Genome-wide associations reveal human-mouse genetic convergence and novel modifiers of myogenesis, *CPNE1* and *STC2*

Authors

Ana I. Hernandez Cordero[1], Natalia M. Gonzales[3], Clarissa C. Parker[5], Greta Sokoloff[10], David J. Vandenbergh[9], Riyan Cheng[6], Mark Abney[3], Andrew Skol[4], Alex Douglas[2], Abraham A. Palmer[7,8], Jennifer S. Gregory[1], and Arimantas Lionikas[1*]

Affiliations:

1. School of Medicine, Medical Sciences and Nutrition, College of Life Sciences and Medicine, University of Aberdeen, Aberdeen, UK
2. Institute of Biological and Environmental Sciences, University of Aberdeen, Aberdeen, AB24 3FX, UK
3. Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA
4. Department of Medicine, University of Chicago, Chicago, IL 60637, USA
5. Department of Psychology and Program in Neuroscience, Middlebury College, Middlebury, VT 05753, USA
6. Department of Health Sciences, University of California San Diego, La Jolla, CA 92093, USA
7. Department of Psychiatry, University of California San Diego, La Jolla, CA 92093, USA
8. Institute for Genomic Medicine, University of California San Diego, La Jolla, CA 92093, USA
9. Department of Biobehavioral Health, Penn State Institute for the Neurosciences, & Molecular, Cellular & Integrative Sciences Program. The Pennsylvania State University, University Park, PA 16802

10. Department of Psychological and Brain Sciences, The University of Iowa, Iowa City, IA 52242

*Corresponding Author: Dr. Arimantas Lionikas, e-mail: a.lionikas@abdn.ac.uk

Abstract

Muscle bulk in adult healthy humans is highly variable even after accounting for height, age and sex. Low muscle mass, due to fewer and/or smaller constituent muscle fibres, would exacerbate the impact of muscle loss occurring in aging or disease. Genetic variability substantially influences muscle mass differences, but causative genes remain largely unknown. In a genome-wide association study (GWAS) on appendicular lean mass (ALM) in a population of 95,545 middle-age (37-48 years) individuals from the UK Biobank we found 125 loci associated with ALM ($P<5x10^{-8}$). We replicated associations for 64% of these loci ($P<5x10^{-8}$) with ALM in a population of 193,688 elderly (65-74 years) individuals. We also conducted a GWAS on skeletal muscle mass of 1,867 mice from the LGSM advanced intercross line and found 23 quantitative trait loci. Five loci and nine positional candidates overlapped between the two species. *In vitro* studies of potential candidates identified *CPNE1* and *STC2* genes as novel modifiers of myogenesis. Collectively, these findings shed new light on the genetics of muscle mass variability in humans and identified new targets for the development of interventions preventing muscle loss. The overlapping genes between humans and the mouse model will facilitate understanding of the cellular mechanisms underlying muscle variability.

Introduction

Skeletal muscle plays key roles in locomotion, respiration, thermoregulation, maintenance of glucose homeostasis and protection of bones and viscera. The loss of muscle due to aging, known as sarcopenia, affects mobility and can lead to frailty and deterioration of quality of life[1]. The risk of disability is 1.5 to 4.6 times higher in the sarcopenic elderly than in the age matched individuals with normal muscle mass[2]. However, lean mass, a non-invasive proxy for muscle mass, differs by more than two-fold between healthy adult individuals of same sex, age and height[3]. Therefore, we hypothesize that differential accretion of muscle mass by adulthood may influence the risk of sarcopenia and frailty later in life.

Genetic factors contribute substantially to the variability in lean mass in humans, with heritability estimates of 40 – 80 %[4]. A continuous distribution of the trait and data obtained from animal models[5-7] indicates a polygenic causality. However, thus far, genome-wide association studies (GWAS) have implicated fewer than a dozen genes, explaining only a small fraction of this heritability[8; 9]. A limited sample size in early studies[10-14] and the effects of confounders such as subject age[8], size of the skeleton and composition of lean mass have hindered detection of genes. The UK Biobank is a resource of demographic, phenotypic and genotypic data collected on ~500,000 individuals[15]. It includes the arm and leg lean mass, body composition and morphometric information, providing a model for improving our understanding of the genetic basis for variability in muscle mass. Skeletal muscle mass, however, changes over the course of individual's lifespan. It reaches a peak around mid-twenties and remains largely stable through mid-fourties, before succumbing to gradual decline, which accelerates after about 70 years of age[16]. There is a substantial degree of individual variability in the slope of muscle change across both the increasing and decreasing phases of the lifespan trajectory[17]. Both the trajectory itself and the slope of individual variability may impede identification of genes.

The indirect estimates of lean mass impose limitations because muscle mass is not an exclusive contributor to this variable. Furthermore, the cellular basis of variability in muscle mass (i.e. if it is caused by the differences in the number of constituent muscle fibres, their size, or both) remains completely latent. Using the laboratory mouse circumvents a number of those limitations.

The mouse shares approximately 90% of the genome with humans[18], and permits analyses of traits not amenable in humans, such as muscle mass[6; 7; 19] and muscle fibre characteristics[20; 21]. The phenotypic differences between the LG/J and SM/J mouse strains make them particularly attractive for complex trait analyses[22-24]. LG/J mice were selected for large body size[25], while SM/J mice were selected for small body size[26]. The F_2 intercross derived from the LG/J and SM/J strains (LGSM)[6; 27] and an advanced intercross line (AIL) of the LGSM (LGSM AIL) developed using a breeding strategy proposed by Darvasi and Soler[28], led to multiple quantitative trait loci (QTLs) for muscle mass[6; 27]. However, these QTLs still encompass tens or even hundreds of genes and require further prioritising. We hypothesized that the detection power of a modest sample size of the LGSM AIL and the superior resolution of a human cohort will facilitate identification of the quantitative trait genes (QTGs) underlying muscle QTLs.

The aim of this study was to identify the genomic loci and the underlying genes for variability in skeletal muscle mass and to assess their effects in the elderly. We addressed this in three stages: (1) we conducted a GWAS in a human cohort of middle-aged individuals from the UK Biobank, and tested the effect of the identified set of loci in an elderly cohort; (2) we conducted a GWAS on hindlimb muscle mass in a population of LGSM AIL mice. (3) In the final stage, we nominated candidate genes by comparing mouse and human loci and validated the myogenic role of selected candidates *in vitro*.

Methods

Stage one: Genome mapping in human populations

UK Biobank cohort

The population in this study consisted of 316,589 adult individuals of 37 to 74 years of age (project ID: 26746). We drew this cohort from the UK Biobank (UKB) project[15]; all participants recruited were identified from the UK National Health Service (NHS) records and attended a baseline visit assessment between 2006 and 2010. During the assessment, participants gave written consent, answered a questionnaire, and were interviewed about their health and lifestyle.

Blood samples and anthropometric measurements were collected from each participant. Assessments were conducted at 22 facilities in Scotland, England, and Wales.

We divided the sample into middle-aged and elderly cohorts. The middle-aged cohort consisted of 99,065 adults ranging from 37 to 48 years of age; these individuals were not affected by sarcopenia. We excluded 3,520 participants that were reported to be ill with cancer, pregnant, or had undergone a leg amputation procedure, as well as individuals with discordant genetic sex and self-reported sex records. We retained a total of 95,545 adult individuals (51,394 females and 44,151 males) for further analyses.

The elderly cohort consisted of 217,524 adults ranging from 63 to 74 years of age. We selected this cohort to test if the effect of the genetic variants identified on middle-aged individuals could also influenced phenotypes later in life. We excluded 23,836 individuals based on the same criteria used for the middle-aged cohort. After exclusions, the elderly cohort included 193,688 individuals of 63 to 74 years of age (100,463 females and 93,225 males) (Table S1).

UK biobank traits

We used the data for standing height (UKB field ID: 50), sitting height (UKB field ID: 20015), whole body fat (UKB field ID: 23100), arm lean mass (UKB field ID: 23121 and 23125), and leg lean mass (UKB field ID: 23113 and 23117) measured as part of the UK Biobank project. Body composition measurements were taken using bioelectric impedance. We calculated leg length by subtracting sitting height from standing height (all measurements were recorded in cm). Because lean mass in the limbs primarily consists of skeletal muscle tissue, we used ALM as a proxy for muscle mass. We calculated ALM as the sum of the muscle mass of two arms and two legs. We checked that all traits were normally distributed by examining the QQ-plot and histogram of residuals from a simple linear model that included sex as a covariate. Residuals were normally distributed and we did not transform any of the traits.

UK biobank genotypes

We obtained genotype data for all participants from the UKB v3 genotypes release[29], which includes genotype calls from the Affymetrix UK BiLEVE Axiom array and the Affymetrix UK Biobank Axiom array. IMPUTE2[29] was used to impute genotypes from the UK10K and 1000 Genomes Phase 3 reference panels [30], as described by Brycroft *et al*[29]. We kept all imputed genotype data (93,095,624 genetic variants (SNPs, Indels and structural variants)) for subsequent analyses in order to capture the effects of both common (MAF > 0.001) and rare variants (MAF < 0.001). The software (BOLT-LMM v2.3.2) [31] we used to perform GWAS was developed for large data sets (i.e.: UK Biobank cohort) and it was only tested for human cohorts, which have different LD patterns from animals. For these two reasons, we used BOLT-LMM v2.3.2 for the analyses of human data only. Although we mainly focused on reporting the effects of common variants, we also reported rare variants associated with ALM as a supplemental table (Table S4).

Appendicular lean mass GWAS

We used BOLT-LMM (v2.3.2)[32] to perform a GWAS for ALM in the middle-aged cohort. The linear mixed model (LMM) approach implemented in BOLT-LMM is capable of analysing large data sets while also accounting for cryptic relatedness between individuals. Specifically, BOLT-LMM calibrates the association statistics using a linkage disequilibrium (LD) score regression approach[33]; this allowed us to evaluate the impact of confounding factors on the GWAS test statistics[33] and calibrate them accordingly. In the absence of confounding factors, p-values should not be inflated, and the LD score regression intercept should be equal to 1[33]. The LD Score regression intercept in this study was 1.043 ± 0.007, suggesting minimal inflation of *P* values due to linkage between markers. After calibrating the test statistics, the mean χ^2 of the ALM GWAS was 1.27 and lambda (λ_{GC}) or genomic control inflation factor was 1.20 (Figure S1), which indicated polygenicity of the trait as described by Bulik-Sullivan and colleagues[33].

We also assessed population structure by running principal component analysis on the genotype calls. We included sex, leg length, whole body fat, and the first four principal components

as fixed effects in the LMM used for the ALM GWAS. Sex was included to account for differences in muscle mass caused by increased testosterone levels in males[34]. Leg length and whole body fat were included because they are biologically related to muscle mass: longer bones result in longer muscles, while fat shares part of its developmental origin with skeletal muscle tissue[35]. Furthermore, each of these traits is correlated with muscle mass. An association was considered statistically significant if its $P < 5 \times 10^{-8}$ ($\alpha = 0.05$). This threshold is the standard for GWAS of complex traits[36; 37].

We obtained variance components and SNP heritability estimates of ALM using BOLT-REML[32]. The BOLT-REML method robustly estimates the variance of genotyped SNPs and fixed effects on the LMM. As described by Loh et al. [38], BOLT-REML partitions the SNP heritability across common alleles; hence, the additive variance is calculated as the cumulative variance of genotyped SNPs.

Age effect on ALM GWAS

To examine the influence of varying age on the genomic loci associated with ALM repeated the analysis using a cohort with a wider age range. To achieve this we subset the middle-aged cohort by randomly selecting 44,727 individuals (37 to 48 years), and likewise, we selected a subset of 44,727 older individuals from the elderly cohort (63 to 74 years) to produce a mixed age population (n = 95,454) (Table S2). We used this population to execute a GWAS on ALM using BOLT-LMM, and included sex, age, leg length and whole body fat as fixed effects in our model. We compared the resulting loci (MAF>0.001) (Table S8) from this analysis to the ALM loci that we identified on the GWAS conducted exclusively in the middle-aged population of a similar sample size (n=95,454). This analysis was restricted to one iteration due demand for computational resources (~130h per GWAS).

Phenotypic variance explained by ALM loci

We defined ALM genomic loci using the web-based platform Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA GWAS[39]). A key feature of this tool is the identification of genomic regions based on the provided summary statistic of a GWAS depending on LD structure; this process is automated using pairwise LD of SNPs in the reference panel (1000 genomes project phase 3[40]) previous calculated by PLINK[41]. We provided to FUMA GWAS the summary statistic of our GWAS on ALM with the following parameters: 250kb window (maximum distance between LD blocks), $r^2 > 0.6$ (minimum r^2 for determining LD with independent genome-wide significant SNPs used to determine the limits of significant genomic loci), MAF > 0.001 (minimum minor allele frequency to be included in the annotation), $P < 5 \times 10^{-8}$ (threshold of significance associated variants). In addition, we then performed a stepwise conditional analysis using the software package GCTA[42] to identified extra independent signals within 500kb window; this analysis was conducted only on statistically significant SNPs ($P < 5 \times 10^{-8}$) with MAF > 0.001. We refer to the identified regions and the independent signals as loci throughout the text.

We used the top variant (based on the outcome from FUMA[39] and GTCA[42] software packages) (Table S5) of each locus identified to estimate the proportion of phenotypic variance explained by each locus. We estimated phenotype residuals using a model that included the fixed effects and principal components described above. We then regressed the residuals on the genotype of the top SNP in a linear model. We estimated the coefficients of determination and reported them as the proportion of phenotypic variance explained by each locus.

Genetic effects in the elderly cohort

We tested the combined effect of all 125 genome-wide significant ALM loci identified in the middle-aged cohort in the elderly cohort using the top SNP at each locus. We used PLINK2[41] to impute genotype dosages for each variant identified in the middle-aged GWAS in the elderly cohort. We then estimated a 'genetic lean mass score' for each individual using the following procedure. First, we estimated the contribution of each variant on the phenotype as a product of

204 the SNP effect size obtained from BOLT-LMM (β, calculated based on the reference allele) and the
205 genotype dosage. Second, we calculated the 'lean mass score' for each individual as the sum of
206 the products for all selected variants. We ranked the resulting distribution of lean mass scores in
207 ascending order and partitioned it into five quantiles. We used ALM without any adjustment (raw
208 ALM) because estimates of effects size already accounted for sex, whole body fat and leg length
209 differences. However, since the raw ALM did not meet the assumption of normality, we used a
210 Kruskal-Wallis test (non-parametrical) to evaluate the difference in the median of the phenotypes
211 between the quantiles, and a Wilcoxon test (non-parametrical) for pairwise comparisons between
212 quantiles. We conducted five replicates of a negative control test that consisted on randomly
213 selecting a subset (n = 125) of non-significant SNPs in the middle-aged cohort and generating
214 'lean mass score' as described above for the elderly cohort, this set of SNPs had MAF > 0.001.

215 We also aimed to replicate the individual variants effects on the ALM of the elderly cohort.
216 We checked normality of ALM in the elderly cohort as described for the middle-aged cohort. We
217 tested a subset of genetic variants (n=17,988,060) selected based on their MAF > 0.001 and we
218 used the same LMM, fixed covariates, and genome-wide significance threshold ($P < 5 \times 10^{-8}$) as
219 described for the middle-aged cohort. We conducted a Fisher's exact test to evaluate if overlapping
220 loci between the middle-aged and elderly cohorts were significantly different from random. The null
221 hypothesis was rejected at P < 0.05 (two-tailed).

222

223 Genomic regions tagged by loci

224 We used the 'biomaRT' package in R[43; 44] to retrieve gene and regulatory element
225 annotations at the genomic position of each statistically significant SNP ($P < 5 \times 10^{-8}$) and
226 Polyphen 2 and SIFT[45; 46] to predict the functional consequences of each SNP. We retrieved
227 additional information about the positional candidate genes and their expression levels from
228 Ensembl[47] (release 94 - October 2018) and the Gene Tissue Expression Project (GTEx) portal[48]
229 (See Web Resources).

230

Stage two: LGSM AIL mouse cohort and GWAS

We used three LGSM AIL mouse cohorts for the second stage of this study (n = 1,867). The LGSM AIL was initiated by Dr. James Cheverud at Washington University in St. Louis [49]. Cohort 1 included 490 mice (253 males and 237 females) from LGSM F_{34}. Phenotype data was collected from these mice between 80-102 days of age. Cohort 2 consisted of 506 male mice from F_{50-54}. Cohort 3 includes 887 mice (447 males and 440 females) from F_{50-56}; with age 64 to 111 days of age. Mice were housed at room temperature (70 - 72°F) at 12:12 h light-dark cycle, with 1-4 same-sex animals per cage and with *ad libitum* access to standard lab chow and water.

Mouse traits and genotypes

We collected muscle phenotypes after the animals were sacrificed and frozen. We dissected four muscles and one long bone (tibia or femur) from each mouse at the Pennsylvania State University (n = 584) and the University of Aberdeen (n = 1,283). The dissection procedure involved defrosting the carcasses and removing the muscles (TA, EDL, gastrocnemius and soleus) and tibia from the hind limbs under a dissection microscope. We weighed the muscles to 0.1-mg precision on a Pioneer balance (Pioneer, Ohaus) and measured bone length (mm) using an electronic digital calliper (Powerfix, Profi). We quantile normalized all LGSM AIL traits before mapping QTLs.

Cohort 1 was genotyped using a custom SNP genotyping array[50]. These SNPs (n=2,965) were evenly distributed along the autosomes (mm8, build 36). The median distance between adjacent SNPs was 446 Kb, and the maximum was 18 Mb. Cohort 2 was genotyped at 75,746 SNPs (73,301 on the autosomes and 2,386 on X and Y) using the MEGA Mouse Universal Genotyping Array (MegaMUGA; mm9, build 37); we retained 7,168 autosomal SNPs for subsequent analyses. The median distance between adjacent SNPs was 126.9 Kb and the maximum distance was 15 Mb for all chromosomes except for chromosomes 8, 10, and 14, which had distances of 19, 16, and 16 Mb, respectively. We used a conversion tool in Ensembl to convert SNP positions from mm8 (build 36) and mm9 (build 37) to mm10 (build 38). Cohort 3 genotypes

were obtained from Gonzales and colleges[51]. These genotypes were generated using genotyping by sequencing. This approach has been recently used and described in detail[19]. Only autosomal SNPs known to be polymorphic in the LG/J and SM/J founder strains (n=523,027; mm10, build 38) were retained for subsequent analyses. We combined the genotype data from Cohorts 1-3 using PLINK v.1.9 and imputed missing genotypes using BEAGLE v.4.1[52]. For these steps, we used a reference panel obtained from Heather Lawson's whole genome sequencing data of the LG/J and SM/J strains[53]. Dosage estimates (expected allele counts) were extracted from the output and used for the GWAS; these estimates captured the degree of uncertainty from the imputation procedure. To ensure the quality of the genotype data, we excluded SNP genotypes with MAF < 0.20 and dosage $R^2 < 0.70$ (dosage R^2 corresponds the estimated square correlation between the allele dosage and the "true allele dosage" from the genetic marker, and is used as a measure of imputation quality). After applying these filters, we retained 434,249 SNPs.

Mouse association analyses

The population structure can potentially lead to a rise of false positive associations[54; 55]. The LMM approach is used to map QTLs while dealing with confounding effects due to relatedness[50; 56; 57]. We used the LMM method implemented in the software GEMMA (genome-wide efficient mixed-model association)[58] to analyse the mouse phenotypes. The LMM method implemented in the software GEMMA (genome-wide efficient mixed-model association)[58] was used for the association analysis of all the phenotypes. In our LMM model we include the genotypes, a set of fixed effects described later in this section, and a polygenic effect to deal with the population structure.

The polygenic effect is a random vector which was derived from a multivariate normal distribution with mean zero and a $n \times n$ covariance matrix $\sigma^2 \lambda K$; where n is the number of samples. The relatedness matrix K was defined by the genotypes. The two parameters, σ^2 and λ, were estimated from the data by GEMMA; they represent the polygenic and residual variance components of the phenotypic variance, respectively.

Relatedness matrix and proximal contamination

We used the genotype data to estimate the relatedness matrix K, which was part of the covariance matrix. Although genotype-based and pedigree-based K matrices yield very similar results[59; 60], we have shown that in general, genotype-based estimates are more accurate[59; 61-63]. We constructed the relatedness matrix as $\mathrm{K} = \mathrm{XX}'/\mathrm{p}$, where X is the genotype matrix of entries $\mathrm{x_{ij}}$and $\mathrm{n} \times \mathrm{p}$ dimensions, p is the number of SNPs.

The relatedness matrix K was estimated taking into account the potential problem of proximal contamination[60], which involves loss of power due to including genetic markers in multiple components of the LMM equation. Furthermore, because of LD, markers in close proximity to the genetic marker that is being tested can also lead to deflation of the *P* values[61]. To avoid this problem, the K matrix was estimated by excluding from the calculations the SNPs within the chromosome that was analysed, therefore, K matrix was slightly different for each chromosome.

Genetic and fixed effects

We did not include non-additive effects in the LMMs used for GWAS in the LGSM AIL. Our previous studies[6] suggest that musculoskeletal traits in this population are mostly influenced by additive loci, and by ignoring dominance effects we avoid introducing an additional degree of freedom, hence potentially avoiding a decrease of power to detect QTLs.

To analyse the muscle weights of the combined data, we used four fixed effects in the LMM: sex, dissector of the samples, age, and long bone length. We selected these variables after using a linear model to estimate their effect on the four muscles; only statistically significant effects were included ($P < 0.01$). Sex and dissector were included as binary variables; whereas age and long bone were included as continuous variables. Including long bone length allowed us to capture genetic effects associated with variation in muscle weight *per se* (as opposed to genetic effects on bone length) [19]. In other words, failing to include long bone as a covariate would yield QTLs that are more likely to be genetic contributors to general growth of the skeleton instead of specifically muscle. We used two bones for the long bone variable, for cohort 1 femur, and for cohorts 2 and 3

312 tibia. Based on personal communication with Dr. Cheverud, the femur and tibia bones were found
313 to be positively and highly correlated ($r > 0.8$) in LGSM AIL (F34). We did not include generation (r
314 = 1) and bone type of each cohort (r = 1) as fixed effects since the dissector variable functioned as
315 a proxy for these two variables. Body weight was not used as a fixed effect because muscle weight
316 accounts for a considerable amount of the body weight.

317

318 ## SNP heritability

319 To estimate the SNP heritability or proportion of phenotypic variance explained by all
320 genotypes, we used the $n \times n$ realized relatedness matrix K, which was constructed using all the
321 available genotypes. We extracted the SNP heritability from the QTL mapping outputs; GEMMA
322 provides an estimate of the heritability and its standard error[58]. The SNPs available to estimate the
323 heritability do not capture all genetic causal variants, hence the SNP heritability underestimate the
324 true narrow sense heritability.

325

326 ## Threshold of significance and QTLs intervals

327 The p-values estimated from the likelihood ratio test statistic performed by GEMMA were
328 transformed to $-\log_{10}$ p-values. We calculated a threshold to evaluate whether or not a given SNP
329 significantly contributes to a QTL. We estimated the distribution of minimum p-values under the null
330 hypothesis and selected the threshold of significance to be $100(1 - \alpha)^{\text{th}}$ percentile of this
331 distribution, with $\alpha = 0.05$. In order to estimate this distribution, we randomly permuted phenotypes
332 1,000 times, as described previously [6; 7; 19; 64]. We did not include the relatedness matrix in the
333 permutation tests due to computational restrictions, and because, past studies have found that
334 relatedness does not have a major effect on the permutations test[6; 7].

335 We estimated QTL intervals in three steps. 1) We used Manhattan plots to identify the top
336 SNP within each statistically significant region (SNP with highest $-\log_{10}$ p-values), which we refer to
337 as the peak QTL position. 2) We transformed *P* values from each analysis to LOD scores (base-10
338 logarithm of the likelihood ratio). 3) We applied the LOD interval function implemented in the r/qtl

package[65] to the regions tagged by each peak SNP, and obtained the QTL start and end positions based on the 1.5 LOD score interval. 1.5 LOD intervals are commonly used to approximate the ~ 95% confidence interval of mouse QTLs [5; 66]. The 1.5 LOD interval estimation is comparable to the 95% CI in the case of a dense marker map[67]; hence, its coverage depends on the location of the peak QTL marker relative to the adjacent genotyped markers. We estimated the direction of the QTL effect by calculating the phenotypic mean of each allele based on the peak SNP of each QTL. We adjusted the phenotypic means and standard errors by fitting the fixed effects used in the association analyses to a linear model.

We explored the QTL intervals to identify genes that potentially affect muscle mass and bone length. We retrieved the genomic location of all genes located within the intervals using the BioMart database through the 'biomaRT' package implemented in R[43; 44].

Gene validation using siRNA in C2C12 myoblasts

To validate efficiency of siRNA-mediated gene knockdown, the C2C12 cells were lysed and RNA isolated using RNeasy mini kit (QIAGEN) following manufacturers recommendations. Concentration was assessed using NanoDrop (Thermo Scientific) spectrophotometer and ~1.5 µg of RNA was applied to 1.5% agarose gel to validate its integrity. The cDNA was synthesised using random primers (Invitrogen) and SuperScript II reverse transcriptase (Invitrogen). Quantitative PCR for expression of the target *Cpne1* (F: 5'-GGACTGAACGTGTTCGCAAC-3', R: 5'-ACACGGCTGTCCTTTAGCTC-3'), *Sbf2* (F: 5'-AGCCTGGTGTTGGTATCCAG-3', R: 5'-GTCTCCTGCACCCAAGGAAA-3') and *Stc2* (F: 5'-TGACCCTGGCTTTGGTGTTT-3', R: 5'-GACTTTCCCTGGGCATCGAA-3') genes and the reference *Actb* (F: 5'-GGTGGGAATGGGTCAGAAGG-3', R: 5'-GTACATGGCTGGGGTGTTGA -3') gene was carried out in duplicates on LightCycler 480 II (Roche) using SYBR green Master mix (Roche), 10 ng cDNA and 0.5 µM forward and reverse primer. Quantification of gene expression was performed using the comparative Ct method[68].

C2C12 myoblasts, validated for differentiation, were seeded on 8-chamber slide (Lab-Tek II), batch 1, and 13 mm diameter Thermanox Plastic coverslips (Thermo Fisher Scientific), batch 2, at 100 cells/mm^2 in high glucose growth medium (D5671, Sigma), containing 10% foetal calf serum and 2% glutamine. Next day the cells were washed with PBS and transferred to differentiation medium (D5671, Sigma) supplemented with 10 nM siRNA and Lipofectamine RNAiMAX (Invitrogen) as per manufactures protocol. We used the following siRNAs (Life Technologies): negative control #1, s113938 and 93494 (*Cpne1*), 151885 and 151886 (*Stc2*), s115441 and s115442 (*Sbf2*). The treatment achieved expression knockdown by 55-70%. The differentiation medium with 10nM siRNA and Lipofectamine RNAiMAX were replaced once, after 3 days of incubation. After 6 days of incubation, cells were fixed in 4% paraformaldehyde (PFA). We examined 8 cultures for *Stc2* and 12 for the remaining genes (equally divided between the two siRNAs) and negative control that were generated in two batches on separate occasions.

Cells were washed in PBS, fixed in 4% PFA for 15 min, PBS washed again and permeabilized for 6 min with 0.5% Triton X-100 in PBS. The cells were then blocked for 30 min in blocking buffer (10% foetal calf serum in PBS) and incubated overnight at 4 ºC with primary anti-myosin heavy chains antibody (Monoclonal Anti-Myosin skeletal, Fast, Clone My-32, Mouse Ascities Fluid, M4276, Sigma-Aldrich) diluted (1:400) in PBS. After three washes in 0.025% Tween-20 in PBS at room temperature, secondary donkey anti-mouse IgG H&L antibody (ab150109, abcam) conjugated to fluorescent dye (Alexa Fluor 488) in PBS (1:400) were applied and incubated for 90 minutes. Following three washes in 0.025% Tween-20 in PBS cells were incubated in 300 nM DAPI in PBS for 15 min. After that cells were covered by coverslip using Mowiol 4-88 (Sigma-Aldrich), sealed with nail polish and stored at 4 °C in the dark.

Slides were scanned using Axioscan Z1 slide scanner (Zeiss) using X20 magnification objective. The entire 0.7 cm^2 chamber of a slide or a coverslip was imaged using the wavelength spectrum band of 353-465 nm and 493-517 nm and exposure time 4 ms and 100 ms for DAPI and Alexa Fluor, respectively, at 50% Colibri 7 UV-free LED light source intensity. Alexa Fluor and DAPI channel images of a rectangle area free of artefacts and covering at 14-91% of a chamber of batch 1 and 70% of a coverslip of batch 2 were exported separately for analyses with Fiji[69]. Note that the

rectangle area of the majority of batch 1 samples (88%), covered more than 40% of the cell culture. The exclusion of small coverage images (14-31%) from the statistical analyses described below, showed results comparable to the analysis of all samples; therefore, we reported significance values (*P* values) corresponding to the statistical analysis of all samples.

Three indices characterising the effect of treatment on myogenesis were quantified in an unbiased, automated analyses of the entire exported area: 1) percentage of fluorescent in the Alexa Fluor channel, reflecting the level of myosin expression, and 2) the longest-shortest-path reflecting the length and number of myotubes (Figure S4). The longest-shortest-path analysis was carried out using the analyse skeleton plugin[70] and the shortest path calculation function[71] implemented in Fiji[69]. We carried out the images analyses on a Linux computer and we allocated 190 GB of RAM memory for these analyses. The myotube threshold was set at 103.97 µm for batch 1 and 191.63 µm for batch 2, i.e. the mean (batch 1: 54.34 µm, batch 2:100.95 µm) plus 3 standard deviations (batch 1: SD = 16.54 µm, batch 2: SD = 30.23 µm) of the length of mononucleated and myosin expressing myocytes (n=35) measured in the negative control #1 cells. The myotube length variable did not follow normality, therefore quantile normalization was applied to the variable. All statistical analyses were adjusted for the image area of each sample and batch of cells, by fitting a linear model on the three indices investigated; all subsequent statistical analyses were conducted on the residuals, which met the assumptions of normality and homoscedasticity of residuals. Effect of gene knockdown on these indices was assessed using an ANOVA test to confirm the presence of a statistically significant knockdown effect. After, a T-test was carried out to evaluate the mean differences between the control group and the gene knockdown groups. In addition, we evaluated the myosin expressing area (as percentage of the total) present within in each knockdown versus control groups with an ANOVA test.

Data availability

The human data used for this study can be obtained upon application to the UK biobank project[15].

Results

Over 100 genomic loci associated with appendicular lean mass in humans

The appendicular lean mass (ALM) ranged from 11.8 to 41.6 kg and 15.3 to 42.5 kg in healthy middle age females and males, respectively (Table 1). SNP heritability estimates indicated that 35% of phenotypic variability was due to genetic factors. The GWAS analysis revealed 6,150 autosomal variants (MAF > 0.001) associated ($P < 5 \times 10^{-8}$) with ALM (Figure 1). The associated variants tagged 293 genes and 385 regulatory elements. We used the Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA GWAS[39]) to define genomic regions containing the associated variants, and we identified 77 of them that were on average 0.24 Mb long. Furthermore, we conducted a stepwise conditional analysis that partitioned some of the genomic regions and yielded 48 additional independent signals. We refer to the identified regions and the independent signals as loci throughout the text. In total, we identified 125 loci for ALM (Table S5) which indicates that ALM is influenced by multiple genetic elements. The LD score intercept that we estimated during this ALM GWAS (1.043 ± 0.007) provides further evidence for polygenicity. Cumulative effects of these loci explained 14.28% of SNP heritability.

64% of the same loci affect appendicular lean mass in older adults

Consistent with the aging effect on skeletal muscle, the ALM in the cohort of elderly declined by 4 and 8% in comparison to the middle-age cohort of females and males, respectively ($P < 2\times10^{-16}$). We then used a 'genetic lean mass score' (see Methods for details) to test if the identified 125 loci contributed to ALM variability in the elderly population. The genetic lean mass score had a statistically significant overall effect ($\chi^2 = 583.6$, df = 4, $P = 5.46\times10^{-125}$) on ALM variability in the elderly population (Figure 2). On average, individuals with the highest genetic lean mass score had 0.90 kg, or 4%, more ALM compared to those with the lowest scores (Figure 2). We also found that in some instances negative control iterations resulted in statistically significant ($P < 0.05$) effects (Table S3), however the ALM differences between groups on the negative

controls were modest and in the opposite direction of what would be expected (Type III error) (Figure S3).

We also asked if the variants identified in the middle-aged cohort were associated with ALM in the elderly. A GWAS in the elderly cohort replicated 4,984 variants based on their *P* value ($P < 5 \times 10^{-8}$) and allelic effect (*beta*); moreover, the replicated variants tagged 64% of the ALM loci of the middle-aged cohort (two tailed Fisher test *P* value $< 2.2 \times 10^{-16}$). Overall, the set of genomic loci in the elderly cohort appeared similar to that of the middle-aged adults, with the exception of an approximately 5 Mb region on chromosome 5 (Figure S2). This region showed a very strong association with the ALM variability in older adults (lowest *P* value = 3.10×10^{-55}, *beta* = 0.12 ± 0.01 kg), and had a modest albeit significant association with the ALM of middle-aged individuals (lowest *P* value = 3.30×10^{-11}) with an effect size of *beta* = 0.07 ± 0.01 kg.

23 QTLs contribute to muscle weight variability in LG/J and SM/J strain-derived advanced inter-cross lines.

We examined the weight of four hindlimb muscles of the LGSM AIL (F_{34} and F_{50}-F_{56}): tibialis anterior (TA), extensor digitorium longus (EDL), gastrocnemius and soleus. The LGSM AIL muscles showed extensive individual variability (Table 2); furthermore, the SNP heritabilities of the TA, EDL, gastrocnemius and soleus muscles were 0.39, 0.42, 0.31 and 0.30, respectively (Table 2). The genome mapping of LGSM AIL muscles yielded 23 QTLs ($P < 6.45 \times 10^{-06}$). The TA, EDL and gastrocnemius QTLs explained more than the 50% of the SNP heritability of each trait (Table S6). The soleus muscle phenotypic variability explained by QTLs was 23% of its SNP heritability. Three QTLs were shared between the four muscles (chromosome 7, 11 and 13; (Figure 3); the QTL on chromosome 13 resulted in the strongest association (EDL $P = 2.95 \times 10^{-21}$), with its peak position at 104,435,003 bp, and the percentage of phenotypic variance explained by this locus was 5.2%; the SM/J allele conferred increased muscle mass (Figure 3). Furthermore, six QTLs were

shared between two or three muscles, while fourteen identified QTLs were only associated with one specific muscle (Figure 3).

The mapping resolution was comparable to that attained in the previous study in the LGSM AIL cohort[51]. On average, mouse QTLs were 2.80 Mb long (based on the 1.5 LOD interval) and encompassed 2,259 known genes (Table S7). The median number of genes per QTL was 55; more than half of the mouse QTLs enclosed a modest number of genes, however, seven QTLs contained more than 100 genes each, and a single QTL located on chromosome 7 as many as 644 genes (Table S6). Although all mouse QTLs identified in the LGSM AIL contained polymorphic SNPs, at least seven QTLs covered long genomic regions characterised as identical by descent (IBD) between the LG/J and SM/J strains[53].

Interspecies overlap between appendicular lean mass loci and muscle weight QTLs

The ALM mainly consists of the skeletal muscle of the extremities; however, other tissues also contribute. To test the hypothesis that ALM-associated genetic variants primarily affect the skeletal muscle mass, we overlaid the mouse and human findings. This analysis identified five syntenic regions associated with ALM in humans and muscle mass in mice. This analysis permitted us to shorten the list of positional candidates. Assuming the same causative entity for an overlapping mouse and human locus, these five loci harbour only nine homologous genes. Encouragingly, four of these five genomic loci replicated in the ALM of the elderly cohort.

Selected candidate genes

In selection of the candidate genes we focused on the five most robust loci highlighted by both mouse and human GWAS. Out of the nine genes within these five loci (Table 3), we prioritised the most relevant genes for further testing based on the following information.

STC2

The *STC2* gene had the largest effect size on the ALM in our analyses (beta = 0.877 ± 0.13 kg). The minor allele (A) of a missense SNP (rs148833559 (A/C) was associated with the increase in ALM. Prediction tools (SIFT[46], PolyPhen[72], CADD[73], and REVEL[74]) suggested that the rs148833559 SNP was likely to have a detrimental consequence on *STC2* protein structure. Furthermore, *STC2* is expressed in human skeletal muscle[48].

SBF2

The *SBF2* gene is expressed in skeletal muscle[48] and its expression in skeletal muscle is associated with a cis-eQTL[48]. In addition, within the QTL containing the *SBF2* gene, we found that the majority of genetic variants associated with ALM were located within *SBF2.*

CPNE1

Although little is known about *CPNE1,* it is an intriguing candidate because a premature stop variant (rs147019139) within the gene was associated with an increase in ALM. Furthermore, *Cpne1* was implicated as a positional candidate gene for muscle mass by previous GWAS conducted in outbred (CFW) mice[5].

Novel modifiers of *in vitro* myogenesis

We used siRNA-mediated gene knockdown in C2C12 cells to test if candidate genes affected myogenic differentiation. The *STC2*[75], *CPNE1* (identified as a positional candidate by previous research conducted in CFW mice[5]) and *SBF2* (linked to an aggressive type of Charcot-Marie-Tooth disease[76]) genes were prioritised for this assay. We assessed indices of myogenic differentiation (the number and length of the myotubes, and expression of myosin) of C2C12 cells. In total, 34,989 myotubes were identified and measured in 44 cell cultures (see Methods for details). The gene knockdown had a significant effect on myotube length, with *Cpne1 (P = 0.001, 95% confidence interval = 0.019-0.068, effect size = 0.024)* and *Stc2 (P = 0.015, 95% confidence interval= 0.007-0.066, effect size = 0.017)* showing an increase in length compared to the control

cells (Figure 4). There was no significant difference for the *Sbf2* gene. The pattern of the effect on myosin expressing area was similar to that of myotube length, however, it did not reach statistical significance (P = 0.21). The number of myotubes was also unaffected.

Discussion

The key findings of the present report are as follows: *i*) we identified a set of over 100 loci associated with ALM, a substantial expansion in comparison to previous human studies. *ii*) There is a substantial overlap of the genetic effects between middle aged and elderly subjects. *iii*) Integration of mouse and human GWAS indicates that skeletal muscle is the primary component affected by the ALM loci, facilitates prioritisation of candidate genes, and helps prediction of their effect on cellular mechanisms. *iv*) *In vitro* validation of two genes, *CPNE1* and *STC2*, as novel modifiers of muscle mass in humans.

In total, we mapped 125 loci that collectively explain 14.27% of the SNP heritability of ALM. The most recent report, a meta-analysis of 47 independent cohorts (dbGAP), comparable in sample size but ranging in subjects aged 18-to-100 years, reported five significant associations with lean body mass[8]. Even fewer associations were detected in the earlier, small sample size studies[10; 12-14; 77]. However, our results indicate that ALM is a truly polygenic trait in humans. We hypothesize that multiple factors contributed to the improved locus detection in the present GWAS. We restricted subjects' age to a narrow range, 37 to 48 years, minimising the effects of the developmental and aging-related processes on phenotypic variance. The skeletal muscle is a dynamic tissue reaching its peak mass by late 20s, then a trend of decline emerges after 40s and accelerates about two decades later[1]. An estimated 30-50% decline in muscle mass can be expected between 40 and 80 years of age[78]. These developmental and aging-related changes are not linear in progression and therefore would hamper detection of loci even if accounted for in a linear model. We tested the age effect hypothesis in a randomly generated data set of a similar size (n=95,454) which was equally divided between middle-aged and older individuals. A GWAS on ALM in this dataset identified ~ 13% fewer loci (Table S8) compared to solely middle-aged

548 adults. This analysis also captured the five loci identified by Zillikens and colleagues[8], suggesting
549 that the effects of these loci might be less sensitive to the age differences. Hence, our results
550 support the notion of age effect, which is likely to have a large impact with increasing age range. In
551 addition, unlike Zillikens and colleagues[8], the data set we used was systematically collected as
552 described by the UK Biobank project[15] and we only employed BIA measurements of lean mass.
553 Furthermore, we used a LMM to test the effects of > 17 million variants (MAF > 0.001), and our
554 analysis was adjusted for a different set of fixed effects than in previous research[8; 10; 12; 14]. Hence,
555 a combination of a homogeneous age group, the optimised genomic coverage and the method
556 used to conduct this association analysis contributed to improved detection of loci in the present
557 study.

558 The analyses presented here shed light into the complex genetic mechanisms behind the
559 appendicular muscle mass of humans. In the past, concern was expressed about the
560 reproducibility of association analyses of complex traits; however, an increasing number of human
561 GWAS have shown that their findings are remarkably reproducible[79]. The present study provides
562 further support for the reliability of association studies, demonstrating replication of 64% of ALM
563 loci in the elderly cohort. Furthermore, we show that the genetic profile characterised by depletion
564 of ALM-increasing alleles leads to a lower ALM in elderly individuals (Figure 2). Hence, it is
565 conceivable that genetic architecture predisposing individuals to lower muscle mass may lead to
566 elevated risk of sarcopenia[1].

567 Combining two experimental models, mouse and human, facilitated prioritization of
568 candidate genes for functional validation. To establish the association between the QTGs
569 underlying the identified loci and the muscular phenotype, we focused on the overlapping human
570 and mouse results. Integration of results from these two species permitted circumvention of the
571 limitations imposed by the individual models. While human GWAS often identify associated loci
572 containing single genes, it is often unclear which tissue is most relevant to the phenotype. Although
573 mouse QTLs often contain multiple positional candidate genes, mice can be used as experimental
574 models to identify loci specifically associated with skeletal muscle. In this study, we used a mouse
575 model to show that the association with skeletal muscle mass was specifically related to

differences in the cross-sectional area of the constituent muscle fibres, rather than to the number of muscle fibres in the muscle. This is because between the two founders of the LGSM AIL, the LG/J strain compared to the SM/J strain shows over 50% larger cross-sectional area of muscle fibres, but no difference in the number of fibres in soleus muscle[21]. Hence, it is conceivable that the QTGs of the majority of the overlapping loci affected muscle mass specifically via the hypertrophy of muscle fibres. Such prioritization between the two cellular mechanisms of muscle mass variability is important because genes specifically influencing cross-sectional area of muscle fibres can be targeted pharmacologically to prevent and reverse atrophy of muscle fibres in aging muscle[80].

In an effort to validate the specific QTGs and to establish the causality of their effects on skeletal muscle, we tested the siRNA-mediated knockdown effect on myogenesis *in vitro*. A knockdown of two genes, *CPNE1* and *STC2*, increased the length of the myotubes, implicating an upregulation of myogenic differentiation. We interpret this *in vitro* observation as consistent with the allelic effect of the locus identified in human GWAS. A nonsense allele within *CPNE1* was associated with an increase in ALM in both middle age and elderly populations (Table S9). The gene encodes for Copine 1, a soluble calcium-dependent membrane-binding protein[81] that up to date had not been implicated in morphology or function of skeletal muscle. An allele of the second validated gene, *STC2*, was predicted to have a damaging effect on protein, and was also associated with an increase in ALM. This effect was consistent with overexpression results in a mouse model, showing that transgene animals had substantially reduced muscle mass[75]. The gene encodes Stanniocalcin 2, a homodimeric glycoprotein hormone abundantly expressed in skeletal and cardiac muscle[82], although mechanisms of its effects on skeletal muscle remain unclear. Collectively these analyses revealed two novel modifiers of myogenesis, which were shown for the first time to be associated with muscle mass variability in humans.

In conclusion, the present study capitalised on the advantages of integrating human and mouse GWAS with *in vitro* validation of causative genes. Our results revealed over 100 genomic loci contributing to ALM in middle-aged humans. The effects of the majority of these loci persist in the elderly population. Integration of human and mouse data also highlighted novel candidate

604 genes affecting skeletal muscle mass in mammals. Two genes, *CPNE1* and *STC2* appear to be
605 novel modifiers of *in vitro* myogenesis.

606

607 Supplemental Data

608 Supplemental Data include four figures, nine tables and macro script.

609

610 Declaration of interest

611 The authors declare no competing interests.

612

613 Acknowledgements

614 The authors would like to acknowledge Dr David A. Blizard for his role in the development of the
615 ideas that led to this study and feedback on the manuscript, Professor Helen Macdonald for
616 valuable advice on study design, Dr Leslie R. Noble for help with the UK Biobank data, and Dr
617 Joseph P. Gyekis for help genotyping cohort 2 mice. The authors would like to acknowledge
618 funding from the University of Aberdeen for the Maxwell computer cluster, the Elphinstone and IMS
619 studentship for AIHC; a Schweppe Foundation Career Development Award (AAP), and the NIH
620 (NIAMS (AL: R01AR056280) and NIDA (AAP:R01DA021336, AAP:R21DA024845,
621 AAP:T32MH020065, NMG:F31DA03635803), NIGMS (NMG:T32GM007197), NHGRI (MA:R01
622 HG002899)).

623

624 Web Resources

625 Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA GWAS)[39]. URL:
626 http://fuma.ctglab.nl/

627 Ensembl[47]. URL: https://www.ensembl.org

628 Gene Tissue Expression Project (GTEx) portal[36]. URL: https://gtexportal.org/home/

629

630 # References

631 1. Lauretani, F., Russo, C.R., Bandinelli, S., Bartali, B., Cavazzini, C., Di Iorio, A., Corsi, A.M.,
632 Rantanen, T., Guralnik, J.M., and Ferrucci, L. (2003). Age-associated changes in skeletal
633 muscles and their effect on mobility: an operational diagnosis of sarcopenia. J Appl Physiol
634 95, 1851-1860.
635 2. Janssen, I., Shepard, D.S., Katzmarzyk, P.T., and Roubenoff, R. (2004). The healthcare costs of
636 sarcopenia in the United States. Journal of the American Geriatrics Society 52, 80-85.
637 3. Kim, J., Wang, Z., Heymsfield, S.B., Baumgartner, R.N., and Gallagher, D. (2002). Total-body
638 skeletal muscle mass: estimation by a new dual-energy X-ray absorptiometry method. Am J
639 Clin Nutr 76, 378-383.
640 4. Livshits, G., Gao, F., Malkin, I., Needhamsen, M., Xia, Y., Yuan, W., Bell, C.G., Ward, K., Liu, Y.,
641 and Wang, J. (2016). Contribution of Heritability and Epigenetic Factors to Skeletal Muscle
642 Mass Variation in United Kingdom Twins. J Clin Endocrinol Metab 101, 2450-2459.
643 5. Nicod, J., Davies, R.W., Cai, N., Hassett, C., Goodstadt, L., Cosgrove, C., Yee, B.K., Lionikaite,
644 V., McIntyre, R.E., Remme, C.A., et al. (2016). Genome-wide association of multiple
645 complex traits in outbred mice by ultra-low-coverage sequencing. Nature Genetics 48, 912-
646 918.
647 6. Hernandez Cordero, A.I., Carbonetto, P., Riboni Verri, G., Gregory, J.S., Vandenbergh, D.J., J,
648 P.G., Blizard, D.A., and Lionikas, A. (2018). Replication and discovery of musculoskeletal
649 QTLs in LG/J and SM/J advanced intercross lines. Physiological reports 6.
650 7. Carbonetto, P., Cheng, R., Gyekis, J.P., Parker, C.C., Blizard, D.A., Palmer, A.A., and Lionikas,
651 A. (2014). Discovery and refinement of muscle weight QTLs in B6 x D2 advanced
652 intercross mice. Physiol Genomics 46, 571-582.
653 8. Zillikens, M.C., Demissie, S., Hsu, Y.-H., Yerges-Armstrong, L.M., Chou, W.-C., Stolk, L.,
654 Livshits, G., Broer, L., Johnson, T., Koller, D.L., et al. (2017). Large meta-analysis of
655 genome-wide association studies identifies five loci for lean body mass. Nat Commun 8, 80.
656 9. Urano, T., and Inoue, S. (2015). Recent genetic discoveries in osteoporosis, sarcopenia and
657 obesity. Endocrine journal 62, 475-484.
658 10. Hai, R., Pei, Y.-F., Shen, H., Zhang, L., Liu, X.-G., Lin, Y., Ran, S., Pan, F., Tan, L.-J., and Lei,
659 S.-F. (2012). Genome-wide association study of copy number variation identified gremlin1
660 as a candidate gene for lean body mass. J Hum Genet 57, 33.
661 11. Liu, X.G., Tan, L.J., Lei, S.F., Liu, Y.J., Shen, H., Wang, L., Yan, H., Guo, Y.F., Xiong, D.H.,
662 Chen, X.D., et al. (2009). Genome-wide association and replication studies identified TRHR
663 as an important gene for lean body mass. Am J Hum Genet 84, 418-423.
664 12. Guo, Y.-F., Zhang, L.-S., Liu, Y.-J., Hu, H.-G., Li, J., Tian, Q., Yu, P., Zhang, F., Yang, T.-L.,
665 and Guo, Y. (2013). Suggestion of GLYAT gene underlying variation of bone size and body
666 lean mass as revealed by a bivariate genome-wide association study. Hum Genet 132,
667 189-199.
668 13. Ran, S., Liu, Y.-J., Zhang, L., Pei, Y., Yang, T.-L., Hai, R., Han, Y.-Y., Lin, Y., Tian, Q., and
669 Deng, H.-W. (2014). Genome-wide association study identified copy number variants
670 important for appendicular lean mass. PloS ONE 9, e89776.
671 14. Urano, T., Shiraki, M., Sasaki, N., Ouchi, Y., and Inoue, S. (2014). Large-scale analysis reveals
672 a functional single-nucleotide polymorphism in the 5′-flanking region of PRDM16 gene
673 associated with lean body mass. Aging cell 13, 739-743.
674 15. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P.,
675 Green, J., Landray, M., et al. (2015). UK biobank: an open access resource for identifying
676 the causes of a wide range of complex diseases of middle and old age. PLoS medicine 12,
677 e1001779.

16. Lynch, N.A., Metter, E.J., Lindle, R.S., Fozard, J.L., Tobin, J.D., Roy, T.A., Fleg, J.L., and Hurley, B.F. (1999). Muscle quality. I. Age-associated differences between arm and leg muscle groups. J Appl Physiol (1985) 86, 188-194.
17. Kallman, D.A., Plato, C.C., and Tobin, J.D. (1990). The role of muscle loss in the age-related decline of grip strength: cross-sectional and longitudinal perspectives. J Gerontol 45, M82-88.
18. Guigo, R., Dermitzakis, E.T., Agarwal, P., Ponting, C.P., Parra, G., Reymond, A., Abril, J.F., Keibler, E., Lyle, R., Ucla, C., et al. (2003). Comparison of mouse and human genomes followed by experimental verification yields an estimated 1,019 additional genes. Proceedings of the National Academy of Sciences of the United States of America 100, 1140-1145.
19. Parker, C.C., Gopalakrishnan, S., Carbonetto, P., Gonzales, N.M., Leung, E., Park, Y.J., Aryee, E., Davis, J., Blizard, D.A., Ackert-Bicknell, C.L., et al. (2016). Genome-wide association study of behavioral, physiological and gene expression traits in outbred CFW mice. Nat Genet 48, 919-926.
20. Carroll, A., Cheng, R., Collie-Duguid, E., Meharg, C., Scholz, M.E., Fiering, S.N., Fields, J.L., Palmer, A.A., and Lionikas, A. (2017). Fine mapping of genes determining extrafusal fiber properties in murine soleus muscle. Physiological genomics 49, 141-150.
21. Carroll, A.M., Palmer, A.A., and Lionikas, A. (2012). QTL analysis of type I and type IIA fibers in soleus muscle in a cross between LG/J and SM/J mouse strains. Frontiers in genetics 2, 99.
22. Partridge, C.G., Fawcett, G.L., Wang, B., Semenkovich, C.F., and Cheverud, J.M. (2014). The effect of dietary fat intake on hepatic gene expression in LG/J AND SM/J mice. BMC genomics 15, 1.
23. Cheverud, J.M., Vaughn, T.T., Pletscher, L.S., Peripato, A.C., Adams, E.S., Erikson, C.F., and King-Ellison, K.J. (2001). Genetic architecture of adiposity in the cross of LG/J and SM/J inbred mice. Mammalian genome : official journal of the International Mammalian Genome Society 12, 3-12.
24. Lionikas, A., Meharg, C., Derry, J., Ratkevicius, A., Carroll, A., Vandenbergh, D., and Blizard, D.A. (2012). Resolving candidate genes of mouse skeletal muscle QTL via RNA-Seq and expression network analyses. BMC genomics 13.
25. Goodale, H. (1938). A study of the inheritance of body weight in the albino mouse by selection. Journal of Heredity 29, 101-112.
26. MacArthur, J.W. (1944). Genetics of Body Size and Related Characters. I. Selecting Small and Large Races of the Laboratory Mouse. The American Naturalist 78, 142-157.
27. Lionikas, A., Cheng, R., Lim, J.E., Palmer, A.A., and Blizard, D.A. (2010). Fine-mapping of muscle weight QTL in LG/J and SM/J intercrosses. Physiological genomics 42, 33-38.
28. Darvasi, A., and Soller, M. (1995). Advanced intercross lines, an experimental population for fine genetic mapping. Genetics 141, 1199-1207.
29. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2017). Genome-wide genetic data on ~500,000 UK Biobank participants. bioRxiv.
30. Howie, B., Marchini, J., and Stephens, M. (2011). Genotype imputation with thousands of genomes. G3 (Bethesda, Md) 1, 457-470.
31. Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J., Kutalik, Z., et al. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet 46, 1173-1186.
32. Loh, P.-R., Tucker, G., Bulik-Sullivan, B.K., Vilhjalmsson, B.J., Finucane, H.K., Salem, R.M., Chasman, D.I., Ridker, P.M., Neale, B.M., Berger, B., et al. (2015). Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat Genet 47, 284-290.
33. Bulik-Sullivan, B.K., Loh, P.R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 47, 291-295.
34. Egner, I.M., Bruusgaard, J.C., Eftestol, E., and Gundersen, K. (2013). A cellular memory mechanism aids overload hypertrophy in muscle long after an episodic exposure to anabolic steroids. J Physiol 591, 6221-6230.

35. Sanchez-Gurmaches, J., and Guertin, D.A. (2014). Adipocytes arise from multiple lineages that are heterogeneously and dynamically distributed. Nat Commun 5, 4099.
36. The International HapMap Consortium. (2005). A haplotype map of the human genome. Nature 437, 1299-1320.
37. Pe'er, I., Yelensky, R., Altshuler, D., and Daly, M.J. (2008). Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genet Epidemiol 32, 381-385.
38. Loh, P.-R., Bhatia, G., Gusev, A., Finucane, H.K., Bulik-Sullivan, B.K., Pollack, S.J., de Candia, T.R., Lee, S.H., Wray, N.R., and Kendler, K.S. (2015). Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance components analysis. Nat Genet 47, 1385.
39. Watanabe, K., Taskesen, E., van Bochoven, A., and Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. Nature Communications 8, 1826.
40. Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. Nature 526, 68-74.
41. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4, 7.
42. Yang, J., Ferreira, T., Morris, A.P., Medland, S.E., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., Weedon, M.N., Loos, R.J., et al. (2012). Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet 44, 369-375, s361-363.
43. Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., and Huber, W. (2005). BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics (Oxford, England) 21, 3439-3440.
44. Durinck, S., Spellman, P.T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nature protocols 4, 1184-1191.
45. Flanagan, S.E., Patch, A.M., and Ellard, S. (2010). Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. Genetic testing and molecular biomarkers 14, 533-537.
46. Ng, P.C., and Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 31, 3812-3814.
47. Frankish, A., Vullo, A., Zadissa, A., Yates, A., Thormann, A., Parker, A., Gall, A., Moore, B., Walts, B., Aken, B.L., et al. (2017). Ensembl 2018. Nucleic Acids Research 46, D754-D761.
48. Carithers, L.J., Ardlie, K., Barcus, M., Branton, P.A., Britton, A., Buia, S.A., Compton, C.C., DeLuca, D.S., Peter-Demchok, J., Gelfand, E.T., et al. (2015). A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. Biopreservation and biobanking 13, 311-319.
49. Cheverud, J.M., Routman, E.J., Duarte, F., van Swinderen, B., Cothran, K., and Perel, C. (1996). Quantitative trait loci for murine growth. Genetics 142, 1305-1319.
50. Cheng, R., Lim, J.E., Samocha, K.E., Sokoloff, G., Abney, M., Skol, A.D., and Palmer, A.A. (2010). Genome-wide association studies and the problem of relatedness among advanced intercross lines and other highly recombinant populations. Genetics 185, 1033-1044.
51. Gonzales, N.M., Seo, J., Hernandez Cordero, A.I., St. Pierre, C.L., Gregory, J.S., Distler, M.G., Abney, M., Canzar, S., Lionikas, A., and Palmer, A.A. (2018). Genome wide association analysis in a mouse advanced intercross line. Nature Communications 9, 5162.
52. Browning, Brian L., and Browning, Sharon R. (2016). Genotype Imputation with Millions of Reference Samples. The American Journal of Human Genetics 98, 116-126.
53. Nikolskiy, I., Conrad, D.F., Chun, S., Fay, J.C., Cheverud, J.M., and Lawson, H.A. (2015). Using whole-genome sequences of the LG/J and SM/J inbred mouse strains to prioritize quantitative trait genes and nucleotides. BMC genomics 16, 415.
54. Astle, W., and Balding, D.J. (2009). Population structure and cryptic relatedness in genetic association studies. Statistical Science, 451-471.
55. Price, A.L., Zaitlen, N.A., Reich, D., and Patterson, N. (2010). New approaches to population stratification in genome-wide association studies. Nature Reviews Genetics 11, 459-463.

56. Yang, J., Zaitlen, N.A., Goddard, M.E., Visscher, P.M., and Price, A.L. (2014). Advantages and pitfalls in the application of mixed-model association methods. Nature genetics 46, 100-106.
57. Yu, J., Pressoir, G., Briggs, W.H., Bi, I.V., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., and Holland, J.B. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature genetics 38, 203-208.
58. Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. Nature Genetics 44, 821-824.
59. Parker, C.C., Carbonetto, P., Sokoloff, G., Park, Y.J., Abney, M., and Palmer, A.A. (2014). High-resolution genetic mapping of complex traits from a combined analysis of F2 and advanced intercross mice. Genetics 198, 103-116.
60. Cheng, R., and Palmer, A.A. (2013). A simulation study of permutation, bootstrap, and gene dropping for assessing statistical significance in the case of unequal relatedness. Genetics 193, 1015-1018.
61. Cheng, R., Parker, C.C., Abney, M., and Palmer, A.A. (2013). Practical considerations regarding the use of genotype and pedigree data to model relatedness in the context of genome-wide association studies. G3 (Bethesda, Md) 3, 1861-1867.
62. Speed, D., and Balding, D.J. (2015). Relatedness in the post-genomic era: is it still useful? Nature Reviews Genetics 16, 33-44.
63. Weir, B.S., Anderson, A.D., and Hepler, A.B. (2006). Genetic relatedness analysis: modern data and new challenges. Nat Rev Genet 7, 771-780.
64. Churchill, G.A., and Doerge, R.W. (1994). Empirical threshold values for quantitative trit mapping. Genetics 138, 963-971.
65. Broman, K.W., Wu, H., Sen, S., and Churchill, G.A. (2003). R/qtl: QTL mapping in experimental crosses. Bioinformatics (Oxford, England) 19, 889-890.
66. Manichaikul, A., Dupuis, J., Sen, Ś., and Broman, K.W. (2006). Poor performance of bootstrap confidence intervals for the location of a quantitative trait locus. Genetics 174, 481-489.
67. Dupuis, J., and Siegmund, D. (1999). Statistical methods for mapping quantitative trait loci from a dense set of markers. Genetics 151, 373-386.
68. Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT–PCR. Nucleic acids research 29, e45-e45.
69. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nature Methods 9, 676.
70. Arganda-Carreras, I., Fernández-González, R., Muñoz-Barrutia, A., and Ortiz-De-Solorzano, C. (2010). 3D reconstruction of histological sections: application to mammary gland tissue. Microscopy research and technique 73, 1019-1029.
71. Polder, G., Hovens, H., and Zweers, A. (2010). Measuring shoot length of submerged aquatic plants using graph analysis. In Proceedings of the ImageJ User and Developer Conference 2010, Mondorf-les-Bains, Luxembourg, 27-29 October 2010. pp 172-177.
72. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat Methods 7, 248-249.
73. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J., and Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res 47, D886-d894.
74. Ioannidis, N.M., Rothstein, J.H., Pejaver, V., Middha, S., McDonnell, S.K., Baheti, S., Musolf, A., Li, Q., Holzinger, E., Karyadi, D., et al. (2016). REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. American journal of human genetics 99, 877-885.
75. Gagliardi, A.D., Kuo, E.Y., Raulic, S., Wagner, G.F., and DiMattia, G.E. (2005). Human stanniocalcin-2 exhibits potent growth-suppressive properties in transgenic mice independently of growth hormone and IGFs. American journal of physiology Endocrinology and metabolism 288, E92-105.

847 76. Schneider, B.P., Lai, D., Shen, F., Jiang, G., Radovich, M., Li, L., Gardner, L., Miller, K.D.,
848 O'Neill, A., Sparano, J.A., et al. (2016). Charcot-Marie-Tooth gene, SBF2, associated with
849 taxane-induced peripheral neuropathy in African Americans. Oncotarget 7, 82244-82253.
850 77. Liu, X.G., Tan, L.J., Lei, S.F., Liu, Y.J., Shen, H., Wang, L., Yan, H., Guo, Y.F., Xiong, D.H.,
851 Chen, X.D., et al. (2009). Genome-wide association and replication studies identified TRHR
852 as an important gene for lean body mass. American journal of human genetics 84, 418-423.
853 78. Faulkner, J.A., Larkin, L.M., Claflin, D.R., and Brooks, S.V. (2007). Age-related changes in the
854 structure and function of skeletal muscles. Clinical and experimental pharmacology &
855 physiology 34, 1091-1096.
856 79. Marigorta, U.M., Rodríguez, J.A., Gibson, G., and Navarro, A. (2018). Replicability and
857 Prediction: Lessons and Challenges from GWAS. Trends in Genetics.
858 80. Brocca, L., McPhee, J.S., Longa, E., Canepari, M., Seynnes, O., De Vito, G., Pellegrino, M.A.,
859 Narici, M., and Bottinelli, R. (2017). Structure and function of human muscle fibres and
860 muscle proteome in physically active older men. The Journal of physiology 595, 4823-4844.
861 81. Tomsig, J.L., and Creutz, C.E. (2002). Copines: a ubiquitous family of Ca2+-dependent
862 phospholipid-binding proteins. Cellular and Molecular Life Sciences CMLS 59, 1467-1477.
863 82. Ishibashi, K., Miyamoto, K., Taketani, Y., Morita, K., Takeda, E., Sasaki, S., and Imai, M.
864 (1998). Molecular cloning of a second human stanniocalcin homologue (STC2). Biochem
865 Biophys Res Commun 250, 252-258.

866

867 Figure legends

868 Figure 1. Map of genome associations with the appendicular lean mass (ALM) of humans.
869 Genome wide association study (GWAS) on the ALM of middle-aged adults from the UK Biobank.
870 Significance level is presented on the vertical axis, while the chromosomal position of each genetic
871 marker is shown on the horizontal axis. Red line across the plot represents the genome wide
872 threshold of significance ($P < 5 \times 10^{-8}$). This plot shows the association of variants with MAF >
873 0.001.

874

875 Figure 2. Genetic lean mass score affects the appendicular lean mass (ALM) in elderly humans.
876 The plot shows the ALM (kg) of the elderly cohort on the vertical axis. The elderly cohort was
877 ranked by genetic lean mass score and clustered in five quantiles (Q1 to Q5) (horizontal axis). The
878 average genetic lean mass score (± standard error) of each quantile is shown in parenthesis below
879 the horizontal axis. The overall quantile effect of the genetic lean mass score on ALM was tested
880 with Kruskal-Wallis test and the resulting *P* value is presented on the top of horizontal line above
881 the bars. The ALM median differences between the groups were tested using a Wilcoxon test; the
882 significance level of each comparison is presented above the horizontal lines with a holm adjusted
883 *P* value.

884

885 Figure 3. Muscle weight QTLs identified in mice of the LGSM AIL and density plot of the
886 genotypes. The circle plot (A) shows from the outer to the inner ring the GWAS of the TA, EDL,
887 gastrocnemius and soleus muscle weights. Chromosomal position of each SNP is shown in the
888 outer black circle of the plot; chromosome names are shown outside as “Chr”. Dots within each
889 chromosome space represent the association ($-\log_{10}$ *P value*) of each SNP tested. Dotted blue
890 lines represent the genome-wide threshold ($P < 6.45 \times 10^{-06}$) of significance, and red dots above
891 the genome-wide threshold are significantly associated SNPs. (B) Plots of the allelic effect of the
892 *Skmw34*, *Skmw55* and *Skmw46* QTLs on the EDL muscle mass. These QTLs were identified for
893 the four muscles investigated. Vertical axis represents the residual muscle mass adjusted for sex,
894 age, dissector and long bone length, and the horizontal axis shows the genotypes (LG/J
895 homozygote, heterozygote and SM/J homozygote). Bellow the horizontal axis, the number of
896 individuals with a given genotype is provided. The violin shapes within the plot area represents the
897 distribution of individuals with the genotypes. Box whiskers represent minimum and maximum
898 values distance between a whisker and the top or bottom of the box contains 25% of the
899 distribution, the box captures 50% of the distribution, and the bold horizontal line represents the
900 median. Pairwise comparison *P* value (t-test) is shown above horizontal lines at the top of the
901 plots.

902

903 Figure 4. Gene knockdown effect on C2C12 myotube length.
904 This figure shows the gene knockdown effect of the *Cpne1*, *Sbf2* and *Stc2* genes on myotube
905 length. The overall effect of the gene knockdown on myotube length was tested using ANOVA and
906 the resulting *P* value was 0.00017 ($F_{3,\ 34985} = 6.63$). Vertical axis represents the myotube length
907 (quantile normalised) residuals (adjusted for area analysed and batch of cells), and the horizontal
908 axis shows control and knockdown gene groups. Boxes represent the distribution of the myotube
909 length for each group. Box whiskers represent minimum and maximum values within 1.5-fold
910 interquartile range above the 75th percentile and below the 25th percentile; the box captures 50% of
911 the distribution, and the bold horizontal line represents the median value of the myotube length

normalized residuals distribution for each knockdown group. Each red dot represents a single cell culture sample for each knockdown group. Statistically significant t–test *P*-values between control and knockdown genes are presented above horizontal lines. Effects without a statistically significant difference between the control and gene knockdown are presented as “ns”. *Cpne1* and *Stc2* knockdown groups were not different from each other (*P > 0.05*). *Sbf2* gene knockdown differed from *Cpne1 (P = 0.002)* and *Stc2 (P = 0.043).*

Tables

Table 1. Summary of the middle-aged cohort

Trait	N		MIN	MAX	AVERAGE	SD	SNP heritability ± SE
ALM (kg)	Females =	51,238	11.80	41.60	20.02	2.61	0.36 ± 0.003
	Males =	43,996	15.30	54.50	30.00	3.99	
Arm lean mass (kg)	Females =	51,248	1.00	5.10	2.29	0.32	0.32 ± 0.003
	Males =	44,007	1.40	7.10	3.83	0.58	
Leg lean mass (kg)	Females =	51,258	4.50	16.60	7.76	1.00	0.36 ± 0.003
	Males =	44,020	6.20	20.00	11.25	1.43	
Leg (cm)	Females =	51,228	36.00	113.00	76.56	4.33	0.59 ± 0.010
	Males =	43,967	40.00	122.00	83.80	4.73	
WBF (kg)	Females =	51,239	5.00	109.80	25.68	10.70	0.33 ± 0.006
	Males =	43,793	5.00	88.50	21.08	8.24	

Column description from left to right: 1) Trait, 2) Number of records, 3) Minimum value within the distribution of each trait, 4) Maximum value within the distribution of each trait, 5) Average value of each trait, 6) Standard deviation, 7) SNP heritability of the ALM across sex. All summary statistic values were calculated for each sex group. ALM: appendicular lean mass. WBF: whole body fat.

Table 2. Summary of the LGSM AIL muscle traits

Trait	N		MIN	MAX	AVERAGE	SD	SNP heritability ± SE
TA (mg)	Females =	675	26.60	57.20	42.22	5.34	0.39 ± 0.03
	Males =	1,186	31.60	70.80	50.11	6.73	

EDL (mg)	Females =	675	4.60	10.40	7.52	0.94	0.42 ± 0.03
	Males =	1,184	5.90	13.30	9.31	1.30	
Gastrocnemius (mg)	Females =	675	64.00	133.00	93.15	10.68	0.31 ± 0.03
	Males =	1,187	70.20	174.90	119.32	16.32	
Soleus (mg)	Females =	671	3.20	10.30	6.34	1.18	0.30 ± 0.03
	Males =	1,187	4.00	13.50	7.78	1.64	

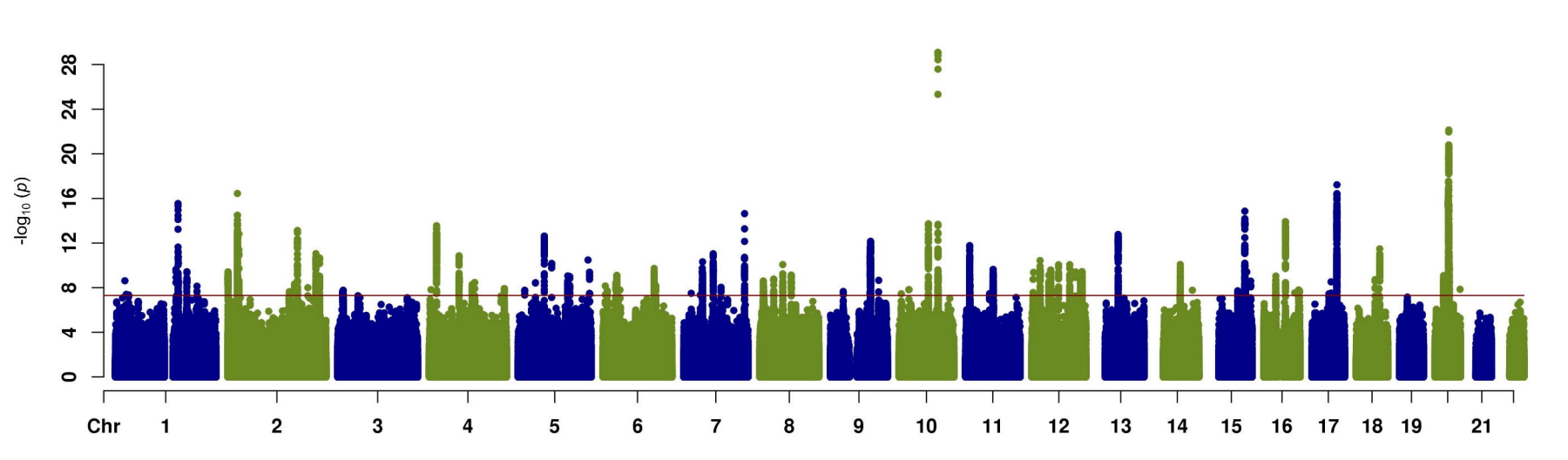
Column description from left to right: 1) Trait, 2) Number of records, 3) Minimum value within the distribution of each trait, 4) Maximum value within the distribution of each trait, 5) Average or mean value of each trait distribution, 6) Standard deviation of the mean, 7) SNP heritability for each trait across sex. Summary statistic values were calculated for each sex group.

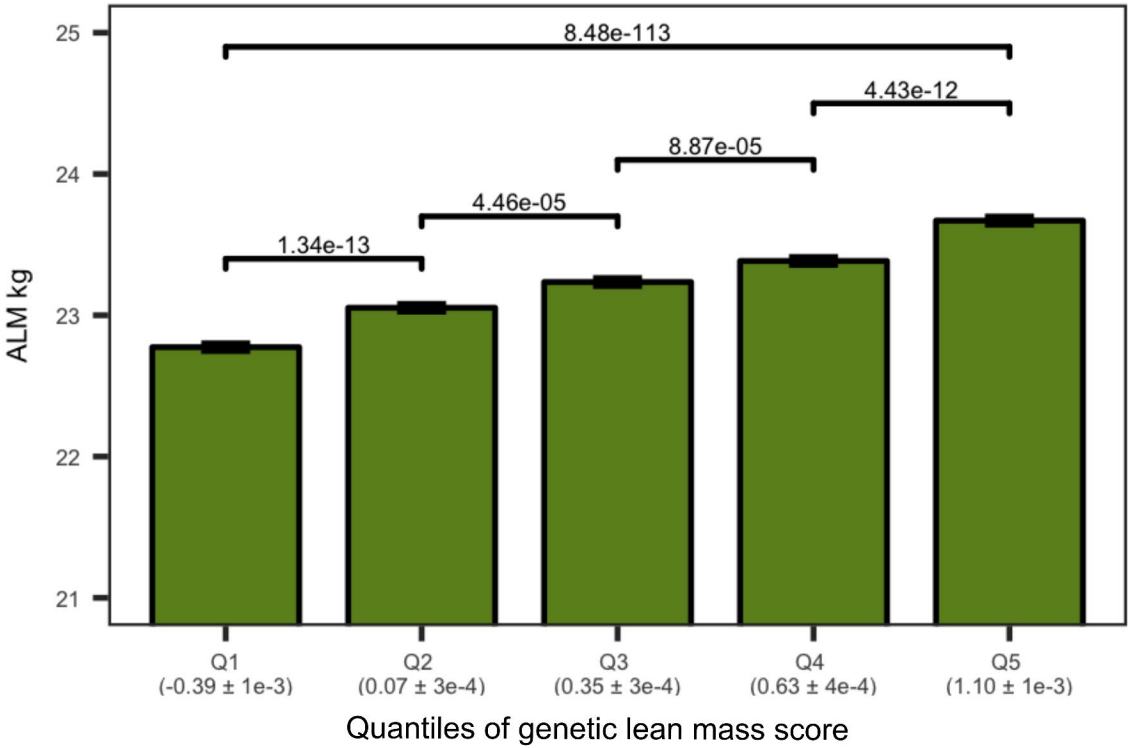
931 Table 3. Syntenic regions between human and mouse QTLs and positional candidate genes

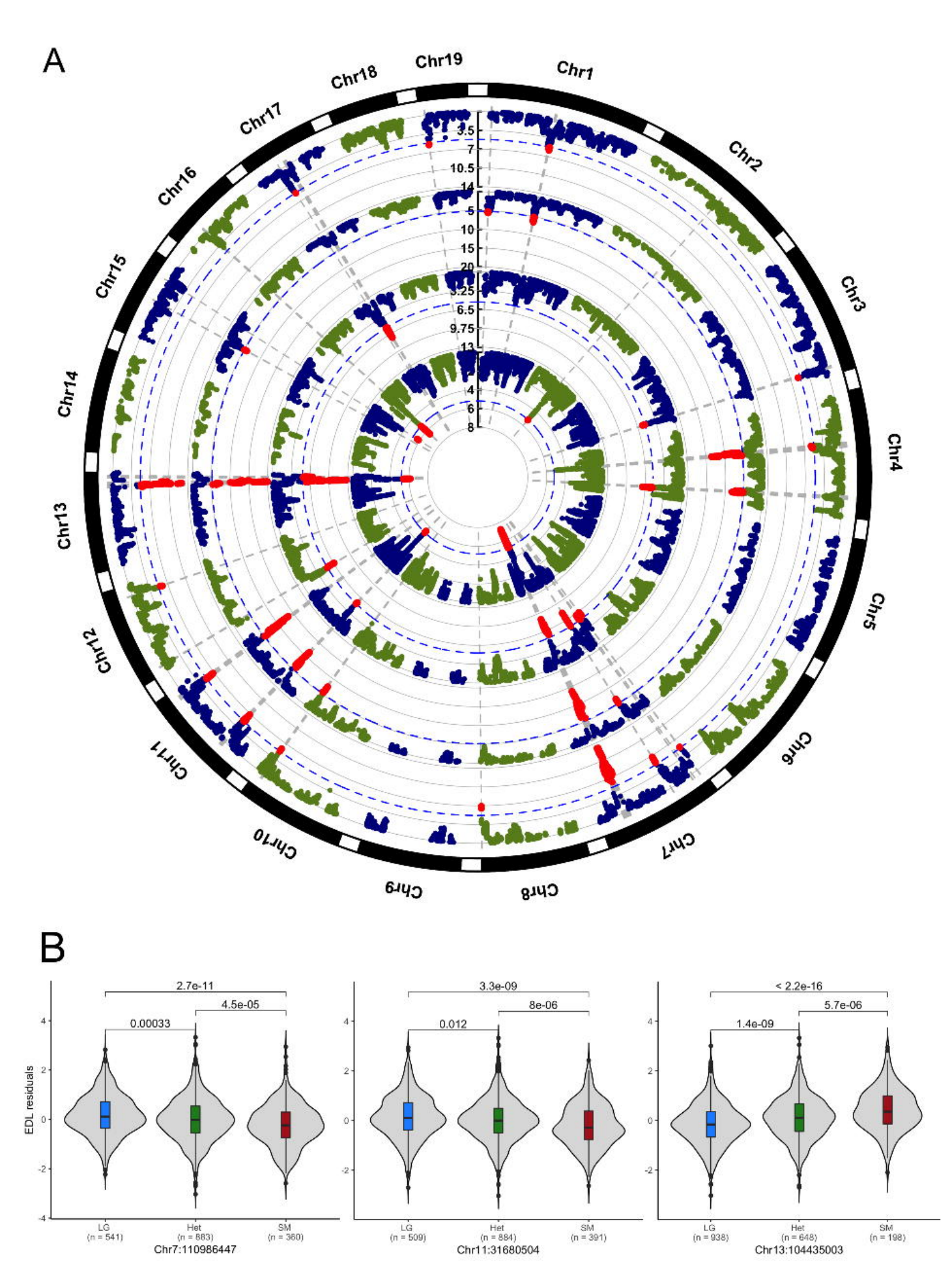
Human locus peak pos	**Mouse QTL peak pos (syntenic to human)**	**Elderly cohort *P***	**Gene symbol**	**Human gene name**	**Differential expression in mouse Soleus**	**Differential expression in mouse TA**
5:64602788	13:104435003	n/a	*ADAMTS6*	ADAM metallopeptidase with thrombospondin type 1 motif 6	0.440	0.641
5:172755066	11:31680504	9.00×10^{-11}	*STC2*	stanniocalcin 2	0.969	0.981
6:32038550	17:34968724	1.90×10^{-10}	*STK19*	serine/threonine kinase 19	0.432	0.319
			TNXB	tenascin XB	0.630	0.541
9:119309525	4:65415188	1.70×10^{-08}	*PAPPA*	pappalysin 1	n/a	0.893
			ASTN2	astrotactin 2	0.014	0.745
11:10303939	7:110986447	3.50×10^{-19}	*SBF2*	SET binding factor 2	0.762	0.893
			ADM	adrenomedullin	n/a	0.280
			AMPD3	adenosine monophosphate deaminase 3	0.064	0.110

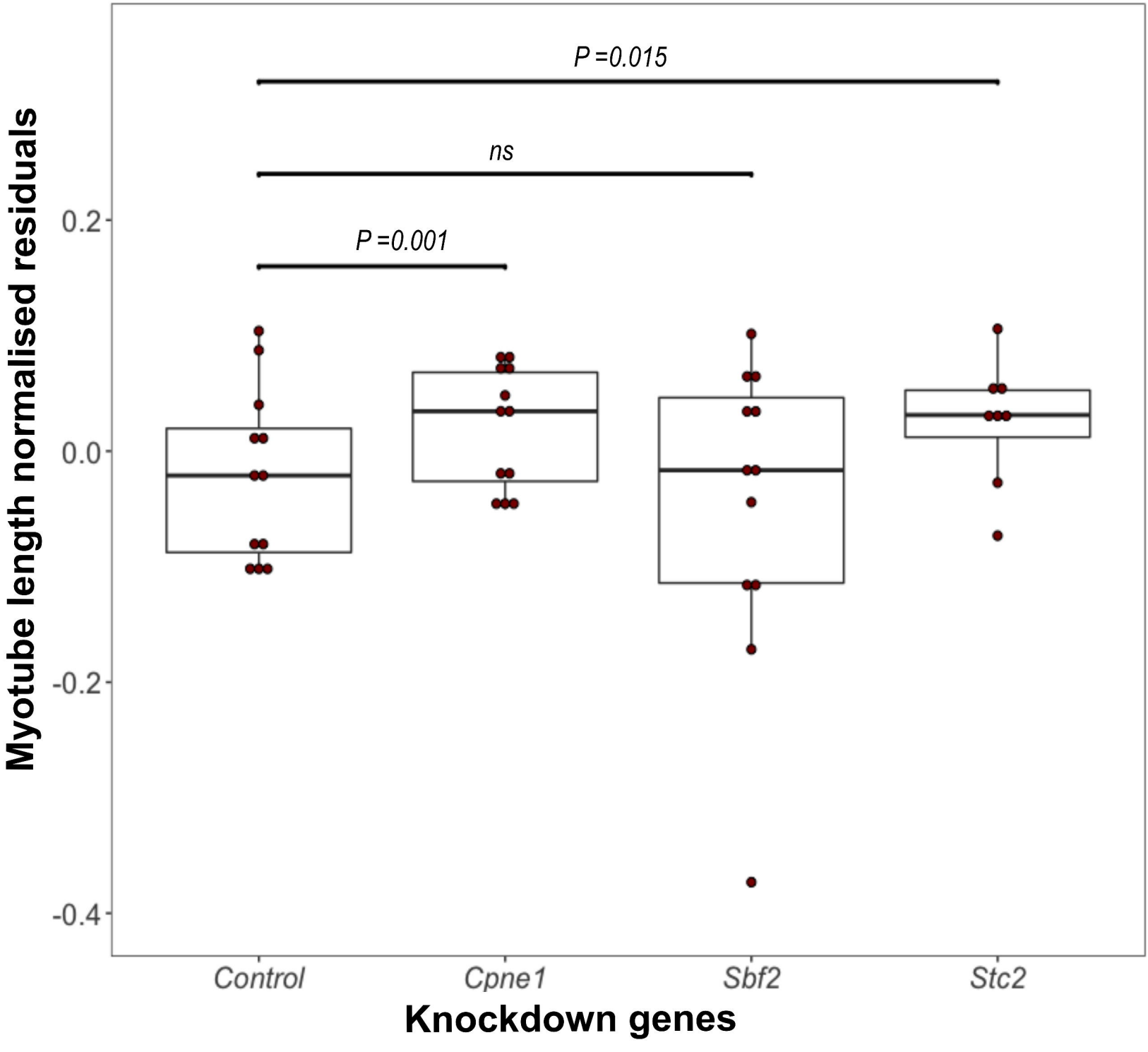
932 Column description from left to right: 1) ALM Human locus peak position as “chromosome: base pair position”, 2) LGSM QTL peak position as
933 “chromosome: base pair position” (syntenic to human), 3) Elderly cohort *P* value, 4) Human gene symbol, 5) Human gene name, 6) Adjusted *P* value
934 of differential expression between the soleus muscle of the LG/J and SM/J mouse strains[24], 7) Adjusted *P* value of differential expression between the
935 TA muscle of the LG/J and SM/J mouse strains[24]

936


-log10 (p)
0
4
8
12
16
20
24
28
Chr
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
21


8.48e-113
4.43e-12
8.87e-05
4.46e-05
1.34e-13
25
24
23
22
21
ALM kg
Q1
(-0.39 ± 1e-3)
Q2
(0.07 ± 3e-4)
Q3
(0.35 ± 3e-4)
Q4
(0.63 ± 4e-4)
Q5
(1.10 ± 1e-3)
Quantiles of genetic lean mass score


A
Chr19
Chr1
Chr18
Chr2
Chr17
Chr3
Chr16
Chr4
Chr15
Chr5
Chr14
Chr6
Chr13
Chr7
Chr12
Chr8
Chr11
Chr9
Chr10
3.5
7
10.5
14
5
10
15
20
3.25
6.5
9.75
13
2
4
6
8
B
EDL residuals
2.7e-11
4.5e-05
0.00033
LG (n = 541)
Het (n = 883)
SM (n = 360)
Chr7:110986447
3.3e-09
8e-06
0.012
LG (n = 509)
Het (n = 884)
SM (n = 391)
Chr11:31680504
< 2.2e-16
5.7e-06
1.4e-09
LG (n = 938)
Het (n = 648)
SM (n = 198)
Chr13:104435003


P =0.015
ns
P =0.001
0.2
0.0
-0.2
-0.4
Myotube length normalised residuals
Control
Cpne1
Sbf2
Stc2
Knockdown genes