Heat Shock Protein Beta-1 (HSPB1) Is Upregulated and Regulates Autophagy and Apoptosis of Renal Tubular Cells in Acute Kidney Injury. *PLoS One* **10**, e0126229, doi:10.1371/journal.pone.0126229 (2015).

65   Yang, Y. *et al.* The ER-localized Ca(2+)-binding protein calreticulin couples ER stress to autophagy by associating with microtubule-associated protein 1A/1B light chain 3. *J Biol Chem* **294**, 772-782, doi:10.1074/jbc.RA118.005166 (2019).

66   Krysko, D. V., Ravichandran, K. S. & Vandenabeele, P. Macrophages regulate the clearance of living cells by calreticulin. *Nat Commun* **9**, 4644, doi:10.1038/s41467-018-06807-9 (2018).

67   Cardoso, A. L. *et al.* Towards frailty biomarkers: Candidates from genes and pathways regulated in aging and age-related diseases. *Ageing Res Rev* **47**, 214-277, doi:10.1016/j.arr.2018.07.004 (2018).

68   Bryan, N. B., Dorfleutner, A., Rojanasakul, Y. & Stehlik, C. Activation of inflammasomes requires intracellular redistribution of the apoptotic speck-like protein containing a caspase recruitment domain. *J Immunol* **182**, 3173-3182, doi:10.4049/jimmunol.0802367 (2009).

69   Sarkar, A. *et al.* ASC directs NF-kappaB activation by regulating receptor interacting protein-2 (RIP2) caspase-1 interactions. *J Immunol* **176**, 4979-4986 (2006).

70   Kalesnikoff, J. *et al.* SHIP negatively regulates IgE + antigen-induced IL-6 production in mast cells by inhibiting NF-kappa B activity. *J Immunol* **168**, 4737-4746 (2002).

71   Tridandapani, S., Wang, Y., Marsh, C. B. & Anderson, C. L. Src homology 2 domain-containing inositol polyphosphate phosphatase regulates NF-kappa B-mediated gene transcription by phagocytic Fc gamma Rs in human myeloid cells. *J Immunol* **169**, 4370-4378 (2002).

72   Taniguchi, N., Kawakami, Y., Maruyama, I. & Lotz, M. HMGB proteins and arthritis. *Hum Cell* **31**, 1-9, doi:10.1007/s13577-017-0182-x (2018).

73   Gobbetti, T. & Cooray, S. N. Annexin A1 and resolution of inflammation: tissue repairing properties and signalling signature. *Biol Chem* **397**, 981-993, doi:10.1515/hsz-2016-0200 (2016).

74   Ehsan, M. *et al.* Adiponectin limits monocytic microparticle-induced endothelial activation by modulation of the AMPK, Akt and NFkappaB signaling pathways. *Atherosclerosis* **245**, 1-11, doi:10.1016/j.atherosclerosis.2015.11.024 (2016).

75   Chen, Y. *et al.* Adiponectin Inhibits TNF-alpha-Activated PAI-1 Expression Via the cAMP-PKA-AMPK-NF-kappaB Axis in Human Umbilical Vein Endothelial Cells. *Cell Physiol Biochem* **42**, 2342-2352, doi:10.1159/000480006 (2017).

76   McDonald, C. *et al.* A role for Erbin in the regulation of Nod2-dependent NF-kappaB signaling. *J Biol Chem* **280**, 40301-40309, doi:10.1074/jbc.M508538200 (2005).

77   Hoch, W. *et al.* Auto-antibodies to the receptor tyrosine kinase MuSK in patients with myasthenia gravis without acetylcholine receptor antibodies. *Nat Med* **7**, 365-368, doi:10.1038/85520 (2001).

78   McGeer, P. L., Lee, M. & McGeer, E. G. A review of human diseases caused or exacerbated by aberrant complement activation. *Neurobiol Aging* **52**, 12-22, doi:10.1016/j.neurobiolaging.2016.12.017 (2017).

79   Distefano, G. & Goodpaster, B. H. Effects of Exercise and Aging on Skeletal Muscle. *Cold Spring Harb Perspect Med* **8**, doi:10.1101/cshperspect.a029785 (2018).

80   Schrack, J. A., Knuth, N. D., Simonsick, E. M. & Ferrucci, L. "IDEAL" aging is associated with lower resting metabolic rate: the Baltimore Longitudinal Study of Aging. *J Am Geriatr Soc* **62**, 667-672, doi:10.1111/jgs.12740 (2014).

81   Brach, J. S. *et al.* The association between physical function and lifestyle activity and exercise in the health, aging and body composition study. *J Am Geriatr Soc* **52**, 502-509, doi:10.1111/j.1532-5415.2004.52154.x (2004).

82   Taylor, H. L. *et al.* A questionnaire for the assessment of leisure time physical activities. *J Chronic Dis* **31**, 741-755 (1978).

83   Bligh, E. G. & Dyer, W. J. A rapid method of total lipid extraction and purification. *Can J Biochem Physiol* **37**, 911-917, doi:10.1139/o59-099 (1959).

84   Wang, Y. *et al.* Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. *Proteomics* **11**, 2019-2026, doi:10.1002/pmic.201000722 (2011).

85   Keller, A., Nesvizhskii, A. I., Kolker, E. & Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* **74**, 5383-5392 (2002).

86   Nesvizhskii, A. I., Keller, A., Kolker, E. & Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* **75**, 4646-4658 (2003).

87   Team, R. C. R: A Language and Environment for Statistical Computing. (2016).

88   Gaujoux, R. & Seoighe, C. A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* **11**, 367, doi:10.1186/1471-2105-11-367 (2010).

89   Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504, doi:10.1101/gr.1239303 (2003).

90   Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**, 417-419, doi:10.1038/nmeth.4197 (2017).

Figures & Figure Legends

Figure 1 Quantitative Analysis of the Skeletal Muscle Proteome with Healthy Aging

(A) Study design, TMT assessment, and bioinformatics platform for protein quantification and age-associated protein analysis. (B) Correlation among all TMTs after using COMBAT experimental bias correction. Note that the R-values are all >0.99. (C) Principal Component Analysis (PCA) of 12 TMT batches, each circle indicating a sample from a TMT batch, 5 samples in each batch, overall 58 samples represented, and no substantial batch effect detectable. (D) Total number of proteins quantified from 12 TMT experiments from 58 donors, according to age group. 4,036 proteins were detected in at least 3 samples in all age groups and were used for further statistical analysis. (E) Partial Least Square (PLS) analysis revealed variance in the component1, 2, and 3 for all donors. All 4,036 proteins were used for PLS analysis, where each circle is a donor and the age groups are color-coded.

Figure 2 Classification of Age-associated Proteins

(A) Effect of age on protein expression levels. The x axis represents the size and sign of the beta coefficient of the specific protein regressed to age (adjusted for covariates) and the y axis represents the relative -log10 *p*-value. Each dot is a protein and all significant proteins are indicated in blue and red (age-associated 1,265 proteins, *p*<0.05). (B) The heatmap of the 1,265 significantly age-associated proteins reveals changing expression profiles across aging. (C) PLS analysis of age-associated proteins were classified into three age groups: 20-49 (young), 50-64 years (middle age), and 65+ (old) years old. (D) Percent distribution of categories of all quantified proteins, percent distribution of the same categories among proteins that were significantly downregulated and upregulated with aging. Proteins which are not considered directly related to mechanisms of aging are annotated as others and their subclassification is shown in the bar plot. (E-F) Log2 protein abundance of contractile, architectural and NMJ proteins. Simple linear regression was shown for age (x axis) and protein (y axis) correlation, confounders were not adjusted, and raw p-values were shown.

Figure 3 Functional Decline of Mitochondrial Proteins with Age

A) Percent coverage within categories of skeletal muscle proteins compared to the Uniprot database. The top section shows various energetics categories, while the z axis indicates the number of proteins identified for each protein category and in parenthesis the number of proteins reported in Uniprot for the same category. (B) Subcellular location of age-associated mitochondrial proteins based on up- or downregulation. Of note, most of the mitochondrial proteins are downregulated. (C) Age-dependent decline of respiratory and electron transport chain proteins. All mitochondrial proteins in the respiratory and electron transport chain that are significantly associated with age are downregulated ($p < 0.05$) except SDHAF2. The inset panel reports data on the proteins that are significantly upregulated with aging, SDHAF2 (mitochondrial) and the membrane protein CD73.

Figure 4 Implications of Proteins that Modulate Transcription and Splicing

A) Log2 protein abundance of age-associated transcriptional regulation proteins. Simple linear regression was shown for age (x axis) and protein (y axis) correlation, unadjusted p-values were shown. (B) Spliceosome major complex pathway protein expression abundance and dsysregulation. KEGG major spliceosome complex pathway representation and spliceosome complex proteins quantified (associated with splicing RNAs *U1*, *U2*, *U4*/*U6*, and *U5*) as plotted in the side square boxes. (C) The log2 abundance expression of 57 spliceosome complex proteins associated with age (*p*<0.05) are depicted as magenta circles, while all other quantified proteins are black circles. All snRNPs and spliceosome regulatory proteins are upregulated with age. (D) The average of all age-associated spliceosome proteins within each age group reveals an upregulation of spliceosome proteins with age. (E) Effect of age (one-year difference) on the 57 proteins of the spliceosome major complex and color coded based on spliceosome domains. Inset (left) is a legend for the complex domains and inset (right) shows that PRPF8 protein is robustly overrepresented with age.

Figure 5 Age-associated Alternative Splicing

(A) The number of participants with detected splicing variants is substantial, with >20% of the participants showing <5% splicing variants for the detected gene. Overall, 3.7% of all identified skeletal muscle transcripts (3,791 genes) show evidence of differential alternative splicing with aging. (B) Comparison of skeletal muscle age-associated splicing events (positive and negative). Negative events are downregulated with age and positive events are upregulated with age. The category of event is plotted on the X axis, and the Y axis shows the number of splicing events for each category. (C) Average Percent Spliced-in (PSI) from 6,255 events for each donor is depicted as a red circle. Average PSI across 53 donors ranging in age from 20 to 84 suggests an increase in alternative splicing with aging. (D) Comparison between skeletal muscle splicing events detected in GESTALT (solid bars) and splicing events reported in the Ensemble transcript splice events (shaded bars). (E) Comparison of age-associated proteins and age-associated alternatively splicing genes suggesting 30% (385) of the age-associated proteins undergo alternative splicing. Analysis of PANTHER database shows classes of enriched genes from different protein categories, and the number of genes representing each class is shown with #.

Figure 6 Age-associated Proteostasis Proteins

(A) Log2 protein abundance of all 74 age-associated proteostasis proteins across all 58 donors. Rows represent proteostasis proteins and columns represents donors. The average expression of all age-associated proteostasis proteins from each donor is plotted at the top and bottom (gray circles) with loess curves. The top section shows downregulated clusters of proteins (average of 24 proteins) and the bottom shows upregulated clusters of proteins (average of 54 proteins). The protein rows are ordered based on hierarchical clustering and displayed by dendrograms. (B) Confounders-adjusted β age coefficient of age-associated proteins, showing age-differentially regulated proteostasis proteins, over representation of proteostasis category proteins and the log2 magnitude of protein change with each year of age. (C) The increase of autophagy protein sub category with age is shown. Except HSPA8 and EIF4G1 all other autophagy proteins are positively correlated with age. Each bar plot shows each autophagy protein sub category and the average change over a year of age. The gene/proteins names are organized from lowest to highest log2 expression change per year of age. (D) Raw log2 abundance of autophagy proteins TDP and CALR were shown, simple linear regression method was used for age (x axis) and protein (y axis) correlation, of note unadjusted p-values were shown.

Figure 7 Age-associated Immune Proteins

(A) Immune-related proteins are depicted; the x axis shows the genes that code for age-differentially regulated proteins, while the y axis shows the log2 fold expression difference associated with age. The increase in innate immune signaling and lipid responses may indicate a reaction to adipocytes muscle infiltration, which in turn causes activation of innate immune signaling. (B-D) Examples of dysregulated proteins are shown from (B.1-B.4) pro-inflammatory. (C) Anti-Inflammatory. (D) Complement pathway proteins. Inflammasome adaptor protein PYCARD is positively associated with age, and the abundance of this protein is a key mechanism by which IL-1β pathway activation is regulated. In B-D raw log2 protein abundance and unadjusted *p*-values are shown.
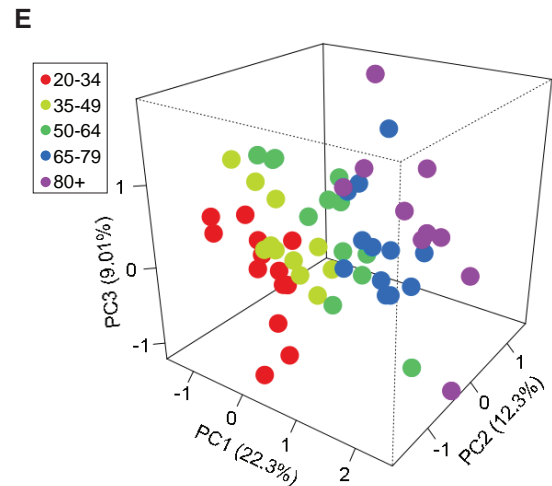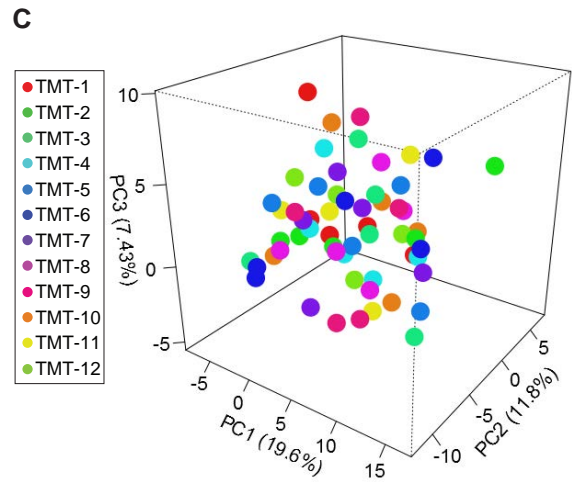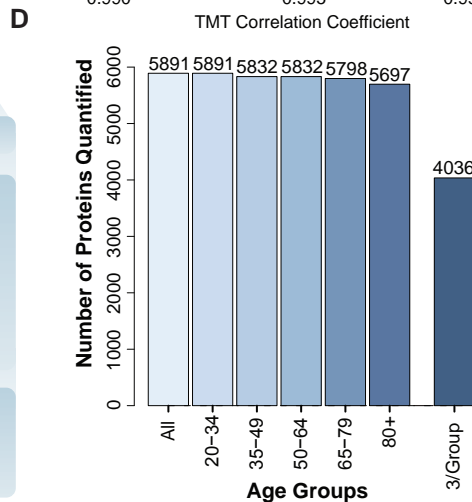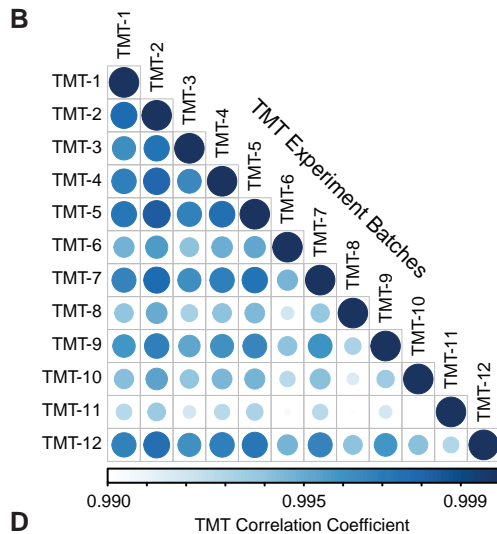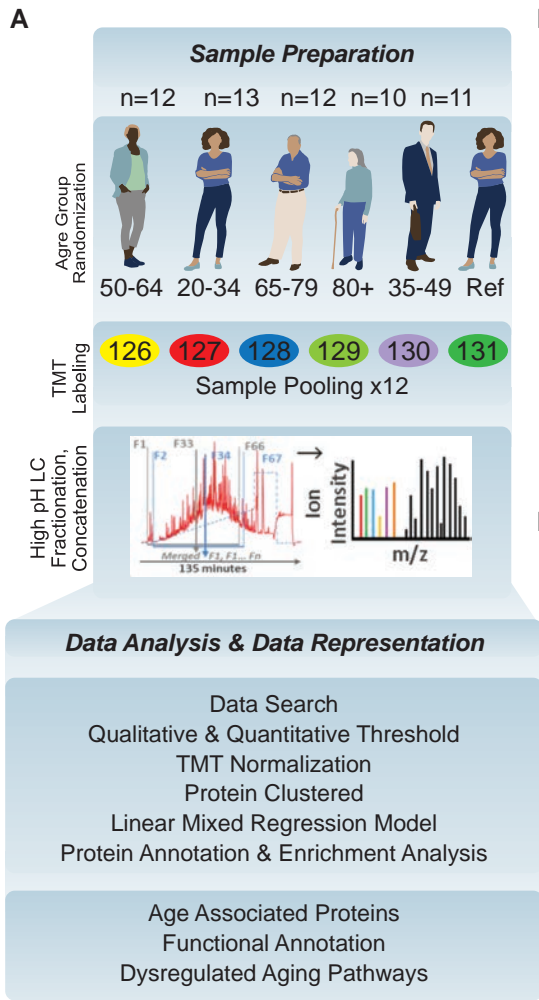
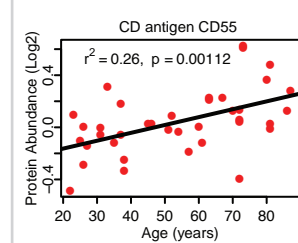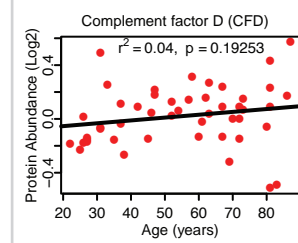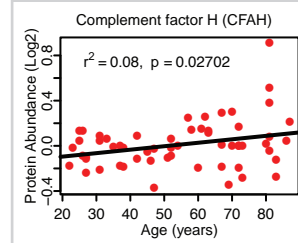Figure 1
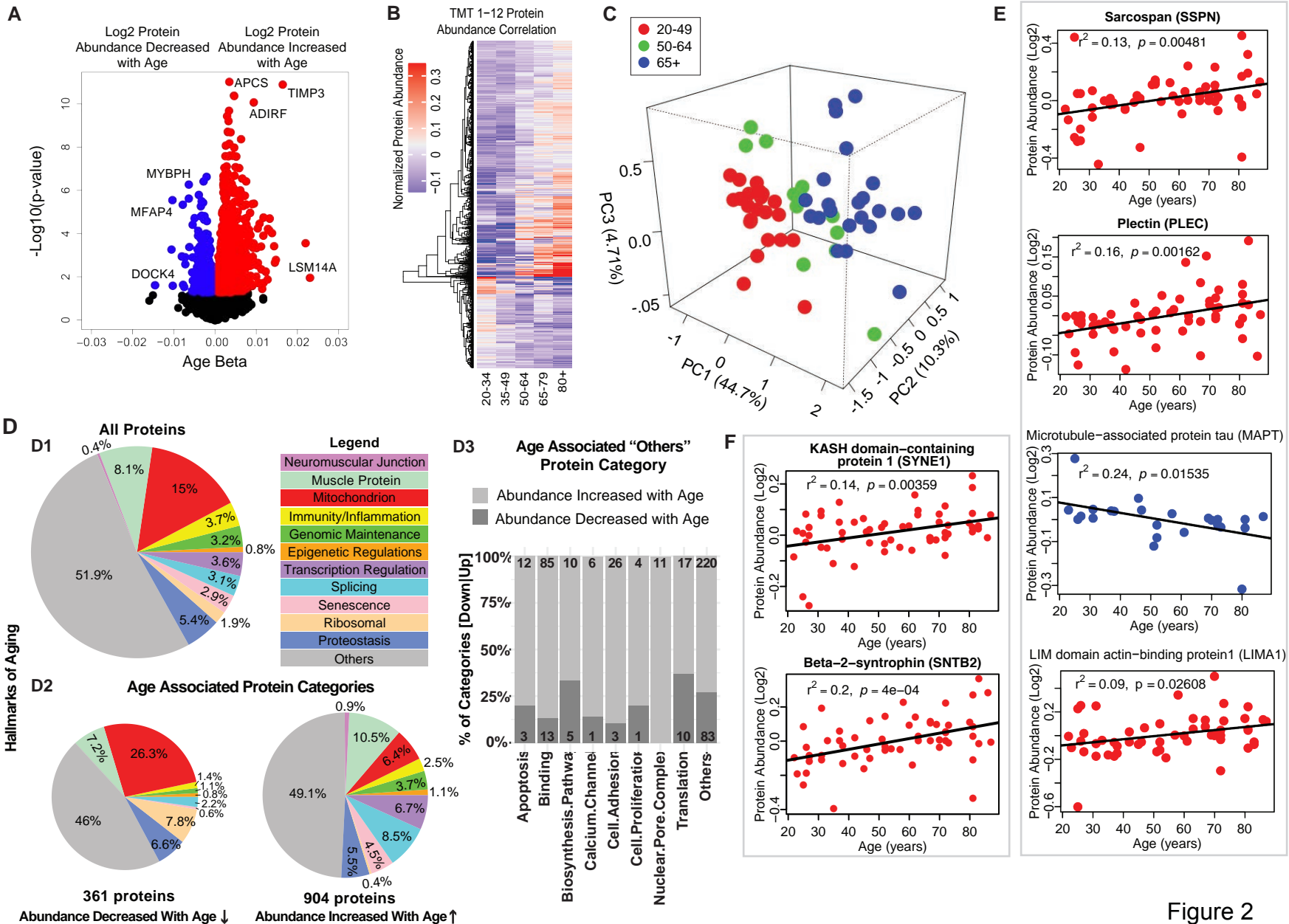
Figure 2
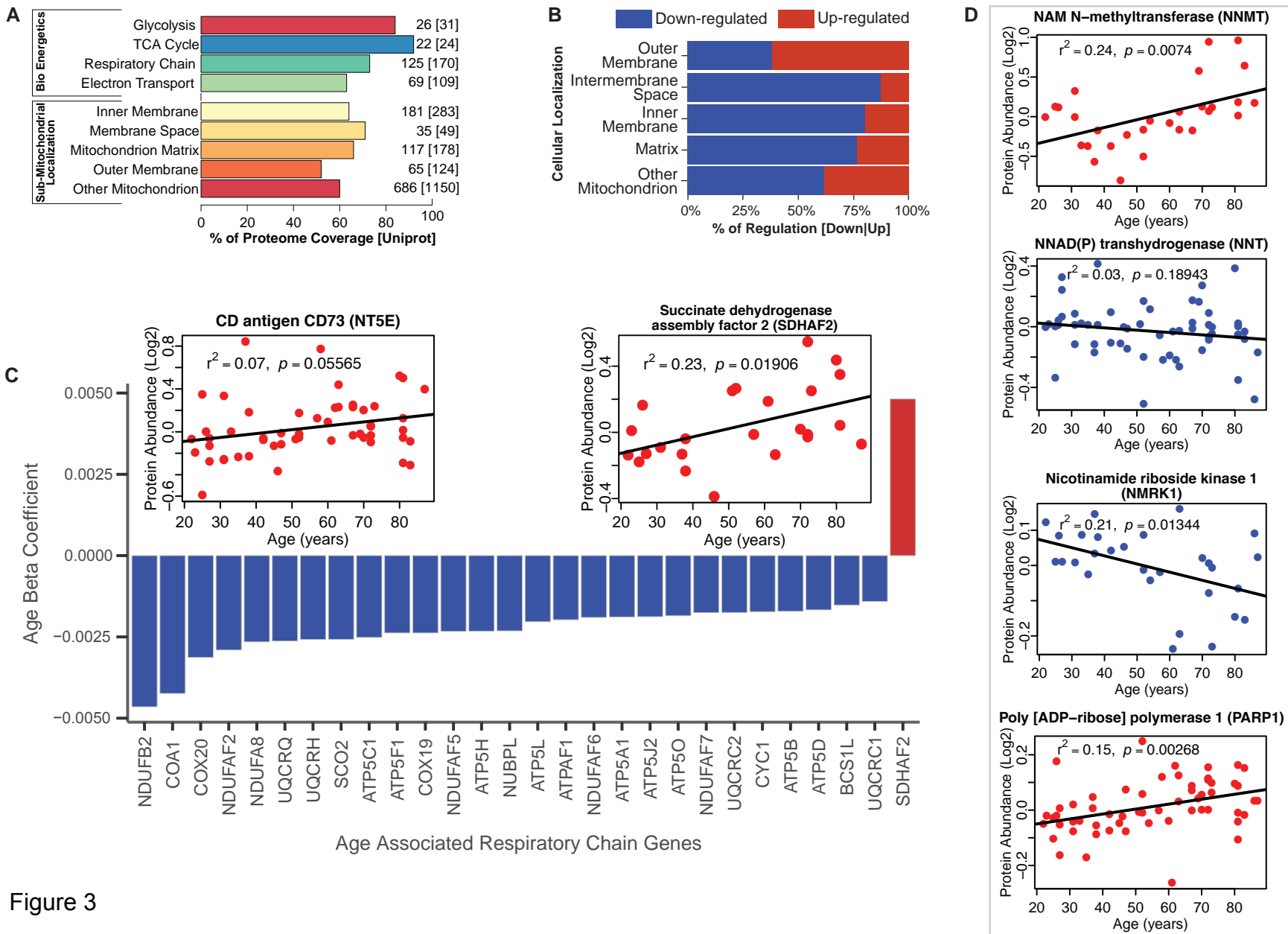
Figure 3
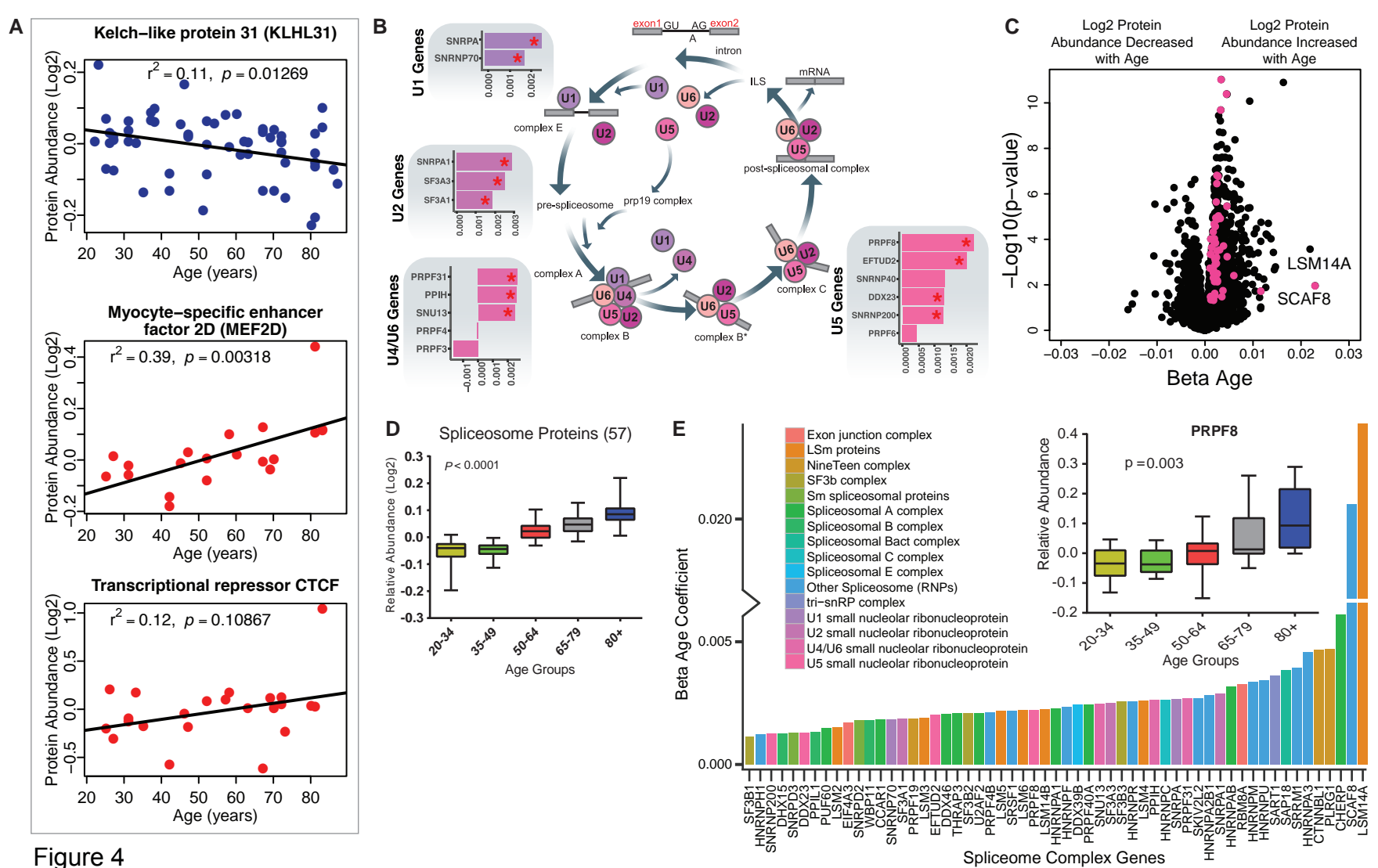
Figure 4
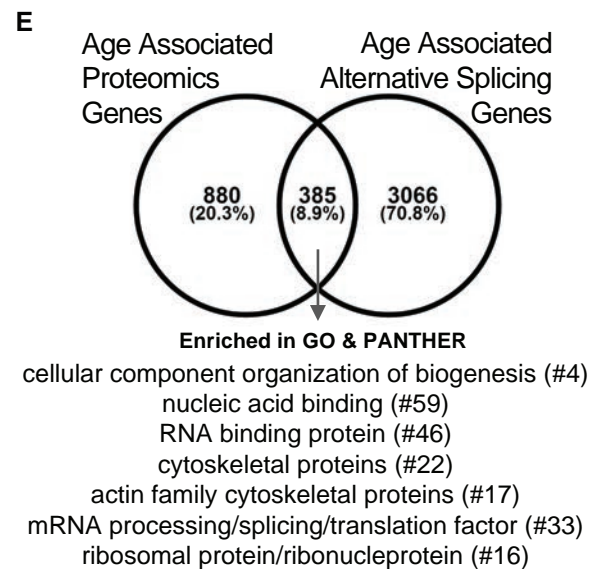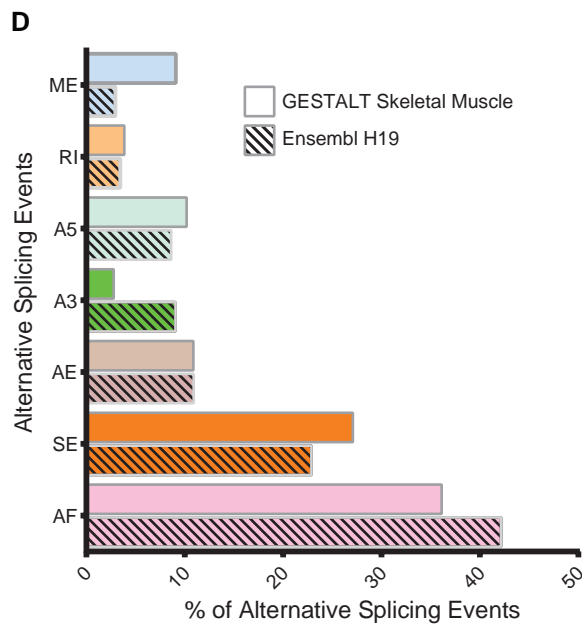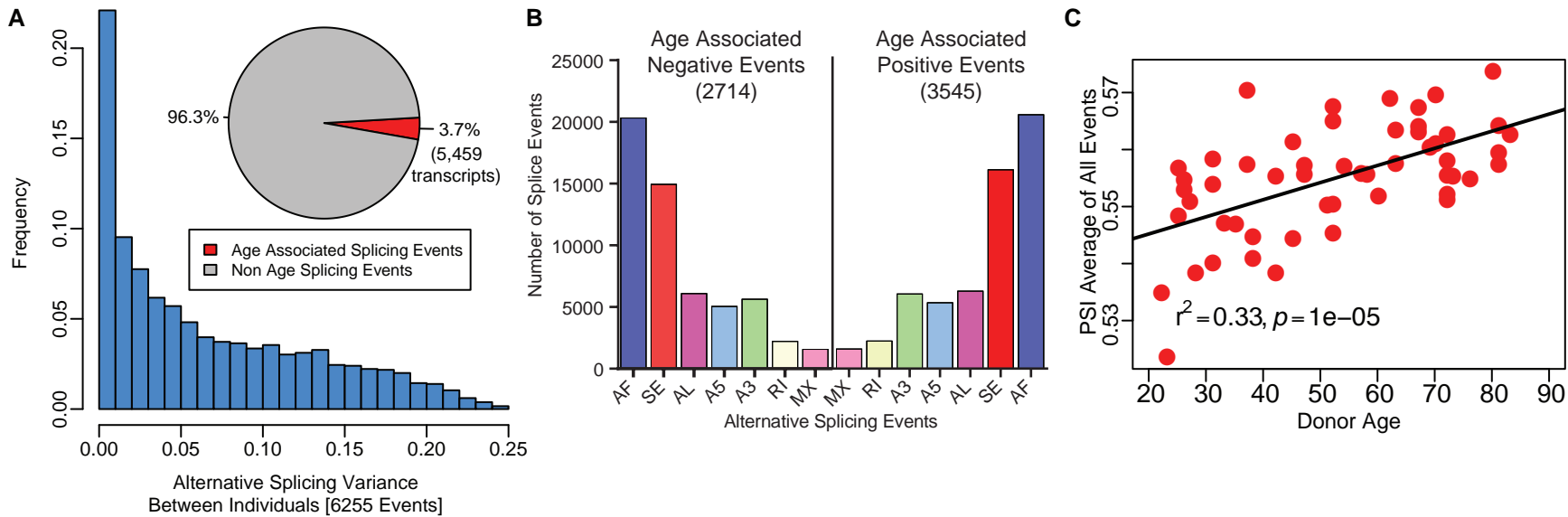
Figure 5
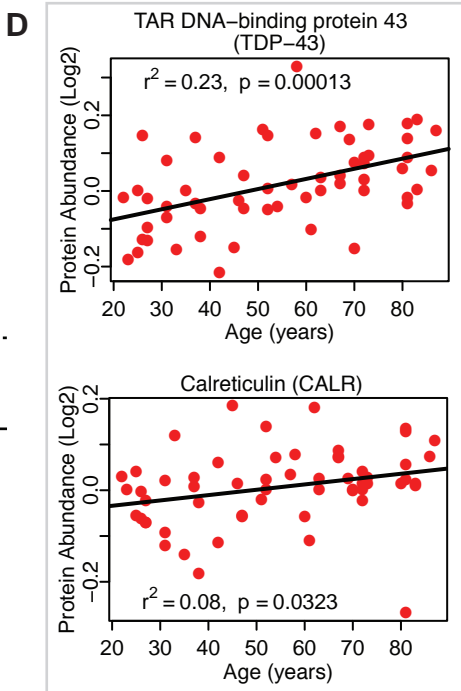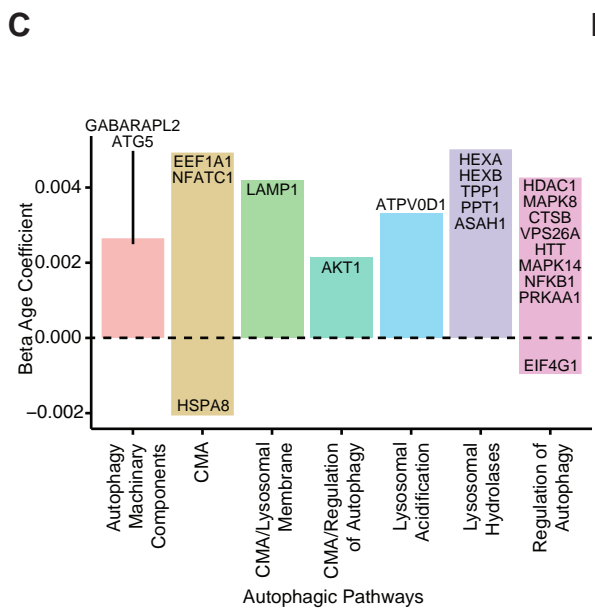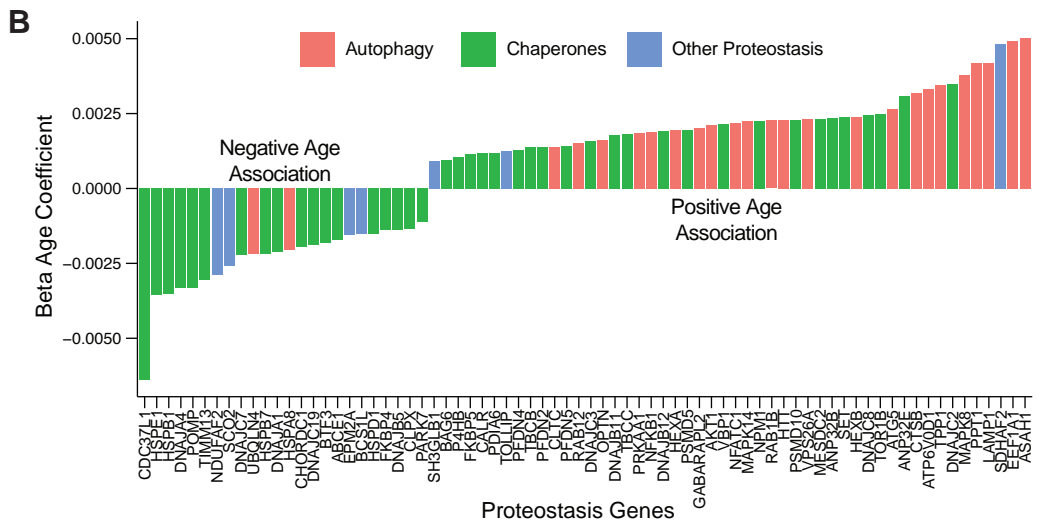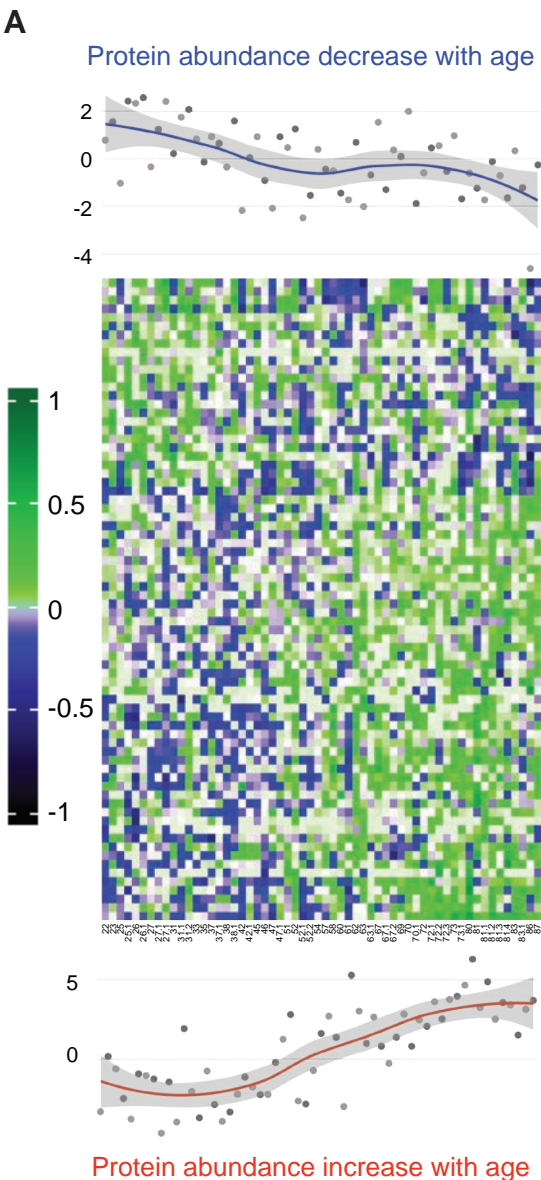
Figure 6
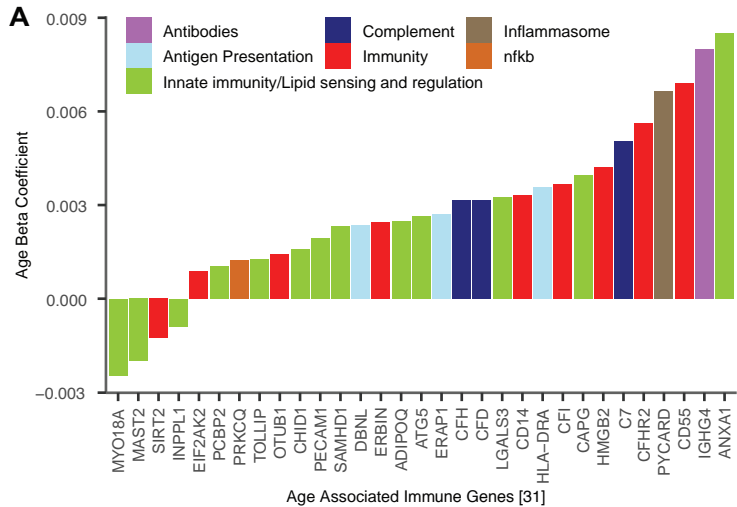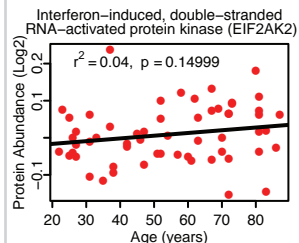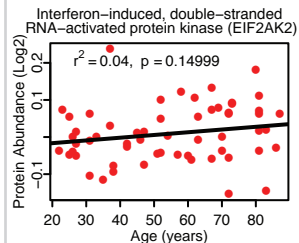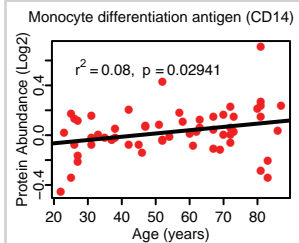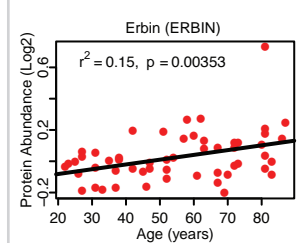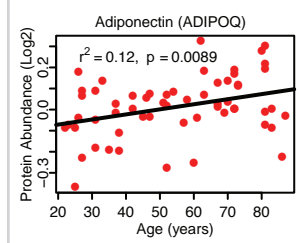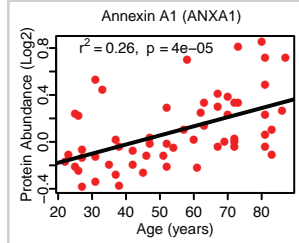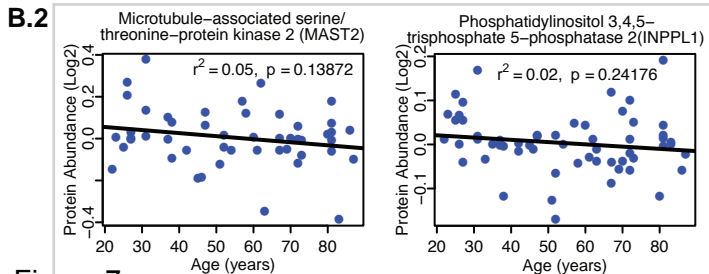
Figure 7

Supplemental Information

Aging Skeletal Muscle Proteomics Finds Changes in Spliceosome, Immune factors, Proteostasis and Mitochondria

Ceereena Ubaida-Mohien[$], Alexey Lyashkov[$], Marta Gonzalez-Freire[$], Ravi Tharakan[$], Michelle Shardell[$], Ruin Moaddel[$], Richard D. Semba[#], Chee W. Chia[$], Myriam Gorospe[$], Ranjan Sen[$] & Luigi Ferrucci[$]*

Author Affiliations

[$]*Intramural Research Program, National Institute on Aging, National Institutes of Health, Baltimore, MD 21224, USA*

[#]*Johns Hopkins Medical Institute, Baltimore, MSD 21205, USA*

Contact Info

*Correspondence**: FerrucciLu@grc.nia.nih.gov

Table S1 (related to Figure 1)

Age Group	20-34	35-49	50-64	65-79	80+	*P*-value	R^2
	(*n*=13)	(*n*=11)	(*n*=12)	(*n*=12)	(*n*=10)	--	--
Age (yr)	27.2 ± 3.3	41.3 ± 4.5	57.1 ± 4.7	70.3 ± 2.3	82.4 ± 2.4	--	--
Gender	*M*8, *F*5	*M*7, *F*4	*M*7, *F*5	*M*8, *F*4	*M*6, *F*4	--	--
Education (yr)	16 ± 3	14 ±3	14 ± 2	16 ± 2	17 ± 2	0.3305	--
Race	9*C*, 2*AA*, 2*A*	5*C*, 6*AA*	8*C*, 4*AA*	10*C*,1*AA*,1*A*	9*C*,1*AA*	0.0958	--
***BMI, kg/m^2**	25.9 ± 2.8	26.4 ± 2.6	26.6 ± 3.2	26.4 ± 2.4	25.2 ± 3.9	0.3458	0.007
Height (cm)	172 ± 11	177 ± 10	169 ± 4	172 ± 11	172 ± 6	0.3985	--
***Weight (kg)**	76 ± 10	81 ± 9	77 ± 12	75 ± 13	73 ± 16	**1.74E-05**	0.34
***Waist Circumference (cm)**	82 ± 7	87 ± 7	90 ± 11	92 ± 11	92 ± 13	**6.32E-06**	0.39
***KEIS (left) ±**	192 ± 31	208 ±55	200 ± 71	165 ± 62	130 ± 42	**4.29E-07**	0.40
***KEIS (right) ±**	194 ± 38	220 ± 65	194 ±78	169 ± 53	147 ± 57	**2.41E-07**	0.41
†Physical Activity	1.8 ± 1.4	1.8 ± 1.3	2 ± 1.1	2.3 ± 1	1.5 ± 1.1	0.5145	--

Table S1 (related to Figure 1). Baseline Characteristics of the GESTALT Skeletal Muscle Participants. Participants are classified into 5 different age groups. Gender: *M* is Male, *F* is Female; the number of participants is indicated. Age is indicated in years as mean and standard deviation (SD ±) for each age group. Race: number of participants is shown on the left and race is shown in italics; *C* is Caucasian, *AA* is African American, and *A* is Asian. Body Mass Index (BMI) is expressed as mean and SD (±) for each group. *P*-value is calculated by 1-way ANOVA with Kruskal-Wallis test. Race is analyzed by Chi-square test.

**P*-value calculated from linear regression model, gender adjusted.

± Knee Extension Isokinetic Strength (KEIS) (30^0/sec; Nm).

†Physical activity is calculated from self-report involvement in weight circuit, vigorous exercise, brisk walking and casual walking and summed as high-intensity physical activity hours per week. This is further categorized into 0 (not active),1 (moderately active), 2 (active), and 3 (highly active) and expressed as mean of categorical variables (0,1,2,3) ± SD.
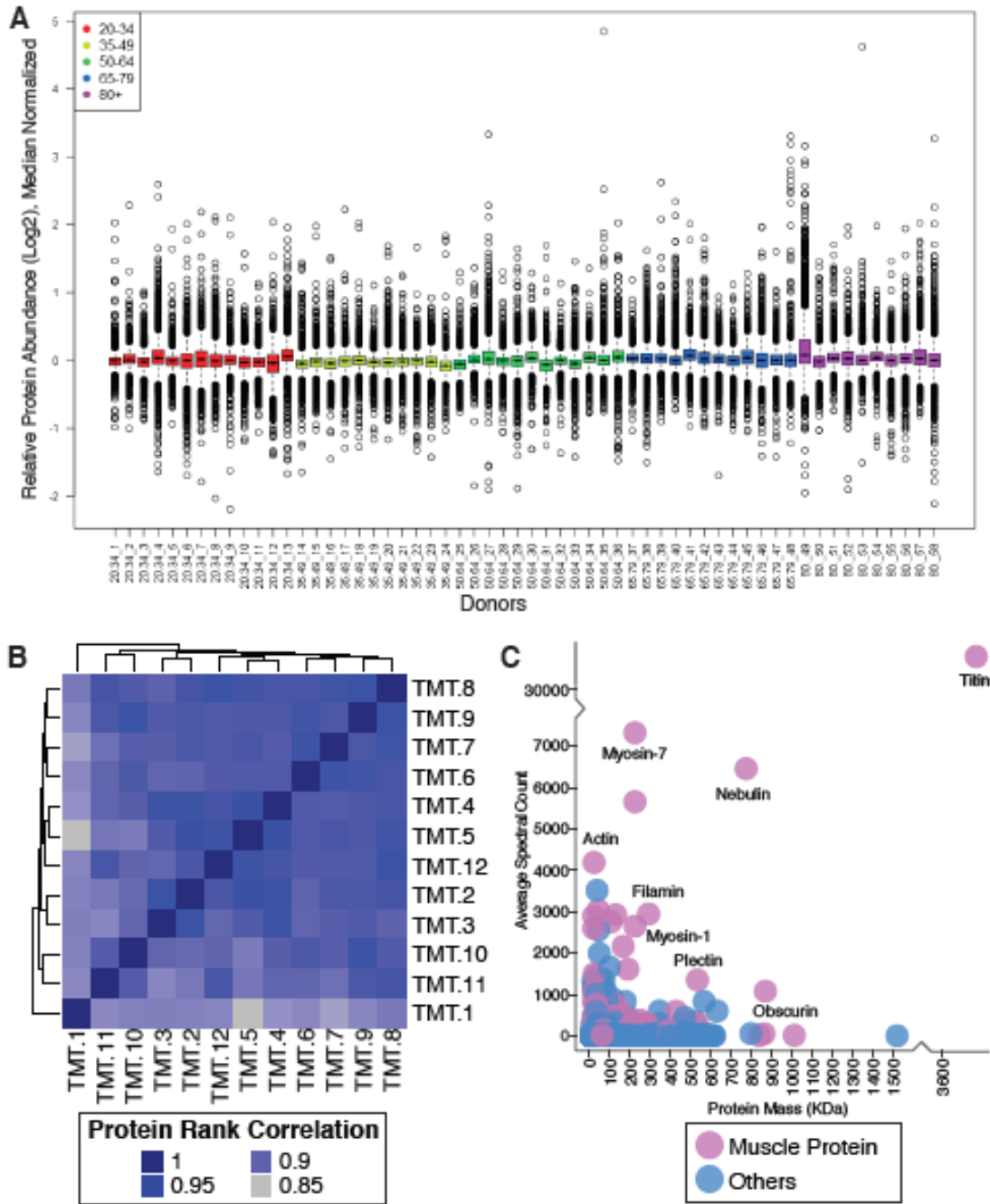
Figure S1 (related to Figure1)

A. Normalized Relative Protein Expression from All Participants. Bar plot of log2 normalized protein relative expression abundance from all 58 participants. Each circle is a protein. Median levels for all proteins (5891) from all samples are plotted, 25th and 75th percentiles are represented with first and third quartiles. Age groups are color-coded. This

analysis was followed by a median polish and median sweep normalization by subtracting the spectrum median log2 intensity from the observed log2 spectrum intensities.

B. Spectral Count Correlation Across 12 TMT Experiments. The heatmap figure shows the protein correlation across all 12 TMT experiments. Since we have 12 TMT batches across all samples, we looked at the data robustness across all 12 TMTs. The proteins identified in each TMT are ranked based on spectral abundance, and the ranks between TMTs are correlated. Correlation coefficients between 12 TMT experiments are from 0.99 to 0.85, demonstrating that the relative abundance of proteins is robustly replicated across TMT sets.

C. Skeletal Muscle Proteins Detected According to Abundance and Mass. The abundance of proteins quantified, the X axis is the mass distribution of the protein and the Y axis is the average spectral abundance of the proteins. Proteins are color-coded to visualize enrichments of muscle protein.
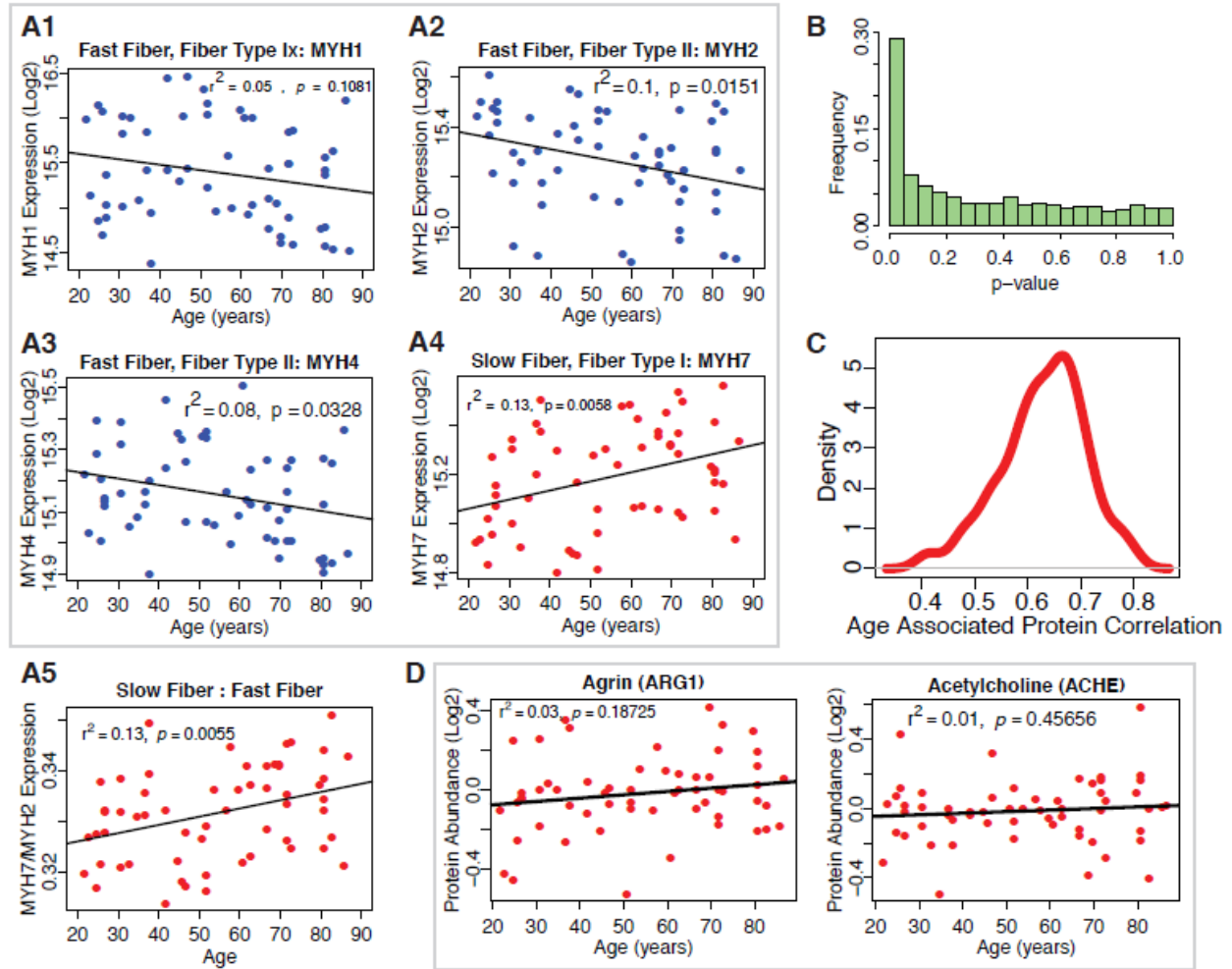
Figure S2 (related to Figure 2).

A1-A5. Protein Biomarkers of Slow-Twitch and Fast-Twitch Fiber Types and Estimated Muscle Fiber Ratios. Log2-normalized fiber expression intensity from all the 58 donors are shown for fast fiber-twitch type proteins MYH4, MYH1, MYH2 and slow-twitch fiber type protein MYH7 (A1-A4). The estimated ratios of slow-twitch fiber to fast-twitch fiber (A5) suggest that slow-twitch fiber type expression increases with age.

B. Statistical Significance of the Proteins and Robustness of Age-Association. *p*-value distribution for quantified proteins and their association with age. The *p*-values are calculated from t-tests use Satterthwaite's approximations of the F-statistics using lmerTest. To account for multiple comparisons, we performed a Benjamini-Hochberg (BH) correction, and we still found 917 proteins that have a BH corrected *p*-value <0.1.

C. Robustness of the Estimated Beta Coefficients for Age for Quantified Proteins Across all 12 TMT Experiments. To assess the robustness in the identification of the 1265 age-associated proteins across all 12 TMT experiments, we estimated the effect of age (beta coefficient) on single proteins in regression analyses run in two separate sets of 6 TMT sets selected at random from the 12 available and estimated the average correlation coefficient between proteins quantified in the two sets. Then, we repeated this analysis 100 times and plotted the distribution of correlation coefficients. The results of the linear mixed models appear to be robust against TMT variability.

D. Clusters of Proteins Dysregulated with Aging- NMJ Related Proteins. Neuromuscular protein abundance with age. Agrin and Acetylcholineesterase proteins are not significantly associated with age.
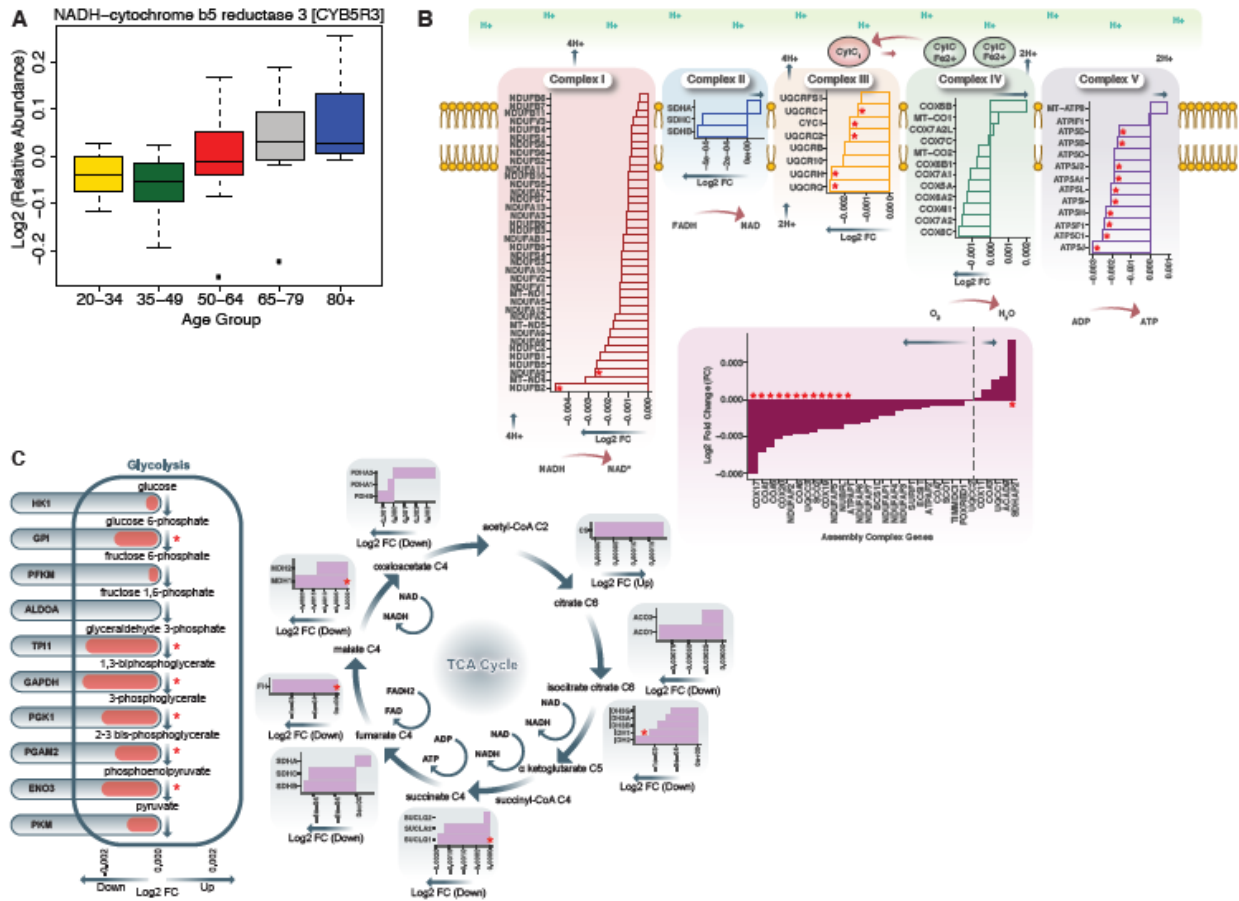
Figure S3 (related to Figure 3).

A. Age-associated protein abundance of CYB5R3 protein. The abundance of this protein is significantly higher at older age; age groups are color-coded.

B. Respiratory Chain Complex I-V and Aging. Electron Transport Chain Protein Quantification. Proteins quantified from Complex I, Complex II, Complex III, Complex IV, Complex V and Assembly complex proteins are represented. Age-associated proteins are marked by a red asterisk (*). Log2 fold ratios of the gene are on x axis; arrows pointing to left shows underrepresented proteins and arrows pointing to the right are overrepresented proteins.

C. Dysregulation of Bioenergetic Pathway. Proteins quantified from glycolysis and TCA cycle are shown. (Left) Of the 26 glycolysis proteins quantified, 6 are significantly underrepresented with age. (Right) Of the TCA cycle gene products shown, 4 are significantly decreased with aging. A red asterisk indicates genes significantly changed with age ($p < 0.05$).
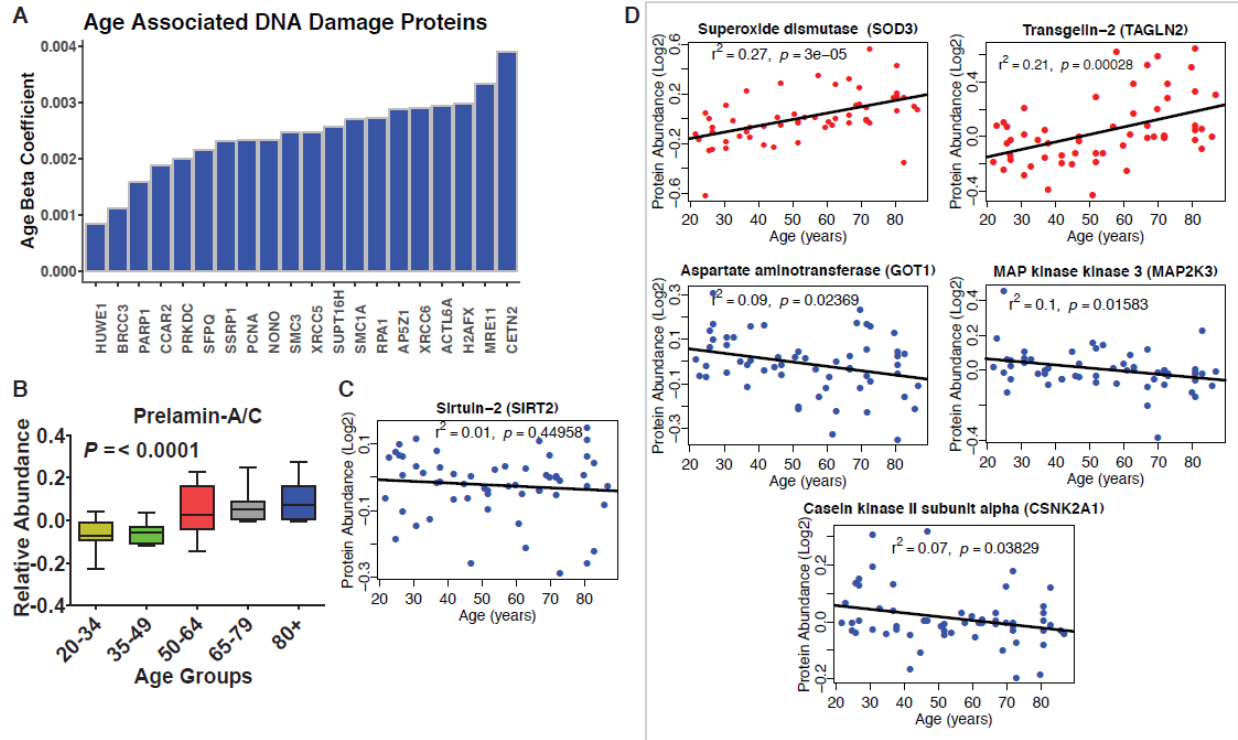
Figure S4 (related to Figure 2).

A. Dysregulation of Proteins Involved in Genomic Maintenance. A large number of proteins involved in genetic maintenance are overrepresented at older ages.

B. Prelamin. Protein levels increase significantly with age.

C. Sirtuin 2. Raw protein abundance of SIRT2 is plotted according to age and shows no age association. However, adjusting for confounder the negative association with age becomes statistically significant (p=0.032).

D. Dysregulation of Proteins Involved in Cellular Senescence. Age-associated proteins of cellular senescence. Log2 raw protein abundance and age correlation is shown from simple linear regression method. Unadjusted p-values are represented with r^2.
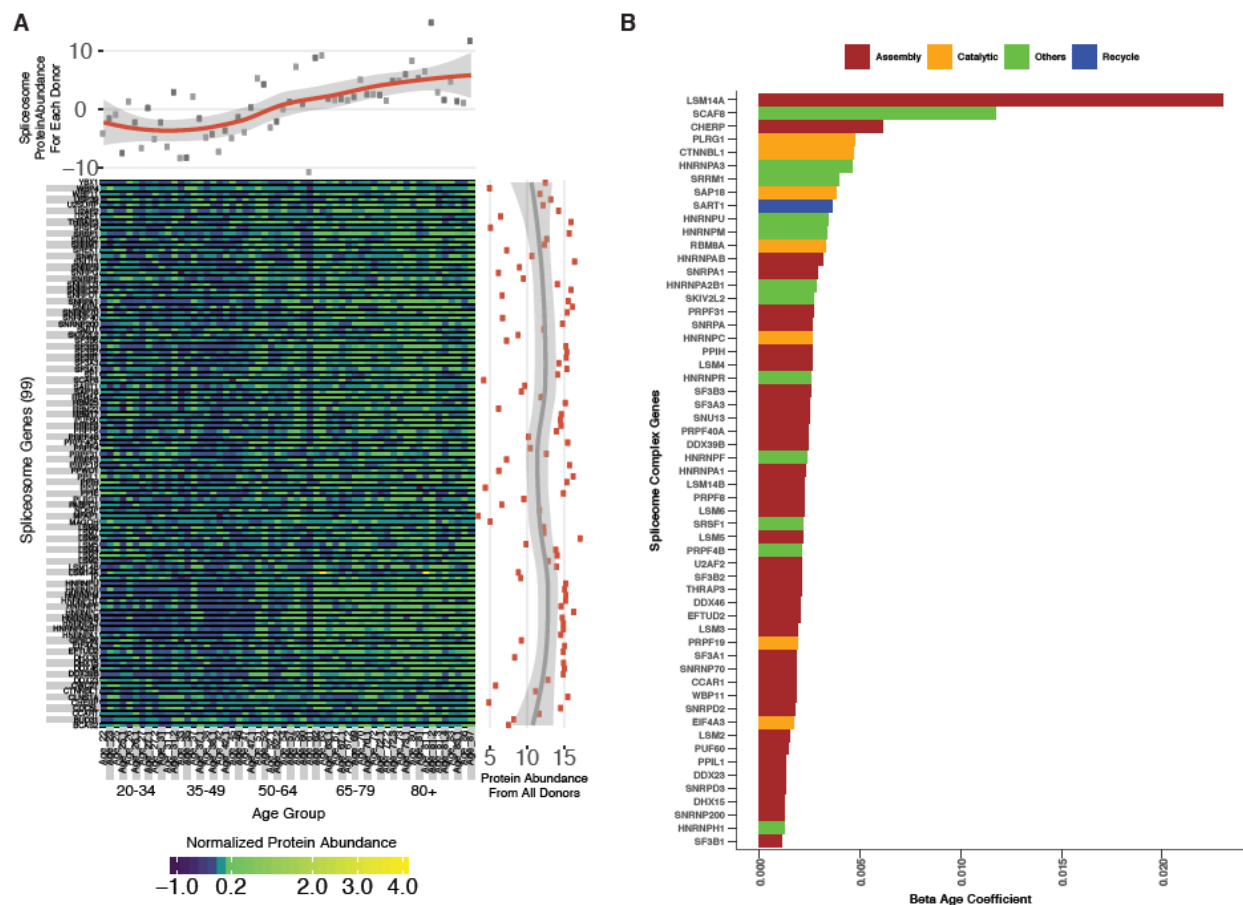
Figure S5 (related to Figure 4).

A. Spliceosome Protein Abundance and Expression Variation Across Ages. A total of 99 spliceosomal proteins are detected across all 5 age groups. The abundance of each protein is represented as a heatmap; each column represents a study participant and each row indicates a specific protein. The x axis scatter plot across the top is the average of spliceosomal protein abundance for each donor, revealing an increase in spliceosome abundance with age. The y axis scatterplot on the right is the average log2 protein abundance from all donors for each protein, showing that most of the spliceosome proteins are highly abundant.

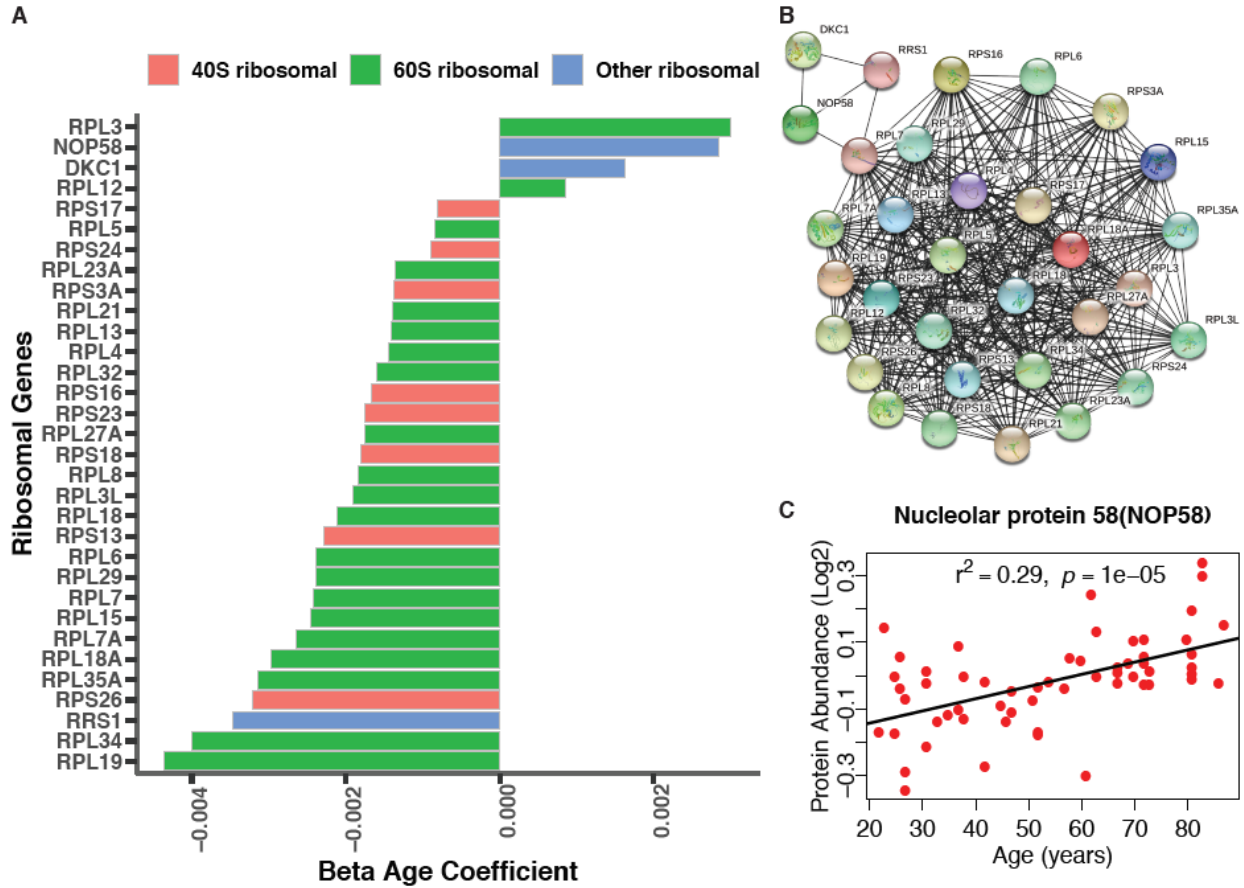**B.** Spliceosomal protein categories and age association.

Figure S6 (related to Figure 2).

A. Depletion of Ribosomal Proteins with Age. Age-associated ribosome proteins and interacting partners. Different categories of ribosome proteins are color-coded.

B. Protein-protein interaction of age-associated ribosome proteins from co-expression. The protein-protein interaction enrichment score *p*-value <1.0e-16. Color node shows query proteins and black edge represent co-expressed protein association. The top left cluster NOP58 is required for the biogenesis of the 60S ribosomal subunit and the biogenesis of box C/D snoRNAs such as *U3*, *U8* and *U14* snoRNA.

C. Log2 raw protein abundance of NOP58, a nucleolar protein important for ribosomal biogenesis, according to age. Unadjusted *p*-value is shown with r^2.
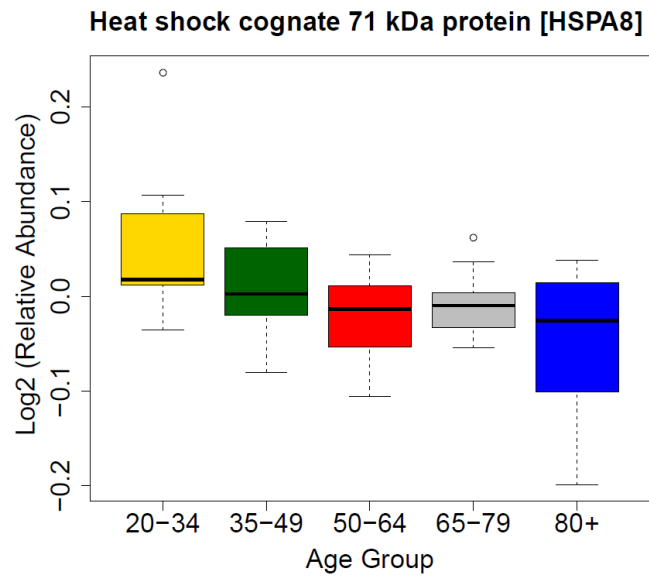
Heat shock cognate 71 kDa protein [HSPA8]

Figure S7 (related to Figure 6).

HSPA8 protein and its association with age.