	Adult		Old	
	Control mean ± sem[1]	Exercise mean ± sem[1]	Control mean ± sem[1]	Exercise mean ± sem[1]
Mass 2 Skew	-0.317 ± 0.913	-1.847 ± 0.913	0.977 ± 0.550	-0.323 ± 0.661
Mass 2 N_{skew}	-0.347	-2.0235	1.776	-0.489
Mass 2 Kurt.	-2.04 ± 2.0	3.927 ± 2.0	0.174 ± 1.063	-0.428 ± 1.249
Mass 2 $N_{kurtosis}$	-1.02	1.9635	0.1637	-0.3346
%Dif Grip Skew	1.341 ± 0.913	0.77 ± 0.913	0.340 ± 0.55	1.618 ± 0.661
%Dif Grip N_{skew}	1.469	0.843	0.618	2.448
% Dif Grip Kurt.	1.062 ± 2.0	3.879 ± 2.0	-1.258 ± 1.063	2.919 ± 1.279
%Dif Grip $N_{kurtosis}$	0.531	1.9395	-1.1834	2.2823
%Dif Rota Skew	0.660 ± 0.913	0.561 ± 0.913	-0.114 ± 0.55	0.779 ± 0.661
%Dif Rota N_{skew}	0.723	0.614	-0.2073	1.1785
% Dif Rota Kurt.	-2.579 ± 2.0	-1.873 ± 2.0	-1.450 ± 1.063	0.298 ± 1.279
%Dif Rota $N_{kurtosis}$	-1.290	-0.937	-1.3641	0.2330
%Dif End. Skew	0.516 ± 0.913	0.594 ± 0.913	1.056 ± 0.55	1.747 ± 0.661
%Dif End. N_{skew}	0.565	0.651	1.92	2.643
% Dif End. Kurt.	0.247 ± 2.0	-1.043 ± 2.0	1.254 ± 1.063	3.59 ± 1.279
%Dif End. $N_{kurtosis}$	0.1235	-0.522	1.1797	2.8069
%Dif Act. Skew	-	2.033 ± 0.913	-	-0.013 ± 0.661
%Dif Act. N_{skew}	-	2.227	-	-0.0197
% Dif Act. Kurt.	-	4.351 ± 2.0	-	-0.286 ± 1.279
%Dif Act. $N_{kurtosis}$	-	2.176	-	-0.2236
FIAV Skew	-0.69±.913	-0.701 ± 0.913	-0.353 ± 0.55	0.213 ± 0.661
FIAV N_{skew}	-0.766	-0.768	-0.6418	0.3222
FIAV Kurt.	-1.417±2	0.282 ± 2.0	-0.717 ± 1.063	-0.992 ± 1.279
FIAV $N_{kurtosis}$	-0.709	0.141	-0.6745	-0.7756

Table S1. Skew and Kurtosis of Mass, Frailty Criteria and FIAV. The cut-off point of normalized skew and kurtosis of greater than 2 was considered as a possible indicator of non-normality in the sample. The data were further investigated using Kolmogorov-Smirnov and Shapiro-Wilk Goodness-of-Fit tests. Using α=.05, only the control old group of the grip test (p=.016, skew=0.340±.6, kurt.=1.3±1.1) and the adult group activity (p=.011, skew=2.033±0.913 ,kurt.=4.351±2.0), violated normality on the Kolmogorov-Smirnov. Using Shapiro-Wilk only the old exercise endurance (p=.017, skew=1.747±0.661, kurtosis=3.59±1.279), adult activity (p=.019, skew and kurtosis in previous sentence), and old exercise group grip test (p=.023, skew=1.6±0.661, kurt.=2.9±1.279) violated normality. The two tests only agreed on the young activity group (n=5). ANCOVA that we used for means testing is considered quite robust in most cases of mild violation of normality (and given that the number of subjects was quite low: n=5 for both adult groups, and n=11 and 17 for the older mice), we did not use non-parametric testing. [1] kurtosis and skew are mean ± sem (standard error of measurement), N_{skew} and N_{kurt} have no confidence interval. - = not recorded. N_{skew} = normalized skew. N_{kurt}. = normalized kurtosis.

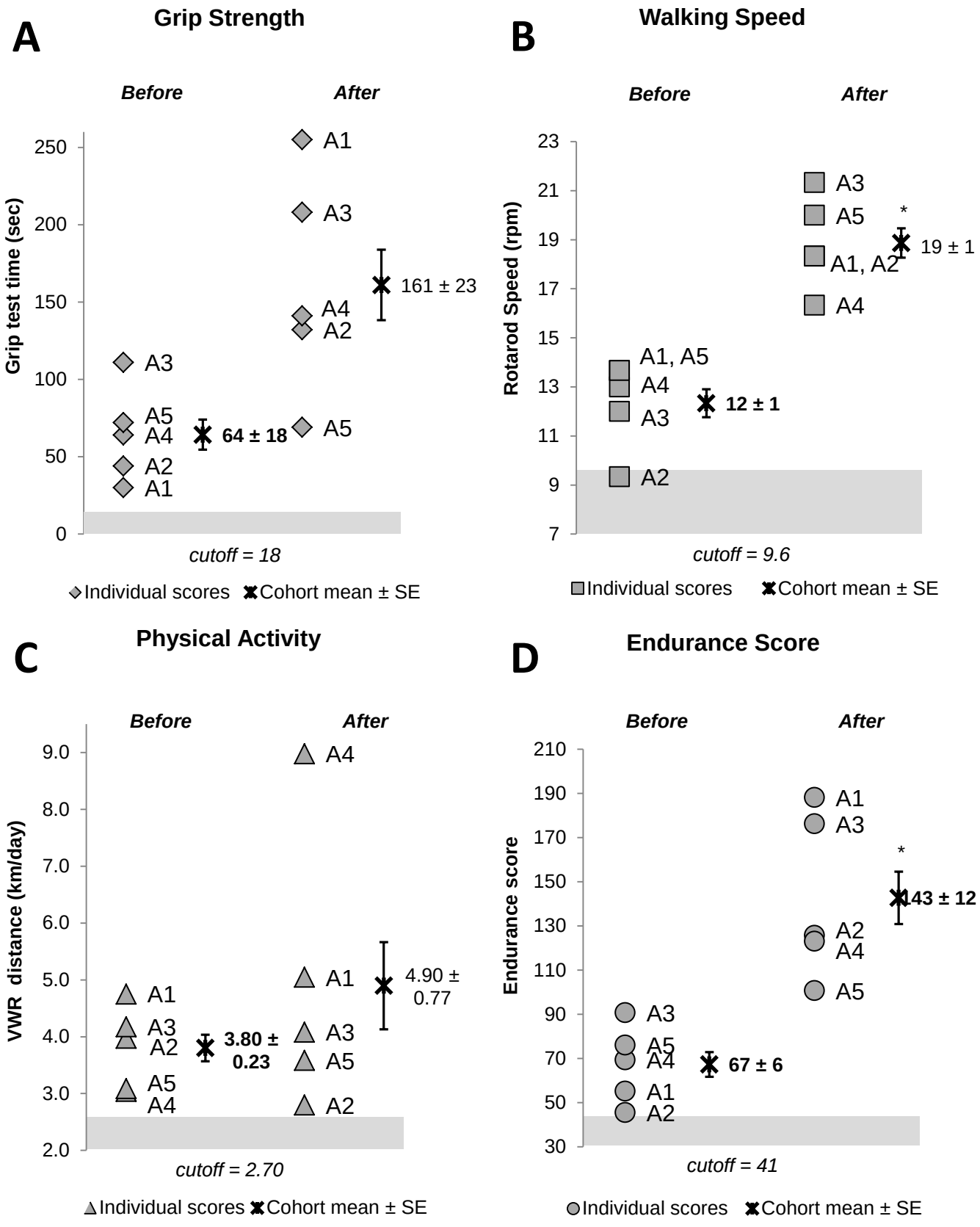
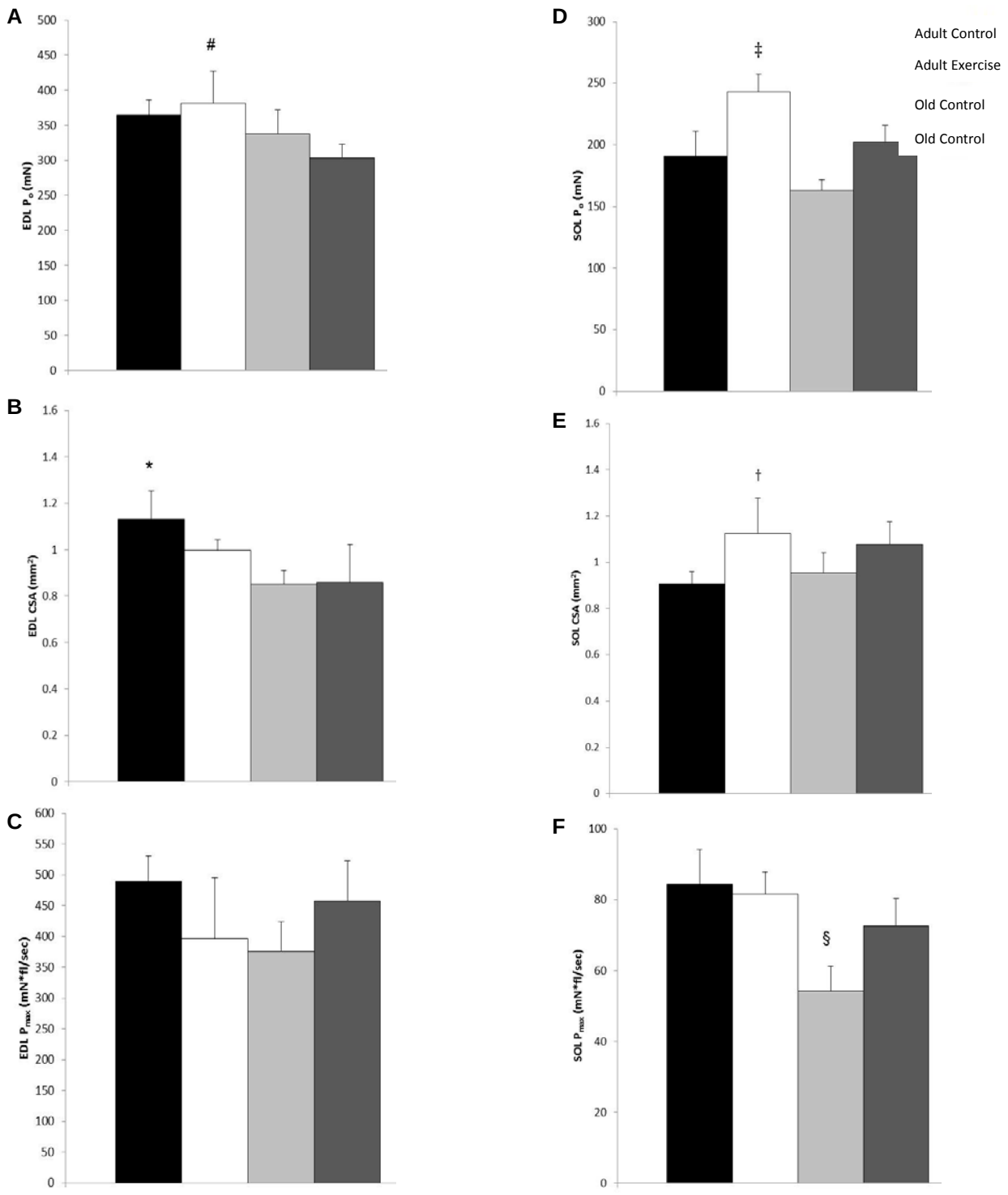
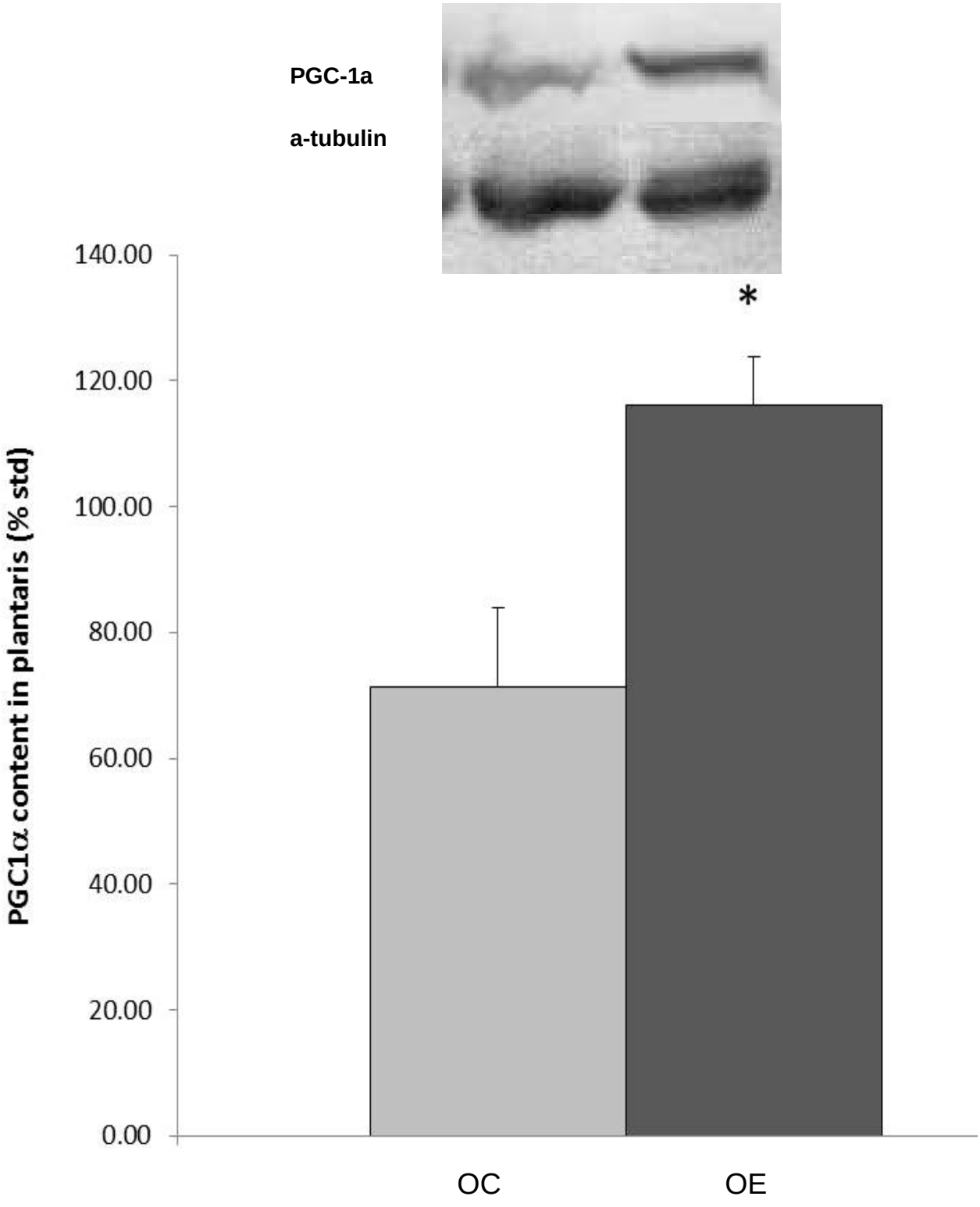


Figure S1. Effect of exercise (VWR) on Frailty Criteria in the adult mice. Panels A-D. show the performance of individual mice in each criterion (**A.** grip strength, **B.** walking speed, **C.** physical activity, and **D.** endurance score) before and after exercise. In each panel, the scores of individual mice and the group mean ± SE before and after exercise intervention are presented. The data points within the shaded regions represent the mice that fell below the cutoff points (1.5 SD below the baseline mean). *: indicates significant difference from the group mean before intervention (paired t-test, $p < 0.05$).



S2. Whole Muscle Contractility. Left (A-C): Isolated whole-muscle contractile properties in EDL. ***A:*** EDL peak tetanic force (P_o) #AE vs OE (*P*<0.05). Main effect of age was significant (*P*<0.05). ***B:*** EDL cross-sectional area (CSA). *AC vs OC (*P*<0.05). ***C:*** EDL P_{max}. No significant differences. **Right (D-F): Isolated whole-muscle contractile properties in SOL.** ***D:*** SOL P_o. ‡AE vs AC *(P*<0.05). Main effects of age and exercise were significant (*P*<0.05). ***E:*** SOL CSA. †AE vs AC (*P*<0.05). Main effect of exercise was significant (*P*<0.05). ***F:*** SOL P_{max}. §OC vs AC (*P*<0.05). Main effects of age and age*exercise were significant (*P*<0.05). Data expressed as means ± SE. Main effect analyses were determined by 2-way ANOVA and post hoc analysis was performed with the T-Test.



S3. Total PGC-1α content in the plantaris. *Content was 61% significantly higher in the plantaris of exercised old mice (OE, n=5) compared to the control old mice (OC, n=5) (p=.016).

SUPPLEMENTAL SECTION

METHODS (Addendum to Main Paper)

Frailty Criteria: Inverted Cling Grip Test, Rotarod, Endurance, Activity Level

Inverted cling grip test

The inverted cling grip test was used for measuring the muscular strength and endurance in the mouse. It was used as one of the criteria in the Frailty Index in our previous study [8]. Each mouse was placed on the top lid of the custom device, the lid was closed, and the mouse was inverted inside of the cage. The time (sec) the mouse held on to the grid before falling was recorded. Two trials were performed with a 20 minute rest period between (if the time of the first trial was 5 seconds or less, the time was considered invalid and another trial would be performed to ensure the initial trial was not simply a “slip” and was representative of weakness). The final score of the grip test was calculated from the times of two valid trials.

Rotarod test

The rotarod test is widely used for testing overall motor functions in mice, including overall motor coordination and endurance. It was one of the Frailty criteria in mice for determining walking speed. A mouse was placed on the cylinder of a Rotarod LSI testing device (LSi Letica Rota-Rod R/S), which was set at on an acceleration protocol (4 revolutions per minute, RPM, to 40 RPM over 5 minutes). The rotating speed (RPM) of the cylinder when the mouse fell was recorded. The time (sec) when the mouse fell was recorded as latency to fall. Before the actual test, the mice were familiarized to the test with 3 trials (of varying protocols) per day over 3 days. The actual rotarod test was performed on the fourth day, which included 3 trials with 20 min rest between each trial. The final score of the rotarod test was calculated as the average of those three trials (RPM).

Endurance score

The inverted cling grip test and rotarod test both include an element of endurance. Thus, we calculated the endurance score from the time (sec) of grip and rotarod test [endurance score (sec) = (time of grip test + time of rotarod test)/2]. This score combines the two different measurements into a third score that describes the average endurance of the mice.

Physical activity Level

As one of the Frailty Criteria, voluntary wheel running was used for evaluating the physical activity level in each animal (exercise group, 5 adult and 11 old). The mice were individually housed in the Lafayette Activity Wheel Cages (#80820F/ 86060, Lafayette Instruments, Lafayette, IN). The daily running distance (km/day) during the first and the last week were recorded for measuring the physical activity. The running distance was recorded at the same time every day, in revolutions. The unit of revolutions was converted into kilometers for data analysis (1 rev = 0.4m). The average daily running distance during one week was calculated for the final score of voluntary wheel running (km/day).

Muscle Physiology, Morphology and Biochemistry

Physiological analysis: Muscle contractility

The maximum isometric contractile force (P_0), maximum velocity (Vmax), and maximal power (Pmax) were analyzed as described previously [12]. Briefly, the muscles were perfused with 95% O_2/5% CO_2 in 25°C Krebs-Ringer buffer in a tissue bath. Each muscle was tied to a force transducer (Aurora 300b, Aurora Scientific, Aurora, Ontario, Canada) with 5-gauge suture and suspended between two platinum electrodes. Using an Aurora High Power Bi-Phase Current Stimulator and a Dual-System Signal Interface controlled by Dynamic Muscle Control software (v.4.1.6, Aurora Scientific), muscles were stimulated to examine peak twitch force (P_t), optimal length (L_o), pre-load tension, maximum isometric contractile force (P_0), force/max force (P/ P_0)

from 10 Hz to 180 Hz stimulus frequency, and velocity at percentages of P_0 (10%-90%). Maximum unloaded velocity was calculated using the Hill Equation [$(V+b)(P/P_0 \pm a/P_0)=b(1+a/P_0)$, V=max. velocity at fractional load], data curve fit (only $r^2>0.98$) using MatLab (MathWorks, Natick, MA). Maximal power (P_{max}) was calculated as the point on the force-power curve where the first derivative (slope) is equal to zero using MatLab (MathWorks, Natick, MA). Whole-muscle cross-sectional area (CSA) of the SOL and EDL was calculated from L_o (mm), muscle weight (g), and muscle density, which was assumed to be 1.06 mg/mm^3 [43,44].

Morphological analysis: Fiber size identification

Cross-sectional areas (CSA) of individual fibers were analyzed for identifying the fiber size in TA muscle using H&E stain. Briefly, the frozen muscles were embedded in OCT (Tissue-Tek, Torrance, CA) and serially sliced from the mid-belly using a Cryostat (Leica CM3050S at -25°C to produce 16 cross-sectional slices (10μm thick). One section was then stained with hematoxylin and eosin (H&E). CSA of each cell was determined by circling individual cells to obtain measurements using ImageJ analysis software (National Institutes of Health, http://rsb.info.nih.gov/ij/).

Morphological analysis: Fiber type composition

To identify the fiber type composition transformation after intervention in the adult and old groups, myosin heavy chain isoform (MHC) was analyzed by silver stain and immunohistochemistry.

MHC analysis (silver stain): After the whole muscle contractile experiments were performed, the muscles were removed from the tissue bath and placed in relaxing buffer as previously described [13]. The bundles were homogenized in 20 mM Tris-HCl. The homogenate was centrifuged at 4°C at 13,793*xg* for 15 min, and the supernatant was collected. Protein

concentration was determined using a Pierce BCA Protein Assay Kit (#23227, *Pierce/Thermo, Rockford, IL*). MHC isoform expression of the SOL and EDL were determined by separating proteins with gel electrophoresis on 5% acrylamide gels and silver staining as previously described [13]. Quantification was performed with BioRad QuantityOne software (Bio-Rad, Hercules, CA).

MHC analysis (Immunohistochemistry): Additional cross-sectional slices were stained for type I (slow), fast, IIa, and IIb fibers using a Mouse-on-Mouse (MoM) kit (Vector Laboratories, Burlingame, CA). Avidin-biotinylated horseradish peroxidase complex was used for visualization of the antibodies. Images were captured using a Nikon Eclipse E400 microscope (Nikon, Melville, NY) with a high-resolution digital camera (Nikon DS-Ri1) interfaced to a computer with Nikon NIS-Elements-F software. The immunostained cross-sections were evaluated for fiber type composition by analyzing the maximum number of cells that could be visualized from the entire muscle. From these images, the cells expressing each of the MHC isoforms were counted and logged into an Excel database for analysis.

Biochemistry analysis: Marker of mitochondrial metabolism – PGC-1α

The mitochondrial metabolism was evaluated by the total content of PGC-1α using Western Blotting. The plantaris from the right hindlimb of 5 OE and 5 OC mice were homogenized in a 10:1 vol/wt ratio of modified radioimmunoprecipitation (RIPA) buffer as described previously [14]. Homogenates were sonicated for 30s (PA4 Ultrasonic, Patterson Dental, St Paul, MN) and centrifuged at 5,000x*g* for 20 min at 4°C. The supernatant was aliquoted and stored at -80°C. Protein concentration was determined from the supernatant by BCA (Pierce kit #23252, Thermo Fisher, Rockford, IL) using BSA as a standard.

Total PGC-1α content was determined by Western blotting. Total α-tubulin content was measured as loading control. Aliquots of homogenized muscle sample supernatants and standards were thawed over ice and diluted 1:1 with sample buffer containing 1.25M Tris, pH 6.8, glycerol, 20% SDS, 2-mercaptoethanol, 0.25% bromophenol blue solution, and deionized water. Samples containing 60 µg of total protein were separated via SDS-PAGE for 105 min at 200V. The gels were electrotransferred via a semi-dry transfer cell (Bio-Rad, Hercules, CA.) using 25V for 20 min to 0.4 mm PVDF membranes (Millipore, Bedford, MA). The membranes were blocked in TTBS (TBS, 50 mM Tris, 150 mM NaCl, containing 0.1% Tween-20), and 10% nonfat dry milk for 1 h at room temperature before incubating overnight at 4°C with antibodies directed against PGC-1α (#515667, EMD Calbiotech), and α-tubulin (#2144, Cell Signaling). The antibodies were diluted 1:1000 for PGC-1α, and 1:500 for α-tubulin in TTBS containing 2% nonfat dry milk. Membranes incubated for 1 h with secondary antibody (goat anti-rabbit IgG, #170-6518, Bio-Rad). Dilutions were 1:1000 for PGC-1α and 1:500 for α-tubulin. The immunoblots were developed using BCIP/NBT (#203790, EMD-Calbiotech) and imaged using a BioRad GS-800 imager. The mean density of each band was quantified using ImageJ software.

Statistics

Differences in baseline grip and rotarod tests were analyzed using independent t-tests. Differences in VWR between adult and old were analyzed using one-way ANOVA (for percent changes) and one-way Repeated Measure ANOVA. Grip, rotarod and endurance percent changes were analyzed with a Univariate 2x2 ANCOVA (body mass as covariate) and the measurements themselves were analyzed with 2x2x2 Repeated measures ANCOVA (post hoc test: least significant difference, LSD). Whole muscle contractility, MHC isoforms, fiber type and CSA of the TA were analyzed using two-way ANOVA with LSD post hoc. Differences between control and exercised old plantaris PGC-1α content, fiber type, and CSA were analyzed using independent t-tests.

The frailty intervention assessment value of the exercise animals was examined using paired samples T-test, with-in each age group, and a 2x2x2 Repeated Measures ANOVA, comparing FIAV1 to FIAV2. The FIAV2 scores and the individual standardized criteria and the differences (between pre- and post-exercise) were each analyzed separately (2x2x2 ANCOVA with body mass as adjuster for rotarod, grip test, and endurance 2x2x2 ANOVA for VWR). The relationship between the age group, body mass and FIAV2 was described using multi-linear regression. The magnitude of the effect was described using the effect size index (mean2–mean1 divided by standard deviation of mean1) and the standardized response mean (mean2–mean1 divided by standard deviation of the change in scores). The coefficient of variation (standard deviation divided by the mean) was used to compare variability. Differences were considered significant at $P<0.05$. Data are expressed as mean±SE, unless otherwise indicated. Statistical analyses were performed using SPSS version 16.0 and 21.0 (IBM, Chicago, IL).

RESULTS (Addendum to Main Paper)

Inverted cling grip test

Age was a strong predictor of grip test performance such that the adult mice perform 237% better than old mice (main effect, percentage change from baseline, mean difference 158±51%; collapsed across the intervention: $p=0.004$). Grip test performance showed a tendency to improve 154% following the exercise intervention (main effect, percentage change from baseline, mean difference 86±50%; collapsed across the age groups: $p=0.09$). The age*intervention interaction was not significant ($p=0.561$).

Our post-hoc analyses of the *percent difference* between groups revealed neither the old exercise group (57±52%, $p=0.184$) nor the adult exercise group (115±85%, $p=0.448$) improved compared to the age-matched controls. However, when comparing the two exercise groups, the

adult exercise group trended to improve more than the old exercise group (187±73%, p=0.09). Finally, the adult exercise group improved more than the old control group (243±70%, p=0.009). (Above statistics from a Bonferroni corrected 2x2 ANCOVA, adjusted for body mass, F=3.4, p=0.02)

Comparing the performance of the grip test of each mouse (within subjects) at the baseline and at the end of the exercise intervention, we found the age of the mouse was the major contributing factor for performance (2x2x2 Repeated Measures Bonferroni corrected ANCOVA, F=18.4, p<0.001). The only significant pairwise difference was between the adult and the old control mice (67±22 sec; p=0.026). In contrast, the comparison between mice (between subjects), age group trended toward significance (F=3.5, p=0.071); but, the age group*treatment interaction was significant (F=5.5, p=0.026).

Rotarod test

The main effect of age showed no significant *percentage difference* from baseline (p=0.614). However, the raw performance time was greater in the adult mice compared to the old mice (One-Way ANOVA, 94±11 vs. 56±4sec, respectively, p<0.05). Rotarod performance showed an improvement of 138% following the exercise intervention (main effect, percentage change from baseline, mean difference 135±21%; collapsed across the age groups: p<0.001). There was an interaction between age and the exercise intervention (F=4.3, p=0.045) with the rotarod performance time.

The post-hoc analyses showed the adult and old exercise groups had greater performance times than the adult and old control groups (115±79%, p=0.003; 57±48%, p=.036, respectively.) The effect of the exercise intervention was greater in the adult mice, 187±68% (p=0.001) more

than the old exercise group. (Above statistics from Bonferroni corrected 2x2 ANCOVA, adjusted for body mass and collapsed across treatment, F=5.1, p=0.003).

Examining each mouse (pre- and post-intervention, within subjects), we found that both treatment and treatment*age interaction influenced performance (treatment F=24.4, p<0.001; treatment*age F=10.5, p=0.003, 2x2x2 Repeated Measures ANCOVA adjusted for body mass).

Endurance score

Age was not a significant factor on endurance (main effect, adult mice perform 139% better than old mice; percentage change from baseline had mean difference 30±20%; collapsed across the intervention: p=0.165). The endurance showed an improvement of 300% following the exercise intervention (main effect, percentage change from baseline, mean difference 136±21; collapsed across the age groups: p<0.001). There was an interaction between age and the exercise intervention with this frailty criterion (F=4.55, p=0.04).

When examining the post-hoc analyses the adult exercise group increased endurance 480% more than the adult control (mean difference 179±35%, p<0.001) and improved 148% more than the old exercise group (mean difference 73±30%, p=0.021). The effect of the exercise intervention on difference percentage was greater in the adult group compared to the old group (179 ± 21%, p=0.004 and 92 ± 21%, p<0.001, respectively). (Above statistics from a Holm-Bonferroni corrected 2x2 ANCOVA, adjusted for body mass, F=12.04, p<0.001).

Examining each mouse (within subjects), we found the intervention increased endurance (F=87.3, p<0.001) and the intervention*age interaction was significant (F=14.3, p=0.001) (2x2x2 Repeated Measures ANCOVA, adjusted for body mass). Between subjects, age group approached significance (F=4.0, p=0.053) and treatment was significant (F=11.3, p=0.002).

Physical activity level

There was no significant *percentage difference* in physical activity with age or intervention (137±40% and 91±13%, adult and old, respectively). However, there was a difference in the VWR volume (degree of exercise intervention)**.** Average VWR, measured in km/24-hr period, was significantly greater in adult than old mice (5.0±0.64 km vs. 1.8±0.14 km, $p<0.05$). The average daily distance per week over the 4wk training period showed that the significant difference was maintained across the 4wks. In order to determine whether the difference in exercise volume influenced the other outcome measurements correlations and regressions were performed and we found no significant relationship (difference before and after treatment in km).