



# Skeletal muscle transcriptome in healthy aging

Robert A. Tumasian III <sup>1</sup>, Abhinav Harish<sup>1</sup>, Gautam Kundu<sup>1</sup>, Jen-Hao Yang<sup>1</sup>, Ceereena Ubaida-Mohien<sup>1</sup>, Marta Gonzalez-Freire<sup>1</sup>, Mary Kaileh<sup>1</sup>, Linda M. Zukley<sup>1</sup>, Chee W. Chia<sup>1</sup>, Alexey Lyashkov<sup>1</sup>, William H. Wood III<sup>1</sup>, Yulan Piao<sup>1</sup>, Christopher Coletta<sup>1</sup>, Jun Ding<sup>1</sup>, Myriam Gorospe<sup>1</sup>, Ranjan Sen<sup>1</sup>, Supriyo De<sup>1</sup> & Luigi Ferrucci <sup>1</sup>✉

Age-associated changes in gene expression in skeletal muscle of healthy individuals reflect accumulation of damage and compensatory adaptations to preserve tissue integrity. To characterize these changes, RNA was extracted and sequenced from muscle biopsies collected from 53 healthy individuals (22–83 years old) of the GESTALT study of the National Institute on Aging–NIH. Expression levels of 57,205 protein-coding and non-coding RNAs were studied as a function of aging by linear and negative binomial regression models. From both models, 1134 RNAs changed significantly with age. The most differentially abundant mRNAs encoded proteins implicated in several age-related processes, including cellular senescence, insulin signaling, and myogenesis. Specific mRNA isoforms that changed significantly with age in skeletal muscle were enriched for proteins involved in oxidative phosphorylation and adipogenesis. Our study establishes a detailed framework of the global transcriptome and mRNA isoforms that govern muscle damage and homeostasis with age.

<sup>1</sup>National Institute on Aging–Intramural Research Program, National Institutes of Health, Baltimore, MD, USA. ✉email: [FerrucciLu@grc.nia.nih.gov](mailto:FerrucciLu@grc.nia.nih.gov)

As older populations continue to expand worldwide, there is growing urgency to discern the underlying biological mechanisms that drive the deleterious manifestations of aging, including disease susceptibility and functional impairments. The study of biomarkers associated with healthy aging in the absence of overt disease may help uncover mechanisms of biological aging, including the accumulation of damage and intervening homeostatic mechanisms. A first step in this process is to produce detailed catalogs of molecular parameters that change systematically over the lifespan in a population of very healthy individuals.

The Genetic and Epigenetic Signatures of Translational Aging Laboratory Testing (GESTALT) study of the National Institute on Aging–National Institutes of Health (NIH) was designed to collect and analyze epigenetic, transcriptomic, and proteomic biomarkers from multiple tissues (including the muscle) from healthy individuals dispersed over a wide age range<sup>1</sup>. The specific goal of the study reported here is to elucidate differences in the transcriptomic network of the skeletal muscle as a function of age. Earlier large-scale studies identified changes in gene expression associated with skeletal muscle aging and acute exercise in well-defined populations using microarrays<sup>2,3</sup>. Other studies have also identified genomic and proteomic signatures associated with skeletal muscle aging<sup>4–7</sup>. However, no previous study has collected this information from very healthy individuals over a wide age range enrolled according to strict clinical and functional criteria.

To overcome some of these limitations, and to replicate and complement information collected in previous studies, we performed an RNA-sequencing (RNA-seq) analysis from skeletal muscle biopsies gathered from 53 healthy individuals of ages ranging from 22 to 83 years old. We used the data to comprehensively analyze changes in gene expression and isoforms that occur with aging. Some of our results identify basic biological mechanisms of aging, such as cellular senescence and insulin signaling. Other results confirm previous findings in the literature, in some cases, for genes whose function in the muscle and in aging is not well understood. Elucidation of the mechanisms that drive changes in the expression of these genes with aging and a stronger knowledge of their physiological functions are critical for understanding both damage accumulation as well as strategies of biological resilience with aging, which could serve as targets for interventions. We postulate that imbalances in these same mechanisms and compensatory strategies are likely to cause debilitating pathologies of aging muscle, including frailty and sarcopenia.

## Results

**Participant characteristics.** Skeletal muscle biopsies were obtained from 53 very healthy GESTALT participants (22–83 years, median = 52 years), who were defined as very healthy based on strict inclusion criteria developed by the Clinical Research Unit of the National Institute on Aging<sup>8</sup> (see “Methods”). The majority of our cohort was Caucasian ( $n = 38$ ) and male ( $n = 33$ ), and completed at least a high school education ( $n = 30$ ) (Supplementary Fig. 1a). In spite of the strict inclusion criteria, there were significant ( $p < 0.05$ ) differences in waist circumference (cm), fitness ( $\text{VO}_2$  max, ml/kg/min), fasting glucose (mg/dL), 400 m walking time (s), and knee strength (Nm) between younger ( $\leq 52$  years) and older ( $> 52$  years) participants (Supplementary Fig. 1a).

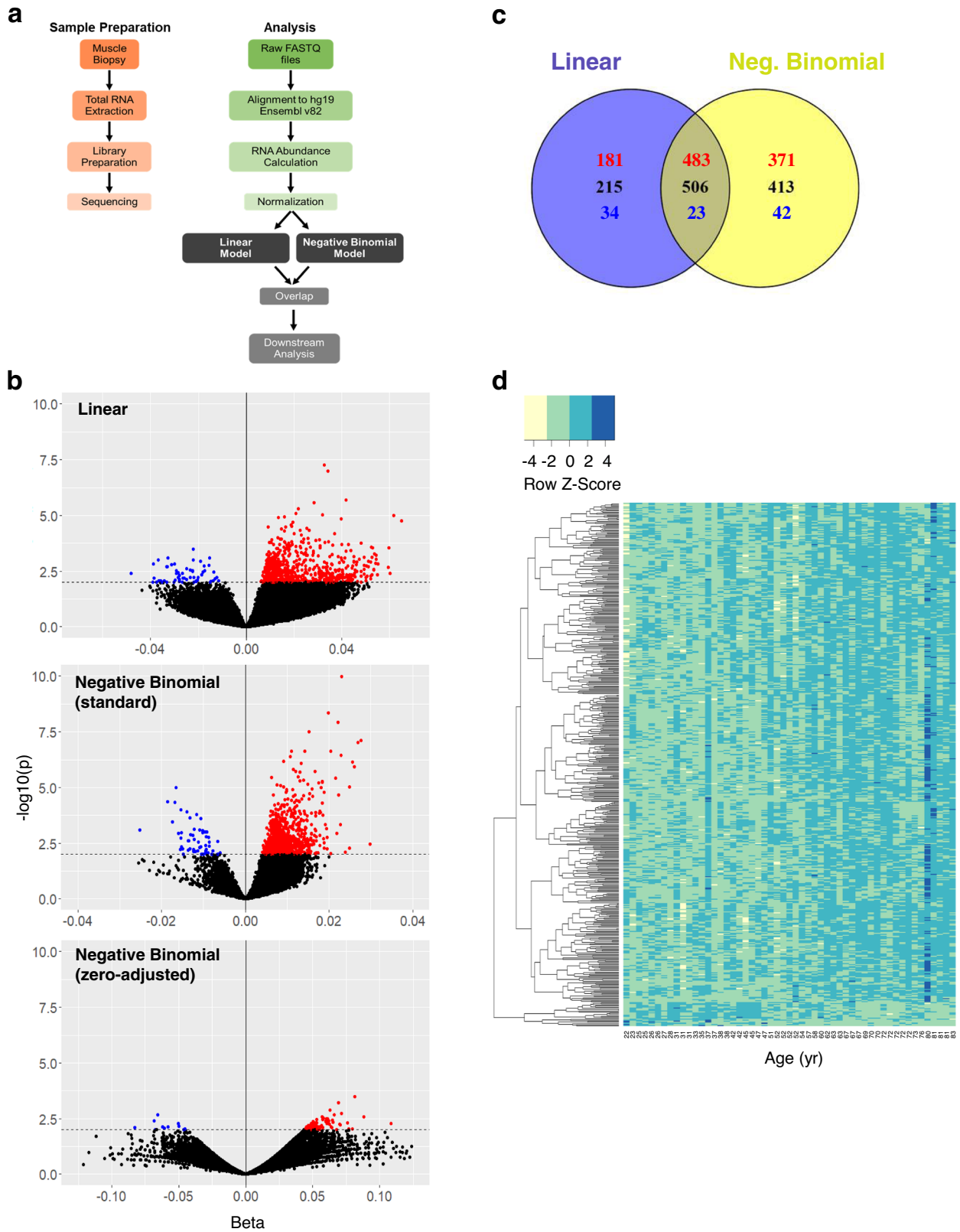
**RNA (human Ensembl genes (ENSG)) and isoform (human Ensembl transcripts (ENST)) detection.** Prior to studying age-related changes in the human skeletal muscle transcriptome, we

determined the average number of protein-coding and noncoding RNAs (ENSGs) that were reliably detected using the Ensembl hg19 v82 (September 2015) database. On average, 24,453 (ranging from 16,203 to 36,119) RNAs were detected per sample in our study population, among which 15,291 (ranging from 11,517 to 18,966) were protein-coding RNAs and 9162 (ranging from 4686 to 17,339) were noncoding and other types of RNAs (Supplementary Fig. 1b). By “other types,” we mean the biotypes that are not specifically mentioned as bona fide “protein-coding,” and which cannot be categorized well. Similarly, on average, 66,633 (ranging from 42,928 to 110,124) mRNA isoforms were detected per sample in our study population, among which 27,908 (ranging from 20,954 to 40,371) were protein-coding isoforms and 38,075 (ranging from 21,768 to 69,753) were noncoding and other types of isoforms (Supplementary Fig. 1c). Overall, roughly 85% of all detected RNAs and 84% of all detected isoforms were identified in the Ensembl hg19 v82 (September 2015) database, underscoring the reliability of our data (Supplementary Fig. 1b, c). In this study, we define ENSGs as genes/RNAs and ENSTs as transcripts/isoforms.

**Analysis of RNA (ENSG) expression profiles in GESTALT participants.** An overview of the sample preparation and analytical approaches used in this study is provided (Fig. 1a). Linear regression analysis identified 664 RNAs that were overrepresented ( $p < 0.01$ , positive  $\beta$ -values for age) and 57 RNAs that were underrepresented ( $p < 0.01$ , negative  $\beta$ -values for age) in older compared to younger individuals (Fig. 1b). Together, the standard and zero-adjusted negative binomial models (see “Methods”) identified 854 RNAs that were overrepresented and 65 RNAs that were underrepresented with older age (Fig. 1b). Of note, the directionality of RNA age associations was highly consistent between the two models. Overall, 506 RNAs were significantly ( $p < 0.01$ ) associated with age from both the linear and negative binomial models; of these, 483 RNAs had higher expression with older age and 23 RNAs had lower expression with older age (Fig. 1c). Moreover, eight of the top ten RNAs with the lowest  $p$ -values overlapped between the two models, underscoring the robustness of these findings. A heat map of the expression levels of these 506 shared RNAs with age is shown (Fig. 1d), with the 23 underrepresented RNAs clustered at the bottom (blue to yellow rows with increasing age). Only a 37-year-old donor and an 80-year-old donor did not cluster with the study population. As the medical records of these participants did not suggest any possible cause for exclusion and the overall trends of expression level with age were conserved, all data for these participants were maintained in our analysis.

**Top RNAs (ENSGs) displaying differential gene expression in the skeletal muscle with age.** For in-depth analysis and validation, we focused on the top 20 RNAs most related to aging, ranked by the lowest  $p$ -value, which included a few noncoding RNAs, including pseudogene-encoded RNAs and antisense RNAs. The names, biotypes (protein-coding or noncoding), associated functions, and  $\beta$ -values for age identified by the linear model are listed (Figs. 2a and 3a). Linear regression plots for the top ten RNAs within these groups with positive and negative  $\beta$ -values for age are shown in Figs. 2b and 3b, respectively.

Our analysis found age-associated transcripts that have never been reported in the literature. Specifically, cyclin-dependent kinase inhibitor 2B (*CDKN2B*) mRNA (cyclin-dependent kinase inhibitor p15<sup>INK4b</sup>, which arrests the cell cycle by inhibiting cyclin-dependent kinase (CDK) 4 activity and is implicated in cellular senescence<sup>9</sup>) was the top age-related transcript, followed by *IRS2* mRNA (insulin receptor substrate 2, involved in insulin signaling



and lipid metabolism in the skeletal muscle<sup>10,11</sup>), *NR2F2* mRNA (nuclear receptor subfamily 2 group F member 2, important for myogenesis and skeletal muscle development<sup>12</sup>), and *LPP* mRNA (LIM domain containing preferred translocation partner in lipoma, implicated in smooth muscle differentiation<sup>13</sup>), all of which had significantly ( $p < 0.01$ ) higher expression in older age.

Our analysis also confirmed RNA (ENSG) expression patterns with aging, previously identified in the literature, including *FAM83B* (Family with Sequence Similarity 83 Member B, involved in hypoxia response<sup>14</sup>), *C12orf75* (chromosome 12 Open Reading Frame 75, implicated in insulin signaling and energy metabolism<sup>15</sup>), *SKAP2* (Src Kinase-Associated Phosphoprotein 2, which regulates

**Fig. 1 Methodology overview, model overlap, and heat map of significant changes in RNA (ENSG) levels.** **a** Flowchart of sample preparation (orange) and analytical methods (green to gray) used in this study. **b** Volcano plots capture all significant ( $p < 0.01$  from two-sided Wald test, unadjusted) RNAs identified by our models: linear model including all RNAs, standard negative binomial model including only RNAs with all non-zero CPB values, and zero-adjusted negative binomial model including only RNAs with at least one value of zero CPB. Red and blue points indicate significant RNAs with positive and negative  $\beta$ -values for age, respectively. Black points denote RNAs with nonsignificant  $\beta$ -values for age. **c** Venn diagram displaying the overlap between all RNAs significantly associated with aging identified by the linear (blue) and negative binomial (yellow) models (overlap shown in brown). Black denotes total significant RNAs, whereas red and blue values represent significant RNAs with positive and negative  $\beta$ -values for age, respectively. **d** Heat map representing the expression levels of the 506 significant RNAs shared between the linear and negative binomial models with age. Row-wise z-scores were calculated and color-coded to display changes in RNA expression with increasing age (22–83 years old). Light yellow and dark blue bars indicate RNAs with very low and very high expression levels, respectively, at a particular age. In addition, a correlation-based distance method was used to cluster the RNAs obtained from the linear model with positive and negative  $\beta$ -values for age. RNAs with positive  $\beta$ -values are shown in the top of the heat map, containing RNAs with low to high expression (yellow to blue rows) with increasing age. Similarly, RNAs with negative  $\beta$ -values are shown on the bottom of the heat map, containing RNAs with high to low expression (blue to yellow rows) with increasing age.

sarcomere function<sup>16</sup>), *CRIMI* (Cysteine-Rich Transmembrane Bone Morphogenetic Protein Regulator 1, implicated in smooth muscle contractility<sup>17</sup>), *CFAP61* (Cilia and Flagella-Associated Protein 61, highly abundant in old skeletal muscle<sup>18</sup>), *FEZ2* (Fasciculation and Elongation Protein Zeta 2, participating in fasciculation and axonal elongation<sup>6</sup>), *LGII* (Leucine-Rich Glioma-Inactivated 1, differentially expressed in frail compared to healthy individuals<sup>18</sup>), and *MYLK4* (Myosin Light Chain Kinase Family Member 4, necessary for contraction, motility, and cell growth<sup>19–21</sup>) mRNAs (Figs. 2a, b and 3a, b, and Supplementary Figs. 2a, b and 3a,b). These mRNAs were found to be associated with skeletal muscle aging (including *C12orf75* and *LGII* mRNAs) and with exercise (including *C12orf75* and *CFAP61* mRNAs) in previous studies performed in different populations, and using different technologies<sup>6,9,22,23</sup>.

Linear regression plots for the RNAs (ENSGs) ranked 11–20 with positive and negative  $\beta$ -values for age (Supplementary Fig. 4), the top 20 RNAs identified by the negative binomial model with positive and negative  $\beta$ -values for age (Supplementary Figs. 2a and 3a), and negative binomial regression plots for the top 20 RNAs with positive and negative  $\beta$ -values for age are included (Supplementary Figs. 2b, 3b, and 5). Among the top 20 overrepresented RNAs identified by the linear and negative binomial models (Fig. 2a and Supplementary Fig. 2a), there was an overlap of 13 RNAs, of which 12 were protein-coding RNAs. Similarly, among the top 20 underrepresented RNAs obtained by both models (Fig. 3a and Supplementary Fig. 3a), there were 15 RNAs shared, of which 12 were protein-coding RNAs. Overall, 39 distinct protein-coding RNAs were identified among our four top 20 lists (Supplementary Fig. 6).

To validate our most significant results, we used reverse transcription followed by real-time, quantitative PCR (RT-qPCR) analysis. Validation was limited to a few individuals at the two ends of the age distribution (20–34 years vs. 65–79 years) for whom muscle RNA was still available. Eight RNAs that were highly associated with aging in our study, as well as in previous literature, were selected for validation. Four of the eight RNAs (ENSGs) selected for validation [*FAM83B* ( $p = 0.0472$ ), *LGII* ( $p = 0.0283$ ), *NR2F2* ( $p = 0.0278$ ), and *C12orf75* ( $p = 0.0163$ ) mRNAs] had significantly higher expression in the old (80+ years) compared to the young (22–34 years) (Fig. 4 and Supplementary Table 1). Expression levels of *CDKN2B*, *CRIMI*, *IRS2*, and *SKAP2* mRNAs were also higher in the older age group, although not significantly, possibly due to the limited sample size ( $n = 5$ ).

**Proteomic and pathway analyses.** Next, we considered whether differential abundance of these RNAs (ENSGs) in older skeletal muscle was reflected in differences at the protein level. For this

analysis, we selected mRNAs for which we had proteomic data, obtained from the same skeletal muscle using a shotgun liquid chromatography-mass spectrometry method<sup>1</sup> that yielded 4281 proteins. In agreement with previous reports, we found that the correlation coefficients of mRNAs and their corresponding expressed proteins showed a symmetrical, quasi-normal distribution around approximately zero, with more values above than below zero<sup>24</sup> (Fig. 5a). Similarly, adjusted regression models were used to estimate the  $\beta$ -coefficients for age for the 4281 mRNAs, for which corresponding proteomic data were available (age-protein- $\beta$  vs. age-mRNA- $\beta$ ; Fig. 5b gray dots), and separately for 122 mRNAs whose  $\beta$ -coefficients showed significant association ( $p < 0.01$ ) with age (Fig. 5b red dots). Interestingly, there was a significant positive correlation between the  $\beta$ -coefficients, although with substantial noise. Then, the abundance of 122 mRNAs and their respective proteins was explored in the heat map analysis shown in Fig. 5c, with the correlation distribution in Supplementary Fig. 7a. As a comparison, we repeated these analyses for a randomly selected sample of 122 mRNAs. This analysis showed far less correlation with their corresponding proteins (Supplementary Fig. 7b).

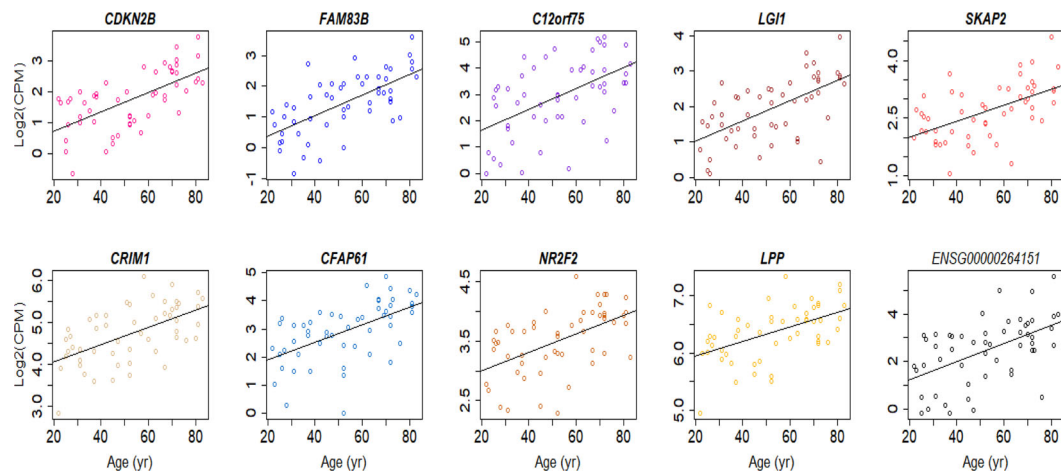
By Kyoto Encyclopedia of Genes and Genomes (KEGG) annotation, the top positively correlated protein–mRNA pairs were found involved in glycolysis/gluconeogenesis (e.g., PGK1, TPI1 (Triose-phosphate Isomerase 1), PFKFB2 (6-Phosphofructo-2-Kinase/Fructose-2,6-Biphosphatase 2), and MYOZ2) and metabolism (e.g., PGK1 and LDHA (Lactate Dehydrogenase A)). Sarcomere proteins such as ANK3, PDLIM, and TNNT3 (troponin T3, fast skeletal type) also showed high protein–mRNA correlations, whereas Histone protein H2A.Z showed the strongest negative correlation (Fig. 5c inset). By functional annotation, 16 proteins were annotated as a muscle function, of which 15 were positively correlated with their respective mRNA levels (Supplementary Fig. 7c). Finally, quantitative proteomic data were available for 14 out of the top 39 mRNAs (ENSGs) showing differential expression with age; among these, four were overrepresented and 10 were underrepresented with age (Supplementary Fig. 6). Of the 14 mRNAs (ENSGs), 10 displayed significant ( $p < 0.05$ ) correlations between protein abundance and age (Supplementary Fig. 8). Nine of these ten proteins displayed significant ( $p < 0.05$ ) correlations with mRNA abundance. Four of these nine proteins are involved in carbohydrate metabolism, namely OXCT1 (3-Oxoacid CoA-Transferase 1, implicated in energy production from ketone bodies<sup>25</sup>), LDHA (involved in anaerobic glycolysis<sup>26</sup>), PFKFB2 (encoding a heart isozyme<sup>27</sup>), and TPI1 (which catalyzes the isomerization of glyceraldehyde 3-phosphate and dihydroxyacetone phosphate<sup>28</sup>). Three of the nine proteins are also direct constituents of the contractile system, namely MYL1 (Myosin Light Chain 1, expressed in fast skeletal muscle and associated with myopathy<sup>29</sup>), TNNT3 (which influences muscle contractility and



**a**

	Rank	RNA Ensembl ID (common name)	P-value	Beta	Associated Functions & Characteristics
Protein-Coding	1	ENSG00000147883 ( <b>CDKN2B</b> )	5.55E-08	0.0328	Aging <sup>9</sup> , p15 tumor suppressor <sup>22</sup> , aneurysm formation <sup>63</sup>
	2	ENSG00000168143 ( <b>FAM83B</b> )	9.92E-08	0.0343	Hypoxia response pathway <sup>14</sup> , aging, skeletal muscle weakness <sup>18</sup>
	3	ENSG00000235162 ( <b>C12orf75</b> )	1.99E-06	0.0419	Energy metabolism, insulin signaling <sup>15</sup> , aging, skeletal muscle weakness <sup>18</sup>
	4	ENSG00000108231 ( <b>LG11</b> )	2.63E-06	0.0284	Aging, skeletal muscle weakness <sup>18</sup> , limbic encephalitis <sup>84</sup> , epilepsy <sup>85</sup>
	5	ENSG0000005020 ( <b>SKAP2</b> )	5.01E-06	0.0218	Sarcomere function and regulation, higher expression in older subjects <sup>16</sup>
	6	ENSG00000150938 ( <b>CRIM1</b> )	8.4E-06	0.0208	Skeletal muscle aging <sup>16</sup> , smooth muscle contractility <sup>17</sup> , CNS development <sup>86</sup>
	7	ENSG00000089101 ( <b>CFAP61</b> )	9.63E-06	0.0320	High expression in old skeletal muscle, aging, skeletal muscle weakness <sup>18</sup>
	8	ENSG00000185551 ( <b>NR2F2</b> )	1.15E-05	0.0160	Myogenesis, skeletal muscle development <sup>12</sup> , aging <sup>23</sup> , muscular dystrophy <sup>87</sup>
	9	ENSG00000145012 ( <b>LPP</b> )	1.29E-05	0.0135	Smooth muscle differentiation <sup>13</sup> , cell migration and proliferation <sup>88</sup>
	11	ENSG00000196482 ( <b>ESRRG</b> )	2.09E-05	0.0196	Type I muscle fiber development, estrogen signaling <sup>34</sup> , obesity susceptibility <sup>45</sup>
	12	ENSG00000185950 ( <b>IRS2</b> )	2.71E-05	0.0189	Insulin signaling <sup>10</sup> , lipid metabolism in skeletal muscle <sup>11</sup> , longevity in mice <sup>89</sup>
	13	ENSG00000171055 ( <b>FEZ2</b> )	3.07E-05	0.0190	Fasciculation, axonal elongation, skeletal muscle aging <sup>6</sup>
	14	ENSG00000168769 ( <b>TET2</b> )	3.3E-05	0.0092	Smooth muscle plasticity <sup>90</sup> , skeletal myogenesis <sup>91</sup> , DNA demethylation <sup>92</sup>
	15	ENSG00000185760 ( <b>KCNQ5</b> )	3.35E-05	0.0165	Potassium channel subunit <sup>93</sup>
	16	ENSG00000115380 ( <b>EFEMP1</b> )	3.74E-05	0.0254	Early aging in mice <sup>70</sup> , lymph node metastasis, angiogenesis <sup>94</sup>
	17	ENSG00000132464 ( <b>ENAM</b> )	4.03E-05	0.0147	Tooth enamel formation <sup>95</sup>
	18	ENSG00000181690 ( <b>PLAG1</b> )	4.6E-05	0.0263	Higher expression in older subjects <sup>16</sup> , bovine fetal development and body size <sup>39</sup>
	19	ENSG00000123836 ( <b>PFKFB2</b> )	4.74E-05	0.0225	Heart isozyme <sup>27</sup>
	20	ENSG00000138439 ( <b>FAM117B</b> )	5.10E-05	0.0161	Sarcoidosis susceptibility <sup>96</sup> , candidate for early onset myocardial infarction <sup>97</sup>
Non-Coding	10	ENSG00000264151	1.43E-05	0.0397	lincRNA

**b**



**Fig. 2 Top 20 significant ( $p < 0.01$ ) RNAs (ENSGs) identified by linear regression with positive  $\beta$ -values for age. **a** Linear model  $p$ -values (two-sided Wald test, unadjusted) and  $\beta$ -values, and RNA characteristics obtained from literature review, are shown. RNAs that failed to have at least ten samples with CPM values above  $-3$  were removed. **b** Linear plots representing  $\log_2(\text{CPM})$  values with age for the top ten RNAs ranked by the lowest  $p$ -value (two-sided Wald test, unadjusted)<sup>36,37,84–96</sup>.**

development<sup>30,31</sup>), and TPM1 (Tropomyosin 1, which modulates actin–myosin interactions<sup>32</sup>) (Fig. 5d). Similar to previous studies, and in spite of the fact that the effect of age on mitochondrial protein representation is higher than any other class of proteins, mRNAs encoding mitochondrial proteins did not show the highest correlation with their respective proteins<sup>33</sup>.

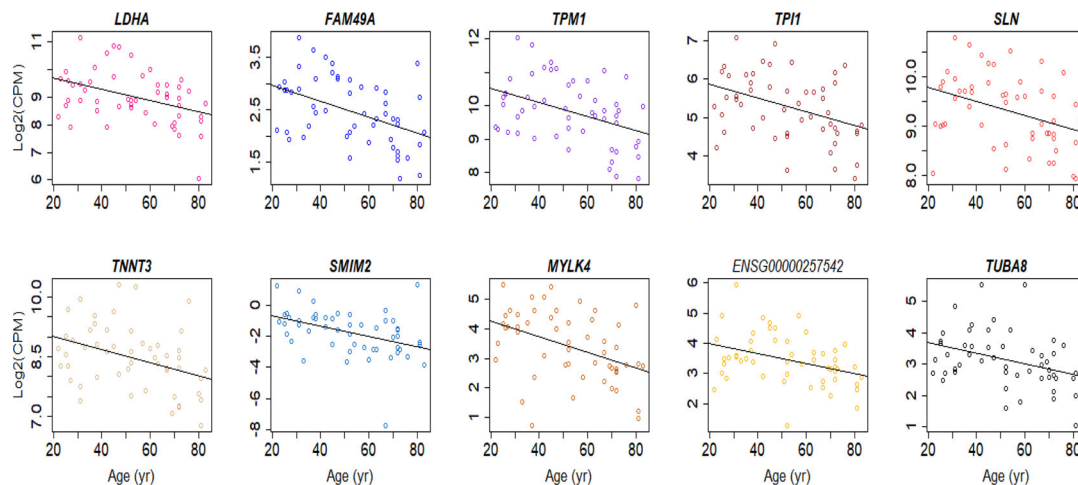
Of the 506 RNAs (ENSGs) significantly associated with age, 478 were found in the Ingenuity Pathway Analysis (IPA) database. IPA identified 27 biological pathways that were significantly ( $p < 0.05$ ) associated with aging, with adipogenesis being the most dysregulated. Next, we investigated the expression patterns of the input genes annotated to this pathway. Expression levels of 12 out of the 132 genes in the adipogenesis pathway were

significantly changed with age (Supplementary Fig. 9, labeled purple).

**Differential transcript usage with age.** To further explore changes in transcripts (ENSTs) that occur in the skeletal muscle with aging, we first identified specific splice variants of skeletal muscle proteins that were differentially altered with aging. For this initial analysis, we focused on genes whose RNAs (ENSGs) were among the four top 20 lists (Figs. 2a and 3a, and Supplementary Figs. 2a and 3a) with significant association with aging. We then tested for associations between differential transcript usage (DTU, see “Methods”) and age as a continuous variable using linear regression, and found changes

**a**

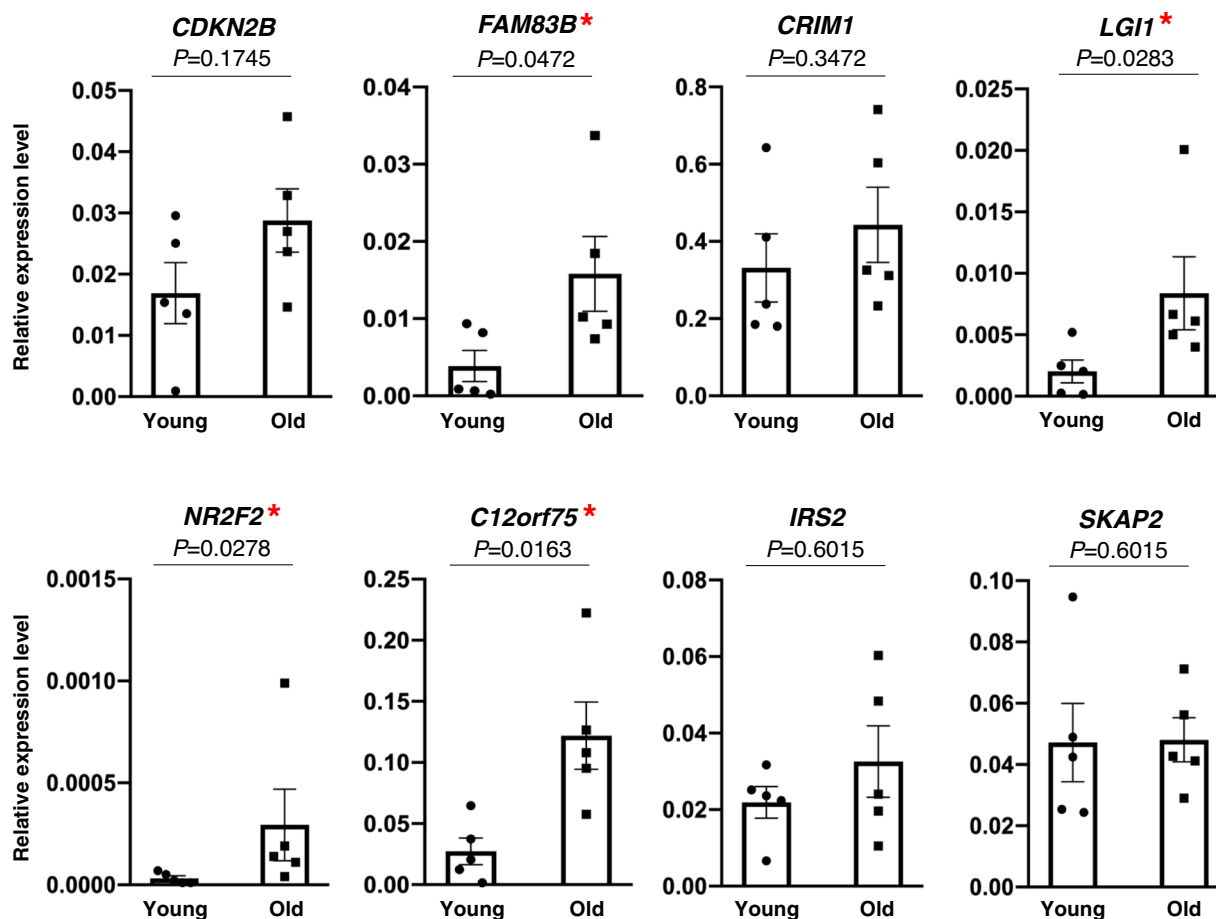
	Rank	RNA Ensembl ID (common name)	P-value	Beta	Associated Functions & Characteristics
Protein-Coding	1	ENSG00000134333 ( <i>LDHA</i> )	0.0003	-0.0222	Porcine energy regulation, glycogen metabolism, and glycolysis of skeletal muscle <sup>98</sup>
	2	ENSG00000197872 ( <i>FAM49A</i> )	0.0008	-0.0154	Increased proteomic expression with age in mice <sup>99</sup> , dementia <sup>100</sup>
	3	ENSG00000140416 ( <i>TPM1</i> )	0.0010	-0.0222	Modulates actin-myosin interactions <sup>32</sup>
	4	ENSG00000111669 ( <i>TPH1</i> )	0.0012	-0.0154	Glycolysis enzyme, energy generation for muscle cells in chickens <sup>101</sup>
	5	ENSG00000170290 ( <i>SLN</i> )	0.0018	-0.0156	Inhibits rat sarcoplasmic Ca <sup>2+</sup> -ATPases <sup>102</sup> , thermogenesis <sup>103</sup> , muscle performance <sup>104</sup>
	6	ENSG00000130595 ( <i>TNNT3</i> )	0.0019	-0.0175	Skeletal muscle contractibility <sup>30</sup> , porcine muscle development, tropomyosin binding <sup>31</sup>
	7	ENSG00000139656 ( <i>SMIM2</i> )	0.0020	-0.0331	Potential indicator of physical activity <sup>105</sup>
	8	ENSG00000145949 ( <i>MYLK4</i> )	0.0024	-0.0250	Aging, weakness <sup>18</sup> , contraction <sup>19</sup> , motility <sup>20</sup> , TGF-beta pathway <sup>21</sup> , circadian rhythms <sup>106</sup>
	10	ENSG00000183785 ( <i>TUBA8</i> )	0.0035	-0.0180	Intracellular transport of macromolecules and cytoskeletal support <sup>107</sup>
	13	ENSG00000180209 ( <i>MYLPF</i> )	0.0047	-0.0185	Fast and slow skeletal muscle development <sup>29</sup> , lower in old caprine skeletal muscle <sup>108</sup>
Non-Coding	14	ENSG00000173436 ( <i>MINOS1</i> )	0.0060	-0.0123	Mitochondrial function and cristae organization <sup>109</sup>
	18	ENSG00000166343 ( <i>MSS51</i> )	0.0071	-0.0131	Skeletal muscle-specific modulator of cellular metabolism <sup>110</sup>
	20	ENSG00000168530 ( <i>MYL1</i> )	0.0088	-0.0115	Expressed in fast-type muscle <sup>14</sup>
	9	ENSG00000257542	0.0031	-0.0173	Olfactory receptor pseudogene
	11	ENSG00000270136	0.0036	-0.0136	MINOS1 readthrough gene (nonsense-mediated decay)
	12	ENSG00000232407	0.0039	-0.0482	Peptidylprolyl isomerase A pseudogene
	15	ENSG00000227258	0.0061	-0.0233	SMIM2 antisense RNA
	16	ENSG00000184844	0.0062	-0.0285	Cytochrome c, somatic pseudogene
	17	ENSG00000234281	0.0062	-0.0202	LANCL1 antisense RNA
19	ENSG00000256364	0.0084	-0.0139	MLEC antisense RNA	

**b**

**Fig. 3 Top 20 significant ( $p < 0.01$ ) RNAs (ENSGs) identified by linear regression with negative  $\beta$ -values for age. **a** Linear model  $p$ -values (two-sided Wald test, unadjusted) and  $\beta$ -values, and RNA characteristics obtained from literature review, are shown. RNAs that failed to have at least ten samples with CPM values above  $-3$  were removed. **b** Linear plots representing  $\log_2(\text{CPM})$  values with age for the top ten RNAs ranked by the lowest  $p$ -value (two-sided Wald test, unadjusted)<sup>97–108</sup>.**

in 1005 RNAs (and their ~1100 splice variants) significantly ( $p < 0.01$ ) associated with age. For each of the identified genes, we compiled a list of splice variants already reported in the literature that were also detected in our RNA-seq analysis. Among these, we searched for evidence of variants that changed systematically with aging and found that five splice variants were identified with significant ( $p < 0.01$ ) changes in DTU with age. These variants were transcribed from four genes: *CFAP61*<sup>18</sup>, *ESRRG* (Estrogen-Related Receptor Gamma, important for type-1 muscle fiber development, estrogen signaling, and obesity susceptibility<sup>34,35</sup>), *TET2* (Tet Methylcytosine Dioxygenase 2, involved in myogenic differentiation of skeletal myoblast cells and muscle age-related decline in mice<sup>36–38</sup>), and *PLAG1* (Pleiomorphic Adenoma Gene 1, known to

be more highly expressed in older humans<sup>16,39</sup>) (Fig. 6a and Supplementary Fig. 10). Linear regression lines for both the significant and nonsignificant age associations of all identified mRNA splice variants in these four genes are plotted in Fig. 6a, b. The expression levels of three out of these five splice variants were significantly ( $p < 0.01$ ) higher with older age (Supplementary Fig. 11). In some cases, we found dissimilar trends in DTU with age on the percentage and  $\log_2(\text{TPM} + 1)$  scales. For example, the DTU of a splice variant of *PLAG1* mRNA (ENST00000423799) on the percentage scale was lower with older age, as indicated by a negative  $\beta$ -value (Supplementary Fig. 10b), whereas the differential levels of this splice variant on the  $\log_2(\text{TPM} + 1)$  scale was virtually constant with older age, indicated by a near-zero slope (Supplementary Fig. 11). To extend

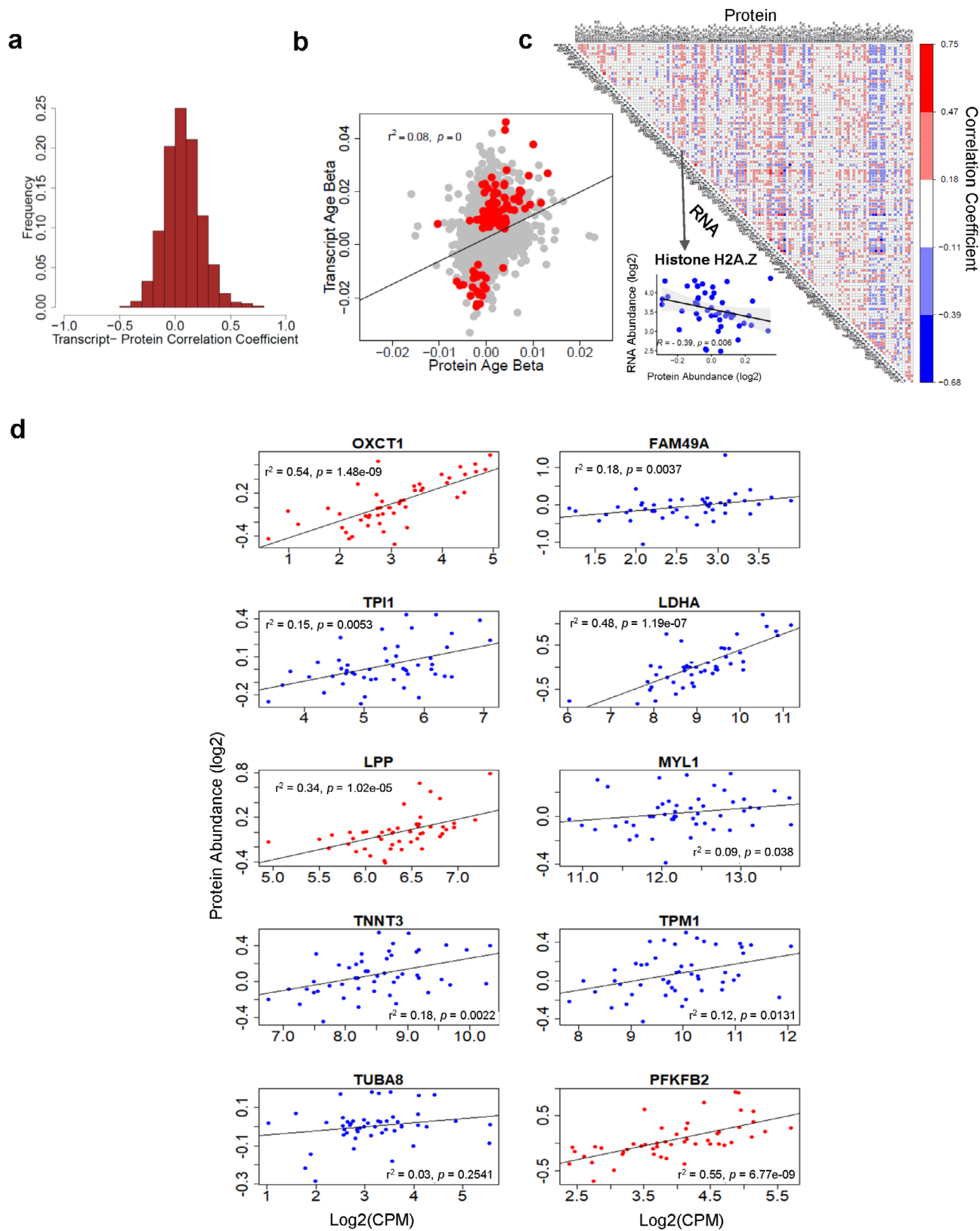


**Fig. 4 Validation of RNA (ENSG) expression patterns with age using RT-qPCR analysis.** Relative expression levels of five young (20–34 years) and five old (65–79 years) participants were calculated relative to *GAPDH* mRNA levels, and were compared using a two-sided Kruskal–Wallis test. Of the eight RNAs selected for analysis, statistically significant changes in expression with age were identified for *FAM83B* ( $p = 0.0472$ ), *LGI1* ( $p = 0.0283$ ), *NR2F2* ( $p = 0.0278$ ), and *C12orf75* ( $p = 0.0163$ ) mRNAs, denoted by a red asterisk. All  $p$ -values are unadjusted. Error bars represent mean  $\pm$  SEM.

this analysis further, we looked at the genes that did not change significantly with age in either the linear or negative binomial model and found two transcripts that changed significantly with age (in opposite directions) from the DTU data (Supplementary Fig. 12a). Out of 890 such genes, we selected *SELENOF* (Selenoprotein F, encoding an mRNA that increases with age and is a target of microRNA mmu-mir-136, which decreases with age in skeletal muscle<sup>40</sup>), and *SLC47A1* (Solute Carrier Family 47 Member 1, encoding a protein that contributes to the tissue distribution and excretion of many drugs and is widely expressed in skeletal muscle<sup>41</sup>) for plotting (Supplementary Fig. 12b–d), selected based on absolute  $\beta$ -values for the transcripts. These findings suggest that the relative contribution of different splice variants to the expression of a certain gene may be as biologically meaningful as the total quantity.

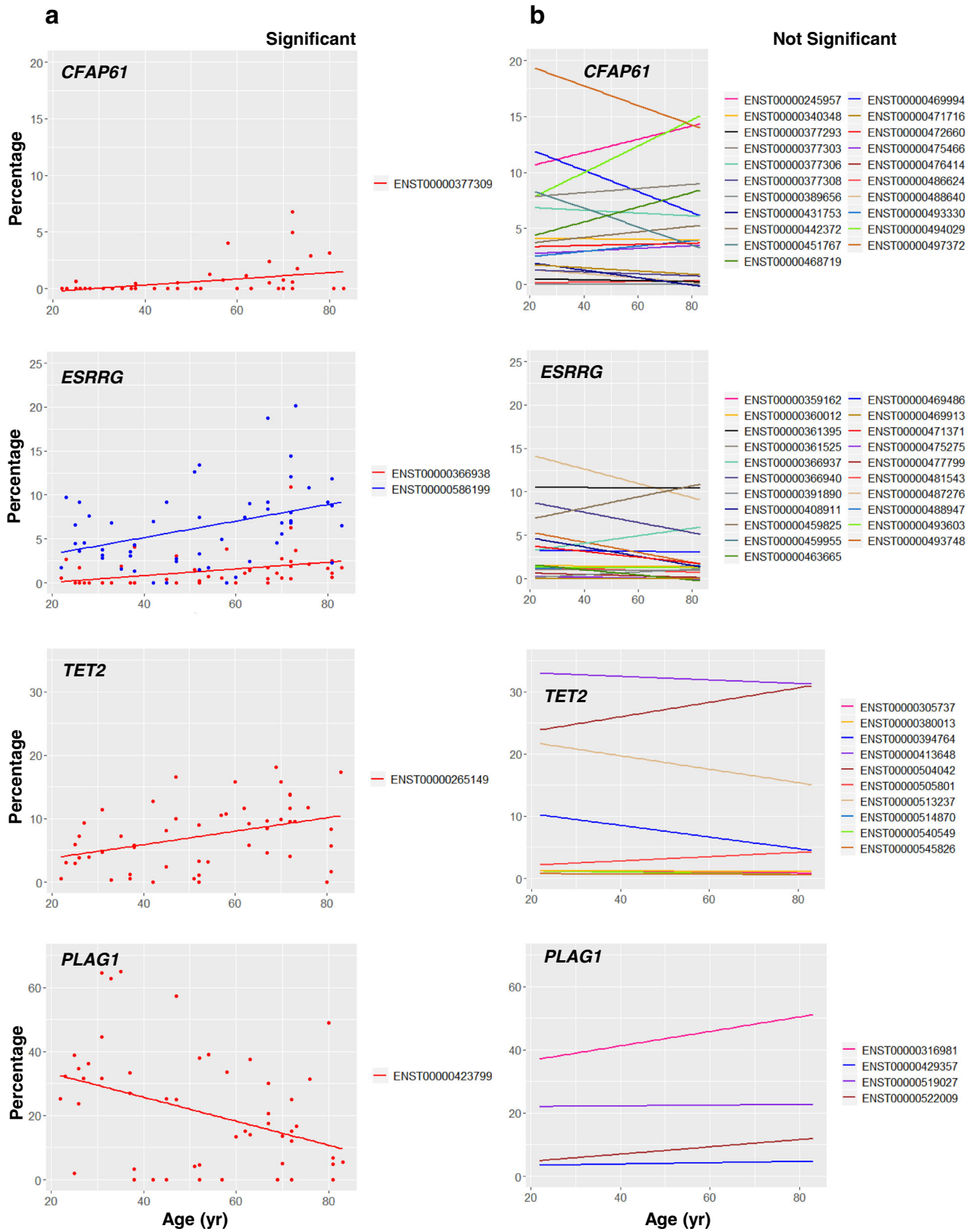
**Differential exon usage with age.** As splice variants are generated by differential exon usage, we determined the level of different exons used by counting the number of reads aligned to each exon, followed by differential exon usage analysis using the DEXSeq package<sup>42</sup>. There were 163 RNAs (ENSGs) with statistically significant (false discovery rate (FDR) < 0.1) differences in exon usage between young (20–34 years) and old (80+ years) individuals. There were 17 RNAs that showed both significant differential exon usage between young and old participants, and significant correlation of DTU with aging (Supplementary Fig. 13a). For these RNAs, a spectral ribbon plot was created to

display changes in exon usage with age (Supplementary Fig. 13b). Next, we created transcriptome-wide UCSC browser tracks and specifically investigated these 17 RNAs. Among them, two examples were selected to illustrate age-related differential exon usage: *SMIM11A* mRNA (Small Integral Membrane Protein 11A, of unknown function) and *RXYLT1* mRNA (Ribitol Xylosyltransferase 1, associated with Walker–Wahlburg syndrome and muscular dystrophy–dystroglycanopathy<sup>43,44</sup>). The UCSC browser coverage plots (read counts per million, CPM) and the age-related exonic expression levels identified by DEXSeq analysis show exons splicing in or out with age (Fig. 7 and Supplementary Fig. 14). We found that exon 11 (labeled purple in the gene model) was completely lost or significantly ( $p < 0.01$ ) reduced in the *SMIM11A* mRNA of older individuals (80+ years) compared to younger (20–34 years) individuals (Fig. 7). Similarly, we found that exon 9 (labeled purple in the gene model) was spliced in or had significantly ( $p < 0.01$ ) higher expression in the *RXYLT1* mRNA when comparing old to young participants (Supplementary Fig. 14). Interestingly, the significantly differentially spliced exon of *SMIM11A* mRNA (ENST00000399292) is a 3′-untranslated region with binding sites for 35 RNA-binding proteins, including ELAVL1 (ELAV-like RNA-Binding Protein 1, also known as Hu antigen R or HuR, which is highly expressed in many cancers and implicated in controlling mRNA turnover and translation), MOV10 (Moloney Leukemia Virus 10, an RNA helicase), UPF1 (Up-Frameshift Suppressor 1 Homolog, a protein involved in mRNA nuclear export and surveillance, and

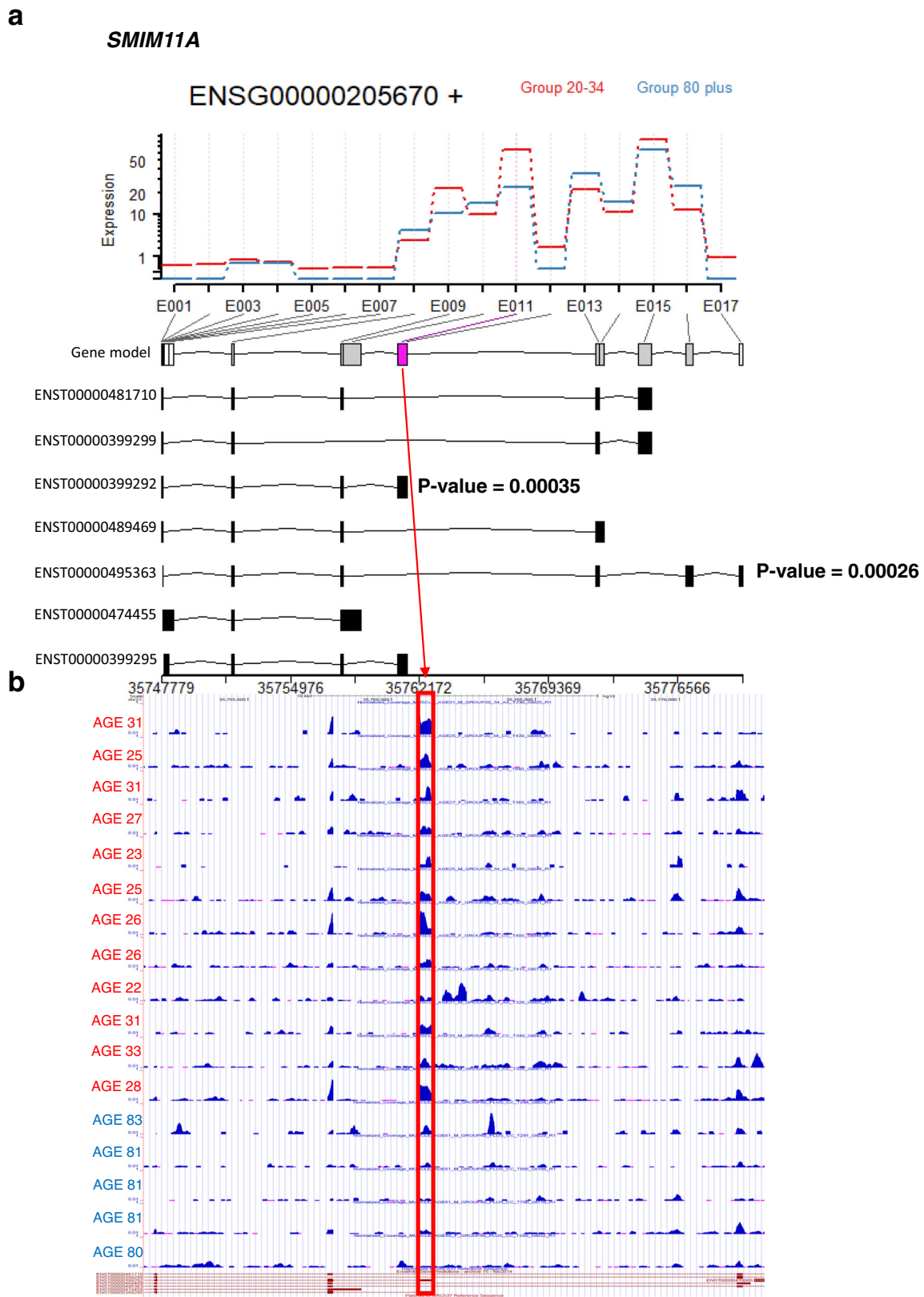


**Fig. 5 Significant ( $p < 0.05$ ) downstream protein abundance and  $\log_2(\text{CPM})$  correlations.** **a** Distribution of 4281 protein-RNA correlation coefficients. **b** Scatterplot of the age  $\beta$ -coefficients for the protein-RNA matches limited to the 122 mRNAs significantly associated with age ( $p$ -value from two-sided Wald test, unadjusted). **c** Heat map of 122 mRNAs significantly associated with age and their correlations with proteins. The mRNA with the strongest negative correlation (H2AZ) is shown in the inset ( $p$ -value from two-sided Wald test, unadjusted). **d** Among all the RNAs in the four top 20 lists, ten RNAs were identified to have statistically significant correlations between relative protein abundance and age (red and blue points indicate mRNAs with positive and negative correlations, respectively). Of these ten RNAs, nine had statistically significant correlations between relative protein [ $\log_2(\text{Protein Abundance})$ ] and RNA abundance [ $\log_2(\text{CPM})$ ] ( $p$ -values from two-sided Wald test, unadjusted).





**Fig. 6 Differential transcript usage analysis for top RNAs (ENSGs).** Among the RNAs in the four top 20 lists, *CFAP61*, *ESRRG*, *TET2*, and *PLAG1* mRNAs were identified to have splice variants with statistically significant changes in differential transcript usage with age. **a** Change in differential transcript usage for all significant variants of the four mRNAs (all  $p < 0.05$  from two-sided, unadjusted Wald tests). **b** Change in differential transcript usage for all nonsignificant variants of the four RNAs, provided for comparison (all  $p > 0.05$  from two-sided, unadjusted Wald tests).



nonsense-mediated mRNA decay), and FUS (Fused in Sarcoma, a protein involved in pre-mRNA processing and mRNA export)<sup>45,46</sup> (Supplementary Fig. 15a). Many of these proteins bind to multiple sites along this exon and are known to regulate RNA abundance and stability (Supplementary Fig. 15b). Further

analysis of *SMIM11A* and *RXYLT1* mRNAs showed that their absolute abundance did not change significantly ( $p < 0.01$ ) with age (Supplementary Fig. 16a–d).

Gene set enrichment analysis (GSEA)<sup>47</sup> was conducted using the 5325 RNAs (ENSGs) with at least one significantly changing splice

**Fig. 7 Changes in the differential exonic usage of *SMIM11A* mRNA in young (20–34 years,  $n = 12$ ) compared to old (80+ years,  $n = 5$ ) age groups.**

**a** Red and blue bars represent the average exon usage for all 17 exons of *SMIM11A* mRNA in the young and old groups, respectively. DEXSeq analysis identified one exonic region (E011) that significantly changed in abundance between the age groups (pink in the gene model). The average exonic usage in this region is higher in the young age group and lower in the old age group. Of the seven variants of *SMIM11A* mRNA, two have statistically significant changes in expression continuously with age according to our linear model. Linear model  $p$ -values (two-sided Wald tests, unadjusted) are displayed for these variants. The horizontal axis below the variant models denotes the positions in the genome. **b** Ages of participants in the young (red) and old (blue) age groups are shown on the vertical axis. UCSC browser tracks indicate that the exons displaying differential expression (location on each transcript identified by red bar) identified in our DEXSeq analysis are more expressed in the young age group than in the old age group, indicated by the presence of larger blue peaks in younger participants. UCSC browser tracks support our DEXSeq findings.

variant with age. GSEA identified six significantly ( $p < 0.05$ ) upregulated pathways in older (80+ years) compared to younger (20–34 years) participants: oxidative phosphorylation, adipogenesis, G2/M checkpoint, MTORC1 signaling, fatty acid metabolism, and ultraviolet response (Supplementary Fig. 17).

Finally, we compared the genes associated with human muscle aging from several past studies, namely Melov et al.<sup>48</sup>, Phillips et al.<sup>2</sup>, Sood et al.<sup>3</sup>, and Timmons et al.<sup>49</sup>, as well as genes induced by exercise, and we found a considerable overlap (Supplementary Fig. 18). In particular, RNAs (ENSGs) expressed from two genes were identified by all five studies including the current study (*LGII* and *FEZ2*), 11 were identified by four out of five studies (*CAMTA1*, *COL4A5*, *KLF5*, *CYP11B1*, *USP54*, *CTH*, *PLAG1*, *NT5C2*, *TP11*, *PFKFB2*, and *UNC13C*), 44 were identified by three out of five studies, and 228 were identified by two out of five studies. Considering all five studies, our findings were replicated for 285 transcripts, whereas 82 transcripts identified as significantly changing with age in our study were replicated in at least one other study (Supplementary Fig. 18a, b). The overlap between transcripts that changed with exercise in Phillips et al.<sup>2</sup> and Keller et al.<sup>50</sup>, and those that changed with aging in our study shows eight RNAs associated with aging in our study, which were impacted by exercise in both prior studies (expressed from genes *TSPAN2*, *SKAP2*, *CAV1*, *NID1*, *JAM2*, *EDNRA*, *ABCG1*, and *GLCC11*) and 56 RNAs impacted by exercise by at least one of these two studies (Supplementary Fig. 18c, d).

## Discussion

In this study, we used RNA-seq data obtained from muscle biopsies collected from a group of individuals dispersed over a wide age range, who were established to be very healthy based on a strict enrollment criteria to explore transcriptomic changes that occur in healthy aging. Our aim was to ascertain transcriptional changes that occur with aging, without the potentially confounding effects of the ongoing diseases.

We found appreciable overlap between the 506 RNAs (ENSGs, containing 476 gene symbols) found in our study and prior microarray analyses that identified transcripts associated with muscle aging and exercise status (Supplementary Fig. 18). This comparison was important, as microarray analysis can be highly sensitive and quantitative, and can complement RNA-seq analysis, which offers other advantages, such as detection of previously unknown transcripts. We found that 82 age-associated transcripts identified in our study had already been linked to muscle aging in previous studies, namely *FAM83B*, *C12orf75*, *LPP*, *SKAP2*, *CRIM1*, *FEZ2*, *LGII*, and *MYLK4* mRNAs<sup>3,6,9,22,23,49</sup>. In addition, eight transcripts also identified in our study were differentially regulated with frailty and were reversed by physical activity (*TSPAN2*, *SKAP2*, *CAV1*, *NID1*, *JAM2*, *EDNRA*, *ABCG1*, and *GLCC11* mRNAs)<sup>2,9</sup>. The replication of these findings using different methods and in a highly selective healthy population strongly suggests that the encoded proteins are profoundly involved in skeletal muscle aging and raises the possibility that when these changes are overt, they may accompany the development of frailty.

Many of the transcripts (ENSTs) identified in our study as associated with aging encode proteins that are implicated in mechanisms of muscle maintenance. *FAM83B* is a member of the *FAM83* protein family, characterized by the presence of a DUF1669 domain, which is implicated in a myriad of cellular processes through the regulation of casein-kinase 1, including wingless-related integration site (Wnt) signaling, which is imperative for muscle repair and maintenance<sup>51</sup>.

*C12orf75* (also known as *OCC1*) has been identified as a primary Wnt signaling regulator, although the specific mechanism is still unclear<sup>52</sup>, and *C12orf75* mRNA was found to be over-represented in the quadriceps muscles of patients with chronic pulmonary diseases compared to controls<sup>15</sup>.

LPP is mostly expressed in the cytoskeleton of striated muscle (heart and skeletal muscle) and contributes to mechanoreception, sensing of cross-binding activity, and mitochondrial signaling. Myocyte contraction regulates the production of muscle LIM proteins (MLPs, including LPP) and their nuclear import. In the nucleus, MLPs activate multiple myogenic transcription factors, such as myoblast determination protein 1 and myogenin. In the cytoplasm, MLPs regulate autophagy, modulate the assembly of macromolecular complexes along sarcomeres and the cytoskeleton, and crosslink actin filaments into bundles<sup>53</sup>. Through these mechanisms, MLPs contribute to muscle mass maintenance and continuous repair. Consistent with these findings, elevated MLP levels were reported in skeletal myopathies, such as facioscapulothoracic muscular dystrophy, nemaline myopathy, and limb girdle muscular dystrophy. The specific cause of LPP dysregulation with aging is unknown, but it may reflect the need to upregulate repair mechanisms in the muscle that has accumulated damage during the aging process.

*SKAP2* modulates immunity<sup>54</sup>, influences sarcomere function and regulation, and was found to be elevated in muscle cells of older human subjects<sup>16</sup>. Upregulation of *SKAP2* was identified in previous studies of muscle aging, both in vivo and in vitro. Microdamage likely accumulates with aging in myofibers and triggers repair by overexpressing integrins at the site of the insult<sup>55,56</sup>. Integrins signal for macrophage adhesion, migration, and chemotaxis, and stimulate the remodeling of the macrophage cytoskeleton, thereby contributing to changes in cell shape, motility, and directionality<sup>57,58</sup>. *SKAP2* modulates macrophage responses to integrin signals and affects the efficiency of repair mechanisms. Therefore, overexpression of *SKAP2* with aging may be a marker of active repair in response to damage accumulation. Of note, skeletal muscle *SKAP2* mRNA was recently found among a handful of transcripts that monitor a midlife metabolic switch in humans unrelated to the Target of Rapamycin complex pathway<sup>49</sup>.

*CRIM1* is essential for embryonic heart and endothelial cell development and homeostasis in the coronary vasculature, but its role in skeletal muscle physiology and aging is unknown<sup>59</sup>. *FEZ2* acts as a cargo transport adaptor in kinesin-mediated movement and, through this mechanism, has important roles in axonal bundling and elongation, autophagy, and apoptosis<sup>6,60</sup>. These

activities are potentially important to support continuous reinnervation in response to denervation that is essential for preserving muscle homeostasis with aging. The rise in FEZ2 levels with age in skeletal muscle suggests an attempt to counteract the loss of motor units and supports the notion that FEZ2 plays a role in age-related sarcopenia<sup>16</sup>. Interestingly, elevations in physical activity and testosterone treatment reduce FEZ2 levels, resulting in an upregulation of IGF-1 and increased muscle anabolism<sup>61</sup>. MYLK4 phosphorylates the regulatory light chain of myosin and plays a vital role in muscle development; polymorphisms of this gene have been associated with increased muscularity in cows<sup>62</sup>. Given that knowledge of the functions of these proteins is limited, especially in the skeletal muscle and aging, their deeper investigation is essential to elucidate the causes leading to age-related sarcopenia.

Our study also confirmed several age-associated mRNAs (ENSGs) encoding proteins with predicted roles in muscle aging reported by Phillips et al.<sup>2</sup>, including *CDKN2B*, *IRS2*, and *CRIM1* mRNAs. The overrepresentation of the *CDKN2B* mRNA with aging is particularly interesting, as the encoded protein, also known as the p15 tumor suppressor<sup>22</sup>, inhibits cell proliferation, has been associated with cellular senescence<sup>23</sup>, is a major hallmark of aging, and has been implicated in p53-dependent smooth muscle cell apoptosis<sup>63</sup>. The traditional marker of cellular senescence, p16 (*CDKN2A*), was not identified in the human muscle, although it increases relatively frequently in intermuscular and intramuscular adipose tissue<sup>64</sup>.

*IRS2* is strongly associated with insulin resistance to glucose uptake and lipid metabolism in the skeletal muscle, as well as type 2 diabetes<sup>11,65</sup>, but is also involved in skeletal muscle growth and metabolism via its downstream effector molecules Akt and AMPK<sup>66</sup>. *IRS2* mRNA expression levels decline in conditions of insulin resistance and, at least in a mouse model, account for the high risk of myocardial dysfunction associated with chronic insulin resistance and diabetes<sup>67</sup>. Whether decline in *IRS2* also negatively affects muscle function, independent of its substrate uptake functions, is unknown. Although there is abundant epidemiological and experimental evidence that insulin resistance increases with aging, humans over the age of 90 years have less insulin resistance than might be expected and centenarians are surprisingly insulin-sensitive, suggesting that healthy aging and extreme longevity are strongly connected with the avoidance of glucose intolerance and diabetes<sup>68</sup>. The consistently high levels of *IRS2* transcripts in older compared to younger participants may be due to the fact that the strict selection criteria for “very healthy” used in this study selected older individuals that had escaped major chronic diseases, including cardiovascular pathology, cancer, and diabetes, and therefore did not become glucose intolerant. High *IRS2* mRNA levels in the older age group of these healthy and highly functional individuals is also consistent with the hypothesis that compromised insulin signaling plays a role in the development of sarcopenia and frailty<sup>66,69</sup>.

*CRIM1* has been implicated in pelvic smooth muscle contractility in mice<sup>17</sup> and skeletal muscle aging in humans, where its expression levels were elevated<sup>6</sup>. Other examples of mRNAs with higher expression in older age include *PLAG1* mRNA<sup>16,39</sup> and *EFEMP1* mRNA (EGF-Containing Fibulin Extracellular Matrix Protein 1, which induces early aging in mice<sup>70</sup>). Based on the strong connection of these known genes to muscle physiology, we propose that all 506 significant ( $p < 0.01$ ) RNAs obtained by both the linear and negative binomial models appear robustly involved in healthy skeletal muscle aging.

IPA was used to identify specific biological mechanisms associated with the 506 significant ( $p < 0.01$ ) RNAs (ENSGs) found by both models. Similar to Phillips et al.<sup>2</sup>, we identified several significantly enriched pathways linked to the top 506 differentially

abundant RNAs. The most significant pathway associated with this subset of genes, adipogenesis, did not show directionality of regulation with age. However, it is widely known that aging is associated with increased adiposity within the muscle tissue. Of interest, some of our top mRNAs implicated in adipogenesis, including *SOX9* mRNA (SRY-Box 9, necessary for avoiding skeletal malformation) and *SIRT1* mRNA (Sirtuin 1, which influences skeletal muscle differentiation, cellular senescence, and aging), encode proteins with essential roles in skeletal development, inflammation, cellular senescence, and aging<sup>45</sup>.

After analyzing changes in total RNA, further analyses were aimed at understanding whether specific mRNA splice variants among the RNAs (ENSGs) in the 4 top 20 lists (Figs. 2a and 3a, and Supplementary Figs. 2a and 3a) were significantly associated with aging. In particular, we aimed at understanding whether the relative contribution of different splice variants in a specific gene changed systematically with aging.

DEXSeq and linear regression analyses identified *SMIM11A* and *RXYLT1* mRNAs as having significant ( $p < 0.01$ ) changes in exon usage and splice variant usage with age. The biological underpinnings of these age-related changes are unclear, but changes in alternative splicing with aging have already been suggested from proteomic studies showing that skeletal muscle proteins that belong to the splicing machinery are massively overexpressed with aging and significantly underexpressed in response to physical activity<sup>1,71</sup>. Consistent with these results, we found that the group of genes identified as having at least one splice variant changing with aging was enriched for proteins related to oxidative phosphorylation, adipogenesis, and MTORC1 signaling. Based on these findings, we propose that alternative splicing is an adaptive mechanism aimed at counteracting the loss of efficiency of oxidative phosphorylation that is often observed with aging and restored by physical activity. This hypothesis should be explored in future studies.

Our study has a number of features that provide valuable contributions to the literature. We cataloged muscle transcriptomes from a human population dispersed over a wide age range, who were confirmed to be healthy based on very strict inclusion criteria. We used state-of-the-art RNA-seq technology that has not been previously used to study skeletal muscle in a human population of this size. To identify RNAs that systematically change with aging in healthy skeletal muscle, we performed statistical analyses using both linear and negative binomial regression models, and integrated both analyses. Thus, we could identify transcripts whose abundance change homogeneously or exponentially with aging. It is likely that other meaningful age-related patterns of transcript change exist that could not be detected by our analysis (e.g., early- or late-stage changes). However, detecting these changes with a sufficient degree of confidence would have required a larger sample size and this should be addressed in future studies. Of note, the two statistical models used in this study identified considerably more RNAs with positive  $\beta$ -values than negative  $\beta$ -values for age, which is commonly observed when analyzing proteomic and transcriptomic data. We validated eight highly significant RNAs identified in our study by performing RT-qPCR analysis for a number of key transcripts and confirmed the directionality of the changes observed by RNA-seq and proteomic analyses. Unfortunately, we did not have enough tissue samples to perform a more extensive validation.

In conclusion, our study replicated the findings of multiple prior studies performed in different populations and with different technologies (Supplementary Fig. 18), but also found other transcripts for genes whose implications in the aging process have not been previously investigated. We found evidence that some specific splice variants change systematically with aging, and that



this age-associated trend is not necessarily limited to mRNAs that change overall with aging. We note that changes in splice variants with aging tend to cluster in proteins that are involved in oxidative phosphorylation, lipid metabolism, and MTORC1 regulation, raising the possibility that alternative splicing of these genes may be compensatory to the decline of mitochondrial function with aging. These findings are particularly meaningful, because they are devoid of the potentially confounding effects of disease, as we studied a larger group of very healthy, older individuals. However, recruiting such a study population is challenging and some limitations intrinsic to the limited sample size, such as nonlinear changes in the expression and a more comprehensive identification of age-related splicing events, should be considered in future, larger studies. Elucidating the mechanisms that link these transcripts to aging will provide valuable insights into the pathophysiology of age-related sarcopenia and will uncover targets for preventive and therapeutic interventions.

## Methods

**Study population.** Skeletal muscle biopsies were obtained from very healthy GESTALT participants using the methods outlined in Ubaida-Mohien et al.<sup>1</sup>. Participants were excluded from the study if they consumed medication for chronic illness, trained professionally, had a body mass index  $\geq 30$  kg/m<sup>2</sup>, or were diagnosed with cognitive damage, a physical impairment, or a major disease. The Clinical Research Unit of the National Institute on Aging developed these criteria based on medical history, physical exams, and blood tests evaluated by a trained nurse practitioner<sup>72</sup>. Muscle RNA-seq data obtained from the left vastus lateralis of 53 participants over a large age range (22–52 years,  $n = 28$ ; 53–83 years,  $n = 25$ ) was used for this study. All participants underwent a consenting process that entailed a detailed description of the study, including potential risks. All participants signed an informed consent (provided in Supplementary Information) to participate in the study. The study protocol was approved by the NIH Institutional Review Board (provided in Supplementary Information) and complied with all ethical regulations for research with human subjects.

**Sample preparation and sequencing.** After RNA was extracted from the skeletal muscle biopsies, cDNA libraries were prepared and analyzed (Fig. 1a). Total RNA was isolated in QIAcube, Qiagen, cDNA was synthesized using the NuGen Ovation v2 system, and Illumina libraries were generated with the TruSeq ChIP Library Preparation Kit (Sets A [IP-202-1012] and B [IP-202-1024]). RNA was sequenced using the Illumina HiSeq 2500 sequencing system at a depth of  $> 80$  million single-end reads. Out of the 57,773 RNAs in the Ensembl hg19 v82 (September 2015) database, we identified, on average, 24,453 (ranging from 16,203 to 36,119) RNAs at a coverage depth of  $\geq 10$  reads.

**Alignment and quality control.** After sequencing, output data in the form of FASTQ files (combination of FASTA sequences with corresponding quality data) was cleaned using cutadapt (v2.7)<sup>73</sup>. Quality was checked using the fastqc program (v0.11.8)<sup>74</sup>. Reads were aligned using STAR Aligner (v2.4.0j), a high-throughput alignment software capable of mapping full-length RNA sequences and discovering non-canonical splices and fusion transcripts<sup>75</sup>. BAM files obtained as the output from the STAR alignment were then processed using featureCounts (v1.4.6-p5) from the Subread package (v1.6.4)<sup>76</sup>, to count reads to various genomic features such as exons, transcripts (splice variants), and genes (composite model of all splice variants). Further data filtering and statistical analyses were conducted using RStudio (version 1.2.1335) and multiple R libraries.

**Regression models and visualization.** All RNAs that were not expressed across any of the participants (total 568) were removed from the regression analysis (57,773–568 = 57,205 remaining). To analyze RNA (ENSG) expression levels continuously with age, read counts for 57,205 RNAs were converted to log<sub>2</sub>-transformed CPM using the edgeR package (multiple versions – 3.22.5)<sup>77</sup>. This log<sub>2</sub>(CPM) transformation was implemented for normalization. Next, linear regression models were built for each RNA using the MASS package (v7.3-51.5)<sup>78</sup>. For each RNA, we also built a negative binomial regression model, a method commonly used to analyze count data (non-negative integer values). Moreover, given that RNA-seq data mostly follow a negative binomial distribution<sup>79</sup>, we employed an alternate strategy, as the negative binomial distribution requires input of integer values  $\geq 0$ . To address this obstacle, all read counts were instead converted to normalized counts per billion (CPB)<sup>80</sup> and then rounded to the nearest integer for negative binomial regression using glm.nb from the MASS package. Rounding was completed after converting to CPB, rather than CPM, to minimize noise. As most CPM values were only single- or double-digit integers, rounding these small numbers would have introduced much larger error. In addition, we

incorporated a zero-adjustment in the negative binomial models for RNAs with one or more values of zero CPB using hurdle from the pscl package (v1.5.2)<sup>81</sup>. Standard zero-inflation was not used, as our data included only structural (true) zeros and no sampling (random) zeros. Both the linear and negative binomial models included only age and sex as covariates. Of the 57,205 RNAs, there were 544 RNAs with excess zero counts removed from the negative binomial model, in which 203 of these RNAs contained CPB values of zero in  $> 50$  samples. Over-abundance of zeros prevented model construction. An additional 13 RNAs were removed due to very low expression levels (all median CPB  $< 630$ ). *P*-values and  $\beta$ -values from both models were extracted, sorted, and compared to identify RNAs with the most significant changes in expression with age. Only RNAs with at least ten log<sub>2</sub>(CPM) values of  $\geq -3$  were included in our linear model results. Similarly, only RNAs with at least ten CPB values of  $\geq 125$  were included in our negative binomial model results. Both of these thresholds translate to 10 reads of RNA per 80 million total aligned reads for an individual and were implemented to combat very low expression levels and noise.

Next, volcano plots were created to visualize the distribution of RNAs (ENSGs) with statistically significant ( $p < 0.01$ ) expression patterns with age. A conservative *p*-value threshold of 0.01 was used to obtain only RNAs with the most significant changes in expression with age. Volcano plots were generated for the linear and negative binomial model results, in which red and blue points indicate significant RNAs with positive and negative  $\beta$ -values for age, respectively. Two separate volcano plots were created to display the standard and zero-adjusted negative binomial model results. The shared RNAs between the linear and negative binomial models were used for further analysis. A heat map was created using heatmap.2 from the gplots package (v3.0.1.1)<sup>82</sup> to visualize the expression level changes of these significant ( $p < 0.01$ ) RNAs using the standard complete linkage (hierarchical) clustering method. Row-wise *z*-scores were calculated for each RNA and color-coded based on the expression level across age, in which yellow and blue bars represent RNAs with low and high expression, respectively, at a particular age.

The top 20 significant ( $p < 0.01$ ) RNAs (ENSGs) shared by both models with positive and negative  $\beta$ -values for age were studied at a greater depth. We explored the functions and characteristics of these top RNAs through literature review, to identify musculoskeletal processes and other related biological mechanisms found to be associated with these RNAs in previous major studies. GeneCards (v4.12)<sup>45</sup> was also used to explore RNA and protein functionality. Linear and negative binomial plots were generated to display the expression patterns of these highly significant ( $p < 0.01$ ) RNAs as a function of aging. In addition, median read CPB within the five age groups were calculated and included in each negative binomial plot, denoted by red asterisks.

**Validation by RT followed by real-time qPCR analysis.** We validated a subset of the top 20 significantly changed RNAs identified by both regression models using RT-qPCR analysis. These eight highly significant ( $p < 0.01$ ) RNAs obtained from the linear model with positive  $\beta$ -values for age were selected based on literature review. For each RNA, relative expression levels of five young participants (20–34 years) were compared to five old participants (65–79 years) using the Kruskal–Wallis test. Relative expression levels were normalized to levels of the *GAPDH* mRNA, encoding the housekeeping protein GAPDH. We used a non-parametric test, because our limited sample size ( $n = 5$ ) in each age group prevented us from assuming relative expression levels originated from the same underlying population. We were constrained to a small sample size and limited amounts of RNA sample obtained from each muscle biopsy. Therefore, we were unable to expand the RT-qPCR validation in this study. We also recognize that using *GAPDH* mRNA, which is commonly employed for normalization, is a limitation of this study due to its role in glycolysis, an age-modulated metabolic pathway. This shortcoming may account for our low validation rate. A complete list of primers used for RT-qPCR, including names and sequences, is provided in Supplementary Information.

**Proteomic analysis.** Using the 4281 proteins reported by Ubaida-Mohien et al.<sup>1</sup>, we performed correlation analysis with the corresponding mRNAs (ENSGs). In further analysis, 122 age-associated mRNAs were correlated with their respective proteins and correlation coefficients were measured. A heat map of correlations between these mRNAs and proteins was generated using the Corrrplot (v0.84) library in R (v3.6.1). A random set of 122 mRNAs and their corresponding 122 proteins were chosen for random analysis. We then selected the four top 20 transcript lists to identify RNAs with statistically significant ( $p < 0.05$ ) correlations between relative protein abundance and age. Correlations between relative protein abundance and log<sub>2</sub>(CPM) helped us to evaluate relationships between transcriptional and translational changes in expression. This also allowed us to emphasize the biological importance of the RNAs changing most significantly with age.

**Pathway analysis.** We conducted IPA (summer 2019 release) to explore the biological processes associated with the significant ( $p < 0.01$ ) RNAs (ENSGs) obtained from both regression models. For this analysis, input genes were matched to annotated pathways (e.g., the KEGG adipogenesis pathway is annotated with 132

genes). Input genes that were matched to the pathways were counted and the significance of the pathways was measured by calculating  $z$ -scores based on input RNA  $p$ -values and  $\beta$ -values for age from linear regression, and the activated state, a predetermined subset of the most active genes in a pathway. After correcting for technology and tissue bias, adipogenesis remained the most significantly changed pathway identified by IPA.

**Splice variant and exon usage analyses.** Next, we investigated which specific forms of the top RNAs (ENSGs) identified by both models with positive and negative  $\beta$ -values for age were most significantly associated with healthy skeletal muscle aging. All splice variants (ENSTs) considered were obtained from the Ensembl hg19 v82 (September 2015) database. Kallisto (v0.44) software was used to find transcripts per kilobase million (TPM) values of 196,354 splice variants, associated with the 57,205 total RNAs analyzed previously. These TPM values were converted to log<sub>2</sub>-normalized transcripts per million [ $\log_2(\text{TPM} + 1)$ ]. Instead of assessing absolute splice variant expression patterns with age, splice variant reads for each participant were converted into a percentage of the participant's total variant counts for each top RNA. This percent change method was used to determine how DTU changed with age. Afterwards, linear regression models were built and plotted using the ggplot2 package (v3.2.1)<sup>83</sup> to visualize how the proportions of variant reads changed with respect to all other variants of each significant RNA (ENSG) with age. All  $p$ -values and  $\beta$ -values were extracted to obtain candidate splice variants (ENSTs) involved in healthy skeletal muscle aging. In this exploratory step, a less conservative  $p$ -value threshold of 0.05 was used for all splice variant analysis. In addition, for each significant splice variant identified, a linear model was constructed and plotted to display its expression in read CPM with age.

We also used the DEXSeq package (v1.26)<sup>42</sup> to investigate changes in differential exon usage between young (20–34 years,  $n = 12$ ) and old (80+ years,  $n = 5$ ) participants. This strategy allowed us to identify major skipped exons. Data were converted to bed files for DEXSeq analysis. Here, a more stringent FDR-adjusted  $p$ -value threshold of 0.1 was used to obtain the most significant candidate RNAs showing group-wise differences in exon usage among their splice variants. In parallel, we also assessed splice variant (ENST) expression patterns with age using linear regression and a  $p$ -value threshold of 0.01, and compared these findings to our DEXSeq results. A conservative  $p$ -value threshold of 0.01 was used, as correction for multiple testing (Benjamini–Hochberg) drastically reduced the number of significant transcripts, likely due to the small sample size ( $n = 53$ ). Of all the shared RNAs, two RNAs with highly significantly differentially expressed splice variants were selected for further analysis. These expression levels were visualized using the UCSC Genome Browser, to determine whether age-related changes in exon expression within respective regions corresponded to the regions identified in our DEXSeq analysis.

Lastly, GSEA (MSigDB v6.2)<sup>47</sup> was conducted to explore the biological pathways associated with the RNAs (ENSGs) with at least one significantly changing splice variant with age, obtained from linear regression. For this analysis, we again compared young (20–34 years,  $n = 12$ ) and old (80+ years,  $n = 5$ ) participants.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

All RNA-seq data used to generate Figs. 1–3, 6, and 7, and Supplementary Figs. 2–5 and 9–17 are deposited in GEO (GSE164471), and data obtained from the Ensembl hg19 v82 (September 2015) database can be located here: <ftp://ftp.ensembl.org/pub/grch37/release-82>. The mass spectrometry proteomics data used to generate Fig. 5 and Supplementary Figs. 6–8 have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD011967. Additional data inquiries or requests can be directed to the corresponding author [L.F.]. Source data are provided with this paper.

## Code availability

All R code used for this study is available upon request from the corresponding author [L.F.].

Received: 18 July 2019; Accepted: 2 March 2021;

Published online: 01 April 2021

## References

- Ubaida-Mohien, C. et al. Physical activity associated proteomics of skeletal muscle: being physically active in daily life may protect skeletal muscle from aging. *Front. Physiol.* **10**, 312 (2019).
- Phillips, B. E. et al. Molecular networks of human muscle adaptation to exercise and age. *PLoS Genet.* **9**, e1003389 (2013).
- Sood, S. et al. A novel multi-tissue RNA diagnostic of healthy ageing relates to cognitive health status. *Genome Biol.* **16**, 185 (2015).
- Drummond, M. J. et al. Aging and microRNA expression in human skeletal muscle: a microarray and bioinformatics analysis. *Physiol. Genomics* **43**, 595–603 (2011).
- Thalacker-Mercer, A. E. et al. Differential genomic responses in old vs. young humans despite similar levels of modest muscle damage after resistance loading. *Physiol. Genomics* **40**, 141–149 (2010).
- Su, J. et al. A novel atlas of gene expression in human skeletal muscle reveals molecular changes associated with aging. *Skelet. Muscle* **5**, 35 (2015).
- Patel, H. P. et al. Lean mass, muscle strength and gene expression in community dwelling older men: findings from the Hertfordshire Sarcopenia Study (HSS). *Calcif. Tissue Int* **95**, 308–316 (2014).
- Tanaka, T. et al. Plasma proteomic signature of age in healthy humans. *Aging Cell* **17**, e12799 (2018).
- Jiang, S. S. et al. Gene expression profiling suggests a pathological role of human bone marrow-derived mesenchymal stem cells in aging-related skeletal diseases. *Aging (Albany NY)* **3**, 672–684 (2011).
- Kerouz, N. J. et al. Differential regulation of insulin receptor substrates-1 and -2 (IRS-1 and IRS-2) and phosphatidylinositol 3-kinase isoforms in liver and muscle of the obese diabetic (ob/ob) mouse. *J. Clin. Invest* **100**, 3164–3172 (1997).
- Bouzakri, K. et al. siRNA-based gene silencing reveals specialized roles of IRS-1/Akt2 and IRS-2/Akt1 in glucose and lipid metabolism in human skeletal muscle. *Cell Metab.* **4**, 89–96 (2006).
- Lee, H. J. et al. Dysregulation of nuclear receptor COUP-TFII impairs skeletal muscle development. *Sci. Rep.* **7**, 3136 (2017).
- Gorenne, I. et al. LPP expression during in vitro smooth muscle differentiation and stent-induced vascular injury. *Circ. Res.* **98**, 378–385 (2006).
- Wang, X. Z. et al. Multiple across-strain and within-strain QTLs suggest highly complex genetic architecture for hypoxia tolerance in channel catfish. *Mol. Genet. Genomics* **292**, 63–76 (2017).
- Willis-Owen, S. A. G. et al. COPD is accompanied by coordinated transcriptional perturbation in the quadriceps affecting the mitochondria and extracellular matrix. *Sci. Rep.* **8**, 12165 (2018).
- Mamoshina, P. et al. Machine learning on human muscle transcriptomic data for biomarker discovery and tissue-specific drug target identification. *Front Genet* **9**, 242 (2018).
- Wilkinson, L. et al. Association between congenital defects in papillary outgrowth and functional obstruction in Crim1 mutant mice. *J. Pathol.* **227**, 499–510 (2012).
- Hangelbroek, R. W. J. et al. Expression of protocadherin gamma in skeletal muscle tissue is associated with age and muscle weakness. *J. Cachexia Sarcopenia Muscle* **7**, 604–614 (2016).
- Herrer, I. et al. RNA-sequencing analysis reveals new alterations in cardiomyocyte cytoskeletal genes in patients with heart failure. *Lab. Investig.* **94**, 645 (2014).
- Tsompanidis, A. et al. Ciliary neurotrophic factor upregulates follistatin and Pak1, causes overexpression of muscle differentiation related genes and downregulation of established atrophy mediators in skeletal muscle. *Metabolism* **65**, 915–925 (2016).
- Huang, W. et al. Gene expression patterns in transgenic mouse models of hypertrophic cardiomyopathy caused by mutations in myosin regulatory light chain. *Arch. Biochem. Biophys.* **601**, 121–132 (2016).
- Krimpenfort, P. et al. p15(Ink4a) is a critical tumour suppressor in the absence of p16(Ink4a). *Nature* **448**, 943–U11 (2007).
- Sifakis, E. G. et al. Identifying gender independent biomarkers responsible for human muscle aging using microarray data. In: *2013 IEEE 13th Int. Conference on Bioinformatics and Bioengineering (BIBE)* (IEEE, 2013).
- Wang, D. et al. A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol. Syst. Biol.* **15**, 2 (2019).
- Nam, G. H. et al. Identification of ORF sequences and exercise-induced expression change in thoroughbred horse OXCT1 gene. *Gene* **496**, 45–48 (2012).
- Burgner, J. W. & William, J. R. Jr On the origin of the lactate dehydrogenase induced rate effect. *Biochemistry* **23**, 3636–3648 (1984).
- Manzano, A. et al. Cloning, expression and chromosomal localization of a human testis 6-phosphofructo-2-kinase/fructose-2, 6-bisphosphatase gene. *Gene* **229**, 83–89 (1999).
- Shi, Y. et al. Genetic perturbation of glycolysis results in inhibition of de novo inositol biosynthesis. *J. Biol. Chem.* **280**, 41805–41810 (2005).
- Wang, Y. et al. Fast skeletal muscle regulatory light chain is required for fast and slow skeletal muscle development. *FASEB J.* **21**, 2205–2214 (2007).
- Gurnett, C. A. et al. Skeletal muscle contractile gene (TNNT3, MYH3, TPM2) mutations not found in vertical talus or clubfoot. *Clin. Orthop. Relat. Res.* **467**, 1195 (2009).
- Lobjois, V. et al. A muscle transcriptome analysis identifies positional candidate genes for complex trait in pig. *Anim. Genet.* **39**, 147–162 (2008).

32. Strand, J. et al. Modulation of myosin function by isoform-specific properties of *S. cerevisiae* and muscle tropomyosins. *J. Biol. Chem.* **276**, 34832–34839 (2001).
33. Robinson, M. M. et al. Enhanced protein translation underlies improved metabolic and physical adaptations to different exercise training modes in young and old humans. *Cell Metab.* **25**, 581–592 (2017).
34. Gan, Z. et al. Nuclear receptor/microRNA circuitry links muscle fiber type to energy metabolism. *J. Clin. Invest.* **123**, 2564–2575 (2013).
35. Dong, S. S. et al. Epigenomic elements analyses for promoters identify ESRRG as a new susceptibility gene for obesity-related traits. *Int. J. Obes.* **40**, 1170–1176 (2016).
36. Zhong, Xia et al. Ten-eleven translocation-2 (Tet2) is involved in myogenic differentiation of skeletal myoblast cells in vitro. *Sci. Rep.* **7**, 43539 (2017).
37. Liu, R. et al. Ten-eleven translocation-2 (TET2) is a master regulator of smooth muscle cell plasticity. *Circulation* **128**, 2047–2057 (2013).
38. Gontier, G. et al. Tet2 rescues age-related regenerative decline and enhances cognitive function in the adult mouse brain. *Cell Rep.* **22**, 1974–1981 (2018).
39. Karim, L. et al. Variants modulating the expression of a chromosome domain encompassing PLAG1 influence bovine stature. *Nat. Genet.* **43**, 405 (2011).
40. Kim, J. Y. et al. Genome-wide profiling of the microRNA-mRNA regulatory network in skeletal muscle with aging. *Aging (Albany NY)* **6**, 524 (2014).
41. Wittwer, M. B. et al. Discovery of potent, selective multidrug and toxin extrusion transporter 1 (MATE1, SLC47A1) inhibitors through prescription drug profiling and computational modeling. *J. Med. Chem.* **56**, 781–795 (2013).
42. Reyes, A., Anders, S. & Huber, W. *Inferring Differential Exon Usage in RNA-Seq Data with the DEXSeq Package* (bioconductor.statistik.tu-dortmund.de, 2013).
43. Vajsar, J. & Schachter, H. Walker-Warburg syndrome. *Orphanet J. Rare Dis.* **1**, 29 (2006).
44. Many, H. et al. The muscular dystrophy gene TMEM5 encodes a ribitol  $\beta$ 1, 4-xylosyltransferase required for the functional glycosylation of dystroglycan. *J. Biol. Chem.* **291**, 24618–24627 (2016).
45. Stelzer, G. et al. The GeneCards Suite: from gene data mining to disease genome sequence analyses. *Curr. Protoc. Bioinformatics* **54**, 1.30.1–1.30.33 (2016).
46. Li, J. H. et al. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* **42**, D92–D97 (2013).
47. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
48. Melov, S. et al. Resistance exercise reverses aging in human skeletal muscle. *PLoS ONE* **2**, 5 (2007).
49. Timmons, J. A. et al. Longevity-related molecular pathways are subject to midlife “switch” in humans. *Aging Cell* **18**, e12970 (2019).
50. Keller, P. et al. A transcriptional map of the impact of endurance exercise training on skeletal muscle phenotype. *J. Appl. Physiol.* **110**, 46–59 (2010).
51. Fulcher, L. J. et al. The DUF1669 domain of FAM83 family proteins anchor casein kinase 1 isoforms. *Sci. Signal.* **11**, eaao2341 (2018).
52. Najafi, H. et al. Alternative splicing of the OCC-1 gene generates three splice variants and a novel exonic microRNA, which regulate the Wnt signaling pathway. *RNA* **23**, 70–85 (2017).
53. Hoffmann, C. et al. Human muscle LIM protein dimerizes along the actin cytoskeleton and cross-links actin filaments. *Mol. Cell. Biol.* **34**, 3053–3065 (2014).
54. Boras, M. et al. Skap2 is required for beta(2) integrin-mediated neutrophil recruitment and functions. *J. Exp. Med.* **214**, 851–874 (2017).
55. Geissmann, F. et al. Development of monocytes, macrophages, and dendritic cells. *Science* **327**, 656–661 (2010).
56. Hume, D. A. The mononuclear phagocyte system. *Curr. Opin. Immunol.* **18**, 49–53 (2006).
57. Vicente-Manzanares, M. & Sánchez-Madrid, F. Role of the cytoskeleton during leukocyte responses. *Nat. Rev. Immunol.* **4**, 110 (2004).
58. Worthylake, R. A. & Burridge, K. Leukocyte transendothelial migration: orchestrating the underlying molecular machinery. *Curr. Opin. Cell Biol.* **13**, 569–577 (2001).
59. Iyer, S. et al. Crim1 has cell-autonomous and paracrine roles during embryonic heart development. *Sci. Rep.* **6**, 19832 (2016).
60. Teixeira, M. B., Alborghetti, M. R. & Kobarg, J. Fasciculation and elongation zeta proteins 1 and 2: From structural flexibility to functional diversity. *World J. Biol. Chem.* **10**, 28 (2019).
61. Chen, F. et al. Evaluation of early biomarkers of muscle anabolic response to testosterone. *J. Cachexia Sarcopenia Muscle* **2**, 45–56 (2011).
62. Zheng, L. I. et al. Genetic variant of MYLK4 gene and its association with growth traits in Chinese cattle. *Anim. Biotechnol.* **30**, 30–35 (2019).
63. Leeper, N. J. et al. Loss of CDKN2B promotes p53-dependent smooth muscle cell apoptosis and aneurysm formation. *Arterioscler Thromb. Vasc. Biol.* **33**, e1–e10 (2013).
64. Justice, J. N. et al. Cellular senescence biomarker p16INK4a+ cell burden in thigh adipose is associated with poor physical function in older women. *J. Gerontol. Ser. A* **73**, 939–945 (2017).
65. Withers, D. J. et al. Disruption of IRS-2 causes type 2 diabetes in mice. *Nature* **391**, 900 (1998).
66. Long, Y. C. et al. Insulin receptor substrates Irs1 and Irs2 coordinate skeletal muscle growth and metabolism via the Akt and AMPK pathways. *Mol. Cell. Biol.* **31**, 430–441 (2011).
67. Guo, S. Insulin signaling, resistance, and the metabolic syndrome: insights from mouse models to disease mechanisms. *J. Endocrinol.* **220**, T1 (2014).
68. Barbieri, M. et al. Glucose regulation and oxidative stress in healthy centenarians. *Exp. Gerontol.* **38**, 137–143 (2003).
69. White, M. F. IRS2 integrates insulin/IGF1 signalling with metabolism, neurodegeneration and longevity. *Diabetes Obes. Metab.* **16**, 4–15 (2014).
70. McLaughlin, P. J. et al. Lack of fibulin-3 causes early aging and herniation, but not macular degeneration in mice. *Hum. Mol. Genet.* **16**, 3059–3070 (2007).
71. 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 (2011).
72. Schrack, J. A. et al. “IDEAL” aging is associated with lower resting metabolic rate: the Baltimore Longitudinal Study of Aging. *J. Am. Geriatr. Soc.* **62**, 667–672 (2014).
73. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 3 (2011).
74. Pekkarinen, M. *Preprocessing and Analysis of Single-Cell RNA-Sequencing Data* (Tampere Univ., 2018).
75. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
76. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
77. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
78. Ripley, B. et al. *Package ‘mass’*. *Cran R* (2013).
79. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
80. Muller, L. S. M. et al. Genome organization and DNA accessibility control antigenic variation in trypanosomes. *Nature* **563**, 121–125 (2018).
81. Jackman, S. *Classes and Methods for R Developed in the Political Science Computational Laboratory (version 1.4. 9)* (Department of Political Science, Stanford Univ., 2015).
82. Warnes, M. G. R. et al. *Package ‘gplots’*. *Various R Programming Tools for Plotting Data* (2016).
83. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016).
84. Lai, M. et al. Investigation of LGI1 as the antigen in limbic encephalitis previously attributed to potassium channels: a case series. *Lancet Neurol.* **9**, 776–785 (2010).
85. Kalachikov, S. et al. Mutations in LGI1 cause autosomal-dominant partial epilepsy with auditory features. *Nat. Genet.* **30**, 335 (2002).
86. Kolle, G. et al. CRIM1, a novel gene encoding a cysteine-rich repeat protein, is developmentally regulated and implicated in vertebrate CNS development and organogenesis. *Mech. Dev.* **90**, 181–193 (2000).
87. Xie, X., Tsai, S. Y. & Tsai, M. J. COUP-TFII regulates satellite cell function and muscular dystrophy. *J. Clin. Invest.* **126**, 3929–3941 (2016).
88. Grunewald, T. G. P., Saskia, M. P. & Butt, E. Cell adhesion and transcriptional activity—defining the role of the novel protooncogene LPP. *Transl. Oncol.* **2**, 107–116 (2009).
89. Zhang, J. & Liu, F. Tissue-specific insulin signaling in the regulation of metabolism and aging. *IUBMB Life* **66**, 485–495 (2014).
90. Tsumagari, K. et al. Early de novo DNA methylation and prolonged demethylation in the muscle lineage. *Epigenetics* **8**, 317–332 (2013).
91. Lerche, C. et al. Molecular cloning and functional expression of KCNQ5, a potassium channel subunit that may contribute to neuronal M-current diversity. *J. Biol. Chem.* **275**, 22395–22400 (2000).
92. En-lin, S., Sheng-guo, C. & Hua-qiao, W. The expression of EFEMP1 in cervical carcinoma and its relationship with prognosis. *Gynecol. Oncol.* **117**, 417–422 (2010).
93. Hu, J. C. C. & Yamakoshi, Y. Enamelin and autosomal-dominant amelogenesis imperfecta. *Crit. Rev. Oral Biol. Med.* **14**, 387–398 (2003).
94. Boehm, O. et al. Identification of immune-relevant factors conferring sarcoidosis genetic risk. *Am. J. Respir. Crit. Care Med.* **192**, 727–736 (2015).
95. Yang, X. Use of functional genomics to identify candidate genes underlying human genetic association studies of vascular diseases. *Arteriosclerosis Thrombosis Vasc. Biol.* **32**, 216–222 (2012).
96. Fontanesi, L. et al. Study of candidate genes for glycolytic potential of porcine skeletal muscle: identification and analysis of mutations, linkage and physical



- mapping and association with meat quality traits in pigs. *Cytogenet. Genome Res.* **102**, 145–151 (2003).
97. Koller, A. et al. Quantitative proteomics using 15N SILAC mouse. *J. Proteom. Genomics Res.* **1**, 27 (2013).
98. Huang, P. et al. Differences in the frequency of Alzheimer's disease-associated genomic variations in populations of different races. *Geriatrics Gerontol. Int.* **17**, 2184–2193 (2017).
99. Teltatham, T. & Mekchay, S. Proteome changes in Thai indigenous chicken muscle during growth period. *Int. J. Biol. Sci.* **5**, 679 (2009).
100. Tupling, A. R., Asahi, M. & MacLennan, D. H. Sarcolipin overexpression in rat slow twitch muscle inhibits sarcoplasmic reticulum Ca<sup>2+</sup> uptake and impairs contractile function. *J. Biol. Chem.* **277**, 44740–44746 (2002).
101. Bal, N. C. et al. Sarcolipin is a newly identified regulator of muscle-based thermogenesis in mammals. *Nat. Med.* **18**, 1575 (2012).
102. Anderson, D. M. et al. A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell* **160**, 595–606 (2015).
103. Tikkanen, E., Gustafsson, S. & Ingelsson, E. Fitness, physical activity, and cardiovascular diseases: longitudinal and genetic analyses in the UK Biobank Study. *Circulation* **137**, 2583–2591 (2018).
104. Dyar, K. A. et al. The calcineurin-NFAT pathway controls activity-dependent circadian gene expression in slow skeletal muscle. *Mol. Metab.* **4**, 823–833 (2015).
105. Stanchi, F. et al. TUBA8: a new tissue-specific isoform of  $\alpha$ -tubulin that is highly conserved in human and mouse. *Biochem. Biophys. Res. Commun.* **270**, 1111–1118 (2000).
106. Xu, H. et al. Molecular cloning, sequence identification and expression analysis of novel caprine MYLPF gene. *Mol. Biol. Rep.* **40**, 2565–2572 (2013).
107. Alkhaja, A. K. et al. MINOS1 is a conserved component of mitofilin complexes and required for mitochondrial function and cristae organization. *Mol. Biol. Cell* **23**, 247–257 (2012).
108. Moyer, A. L. & Wagner, K. R. Mammalian Mss51 is a skeletal muscle-specific gene modulating cellular metabolism. *J. Neuromuscul. Dis.* **2**, 371–385 (2015).

## Acknowledgements

We thank all study participants and the teams at the Laboratory of Genetics and Genomics, and the Translational Gerontology Branch of the National Institute on Aging for facilitating our work. This work was supported in its entirety by the National Institute on Aging–Intramural Research Program of the National Institutes of Health (NIH). We also thank Dr. Fred E. Indig (NIA, NIH) and Dr. Maxim Artyomov (Washington University, St. Louis), and their teams, for their contributions to this study.

## Author contributions

R.A.T. III carried out statistical analyses and manuscript preparation. A.H., G.K., C.U.M., and C.C. completed additional statistical analyses. J.H.Y., M.G.F., M.K., A.L., W.H.W. III,

and Y.P. carried out laboratory investigations. L.M.Z. and C.W.C. recruited patients and collected medical data. J.D., M.G., R.S., and S.D. provided guidance for statistical analyses and biological insights, and assisted with manuscript preparation. L.F. planned and designed the study, and served as the principal investigator. All authors commented on the draft manuscript.

## Funding

Open Access funding provided by the National Institutes of Health (NIH).

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-22168-2>.

**Correspondence** and requests for materials should be addressed to L.F.

**Peer review information** *Nature Communications* thanks Shanrong Zhao and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2021