


RESEARCH

Open Access



# Multi-ancestry whole genome sequencing analysis of lean body mass

Xiaoyu Zhang<sup>1\*†</sup>, Kuan-Jui Su<sup>2\*†</sup>, Bodhisattwa Banerjee<sup>3†</sup>, Ittai Eres<sup>4†</sup>, Yi-Hsiang Hsu<sup>5</sup>, Carolyn J. Crandall<sup>6</sup>, Rajashekar Donaka<sup>7</sup>, Zhe Han<sup>8</sup>, Rebecca D. Jackson<sup>9^</sup>, Hanhan Liu<sup>8</sup>, Zhe Luo<sup>2</sup>, Braxton D. Mitchell<sup>8</sup>, Chuan Qiu<sup>2</sup>, Qing Tian<sup>2</sup>, Hui Shen<sup>2</sup>, Ming-Ju Tsai<sup>5</sup>, Kerri L. Wiggins<sup>10</sup>, Hanfei Xu<sup>1</sup>, Michelle Yau<sup>5</sup>, Lan-Juan Zhao<sup>2</sup>, Xiao Zhang<sup>2</sup>, May E. Montasser<sup>8</sup>, Douglas P. Kiel<sup>5†</sup>, Hong-Wen Deng<sup>2†</sup>, Ching-Ti Liu<sup>1\*†</sup> and David Karasik<sup>5,7\*†</sup> 

<sup>†</sup>Xiaoyu Zhang, Kuan-Jui Su, Bodhisattwa Banerjee and Ittai Eres co-first authors.

<sup>†</sup>Douglas P. Kiel, Hong-Wen Deng, Ching-Ti Liu and David Karasik co-last authors.

<sup>^</sup>Rebecca D. Jackson is deceased.

\*Correspondence: xyzhang6@bu.edu; ksu2@tulane.edu; ctliu@bu.edu; karasik@hsl.harvard.edu

<sup>1</sup> Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118, USA

<sup>2</sup> Center for Biomedical Informatics and Genomics, Tulane University, New Orleans, LA 70112, USA

<sup>5</sup> Hinda and Arthur Marcus Institute for Aging Research, Hebrew SeniorLife, Boston, MA 02131, USA

Full list of author information is available at the end of the article

## Abstract

**Background:** Lean body mass is a crucial physiological component of body composition. Although lean body mass has a high heritability, studies evaluating the genetic determinants of lean mass (LM) have to date been limited largely to genome-wide association studies (GWAS) and common variants. Using whole genome sequencing (WGS)-based studies, we aimed to discover novel genetic variants associated with LM in population-based cohorts with multiple ancestries.

**Results:** We describe the largest WGS-based meta-analysis of lean body mass to date, encompassing 10,729 WGS samples from six TOPMed cohorts and the Louisiana Osteoporosis Study (LOS) cohort, measured with dual-energy X-ray absorptiometry. We identify seven genome-wide loci significantly associated with LM not reported by previous GWAS. We partially replicate these associations in UK Biobank samples. In rare variant analysis, we discover one novel protein-coding gene, *DMAC1*, associated with both whole-body LM and appendicular LM in females, and a long non-coding RNA gene linked to appendicular LM in males. Both genes exhibit notably high expression levels in skeletal muscle tissue. We investigate the functional roles of two novel lean-mass-related genes, *EMP2* and *SSUH2*, in animal models. *EMP2* deficiency in *Drosophila* leads to significantly reduced mobility without altering muscle tissue or body fat morphology, whereas an *SSUH2* gene mutation in zebrafish stimulates muscle fiber growth.

**Conclusions:** Our comprehensive analysis, encompassing a large-scale WGS meta-analysis and functional investigations, reveals novel genomic loci and genes associated with lean mass traits, shedding new insights into pathways influencing muscle metabolism and muscle mass regulation.



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

## Background

Lean body mass (LM), which includes but is not limited to the body's muscle mass, serves as an important physiological component of body composition. Low lean body mass reflects a lower amount of muscle tissue, which is associated with functional impairment and physical disability and is a major modifiable cause of frailty in the elderly population [1]. Low lean body mass is also associated with higher surgical and post-operative complications, longer length of hospital stay, lower physical function, poorer quality of life [2], malnutrition [3], and mortality [2, 4], which makes it an important measure in clinical practice. Lean mass can be measured by dual-energy X-ray absorptiometry (DXA) or bioelectrical impedance analysis (BIA). Two measurements of LM are usually reported: whole-body lean mass (WB-LM) and appendicular LM (A-LM), the latter being the lean mass in the arms and legs. While A-LM consists largely of skeletal muscle as well as some other connective tissues, WB-LM is determined by skeletal muscle, parenchymatous organs, cardiac muscle, and blood vessels. Studying both WB-LM and A-LM provides a comprehensive understanding of image-based markers of muscle health by capturing systemic and limb-specific changes. Specifically, A-LM may better reflect muscle mass because the limbs have fewer organs included in the lean mass derivation. Additionally, WB-LM provides a more complete picture of overall body composition. This dual approach enhances the assessment of body composition and offers robust insights into sarcopenia and related health outcomes.

Lean body mass has a significant genetic component, as evidenced by a high heritability of 50–80% observed in twins and in families studies [5]. To identify variants associated with this phenotype, we previously performed a large meta-analysis of GWASs that amassed 20 cohorts of European ancestry with a total sample size of > 38,000 for WB-LM and > 28,000 for A-LM [6]. Despite the large sample used, the percentage of phenotypic variance explained by the identified SNPs was very small, suggesting that common variants do not explain most of the heritability of lean body mass. A more recent GWAS of A-LM, conducted in 85,750 middle-aged (aged 38–49 years) individuals from the UK Biobank (UKBB) [7], identified a total of 182 loci, 78% of which were replicated in an independent set of 181,862 older (aged 60–74 years) individuals from the same UKBB cohort [1]. Pei et al. [1] performed a GWAS of A-LM in the full UKBB European-descent cohort and identified > 1000 independent variants meeting genome-wide significance. Although GWAS have identified loci associated with WB-LM and A-LM, they are not designed to identify novel rare variations that may have larger effect sizes than common variants. Also, most GWAS were conducted in populations of European-descent and have limited coverage of variants uncommon in European ancestry populations. Furthermore, sex-specific analyses have not been extensively explored in these studies, potentially overlooking important genetic differences between sexes.

Here we aimed to achieve a better understanding of the genetic etiology of muscle mass by attempting to discover novel genetic variants associated with lean mass. We postulated that deep whole genome sequencing performed in population-based cohorts with multiple ancestries may provide new insights into pathways influencing muscle

metabolism and muscle integrity regulation. This knowledge in turn is important to identify druggable targets and/or predict adverse effects of treatments [8].

**Results**

**Single-variant association, conditional, and replication analysis**

Meta-analysis was performed for ~45,000,000 SNPs in the discovery sample ( $n=10,726$  for WB-LM and 10,672 for A-LM), as well as for sex-specific subgroups (Table 1). The genomic inflation factors ranged from 0.9953 to 1.0023, indicating sufficient control of population stratification and relatedness (Additional file 2: Fig. S1). In total, we identified seven genomic loci that were genome-wide significantly associated with LM ( $p < 5 \times 10^{-8}$ ), five (*ENSG00000233359*, *LINC01661/PRMT6*, *EMP2*, *ZCCHC14-DT*, *PA2G4P2/LINC01722*) for WB-LM and two (*SSUH2*, *RCC2P8/COL25A1*) for A-LM (Table 2, Additional file 2: Figs. S2 and S3). Two of these loci (*LINC01661/PRMT6* and *RCC2P8/COL25A1*) were sex-specific signals, *LINC01661/PRMT6* was identified from WB-LM female analysis, and *RCC2P8/COL25A1* was identified from A-LM male analysis. Among these seven identified loci, there were 12 genome-wide significant variants for WB-LM and 3 genome-wide significant variants for A-LM (Additional file 3: Table S1). All of our identified genome-wide significant loci were conditionally distinct based on our stepwise conditional analysis (Additional file 3: Tables S2–S7). Only one of our identified loci, *SSUH2*, had a previously reported variant (rs6763944, associated with A-LM in a GWAS study [1]) nearby within a  $\pm 0.5$  Mb range. After conditioning on the previously known signal, *SSUH2* remained genome-wide significant (Additional file 3: Table S8). Therefore, all of our identified genome-wide significant loci were novel. For the UKBB replication samples, one lead variant (rs140266099) from our identified locus, *RCC2P8/COL25A*, reached nominal significance (i.e.,  $p < 0.05$ ) using the imputed SNP-Chip data (Table 2). After meta-analysis with UKBB, all variants (rs116652927, rs79764157, rs77796060, rs182466396,

**Table 1** Study characteristics

Study	Discovery						Replication	
	Amish	CHS	FHS	SAFOS	WHI	LOS	UKBB (WGS)	UKBB (SNP-Chip)
N	484 (478)	1054	2863	409 (361)	934	4982	1115	4720
Female (%)	0.523	0.59	0.592	0.619	1	0.502	0.529	0.522
Age <sup>#</sup>	55.2 (14.9)	76.4 (4.7)	56.7 (13.0)	41.1 (15.7)	66.4 (6.8)	39.2 (11.2)	55.7 (7.6)	55.73 (7.6)
Height <sup>#</sup>	165.0 (8.9)	164.3 (9.4)	166.9 (9.6)	162.1 (9.1)	161.3 (6.0)	169.8 (9.2)	169.2 (9.2)	169.4 (9.2)
Weight <sup>#</sup>	75.1 (14.1)	72.9 (13.9)	75.6 (16.2)	78.8 (18.5)	74.1 (15.6)	79.8 (20.1)	77.1 (15.7)	77.3 (15.2)
Total fat <sup>#</sup>	21.8 (9.5)	26.9 (9.8)	26.7 (9.9)	30.9 (12.7)	32.8 (11.4)	24.8 (12.2)	25.9 (9.8)	26.0 (9.4)
WB-LM <sup>#</sup>	50.8 (10.5)	44.4 (10.1)	45.5 (11.2)	47.8 (11.8)	38.0 (5.6)	55.0 (12.3)	47.2 (10.0)	47.3 (9.8)
A-LM <sup>#</sup>	22.4 (5.7)	18.1 (5.0)	19.9 (5.7)	19.2 (6.0)	14.9 (2.9)	24.7 (6.3)	21.3 (5.3)	(5.2)

N: sample size of WB-LM (sample size of A-LM)

<sup>#</sup>: mean (standard deviation); results are based on the larger sample size which is WB-LM

Age is in years, height is in cm, and weight, total fat, WB-LM, and A-LM are in kg

Amish: Amish Cohort; CHS: Cardiovascular Health Study; FHS: Framingham Heart Study; SAFOS: San Antonio Family Osteoporosis Study; WHI: Women's Health Initiative; LOS: Louisiana Osteoporosis Study; UKBB (WGS): UK Biobank for WGS samples; UKBB (SNP-Chip): UK Biobank for imputed genotype samples

**Table 2** Single-variant association results for WB-LM and A-LM

Chr:pos (hg38)	rsid	Gene name	Region ensembl	Minor/ major allele	MAF % (MAC)	Sample size <sup>D</sup>	Beta	SE	p value <sup>D</sup>	Test cohort	p value <sup>R</sup>	Sample size <sup>D+R</sup>	p value <sup>D+R</sup>
WB-LM	1:102,327,011	rs116652927	ENSG00000233359	Intronic	C/T	0.36 (77)	10,723	0.78	0.14	1.66E−08	TOPMed + LOS pooled	15,443	3.75E−08
	1:106,988,891	rs79764157	LINC01661, PRMT6	Intergenic	C/T	0.41 (121)	6256	−0.58	0.11	4.92E−08	TOPMed + LOS female	6846	4.45E−08
	16:10,532,981	rs377263891	EMP2	Exonic	A/T	3.55 (354)	4982	−0.37	0.06	7.95E−11	LOS pooled	NA	NA
	16:87,504,211	rs148735123	ZCCHC14-DT	Intronic	T/G	0.27 (58)	10,726	−0.99	0.16	1.47E−09	TOPMed + LOS pooled	NA	NA
A-LM	20:12,416,504	rs77796060	PA2G4P2, LINC01722	Intergenic	G/T	0.71 (152)	10,721	−0.59	0.1	1.20E−08	TOPMed + LOS pooled	15,441	7.63E−09
	3:8,724,973	rs182466396	SSUH2	Intronic	G/C	1.54 (329)	10,672	−0.39	0.07	2.91E−08	TOPMed + LOS pooled	11,787	2.44E−08
	4:108,798,890	rs140266099	RCC2P8, COL25A1	Intergenic	T/A	0.93 (46)	2479	−0.86	0.15	4.82E−09	LOS male	4733	1.19E−09

Beta: the estimated effect size of the effect allele on the trait; SE: the standard error of the beta estimate; p value: statistical significance of the meta-analysis for the test cohorts; D: discovery cohort; R: replication (UKBB) cohort; D + R: meta-analysis with discovery and replication cohort; \$: UKBB WGS samples; %: UKBB imputed samples; NA: indicates that no data was available for analysis; MAF: minor allele frequency, MAC: minor allele count

and rs140266099) remained genome-wide significant, excluding two variants that were not available in the replication samples (Table 2).

#### Rare variants aggregated association and replication analysis

We applied five strategies for grouping QC-passed rare variants with minor allele frequency (MAF)  $\leq 1\%$  for both lean mass phenotypes to assess the cumulative effect of rare variants within a specific region. A total of three loci were significantly associated with LM ( $p < \text{Bonferroni corrected } p \text{ value}$ ) (Table 3). All of them were identified through sex-stratified analysis. One protein coding gene *DMAC1* was identified from both WB-LM and A-LM female samples. The other lncRNA gene *ENSG00000273183* was selected based on A-LM male samples. The two identified genes were also highly expressed in skeletal muscle tissue based on GTEx portal data [9]. However, despite this observation, they were not replicated in the UKBB samples.

#### Functional annotation using multi-omics data with bioinformatics approaches

Sixty variants were found with LD greater than 0.7 as outlined in Additional file 3: Table S6. Among them, five variants were located in regions relevant to enhancers specific to hMSC, myotubes, and myoblasts cell types, including rs138235889, rs138353434, rs181902470, rs1169376222, and rs187757389 (Additional file 3: Table S9). Through cell-type specific Hi-C data, enhancer-promoter interactions among these five variants were explored, which led to the identification of five genes likely regulated by the five variants, namely *FBXO31*, *MAP1LC3B*, *ZCCHC14*, *ETNPPL*, and *C16orf95* (Additional file 3: Table S10).

For the Gene Ontology (GO) and pathway enrichment analysis, four genes, namely *EMP2*, *PRMT6*, *COL25A1*, and *SSUH2*, were mapped to a total of 86 GO terms. Out of these, 29 GO terms and two pathways showed a statistically significant association with three of the four genes, *EMP2*, *PRMT6*, and *COL25A1*, with an adjusted  $p$  value  $\leq 0.05$  (Additional file 2: Fig. S4, Additional file 3: Tables S11 and S12). For the transcription factor binding sites (TFBS) annotation, we have observed that rs183684601, which is in high LD with rs182466396 (annotated as near gene *SSUH2*), is overlapping with open chromatin histone protein (H3) lysine (K) 27 acetylation (H3K27ac), enhancer chromatin H3K4-monomethylation (H3K4me1), and CCCTC-binding factor (CTCF) binding in muscle-relevant cells [10] among the seven variants related to LM (Additional file 2: Fig. S4, Additional file 3: Tables S11 and S12). Moreover, TRANSFAC identified two predicted allele-specific TFBSs for SNP rs183684601, where Kruppel-like factor 6 (KLF6) is predicted to bind at its reference allele, G, and helicase-like transcription factor (HLTF) is predicted to bind at its alternative allele, A [11]. Also, another SNP rs113293310, which is a proxy for rs182466396, is also annotated with histone modification marks and chromatin states in muscle-relevant cells, but with no predicted TFBS for either allele.

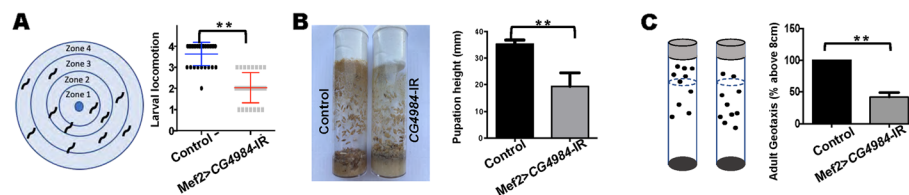
#### Functional validation through animal models

We prioritized two genes based on the following reasoning: one, *EMP2* (modeled in *Drosophila*), was selected because the sentinel SNP was a missense coding variant; the

**Table 3** Rare variants association results

Test gene name	Ensemble ID	Chr:Start–End (hg38)	Gene type	Trait	Sample size <sup>D</sup>	<i>p</i> value <sup>D</sup>	Adjusted <i>p</i> value <sup>D</sup>	Count of rare variants (TOPMed/LOS/UKBB)	Sample size <sup>R</sup>	<i>p</i> value <sup>R</sup>
DMAC1 <sup>#</sup>	ENSG00000137038	Chr9:7,796,500–7,888,380	Protein coding	WB-LM	6260 <sup>F</sup>	5.33E–07	1.20E–02	1/12/1 (TOPMed freeze8 lofmissense variants)	590	0.9478
				A-LM	6229 <sup>F</sup>	1.42E–06	3.20E–02	1/12/1 (TOPMed freeze8 lofmissense variants)	590	0.5781
ENSG00000273183 <sup>§</sup>	ENSG00000273183	Chr7:154,956,429–154,957,107	Antisense	A-LM	4443 <sup>M</sup>	6.49E–07	2.20E–02	1/24/6 (TOPMed freeze8 coding and noncoding variants)	525	0.2590

*p* value: statistical significance of the rare variants analyses for the test cohorts; adjusted *p* values: calculated using Bonferroni correction of the *p* values; count of rare variants (TOPMed/LOS/UKBB): number of functional rare variants per cohort, with functions annotated at the end of each entry; D: discovery cohort; R: replication cohort; F: test cohort includes only females from TOPMed + LOS. M: test cohort includes only males from TOPMed + LOS. #: predicted potential function as a distal membrane arm assembly component 1 [source: HUGO Gene Nomenclature Committee (HGNC) symbol; Acc: HGNC:30536]. §: predicted potential function as an antisense to PAXIP1



**Fig. 1** Silencing of *CG4984* in *Drosophila* muscles led to functional decline and myofibrillar defect. **A–C** *Mef2*>*CG4984*-IR (*Mef2*-Gal4:UAS-*CG4984*-RNAi) flies compared to control (*Mef2*-Gal4) flies. *Mef2*, myocyte enhancer factor 2. **A** Illustration of the concentric circles on the Petri dish that form the zones to determine locomotion. Graph displays the locomotion data based on the number of larvae that crossed each zone within a 1-min time interval. Each data point represents one larva. For each genotype, data are shown for three experiments of 10 larvae each ( $n = 30$  larvae in total). Error bars correspond to SD. \*\* denotes significance of  $p < 0.01$ . **B** The vials show the difference in pupation height between the control and *CG4984*-RNAi flies. Quantitation in the bar graph shows the average pupation height in mm. Error bars correspond to SD. \*\* $p < 0.01$ . **C** Adult mobility measured by negative geotaxis. Three groups of ten female adults from each genotype were tested. The flies were tapped to the bottom. After 10 s, the number of flies above the 8 cm line was recorded. Averaged data are represented by percentages where the number of flies above the 8 cm mark is divided by the total number of flies tested within each group. Error bars correspond to SD. \*\* $p < 0.01$

other, *SSUH2* (modeled in zebrafish), was a causal gene for myopathy, distal, Tateyama type, and rippling muscle disease 2. While the conservation score was low (2 on DIOPT v8.0, sequence similarity = 43%), the protein encoded by *Drosophila* *CG4984* is considered to be the major ortholog of human EMP2 by containing the claudin (=PMPP2) domain of EMP2, supported by two protein databases (Panther [12] and PhylomeDB [13]).

### **Drosophila model**

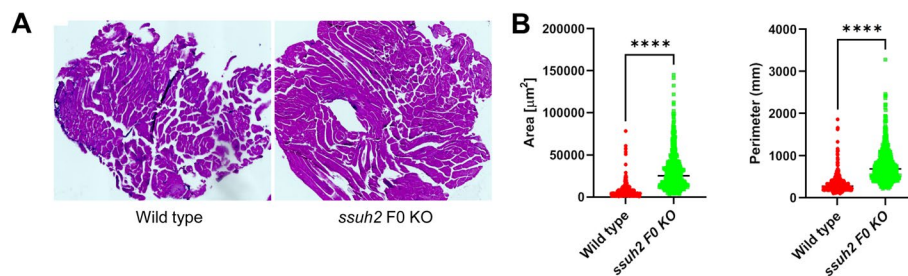
#### ***Silencing CG4984, homolog of EMP2, in Drosophila muscle reduced mobility at the larval-pupal-adult stages.***

Mobility is an important feature of both larval and adult flies. Typical *Drosophila* larvae move away from each other, known as horizontal movement. However, this movement was reduced in the *CG4984*-silenced flies (Fig. 1A). At the end of the 3rd instar, typical larvae will climb up the vial wall and spread out to become pupae. However, the *CG4984*-silenced flies could not climb to the higher vial wall, thus all the pupae located at the lower region. Pupation height as a reflection of the climbing muscle function; therefore, quantification of the height between each pupa and the food surface level in the vial showed that the average pupation height was significantly reduced (Fig. 1B). In adult flies, flying ability can be tested by a geotaxis assay, which revealed a significant reduction in flight ability in the *CG4984*-silenced flies compared to control flies (Fig. 1C). This indicated that the flight muscle function was dramatically restricted by silencing *CG4984*.

#### ***Silencing CG4984, homolog of EMP2, in Drosophila did not affect muscle volume or morphology, nor lipid mass***

Next, we looked at muscle morphology. Given the association of *EMP2* with lean muscle mass, we then used flies with muscle-specific deficiency for *CG4984* (*Mef2*-Gal4:UAS-*CG4983*-RNAi) to study its effect on muscle morphology. The flies showed no significant





**Fig. 2** Multi-locus targeted *ssuh2* gene mutation causes excessive skeletal muscle fibers growth in zebrafish. **A** Histological images of skeletal muscle of adult zebrafish obtained by hematoxylin and eosin staining. Wild type (WT) and *ssuh2* F0 KO genotypes (left and right panels), respectively. **B** Muscle fiber area and perimeter measured in WT and *ssuh2* F0 with  $n=3$  animals per genotype at 2.5 months post fertilization. Each dot on the graph corresponds to the muscle fiber area (491 vs 502 fibers per genotype) and perimeter (493 vs 501 fibers per genotype). Statistical analysis performed using Mann–Whitney test, \*\*\*\* $p=0.0001$ . Scale bar: 20  $\mu\text{M}$

morphological defects in muscle of the second segment when comparing with controls (*Mef2*-Gal4) and *CG4984*-deficient (*Mef2*>*CG4984*-RNAi) larvae, including (1) the organization of muscle bundles (Additional file 2: Fig. S5A); (2) the sarcomere structure of body wall muscle (Additional file 2: Fig. S5B); and (3) the volume of the muscle bundle (Additional file 2: Fig. S5C). To study a potential effect of *CG4984* deficiency on lipid mass, we looked at the fat body, which is the major tissue of lipid and energy storage in *Drosophila*. At the 3rd instar larval stage, the fat body is a single layer of cells, which facilitates the detection of changes in lipid storage. Using Nile red staining of neutral lipid droplet within the fat body, we found no significant difference in either fat body mass or in lipid droplets relative to the area between control (*Lsp2*-Gal4/+ ) and *CG4984*-RNAi (*Lsp2*>*CG4984*-RNAi) flies (Additional file 2: Fig. S5D–F).

These findings indicate no obvious role for EMP2-homolog *CG4984* on muscle morphology or fat body in flies. In summary, these findings demonstrate that while muscle volume did not change in the *CG4984*-RNAi flies, deficiency for EMP2-homolog *CG4984* did cause significant defects in muscle function evident in reduced mobility, from larval to adult.

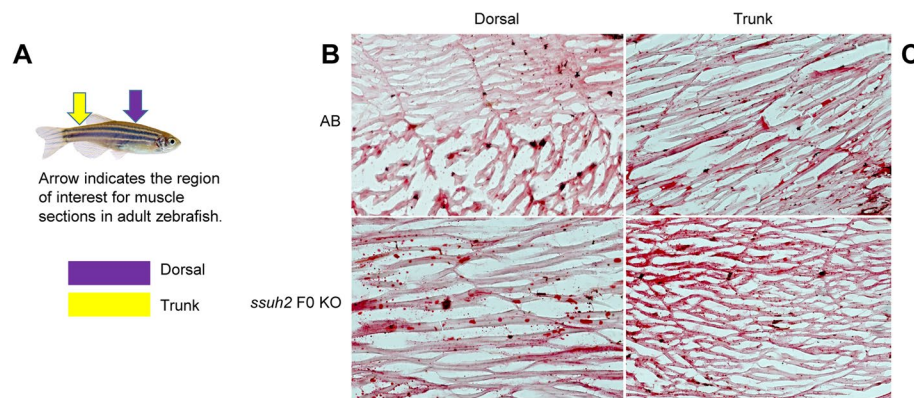
### Zebrafish model

#### **Multi-locus targeted *SSUH2* gene mutation causes excessive skeletal muscle in F0**

We sought to decipher the effect of a multi-locus targeted KO of the *SSUH2* gene in skeletal muscle morphology of zebrafish (F0 generation). The CRISPR/Cas9-based multi-locus targeted *ssuh2* F0 mutant was established based on a high-efficiency method proposed for large genetic screens in zebrafish [14].

To assess the skeletal muscle morphology, we performed hematoxylin and eosin staining. We discovered that *ssuh2* F0 mutant zebrafish demonstrated larger muscle fiber area and perimeter (both  $p<0.00001$ ) in contrast to the wild type controls at the age of 2.5 months post fertilization (Fig. 2A–B, Additional file 2: Fig. S6). Our results indicate that multi-locus targeted KO of the *SSUH2* gene could activate the growth of muscle fibers. We also tested muscle for excess fat by visualization of neutral lipids using Oil Red O staining (Fig. 3A–B, Additional file 2: Fig. S7A–B). The total amount of neutral lipids was apparently not different at dorsal and trunk muscle regions in *ssuh2* F0 mutant in contrast to WT fish.





**Fig. 3** CRISPR/Cas9-based multi-locus targeted *ssuh2* gene in zebrafish revealed no effect in the total amount of neutral lipids in skeletal muscle. **A** Schematic representation of dorsal and trunk muscle fibers' source in adult zebrafish for lipid staining. Region of interest of dorsal and trunk muscle fibers was marked as indicated in a color code. Visualization of lipid content in dorsal muscle (**B**) and trunk muscles (**C**) of zebrafish by Oil Red O (ORO) staining, wild type (top) and *ssuh2* F0 KO (bottom), respectively. The intensity of red color marks the amount of lipids in skeletal muscle. Histological sections of skeletal muscle representative of  $n = 3$  animals per group at 2.5 months post fertilization. Scale bar: 20  $\mu$ m

In regard to the effect of the *ssuh2* gene knockout in zebrafish mobility, we investigated swimming activity in the WT and crispants, with and without acid exposure (stress), using the distance moved (cm) as the dependent variable in a cumulative link mixed effects model approach. Acute change in pH (such as acidity) is among other mild environmental stressors applied to make fish rapidly change their swimming pattern by trying to escape the stressor [15]. Observational time was utilized as a time bin in our cumulative link mixed effects model to account for repeated measurements and temporal variations in swimming activity. Three different models were employed to account for various factors and to assess the robustness of our findings (Additional file 2: Fig. S8): The reference model (model 1) was employed to assess the impact of gene knockout on mobility, accounting for acid exposure and observational time. Models 2 and 3, excluding acid stress, were utilized to investigate the effect of gene knockout across different observational periods. The results indicate that the *ssuh2* gene knockout did not have a significant impact on mobility across all models (model 1:  $p = 0.45$ ; model 2:  $p = 0.26$ ; model 3:  $p = 0.25$ ). In the reference model (model 1), the effect of acid exposure was found to be significant and negatively associated with mobility (estimate =  $-0.21$ ,  $p = 0.001$ ), while the observational time also showed a highly significant negative effect (estimate =  $-0.24$ ,  $p < 2 \times 10^{-16}$ ). When excluding the effect of acid and considering selected fish (model 2) or selected time bins (model 3), the *ssuh2* gene knockout still had no significant effect on mobility. Overall, our findings suggest that the *ssuh2* gene knockout does not have a significant positive impact on zebrafish mobility, while acid exposure and observational time exhibit significant negative associations with effect sizes of  $-0.21$  and  $-0.24$ , respectively.

## Discussion

In this WGS study, we have identified seven distinct genomic loci (ENSG00000233359, LINC01661/PRMT6, EMP2, ZCCHC14-DT, PA2G4P2/LINC01722, SSUH2, and RCC2P8/COL25A1) that were significantly associated with LM. After meta-analysis with the UK Biobank replication samples, all available variants remained genome-wide significant. Although rare variants aggregated analysis is primarily challenged by sample size limitations [16], we were able to discover two genes (DMAC1 and ENSG00000273183) through our grouping strategies. These two identified genes were also highly expressed in the skeletal muscle tissue according to public databases. In addition, the identified genes were further investigated through our functional follow-up using bioinformatic approaches and animal models. Such genetic knowledge is important to identify druggable targets for sarcopenia, since on one hand, targets supported by genetic associations for the drug's lead indication are  $2\text{--}3 \times$  more likely to pass through clinical development than the target of a drug without this genetic backing [8, 17]. On the other, early discontinuation of clinical trials is attributed to the absence of genetic/omic evidence for a drug target in question [18].

Functional annotation for novel discoveries suggested potential roles for the newly identified genes. We have observed that KLF6, a transcriptional activator, plays a role in myoblast/muscle function in the TFBS annotation. Studies have shown that KLF6 and MEF2D co-localize in the nuclei of myogenic cells and that the MEF2 cis element is an important component of the *KLF6* promoter region. TGF $\beta$  has been found to enhance KLF6 protein levels in myoblasts, and inhibition of Smad3 represses this effect. Depletion of KLF6 has been shown to enhance myogenic differentiation and reduce myoblast proliferation in response to TGF $\beta$ . The findings have important implications for understanding muscle development and various muscle pathologies [19]. HLTF (helicase-like transcription factor) is a DNA helicase and a member of the SNF2 family of chromatin-remodeling proteins. While there is currently limited research on the specific role of HLTF in muscle, it has been found to play a role in DNA damage response and repair [20], which may be relevant in the context of muscle regeneration and degeneration.

COL25A1 (collagen type XXV alpha 1 chain) is involved in congenital fibrosis of the extraocular muscles and in arthrogryposis multiplex congenita; recently, a single-cell transcriptomic atlas of human skeletal muscle aging found its expression to be associated with myofiber typing [21]. *PRMT6* (protein arginine methyltransferase 6) encodes a protein that methylates arginine residues in proteins, resulting in specific epigenetic tags for transcriptional repression. As shown in Additional file 2: Fig. S4, *PRMT6* is involved in various methyltransferase activities and has a profound effect on gene regulation through DNA methylation. A recent study has highlighted the importance of *PRMT6* in regulating muscle phenotypes in the context of spinobulbar muscular atrophy (SBMA). Specifically, *PRMT6* is overexpressed in an androgen-dependent manner in the skeletal muscle of patients and mice with SBMA [22].

EMP2 (epithelial membrane protein 2) is an encoded protein that regulates cell membrane composition and is involved in various functions such as endocytosis, cell signaling, and cell proliferation [23]. As shown in Additional file 2: Fig. S4, EMP2 is associated with 19 biological processes and one pathway. It is also worth noting that both EMP2 and COL25A1 are significantly involved in the GO pathway referred to as “supramolecular

fiber organization" (GO: 0097435), resembling a structure of actin filament and myosin [24]. *SSUH2* (Ssu-2 homolog (*Caenorhabditis elegans*)) is a gene associated with distal myopathy, Tateyama type, as well as rippling muscle disease 2, which is a form of limb-girdle muscular dystrophy. It is also involved in the dentin dysplasia type I, thus is suspected to play a role in odontogenesis. The integration of these top genes, known or suspected to have a function in muscle mass or function, reveals that there is currently no direct evidence linking them to muscle mass phenotypes. Particularly, two of these genes, *EMP2* and *SSUH2*, represent novel signals and are associated with the LM in sex-combined samples. Therefore, they are believed to play a role in the general population and have been prioritized for functional follow-up in animal models.

The PMP22\_Claudin domain makes up nearly the entire epithelial membrane protein 2, *EMP2* (167 amino acids) and is conserved from flies to humans. The domain regulates many processes, including the formation of tight junctions, cell–cell adhesion, and cellular contraction [25, 26]. Notably, the PMP22\_Claudin domain only makes up about a third of the fly homolog (CG4984, 447 amino acids), raising the possibility that CG4984 conveys additional functions. This notion is further supported by its homology to two calcium channels in humans—CACNG5 and CACNG7. Furthermore, our *Drosophila* assays showed significantly decreased mobility during larval development and in adult flies deficient for CG4984. Since we could not detect morphological differences in either the muscle tissue or the fat body, the mobility defect could be of neuronal origin. Further research into this gene is warranted, to determine any effects of CG4984/*EMP2* deficiency on neurons or the neuromuscular junctions.

In parallel to the *EMP2* mutagenesis, using crispant (F0 generation mutant) zebrafish, we sought to decipher the effect of targeted *SSUH2* gene knock-out on skeletal muscle morphology and function. Thus, we discovered that *ssuh2* F0 mutant zebrafish demonstrated larger muscle fiber area and perimeter (both  $p < 0.00001$ ) in contrast to the wild type controls. Our results thus indicate that *SSUH2* gene mutation could activate the growth of muscle fibers. We also tested muscle for excess fat by visualization of neutral lipids in dorsal and trunk muscle regions, but found no difference in this parameter, suggesting the larger fibers are not due to excess fat. To evaluate the functional consequence of gene knockout, we investigated swimming performance in the WT and crispant fish, in normal conditions and under stress (acid exposure). The *ssuh2* gene knockout had no significant effect on mobility, either speed or distance moved. Overall, our findings suggest that the *ssuh2* gene knockout affects muscle fiber bulk (a phenotype similar to human A-LM the gene was associated with) but does not have an impact on zebrafish mobility. Further research is necessary to elucidate the underlying biological mechanisms by which *SSUH2* regulates skeletal muscle morphology, thereby enhancing our understanding of its causal and functional role in muscle development.

Our comparison of results with previous GWAS signals reveals significant challenges in replication. Previous GWAS studies predominantly relied on imputed genotypes and focused only on individuals of European ancestry. This lack of replication may be attributed to variations in study designs (Additional file 3: Table S13) and discrepancies in statistical models and sample sizes. Additionally, many of the variants we identified are relatively rare, and our replication cohort, the UK Biobank, employed WGS data for both WB-LM and A-LM, which may account for the modest sample size and limit our ability

to replicate results for gene-centered aggregated rarer variants. While we observed several previously identified variants near our significant loci with  $p$  values  $< 10^{-6}$  (Additional file 3: Table S8), only one variant (rs6763944) is in proximity to our genome-wide significant variants ( $p$  value  $< 5 \times 10^{-8}$ ) [1]. Moreover, it is worth noting that most GWAS employed BIA to assess lean mass. In contrast, our study utilized DXA machines, which may contribute to the observed differences in results.

Our analyses have identified five lncRNAs associated with lean mass: ENSG00000233359, LINC01661, ZCCHC14-DT, LINC01722, and ENSG00000273183. Although these lncRNAs are relatively unexplored and their functional roles are not well understood, the lack of high genomic conservation in animal models limits our ability to validate these findings. Despite this challenge, the identification of these lncRNAs in our study represents a significant discovery. Future research involving loss-of-function and gain-of-function experiments, along with transcriptomic and proteomic analyses, will be crucial for elucidating their precise biological functions.

This study has multiple strengths. First, in contrast with most GWAS published to date, which were conducted in populations of European-descent and have limited coverage of variants uncommon in European ancestry populations, our WGS samples had greater diversity in sample ethnicity, therefore being generalizable to a diverse population of the USA. Second, through our analysis we were able to identify sex-specific signals associated with LM. The chromatin states are not the same in male and female skeletal muscles, as recent methylome and transcriptome integration analysis [27] revealed that skeletal muscle omics manifest profound sex differences. Thus, sex-specificity in both muscle gross phenotype and muscle physiology is well appreciated.

Several aspects of our study can be improved. First, in this study we only analyzed muscle mass (WB-LM and A-LM). But muscle strength and physical performance are important components for understanding of musculoskeletal diseases, such as sarcopenia. The diagnosis of sarcopenia, defined as age-related loss of skeletal muscle, is based on an assessment of skeletal muscle mass together with low muscle strength and low physical performance. Further association study on those phenotypes or even bivariate analysis is worthwhile. Second, our replication cohort (UK Biobank) has a relatively modest sample size and is primarily composed of individuals of European ancestry, especially for the WGS data. This may explain why the rare variant results were not replicated. A larger multi-ethnicity cohort for replication could be needed. Third, although a large sample size in our studies helps mitigate the impact of unmeasured confounders by randomly distributing them across genotypes, we acknowledge a limited ability to adjust for some important confounders, such as physical activity, diet, or sex steroid hormonal concentrations, which have a significant influence on lean mass. Within the range of participating studies, some either lack these measurements or use different methods for collecting the data. This need for homogenizing data on these potential covariates requires follow-up studies to arrange such analyses. Last, due to limited resources, we only validate two identified genes through our animal models. More comprehensive validation studies can be done in the future.

## Conclusions

Through deep WGS data, bioinformatic tools, and functional follow-up in animal models, this study provides new insights into pathways influencing muscle metabolism and muscle mass regulation and informs future studies dedicated to this important metric of the organism.

## Methods

### Discovery cohorts: TOPMed and LOS

This study included a total of 10,726 participants with WB-LM ( $n=10,726$ ) or A-LM ( $n=10,672$ ) measurements who were from the Trans-Omics for Precision Medicine (TOPMed) Consortium and the Louisiana Osteoporosis Study (LOS). The participating TOPMed cohorts included the Genetics of Cardiometabolic Health in the Amish [28] (Amish,  $n=487$  for WB-LM [478 for A-LM]), Cardiovascular Health Study [29] (CHS,  $n=1054$  [1054]), Framingham Heart Study [30, 31] (FHS,  $n=2863$  [2863]), San Antonio Family Osteoporosis Study [32, 33] (SAFOS,  $n=409$  [361]), and Women's Health Initiative [34] (WHI,  $n=934$  [934]) participants. The LOS included 4982 participants, whose ethnic composition is ~72% White, ~23% Black, and ~4% Hispanic/Latino (Additional file 3: Table S14) [35]. Basic characteristics of each study, including sample size, sex composition, and mean (standard deviation, SD), of age, weight, height, total fat, WB-LM, and A-LM, are shown in Table 1. All participants provided informed consent and the appropriate institutional review boards approved all studies (Additional file 2: Table S15). Please refer to Additional file 1: SI 1, SI 2, and SI 3, for more information about each study.

### Replication cohort: UK Biobank

In total, 4720 participants (4720 genotyped on the SNP-Chip and an overlapping 1115 with available whole genome sequencing (WGS) data) were included with measurements for both WB-LM and A-LM from the UK Biobank (UKBB) [36]. The ethnic composition of these participants is ~97% White, 2.2% Asian, 0.9% Black, and 0.2% Hispanic/Latino. The related basic characteristics of this cohort are also shown in Table 1. UKBB has continuously renewed ethical approval from the North West Multi-center Research Ethics Committee, and all studies were carried out in accordance with the appropriate project's Material Transfer Agreement. Informed consent was obtained from all participants. The UKBB application license number associated with the data and research in this study is 69,804.

### Lean mass phenotype measurements

#### TOPMed

Lean mass was measured in all TOPMed cohorts using dual-energy X-ray absorptiometry (DXA) (Hologic 4500W, Hologic QDR-2000, Lunar/Prodigy) (Additional file 3: Table S15). DXA can be used to measure body composition, including bone mineral, fat, and fat-free soft tissue. For this study, we used fat-free soft tissue (i.e., lean mass) as our phenotype. We included two types of lean mass: WB-LM and A-LM. The latter includes only lean mass in the arms and legs, which has been demonstrated to be a valid estimate



of skeletal muscle mass, especially since the arms and legs do not contain visceral organs [37].

### **LOS**

Both fat mass and lean mass were measured with a DXA machine (Hologic QDR-4500 Discovery DXA scanner, Hologic Inc., Bedford, MA, USA) by trained and certified research staff. The machine was calibrated daily, and software and hardware were kept up-to-date during the data collection process. The two lean mass phenotypes derived using manufacturer's image analysis protocols, including WB-LM and A-LM, as demonstrated elsewhere [6], were used in the following analyses.

### **UKBB**

Lean mass was measured in UKBB participants using the Lunar-GE iDXA dual-energy X-ray absorptiometry device (GE-Lunar, Madison, WI, USA). Scans were analyzed by a radiographer using the iDXA device at or shortly after acquisition, generating numerical measures of body composition split into fat mass and lean mass (fat-free soft tissue). We again included two lean mass phenotypes, WB-LM (UKBB data field number 23280) and A-LM (sum of UKBB field numbers 23275 and 23,258), defined in the same way as described for the TOPMed studies.

## **Sequencing data and quality control**

### **TOPMed**

Whole genome sequencing (WGS) data for the TOPMed program are acquired by multiple sequencing centers over time [38]. The TOPMed Informatics Research Center (IRC) performs joint variant identification and genotype calling on all available samples periodically and the resulting call set is referred to as a genotype "Freeze." In this study, we used TOPMed Freeze 8,  $\sim 30 \times$  WGS data with  $\geq 95\%$  of the genome covered to  $10 \times$  or greater. The reads were aligned to human genome build GRCh38 using a common pipeline across all centers. For the variant quality control (QC), the inferred pedigree of related and duplicated samples was used to calculate the Mendelian consistency and to train a support vector machine (SVM) classifier. Variants with excess heterozygosity or Mendelian discordance are filtered out. The sample QC included concordance between annotated and genetic sex inferred from the WGS data, concordance between prior SNP array genotypes and WGS-derived genotypes, and comparisons of observed and expected relatedness from pedigrees. Discordant samples were either excluded or were resolved through prior genotyping comparisons and/or pedigree checks. And the estimated sample contamination was below 10% in our analysis. Please refer to <https://topmed.nhlbi.nih.gov/topmed-whole-genome-sequencing-methods-freeze-8> and each study's dbGaP accession [39] for more details about TOPMed WGS data and quality control.

### **LOS**

A total of 5002 samples with genomic DNA were collected and underwent WGS using a BGISEQ-500 sequencer (Beijing Genomics Institute (BGI Group), Shenzhen, China)

to generate two sequencing runs of paired-end 350 bp reads with an average sequencing depth of  $\sim 21\times$  and 92.29% of the genome covered to at least  $10\times$  coverage. The aligned and cleaned data of each sample were mapped to the human reference genome (GRCh38/hg38) by the use of the Burrows-Wheeler Aligner (BWA) [40] software following the recommended Best Practices for variant analysis with the Genome Analysis Toolkit (GATK version 3.7, <https://www.broadinstitute.org/gatk/guide/best-practices>) to ensure accurate variant calling [41]. The details of the WGS variant analysis including identification of marker duplicates, recalibration for base quality scores, realignment of indels, and variant calling using variant quality score recalibration are described in Additional file 1: SI 4.

### **UKBB**

For the SNP chip genotyping data, the genotypes of UKBB participants (UKBB field number 22828) were determined for  $\sim 90\%$  of participants via Affymetrix UK Biobank Axiom Array (Santa Clara, CA, USA) and for the other  $\sim 10\%$  of participants using the Affymetrix UK BiLEVE Axiom Array [8]. Genotypes were further imputed with the Haplotype Reference Consortium (HRC) panel [42], to ultimately obtain data on  $\sim 96$  million genotypes mapped to GRCh37, with quality control and imputation details as previously described [43]. LiftOver was used to map the genotypes from GRCh37 to GRCh38 coordinates [44]. WGS of 150,119 UKBB participants was performed by two sequencing centers (deCODE Genetics and Wellcome Trust Sanger Institute) on stored blood samples' DNA, using Illumina NovaSeq machines with an average coverage of  $32.5\times$ , and coverage across all samples ranging from  $23.5\times$  to  $\sim 50\times$  [45]. Sequence reads were mapped to GRCh38 and SNPs were jointly called across all individuals in the dataset with GraphTyper [46]. No further QC filters were applied. More details on WGS of UKBB participants, including sequencing center batch effects, sample concordance QC, and GraphTyper parameters (though data in this study come from an earlier WGS release and were not filtered by AAScore), were previously published [45] or described in Additional file 1: SI 5.

### **Single-variant association analysis**

We performed single-variant association tests using linear mixed models on the two lean mass phenotypes, WB-LM and A-LM separately. A two-stage procedure of association test was conducted. Within each study, we first performed linear regression of lean body mass as a function of age (years), age squared, sex, weight (kg), height (cm), total fat (kg), and study specific covariates (e.g., ethnicity) to generate study-specific residuals. Adjusting for height and total fat in our model ensures that the identified SNPs contribute to lean mass independently of their effects on total fat or height [6, 47]. Second, we performed inverse normal transformation on the generated residuals and fit them as a null model with all PCs for TOPMed samples or significant PCs for LOS samples (Additional file 3: Table S16). The output from this second stage was used to perform genome-wide score tests of genetic association for all QC-passed individual variants. We then meta-analyzed association results from TOPMed and LOS through an inverse variance weighted approach and focused on variants with minor allele frequency



(MAF)  $\geq 0.1\%$  and minor allele count (MAC)  $> 40$ . The genome-wide significance threshold  $p < 5 \times 10^{-8}$  was used as our significant level. We also conducted a sex-stratified analysis with the same procedure for WB-LM and A-LM after excluding the sex covariate in the regression model at the first stage. The same procedure was applied to the UKBB cohort using WGS data or imputed SNP-Chip data (if WGS data not available) on significant variants as a replication. Finally, association results between discovery cohorts and the replication cohort were meta-analyzed through inverse-variance approach. The related software we used is provided in Additional file 1: SI 6.

### Conditional analysis

We performed two sets of conditional analysis to pinpoint a short list of important variants. The first approach was conditional on our own findings to identify a set of distinct signals. A stepwise conditional analysis was performed on our variants with  $p < 1 \times 10^{-6}$ . Taking WB-LM as an example, for each chromosome, we identified the most significant variant as the “peak variant” and then fit a new model adjusted for both the covariates as well as this peak variant and calculated new  $p$  values for the rest of variants on that chromosome. If more than one variant was significant at the  $1 \times 10^{-6}$  level in the new result, we performed a second round of conditional analysis, re-fitting the model, adjusting for the new peak variant along with the first peak variant and the original covariates. We continued this procedure iteratively until no additional variants were significant at a  $p$  value threshold of  $< 1 \times 10^{-6}$ . The second approach was conditional on previously reported signals to determine whether our identified signals were novel. We considered previously known signals within a  $\pm 0.5$  Mb range for each of our signals to perform conditional analysis. Previously identified associated variants were downloaded from the GWAS catalog (version: All studies v1.0.2, [https://www.ebi.ac.uk/gwas/api/search/downloads/studies\\_alternative](https://www.ebi.ac.uk/gwas/api/search/downloads/studies_alternative)) and were matched to our variants using either rsid or genomic locus and alleles. We included five GWAS studies for WB-LM or A-LM phenotypes (Additional file 3: Table S17) [1, 6, 7, 48, 49]. We also performed these two types of conditional analysis on males and females separately.

### Rare variants aggregated association analysis

We applied five strategies for grouping rare variants for both lean mass phenotypes to assess the cumulative effect of rare variants within a specific region. Two of them are based on all genes, one including loss of function and missense variants and the other including coding and non-coding regulatory variants. The other three strategies are specific to muscle-related regions. These include muscle-specific lncRNAs from lncRNAKB (<http://psychiatry.som.jhmi.edu/lncrnakb/tissues/index.php?tissue=Muscle>), differentially methylated regions (DMRs) in human skeletal muscle and muscle cell related transcription start sites (TSSs). More details about how we defined these muscle-related regions are described in Additional file 1: SI 6. We included variants with MAF  $\leq 1\%$  that passed QC to aggregate based on the above strategies. The association model was the same as the single variant analysis except we tested each set of aggregated variants here instead of each variant. We applied the same SKAT test for aggregated variants association analysis to TOPMed and LOS participants. We used the probit method to meta-analyze association  $p$  values from TOPMed and LOS studies. We used a Bonferroni

correction to determine the significance, adjusting for the number of aggregated regions. The same approach was also performed on males and females separately. For the replication, we performed the same procedure using WGS data of UKBB except that we only focused on identified regions from discovery cohorts and checked if those regions were nominal significant (i.e.,  $p < 0.05$ ).

#### Functional annotation using multi-omics data with bioinformatics approaches

We utilized a linkage disequilibrium (LD) value of greater than 0.7 with the lead SNPs as the cutoff to identify potential causal variants from our single-variant association results. To ascertain the cell-type specific regulatory function of these potential causal variants, we employed Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) on human mesenchymal stem cells (hMSC), myotubes and myoblasts ATAC-seq [50] and DNase-seq from ENCODE [51]. Our aim was to identify the open chromatin regions, transcription factor binding sites, and enhancer regions, with the aid of ChromHMM [52], specific to hMSC, myotubes, and myoblasts cell types. Moreover, cell-type specific Hi-C data were applied to discern enhancer-promoter interactions.

To comprehend the functions of LM-related SNPs, we conducted gene enrichment analysis using Enrichr [53] and multiple pathway databases such as WikiPathways [54], KEGG [55], and Reactome [56], as well as the GO database [57]. The threshold for significance in GO term and pathway analysis is set at an adjusted  $p$  value of  $\leq 0.05$  using the Benjamini-Hochberg (BH) method. Additionally, we employed muscle-relevant functional annotation and regulatory information from HaploReg [58] and TRANSFAC [11] databases. To identify potential causal variants, we performed HaploReg to select variants within an LD block based on the 1000 Genome Project. Subsequently, we focused on candidate SNPs with muscle-related annotations in histone modification and chromatin status [10] and searched for predicted transcription factor binding sites (TFBS) using the TRANSFAC database [11]. More details can be found in Additional file 1: SI 7.

#### Functional validation through animal models

We considered two types of animal models, including *Drosophila* and zebrafish, to validate our findings. To consider for functional validation, we prioritized genes in which we identified missense variants and genes more likely to be functionally related to musculoskeletal phenotypes. We focused on non-sex specific genes. Furthermore, we utilized Open Targets Genetics web portal (<https://genetics.opentargets.org/>) to verify the association of our identified variants with the genes we selected for knocking out. Approximately 61% of disease-causing genes in humans have functional homologs in *Drosophila* [59], and the fly can be used to study the function of these genes and the consequences of mutations. Zebrafish also have become an attractive animal model for musculoskeletal-specific genetic modeling [60, 61], given both a conserved genetics of the zebrafish as well as their muscle's similarity to human's [62]. Effective gene editing with CRISPR/Cas9 system has been a major tool for the functional study of genes in zebrafish, making it more rapid, by enabling phenotypic analysis on the first generation (G0, also known as crispants) [14, 63], saving valuable time and resources. We briefly describe the animal models we used below, and more details about these two animal models are provided in Additional file 1: SI 8 and SI 9.

### ***Drosophila* model**

The *Drosophila* model was employed because of its well-established genetic tools and conserved mechanisms of muscle development, providing valuable insights into the regulation of muscle mass that can be extrapolated to human biology [64, 65]. Protein encoded by *Drosophila* CG4984 is orthologous to the PMPP2\_Claudin domain of EMP2. The UAS-CG4984-RNAi stock flies were crossed with the *myocyte enhancer factor 2* (*Mef2*)-Gal4 (muscle-specific) or with the *larval serum protein 2* (*Lsp2*)-Gal4 fly lines for fat body-specific silencing of CG4984. Larvae (3 days old) underwent phalloidin staining; body wall muscles were imaged and the number of myofibers obtained. Neutral lipid quantitation was obtained by Nile red staining; lipid droplet size and number were measured. Pupation height and larval locomotion was recorded, followed by the adult locomotion measured by negative geotaxis, and compared between the genotype groups.

### **Zebrafish model**

We utilized the zebrafish model due to its high genetic homology to humans, with approximately 80% of human disease-related genes conserved, making it an ideal organism for high-throughput genetic knock-out studies to investigate gene functions related to human lean muscle mass [61, 66]. We generated *ssuh2* G0 knockout mutants (crisprants) by multi-locus targeted CRISPR/Cas9 technology (to assure high efficiency of the mutagenesis). Using *ssuh2* crisprant zebrafish in comparison to the wild type (WT) controls, we then measured skeletal muscle fiber area and perimeter in young adult (2.5 months old) zebrafish. We also tested muscle for excess fat by visualization of neutral lipids in dorsal and trunk muscle regions. Furthermore, we investigated the swimming distance and swimming behavior, in normal conditions and under stress (acid exposure). Data were tested for normality using the Shapiro–Wilk test ( $\alpha = 0.05$ ). Normally distributed data were analyzed by Student's *t*-test (two genotype groups). Non-normal distributed data were analyzed by a Mann–Whitney test (two groups). Statistical significance was defined as  $p < 0.05$ . The cumulative link mixed effect model approach was used to compare the effect of *ssuh2* gene knockout on mobility.

### **Supplementary Information**

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-025-03520-x>.

Additional file 1: Supplemental information and methods. Detailed descriptions of additional methods, protocols, and analyses supporting the main study.

Additional file 2: Supplemental figures. Figures S1 to S8 providing additional visual data and supporting information referenced in the main text.

Additional file 3: Supplemental tables. Contains supplemental tables (S1–S19) providing detailed datasets, statistical results, and additional analyses supporting the findings of this study.

Additional file 4: Review history.

### **Acknowledgements**

Molecular data for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the National Heart, Lung, and Blood Institute (NHLBI). Genome sequencing for “NHLBI TOPMed: Genetics of Cardiometabolic Health in the Amish” (phs00956.v1.p1) was performed at Broad Institute Genomics Platform (3R01HL121007-01S1). Genome sequencing for “NHLBI TOPMed: Cardiovascular Health Study” (phs001368.v2.p2) was performed at Baylor College of Medicine Human Genome Sequencing Center (HHSN268201600033I), Broad Institute Genomics Platform (HHSN268201600034I). Genome sequencing for “NHLBI TOPMed: Whole Genome Sequencing and Related Phenotypes in the Framingham Heart Study” (phs000974.v1.p1) was performed at the Broad Institute Genomics Platform (3R01HL092577-06S1, 3U54HG003067-12S2). Genome sequencing for “NHLBI TOPMed: San Antonio Family Heart Study” (phs001215.v4.p2) was performed at Illumina

(3R01HL113323-03S1, R01HL113322). Genome sequencing for “NHLBI TOPMed: Women’s Health Initiative” (phs001237.v2.p1) was performed at Broad Institute Genomics Platform (HHSN268201500014C). Core support including centralized genomic read mapping and genotype calling, along with variant quality metrics and filtering, was provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1; contract HHSN268201800002I). Core support including phenotype harmonization, data management, sample-identity QC, and general program coordination was provided by the TOPMed Data Coordinating Center (R01HL-120393; U01HL-120393; contract HHSN268201800001I). We gratefully acknowledge the studies and participants who provided biological samples and data for TOPMed. We extend our gratitude to Dr. Joyce van de Leemput for her invaluable critical reading of this manuscript.

### Review history

The review history is available as Additional file 4.

### Peer review information

Nora Franceschini and Tim Sands were the primary editors of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

### Authors’ contributions

The author contributions for this manuscript are as follows: X.Z. and K.J.S. contributed to methodology, software, formal analysis, data curation, visualization, and original draft writing and review; B.B. participated in validation and original draft writing; I.E. contributed to software, validation, formal analysis, data curation, and original draft writing; Y.H.H. contributed to conceptualization, methodology, validation, resources, and formal analysis; C.C. contributed to conceptualization, resources, supervision, writing—review and editing, and project administration; R.D. contributed to validation and visualization; Z.H. participated in validation; R.D.J. contributed to conceptualization; H.L. contributed to validation and supervision; Z.L. participated in data curation; B.M. contributed to resources and supervision; C.Q., L.J.Z., and Q.T. participated in resources and data curation; H.S. contributed to resources, project administration, supervision, and funding acquisition; M.J.T. contributed to validation, formal analysis, and writing—review and editing; K.L.W. and H.X. contributed to methodology and data curation; M.Y. contributed to investigation, writing—original draft, and writing—review and editing; X.Z. contributed to methodology, validation, investigation, and resources; M.M. contributed to conceptualization and project administration; D.P.K. contributed to conceptualization, methodology, investigation, resources, writing—review and editing, supervision, and funding acquisition; H.W.D. contributed to conceptualization, methodology, resources, data curation, writing—review and editing, supervision, and funding acquisition; C.T.L. contributed to conceptualization, methodology, resources, writing—review and editing, and supervision; D.K. contributed to conceptualization, validation, original draft writing, writing—review and editing, and supervision. All authors have reviewed the manuscript and consented to its content.

### Funding

This research was made possible with generous partial support from various funding sources, including grants from the National Institutes of Health (NIH) (P30AG028747, R01AR46838, R01AR043351, P20GM109036, R01AR069055, U19AG055373, R01AG061917) and the Framingham Heart Study grant AR041398. Additionally, this research received partial support from the Women’s Health Initiative (WHI) program, which is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services, through contracts 75N92021D00001, 75N92021D00002, 75N92021D00003, 75N92021D00004, and 75N92021D00005. This research was also supported by contracts HHSN268201200036C, HHSN268200800007C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086, and 75N92021D00006 and grants U01HL080295 and U01HL130114 from the National Heart, Lung, and Blood Institute (NHLBI), with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided by R01AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org.

### Data availability

The datasets analyzed from TOPMed Freeze 8 are available in the dbGaP repository [accession numbers: phs00956.v1.p1 for Amish (28), phs001368.v2.p2 for CHS (29), phs000974.v1.p1 for FHS (30, 31), phs001215.v4.p2 for SAFOS (32, 33), and phs001237.v2.p1 for WHI (34)]. Instructions for accessing TOPMed data can be found at <https://www.nhlbiwgs.org/topmed-data-access-scientific-community>. The datasets from LOS that support this study’s findings are available from the principal investigator (H.W.D., [hdeng2@tulane.edu](mailto:hdeng2@tulane.edu)) upon reasonable request. Access will be granted for academic research purposes, subject to IRB approval and the completion of a data use agreement. Additionally, the LOS WGS data is in the process of being deposited in the AgingResearchBiobank (<https://agingresearchbiobank.nia.nih.gov/>), where it will be made available to qualified researchers upon application and approval. The UK Biobank data used in this study are available to researchers upon application and approval through the UK Biobank’s access management system (<https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access>). The analytical scripts and experimental data can be accessed at the Mendeley Data (67) [<https://doi.org/10.17632/mk32tnmwwt.1>]. Summary statistics can be found in the Musculoskeletal Knowledge Portal [<https://msk.hugeamp.org/>].

### Declarations

#### Ethics approval and consent to participate

All participants provided informed consent, and all studies were approved by the appropriate institutional review boards. Detailed ethics approval information, including the specific institutional review boards and approval statuses for the TOPMed and LOS cohorts, is provided in Additional file 1: SI 1 and SI 2, as well as Additional file 3: Table S2.

#### Consent for publication

All participants provided informed consent for publication.

### Competing interests

The authors declare no competing interests. Dr. Ittai Eres is employed by Amgen Inc and confirms that this employment did not influence the design, conduct, or reporting of this study. All other authors have no financial or non-financial interests that could be perceived as influencing the research reported in this manuscript. Dr. Douglas P. Kiel has received grants to his institution from Radius Health, Amgen, and Solarea Bio. He serves on a Data and Safety Monitoring Committee for Agnovos and has served on Scientific Advisory Boards for Radius Health and Solarea Bio. He receives royalties for publication in UpToDate by Wolters Kluwer.

### Author details

<sup>1</sup>Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118, USA. <sup>2</sup>Center for Biomedical Informatics and Genomics, Tulane University, New Orleans, LA 70112, USA. <sup>3</sup>Department of Biochemistry, Larner College of Medicine, University of Vermont, Burlington, VT 05405, USA. <sup>4</sup>Amgen Inc, South San Francisco, CA 94080, USA. <sup>5</sup>Hinda and Arthur Marcus Institute for Aging Research, Hebrew SeniorLife, Boston, MA 02131, USA. <sup>6</sup>David Geffen School of Medicine, University of California, Los Angeles, CA 90024, USA. <sup>7</sup>Azrieli Faculty of Medicine, Bar-Ilan University, 130010 Safed, Israel. <sup>8</sup>Department of Medicine, University of Maryland School of Medicine, Baltimore, MD 21201, USA. <sup>9</sup>Department of Medicine, The Ohio State University, Columbus, OH 43210, USA. <sup>10</sup>Department of Medicine, University of Washington, Seattle, WA 98195, USA.

Received: 28 April 2024 Accepted: 28 February 2025

Published online: 28 April 2025

### References

- Pei YF, Liu YZ, Yang XL, Zhang H, Feng GJ, Wei XT, et al. The genetic architecture of appendicular lean mass characterized by association analysis in the UK Biobank study. *Commun Biol.* 2020;3(1):608.
- Prado CM, Purcell SA, Alish C, Pereira SL, Deutz NE, Heyland DK, et al. Implications of low muscle mass across the continuum of care: a narrative review. *Ann Med.* 2018;50(8):675–93.
- Mareschal J, Achamrah N, Norman K, Genton L. Clinical value of muscle mass assessment in clinical conditions associated with malnutrition. *J Clin Med.* 2019;8(7):1040.
- Lee DH, Keum N, Hu FB, Orav EJ, Rimm EB, Willett WC, et al. Predicted lean body mass, fat mass, and all cause and cause specific mortality in men: prospective US cohort study. *BMJ.* 2018;362: k2575.
- Hsu FC, Lenchik L, Nicklas BJ, Lohman K, Register TC, Mychaleckyj J, et al. Heritability of body composition measured by DXA in the diabetes heart study. *Obes Res.* 2005;13(2):312–9.
- Zillikens MC, Demissie S, Hsu YH, Yerges-Armstrong LM, Chou WC, Stolk L, et al. Large meta-analysis of genome-wide association studies identifies five loci for lean body mass. *Nat Commun.* 2017;8(1):80.
- Hernandez Cordero AI, Gonzales NM, Parker CC, Sokoloff G, Vandenbergh DJ, Cheng R, et al. Genome-wide associations reveal human-mouse genetic convergence and modifiers of myogenesis, CPNE1 and STC2. *Am J Hum Genet.* 2020;106(1):138.
- Minikel EV, Painter JL, Dong CC, Nelson MR. Refining the impact of genetic evidence on clinical success. *Nature.* 2024;629(8012):624–9.
- GTEX Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013;45(6):580–5.
- Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. *Nature.* 2015;518(7539):317–30.
- Wingender E, Chen X, Hehl R, Karas H, Liebich I, Matys V, et al. TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Res.* 2000;28(1):316–9.
- Thomas PD, Ebert D, Muruganujan A, Mushayahama T, Albu LP, Mi H. PANTHER: making genome-scale phylogenetics accessible to all. *Protein Sci.* 2022;31(1):8–22.
- Fuentes D, Molina M, Chorostecki U, Capella-Gutierrez S, Marcet-Houben M, Gabaldon T. PhylomeDB V5: an expanding repository for genome-wide catalogues of annotated gene phylogenies. *Nucleic Acids Res.* 2022;50(D1):D1062–8.
- Kroll F, Powell GT, Ghosh M, Gestri G, Antinucci P, Hearn TJ, et al. A simple and effective F0 knockout method for rapid screening of behaviour and other complex phenotypes. *Elife.* 2021;10:10.
- Lee HB, Schwab TL, Sigafoos AN, Gauerke JL, Krug RG 2nd, Serres MR, et al. Novel zebrafish behavioral assay to identify modifiers of the rapid, nongenomic stress response. *Genes Brain Behav.* 2019;18(2):e12549.
- Young KL, Fisher V, Deng X, Brody JA, Graff M, Lim E, et al. Whole-exome sequence analysis of anthropometric traits illustrates challenges in identifying effects of rare genetic variants. *HGG Adv.* 2023;4(1): 100163.
- Trajanoska K, Bherer C, Taliun D, Zhou S, Richards JB, Mooser V. From target discovery to clinical drug development with human genetics. *Nature.* 2023;620(7975):737–45.
- Razuvayevskaya O, Lopez I, Dunham I, Ochoa D. Genetic factors associated with reasons for clinical trial stoppage. *Nat Genet.* 2024;56(9):1862–7.
- Dionysiou MG, Salma J, Bevzyuk M, Wales S, Zakharyan L, McDermott JC. Kruppel-like factor 6 (KLF6) promotes cell proliferation in skeletal myoblasts in response to TGFbeta/Smad3 signaling. *Skelet Muscle.* 2013;3(1):7.
- van Toorn M, Turkiylmaz Y, Han S, Zhou D, Kim HS, Salas-Armenteros I, et al. Active DNA damage eviction by HLTf stimulates nucleotide excision repair. *Mol Cell.* 2022;82(7):1343–58 e8.
- Kedlian VR, Wang Y, Liu T, Chen X, Bolt L, Shen Z, et al. Human skeletal muscle ageing atlas. 2022.
- Prakasam R, Bonadiman A, Andreotti R, Zuccaro E, Dalfovo D, Marchioretto C, et al. LSD1/PRMT6-targeting gene therapy to attenuate androgen receptor toxic gain-of-function ameliorates spinobulbar muscular atrophy phenotypes in flies and mice. *Nat Commun.* 2023;14(1):603.
- O'Leary NA, Wright MW, Brister JR, Ciufu S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 2016;44(D1):D733–45.

24. Zhang Z, Cheng L, Zhao J, Zhang H, Zhao X, Liu Y, et al. Muscle-mimetic synergistic covalent and supramolecular polymers: phototriggered formation leads to mechanical performance boost. *J Am Chem Soc.* 2021;143(2):902–11.
25. Dong Y, Simske JS. Vertebrate claudin/PMP22/EMP2/MP20 family protein TMEM47 regulates epithelial cell junction maturation and morphogenesis. *Dev Dyn.* 2016;245(6):653–66.
26. Morales SA, Telander DG, Mareninov S, Nagy A, Wadehra M, Braun J, et al. Anti-EMP2 diabody blocks epithelial membrane protein 2 (EMP2) and FAK mediated collagen gel contraction in ARPE-19 cells. *Exp Eye Res.* 2012;102:10–6.
27. Landen S, Jacques M, Hiam D, Alvarez-Romero J, Harvey NR, Haupt LM, et al. Skeletal muscle methylome and transcriptome integration reveals profound sex differences related to muscle function and substrate metabolism. *Clin Epigenetics.* 2021;13(1):202.
28. Streeten EA, Ryan KA, McBride DJ, Pollin TI, Shuldiner AR, Mitchell BD. The relationship between parity and bone mineral density in women characterized by a homogeneous lifestyle and high parity. *J Clin Endocrinol Metab.* 2005;90(8):4536–41.
29. Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA, et al. The cardiovascular health study: design and rationale. *Ann Epidemiol.* 1991;1(3):263–76.
30. Dawber TR, Kannel WB, Lyell LP. An approach to longitudinal studies in a community: the Framingham study. *Ann N Y Acad Sci.* 1963;107:539–56.
31. Dawber TR, Meadors GF, Moore FE Jr. Epidemiological approaches to heart disease: the Framingham study. *Am J Public Health Nations Health.* 1951;41(3):279–81.
32. Mitchell BD, Kammerer CM, Schneider JL, Perez R, Bauer RL. Genetic and environmental determinants of bone mineral density in Mexican Americans: results from the San Antonio Family Osteoporosis Study. *Bone.* 2003;33(5):839–46.
33. Mitchell BD, Kammerer CM, Blangero J, Mahaney MC, Rainwater DL, Dyke B, et al. Genetic and environmental contributions to cardiovascular risk factors in Mexican Americans. The San Antonio Family Heart Study. *Circulation.* 1996;94(9):2159–70.
34. Jackson RD, LaCroix AZ, Cauley JA, McGowan J. The Women's Health Initiative calcium-vitamin D trial: overview and baseline characteristics of participants. *Ann Epidemiol.* 2003;13(9 Suppl):S98–106.
35. He H, Liu Y, Tian Q, Papisian CJ, Hu T, Deng HW. Relationship of sarcopenia and body composition with osteoporosis. *Osteoporos Int.* 2016;27(2):473–82.
36. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature.* 2018;562(7726):203–9.
37. Chen Z, Wang Z, Lohman T, Heymsfield SB, Outwater E, Nicholas JS, et al. Dual-energy X-ray absorptiometry is a valid tool for assessing skeletal muscle mass in older women. *J Nutr.* 2007;137(12):2775–80.
38. Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature.* 2021;590(7845):290–9.
39. Regier AA, Farjoun Y, Larson DE, Krashenina O, Kang HM, Howrigan DP, et al. Functional equivalence of genome sequencing analysis pipelines enables harmonized variant calling across human genetics projects. *Nat Commun.* 2018;9(1):4038.
40. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754–60.
41. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297–303.
42. McCarthy S, Das S, Kretschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet.* 2016;48(10):1279–83.
43. Welsh S, Peakman T, Sheard S, Almond R. Comparison of DNA quantification methodology used in the DNA extraction protocol for the UK Biobank cohort. *BMC Genomics.* 2017;18(1):26.
44. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. *Genome Res.* 2002;12(6):996–1006.
45. Halldorsson BV, Eggertsson HP, Moore KHS, Hauswedell H, Eiriksson O, Ulfarsson MO, et al. The sequences of 150,119 genomes in the UK Biobank. *Nature.* 2022;607(7920):732–40.
46. Eggertsson HP, Jonsson H, Kristmundsdottir S, Hjartarson E, Kehr B, Masson G, et al. GraphTyper enables population-scale genotyping using pangenome graphs. *Nat Genet.* 2017;49(11):1654–60.
47. Karasik D, Zillikens MC, Hsu YH, Aghdassi A, Akesson K, Amin N, et al. Disentangling the genetics of lean mass. *Am J Clin Nutr.* 2019;109(2):276–87.
48. Hubel C, Gaspar HA, Coleman JRI, Finucane H, Purves KL, Hanscombe KB, et al. Genomics of body fat percentage may contribute to sex bias in anorexia nervosa. *Am J Med Genet B Neuropsychiatr Genet.* 2019;180(6):428–38.
49. Tachmazidou I, Suveges D, Min JL, Ritchie GRS, Steinberg J, Walter K, et al. Whole-genome sequencing coupled to imputation discovers genetic signals for anthropometric traits. *Am J Hum Genet.* 2017;100(6):865–84.
50. Tsai MJ, Reppe S, Sato T, Gill R, Wein M, Gautvik K, et al. The musculoskeletal 3D epigenome atlas. *American Society of Human Genetics; Oct 18–22; Virtual. ASHG; 2021.* p. 325. <https://www.ashg.org/wp-content/uploads/2022/01/2021-ASHGMeeting-Abstracts.pdf>.
51. Luo Y, Hitz BC, Gabdank I, Hilton JA, Kagda MS, Lam B, et al. New developments on the Encyclopedia of DNA elements (ENCODE) data portal. *Nucleic Acids Res.* 2020;48(D1):D882–9.
52. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods.* 2012;9(3):215–6.
53. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics.* 2013;14: 128.
54. Kelder T, Pico AR, Hanspers K, van Iersel MP, Evelo C, Conklin BR. Mining biological pathways using WikiPathways web services. *PLoS One.* 2009;4(7): e6447.
55. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28(1):27–30.
56. Gillespie M, Jassal B, Stephan R, Milacic M, Rothfels K, Senff-Ribeiro A, et al. The reactome pathway knowledgebase 2022. *Nucleic Acids Res.* 2022;50(D1):D687–92.



57. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 2000;25(1):25–9.
58. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* 2012;40(Database issue):D930–4.
59. Rubin GM, Yandell MD, Wortman JR, Gabor Miklos GL, Nelson CR, Hariharan IK, et al. Comparative genomics of the eukaryotes. *Science.* 2000;287(5461):2204–15.
60. Kague E, Karasik D. Functional validation of osteoporosis genetic findings using small fish models. *Genes (Basel).* 2022;13(2):279.
61. Daya A, Donaka R, Karasik D. Zebrafish models of sarcopenia. *Dis Model Mech.* 2020;13(3):dmm042689.
62. Talbot JC, Teets EM, Ratnayake D, Duy PQ, Currie PD, Amacher SL. Muscle precursor cell movements in zebrafish are dynamic and require six family genes. *Development.* 2019;146(10):dev171421.
63. Bek JW, Shochat C, De Clercq A, De Saffel H, Boel A, Metz J, et al. Lrp5 mutant and crispant zebrafish faithfully model human osteoporosis, establishing the zebrafish as a platform for CRISPR-based functional screening of osteoporosis candidate genes. *J Bone Miner Res.* 2021;36(9):1749–64.
64. Schnorrer F, Schonbauer C, Langer CC, Dietzl G, Novatchkova M, Schernhuber K, et al. Systematic genetic analysis of muscle morphogenesis and function in *Drosophila*. *Nature.* 2010;464(7286):287–91.
65. Chaturvedi D, Reichert H, Gunage RD, VijayRaghavan K. Identification and functional characterization of muscle satellite cells in *Drosophila*. *Elife.* 2017;6:6.
66. Karuppasamy M, English KG, Henry CA, Manzini MC, Parant JM, Wright MA, et al. Standardization of zebrafish drug testing parameters for muscle diseases. *Dis Model Mech.* 2024;17(1):dmm050339.
67. Zhang X, Su K, Hsu YH, Crandall CJ, Han Z, Jackson RD, et al. Multi-Ancestry Whole Genome Sequencing Analysis of Lean Body Mass, Mendeley Data, V1. 2025. <https://doi.org/10.17632/mk32tnmwwt.1>.
68. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families. The Framingham Offspring Study. *Am J Epidemiol.* 1979;110(3):281–90.
69. Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham Offspring Study. Design and preliminary data. *Prev Med.* 1975;4(4):518–25.
70. Roubenoff R, Baumgartner RN, Harris TB, Dallal GE, Hannan MT, Economos CD, et al. Application of bioelectrical impedance analysis to elderly populations. *J Gerontol A Biol Sci Med Sci.* 1997;52(3):M129–36.
71. Deng HW, Shen H, Xu FH, Deng HY, Conway T, Zhang HT, et al. Tests of linkage and/or association of genes for vitamin D receptor, osteocalcin, and parathyroid hormone with bone mineral density. *J Bone Miner Res.* 2002;17(4):678–86.
72. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am J Epidemiol.* 2017;186(9):1026–34.
73. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011;43(5):491–8.
74. Mbatchou J, Barnard J, Backman J, Marcketta A, Kosmicki JA, Ziyatdinov A, et al. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat Genet.* 2021;53(7):1097–103.
75. Gogarten SM, Sofer T, Chen H, Yu C, Brody JA, Thornton TA, et al. Genetic association testing using the GENESIS R/Bioconductor package. *Bioinformatics.* 2019;35(24):5346–8.
76. Zhan X, Hu Y, Li B, Abecasis GR, Liu DJ. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics.* 2016;32(9):1423–6.
77. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010;26(17):2190–1.
78. Seifuddin F, Singh K, Suresh A, Judy JT, Chen YC, Chaitankar V, et al. lncRNAKB, a knowledgebase of tissue-specific functional annotation and trait association of long noncoding RNA. *Sci Data.* 2020;7(1):326.
79. Voisin S, Jacques M, Landen S, Harvey NR, Haupt LM, Griffiths LR, et al. Meta-analysis of genome-wide DNA methylation and integrative omics of age in human skeletal muscle. *J Cachexia Sarcopenia Muscle.* 2021;12(4):1064–78.
80. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012;489(7414):57–74.
81. Zhou W, Bi W, Zhao Z, Dey KK, Jagadeesh KA, Karczewski KJ, et al. SAIGE-GENE+ improves the efficiency and accuracy of set-based rare variant association tests. *Nat Genet.* 2022;54(10):1466–9.
82. Spletter ML, Barz C, Yeroslaviz A, Zhang X, Lemke SB, Bonnard A, et al. A transcriptomics resource reveals a transcriptional transition during ordered sarcomere morphogenesis in flight muscle. *Elife.* 2018;7:7.
83. Gargano JW, Martin I, Bhandari P, Grotewiel MS. Rapid iterative negative geotaxis (RING): a new method for assessing age-related locomotor decline in *Drosophila*. *Exp Gerontol.* 2005;40(5):386–95.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.