nature genetics

ARTICLES

https://doi.org/10.1038/s41588-022-01165-1

OPEN

Genome-wide association analyses of physical activity and sedentary behavior provide insights into underlying mechanisms and roles in disease prevention

Although physical activity and sedentary behavior are moderately heritable, little is known about the mechanisms that influence these traits. Combining data for up to 703,901 individuals from 51 studies in a multi-ancestry meta-analysis of genome-wide association studies yields 99 loci that associate with self-reported moderate-to-vigorous intensity physical activity during leisure time (MVPA), leisure screen time (LST) and/or sedentary behavior at work. Loci associated with LST are enriched for genes whose expression in skeletal muscle is altered by resistance training. A missense variant in *ACTN3* makes the alpha-actinin-3 filaments more flexible, resulting in lower maximal force in isolated type II_A muscle fibers, and possibly protection from exercise-induced muscle damage. Finally, Mendelian randomization analyses show that beneficial effects of lower LST and higher MVPA on several risk factors and diseases are mediated or confounded by body mass index (BMI). Our results provide insights into physical activity mechanisms and its role in disease prevention.

Low levels of physical activity have a major effect on disease burden and it is estimated that more than 5 million deaths per year might be prevented by ensuring adequate levels[1]. Despite efforts to increase physical activity levels[2], an estimated 28% of the world's population is insufficiently active, and the prevalence of physical inactivity in high-income countries rose from 31.6% in 2001 to 36.8% in 2016 (ref. [3]). Trends of decreasing physical activity levels over time coincide with increases in the time spent sedentary[4], which may pose an independent risk for public health[5,6].

Physical activity and sedentary behavior are affected by public policy and social support, as well as by cultural, environmental and individual factors[7]. Factors like socioeconomic status, built environment and media all influence physical activity at a population level[7]. In parallel, innate biological factors (for example, age, sex hormones, pre-existing medical conditions, epigenetics and genetics) also explain a moderate proportion of the interindividual variability in physical activity and sedentary behavior. Heritability estimates (h^2) range from 31% to 71% in large twin studies[8,9]. Identifying the genetic factors that influence daily physical activity will improve our understanding of this complex behavior, and may (1) facilitate unbiased causal inference; (2) help identify vulnerable subpopulations; and (3) fuel the design of tailored interventions to effectively promote physical activity. A mechanistic understanding of physical activity at a molecular level may even allow its beneficial effects to be attained through pharmacological intervention[10].

Genome-wide association studies (GWAS) have identified thousands of loci associated with cardiometabolic risk factors and diseases[11]. However, similar efforts for physical activity have been sparse and initially had limited success. This likely reflects the comparatively small sample size of these efforts[12], along with heterogeneous assessments of physical activity across studies. More recently, GWAS using data from UK Biobank identified nine loci associated with self-reported moderate and/or vigorous intensity physical activity or sports and exercise participation ($n \approx 377{,}000$ individuals) and eight associated with accelerometry-assessed physical activity and sedentary behavior ($n \approx 91{,}000$)[13,14]. Hence, on the assumption that physical activity is a highly polygenic trait, many common variants influencing physical activity undoubtedly remain to be identified.

Here, we combine data from up to 703,901 individuals (94.0% European, 2.1% African, 0.8% East Asian, 1.3% South Asian ancestries, and 1.9% Hispanic) from 51 studies in a multi-ancestry meta-analysis of GWAS for MVPA, LST, sedentary commuting and sedentary behavior at work. This yields 104 independent association signals in 99 loci, implicating brain and muscle, among others organs. Follow-up analyses improve our understanding of the molecular basis of leisure time physical activity and sedentary behavior, and their role in disease prevention.

Results

Genome-wide analyses yield 99 associated loci. In our primary meta-analysis of European ancestry men and women combined (Supplementary Tables 1, 2), we identify 91 loci that are associated ($P < 5 \times 10^{-9}$) with at least one of four self-reported traits: MVPA (n up to 606,820), LST (n up to 526,725), sedentary commuting (n up to 159,606) and sedentary behavior at work (n up to 372,605) (Supplementary Table 3, Figs. 1 and 2, and Supplementary Fig. 1). The non-European ancestry meta-analyses do not provide new associations themselves and are only used in multi-ancestry meta-analyses. Multi-ancestry and sex-specific meta-analyses yield eight additional loci, resulting in a total of 104 independent association signals in 99 loci (Supplementary Tables 3 and 4). The vast majority of these—89 independent single nucleotide polymorphisms (SNPs) in 88 loci (35 not previously reported[13,15])—are associated with LST, explaining 2.75% of its variance. We also identify 11 loci for MVPA (six not previously reported[13,15,16], four that overlap with LST) and four loci for sedentary behavior at work (all previously reported[13,15]; Supplementary Table 3). No loci are identified for sedentary commuting. To increase statistical power for the discovery of new loci, we perform a multi-trait analysis of GWAS (MTAG) using summary statistics of MVPA and LST. This yields

A full list of authors and their affiliations appears at the end of the paper.

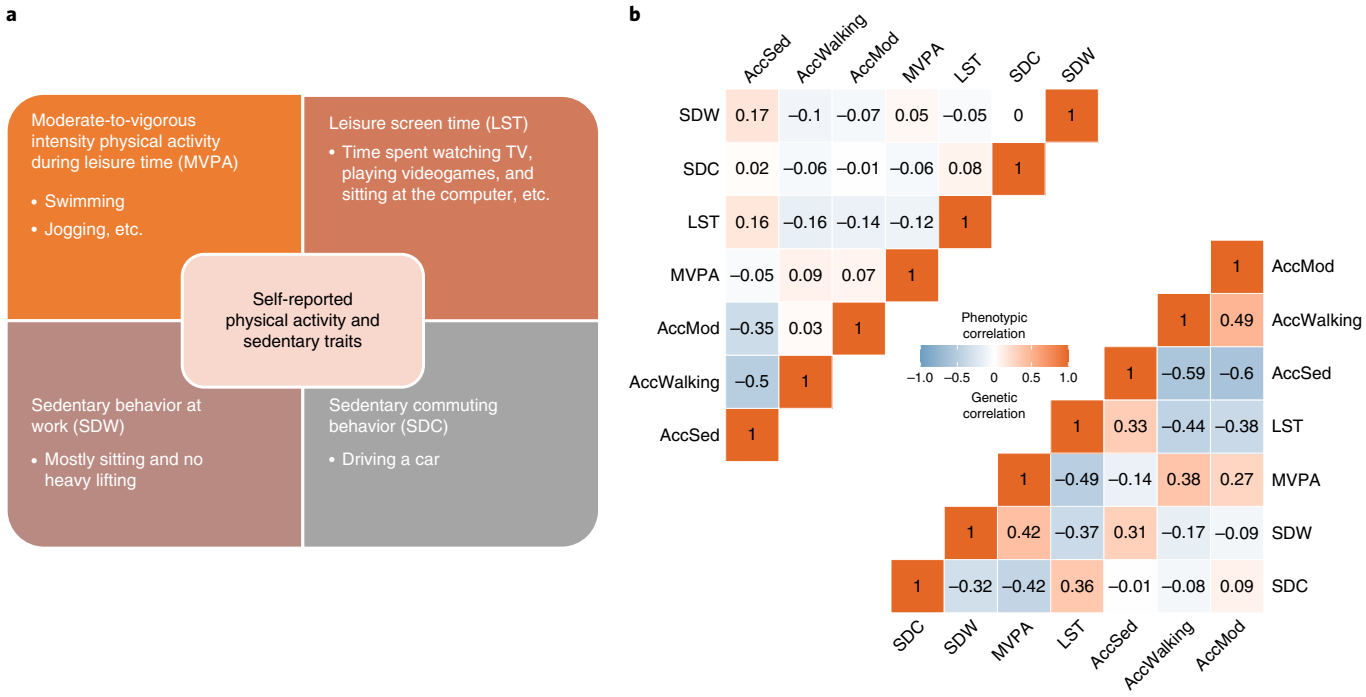


Fig. 1 | Overview of the four self-reported physical activity and sedentary traits and correlations with objectively assessed traits. a, An overview of the four self-reported physical activity and sedentary traits. **b**, Phenotypic (upper left) and genetic (lower right) correlation coefficients between the four self-reported physical activity and sedentary traits studied here and three accelerometer-assessed traits quantified in UK Biobank participants. AccMod, accelerometer-assessed proportion of time spent in moderate intensity physical activity; AccSed, accelerometer-assessed proportion of time spent sedentary; AccWalking, accelerometer-assessed proportion of time spent walking; SDC, sedentary commuting behavior; SDW, sedentary behavior at work.

13 additional loci: eight loci for MVPA and eight for LST, with three loci overlapping (Supplementary Table 5)[17].

SNP-heritability estimates range from 8% for MVPA to 16% for LST (Supplementary Table 6 and Methods). Genetic correlations between the four traits range from −0.32 for sedentary behavior at work and sedentary commuting, to −0.49 for LST and MVPA (Fig. 1b). To ensure adequate statistical power in instrumental variable and enrichment analyses, we focus on LST and MVPA from here onwards.

Genetic correlations of self-reported LST and MVPA with objective, accelerometry-assessed daily physical activity traits in UK Biobank range from 0.14 to 0.44 (Fig. 1b). Importantly, five of the eight loci previously identified for objectively assessed daily physical activity in UK Biobank data[13,14] show directionally consistent associations ($P<0.05$) with self-reported LST and/or MVPA in our study (Supplementary Table 7). By contrast, 39 LST- and 4 MVPA-associated loci observed here show directionally consistent associations ($P<0.05$) with at least one objectively assessed physical activity and/or sedentary trait (using accelerometry) in UK Biobank (Supplementary Table 8). In line with this, each additional LST-decreasing and MVPA-increasing allele in unweighted genetic predisposition scores of the 88 LST- and 11 MVPA-associated loci, respectively, are associated with higher objectively assessed daily physical activity levels in UK Biobank ($P=5\times10^{-23}$ for LST; $P=2\times10^{-3}$ for MVPA, Supplementary Table 8).

As external validation, we use the European ancestry summary statistics of LST and MVPA to construct polygenic scores (PGSs), and examine their associations with MVPA in 8,195 Bio*Me* BioBank participants of European (n = 2,765), African (n = 2,224) and Hispanic (n = 3,206) ancestry. In general, a higher PGS for MVPA is associated with higher odds of engaging in more than 30 min per week of MVPA, and a higher PGS for LST with lower odds of engaging in MVPA. Individuals at the highest decile of the PGS for LST are 26% less likely to spend more than 30 min per week on MVPA compared with individuals at deciles 4 to 6 (odds ratio (OR) [95% confidence intervals (CI)] = 0.74 [0.55–0.99]) (Fig. 3 and Supplementary Table 9).

Shared genetic architecture. Using linkage disequilibrium (LD) score regression implemented in the LD-Hub[18], we observe significant ($P<4.6\times10^{-4}$) genetic correlations of LST and MVPA with adiposity-related traits ($r=-0.41$ to -0.20), especially with body fat percentage ($r_g=0.4$ and -0.3, respectively; Fig. 4, Supplementary Fig. 2 and Supplementary Table 10). In line with moderate genetic correlations, 11 of the 99 self-reported loci for physical activity and sedentary behavior have previously been associated with obesity-related traits[19–25]. In addition, PGSs for lower LST and higher MVPA are associated with lower BMI in up to 23,723 participants from the Bio*Me* BioBank (Supplementary Table 9), and a phenome-wide association study (PheWAS) in 8,959 Bio*Me* European ancestry samples shows a negative association between the PGS for MVPA and morbid obesity ($P=1.1\times10^{-5}$, Supplementary Fig. 3). Strikingly, genetic correlations with body fat percentage are similar for self-reported LST, MVPA (Fig. 4) and accelerometer-assessed physical activity traits[13,14] (Supplementary Fig. 2).

Besides adiposity, less sedentary behavior and higher physical activity levels are also genetically correlated with a more favorable cardiometabolic status, including lower triglyceride, total cholesterol, fasting glucose and fasting insulin levels, and lower odds of type 2 diabetes and coronary artery disease; as well as with better mental health outcomes, a lower risk of lung cancer and with longevity (Fig. 4 and Supplementary Fig. 2).

Causal inference. To assess directions of causality between sedentary behavior/physical activity and BMI, we next perform two-sample Mendelian randomization (MR) analyses using multiple

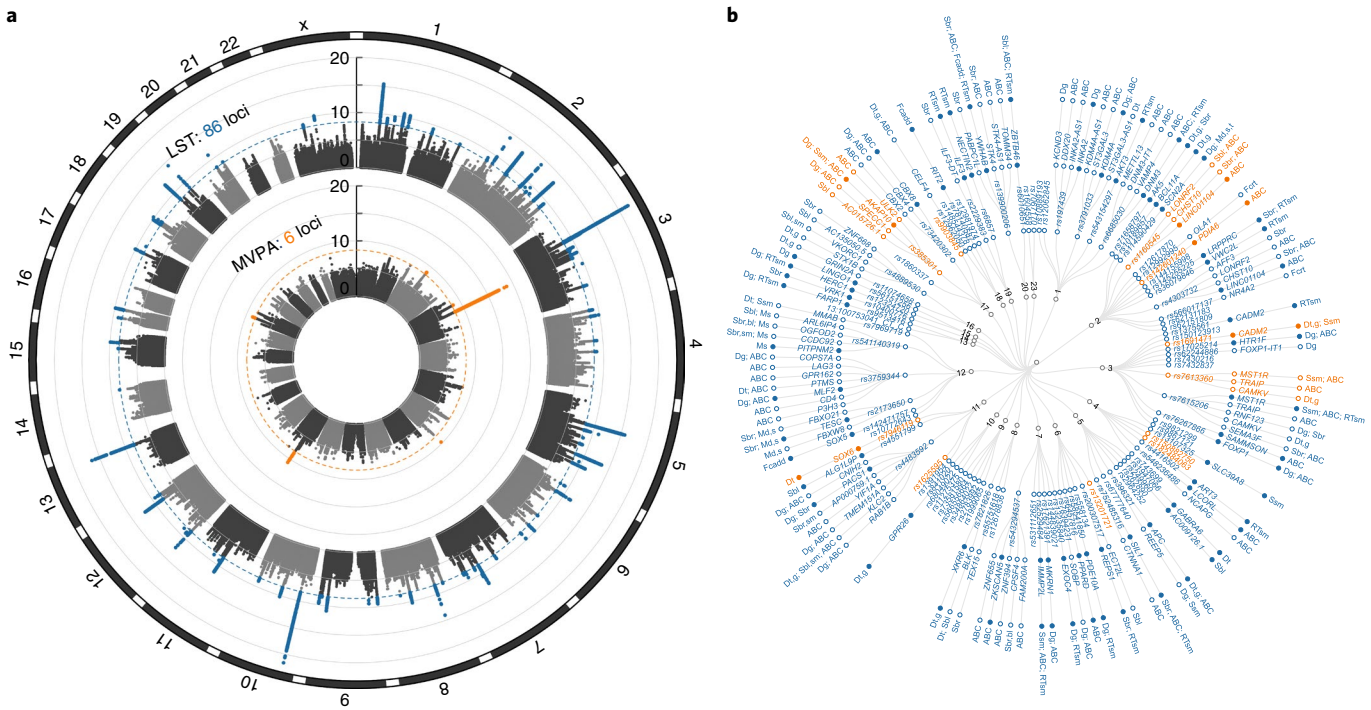


Fig. 2 | Main results of GWAS and downstream gene prioritization for LST and MVPA. a, Circular Manhattan plot summarizing the results from European ancestry meta-analyses for LST and MVPA. Outer track, LST; inner track, MVPA. Genome-wide significant variants ($P < 5 \times 10^{-9}$) are highlighted in orange for loci associated with MVPA and in blue for loci associated with LST. **b**, Dendrogram showing the 101 independent association signals in LST- and MVPA-associated loci from European ancestry or multi-ancestry meta-analyses. Moving outwards from the center are: (1) chromosome; (2) lead SNP identifiers, in orange for loci associated with MVPA, in blue for loci associated with LST; (3) the most promising gene(s) prioritized in the locus (closest genes are highlighted by filled circles); and (4) the approach(es) by which the gene was prioritized, that is, DEPICT gene prioritization (Dg) or tissue enrichment (Dt); SMR of eQTL signals in blood (Sbl), brain (Sbr) or skeletal muscle (Ssm); credible variants identified by FINEMAP that (i) are coding and likely to have a detrimental effect on protein function (Fcadd) or (ii) show evidence of three-dimensional interactions with the candidate gene in central nervous system cell types (Fcrt); activity-by-contact (ABC) in 26 relevant tissues and cell types; a contribution to enrichment for altered expression in skeletal muscle following a resistance training intervention (RTsm); and/or proximity to an association signal for spontaneous running speed (Ms), time run (Mt) or distance run (Md) in a GWAS of 100 inbred mouse strains.

MR methods that utilize genome-wide full summary results or genome-wide significant loci (Supplementary Table 11 and Methods)[26–30]. Causal Analysis Using Summary Effect Estimates (CAUSE)[26] as well as traditional MR methods consistently show that LST and BMI causally affect each other, with the causal effect (the per 1 s.d. unit increase in each trait) of higher LST on higher BMI being two- to threefold larger than the effect of BMI on LST (Fig. 5a, Table 1 and Supplementary Table 11). Results are similar for bidirectional causal inference tests using body fat percentage instead of BMI (Table 2). However, CAUSE cannot distinguish a model of causality from horizontal pleiotropy for body fat percentage and LST (Table 2). CAUSE also illustrates a causal effect of higher LST on higher recalled adiposity and height in childhood (Table 2), supporting our hypothesis that a genetic predisposition for higher LST later in life represents a lifelong predisposition that already influences adiposity through sedentary behavior early in life. We observe similar evidence for causal effects between MVPA and adiposity, with smaller effects when compared with LST.

We next investigate the causal effects of LST and MVPA on common diseases and risk factors, with and without adjusting for BMI (Supplementary Tables 12 and 13). In univariate analyses, we observe effects of lower LST on higher high-density lipoprotein cholesterol levels, higher parental age at death, and on lower odds of type 2 diabetes, attention deficit hyperactivity disorder and depression. The CAUSE model only supports evidence for a causal effect of LST on attention deficit hyperactivity disorder and parental age at death. Importantly, multivariable MR analyses show that all protective causal effects of lower LST are either mediated or confounded by BMI.

Directions of causal effects are consistent across LST and MVPA, but only reach significance for MVPA on parental age at death when using the CAUSE model. As for LST, multivariable MR results suggest that the protective causal effects of higher MVPA are either mediated or confounded by BMI, but results should be interpreted with caution for MVPA because of weak instrument bias (conditional F statistics <10)[31] (Fig. 5b and Supplementary Table 13).

Gene expression in skeletal muscle following training. Although behavior is mainly influenced by signals from the brain, in the case of physical activity, characteristics of skeletal muscle can play a facilitating or restricting role[32]. Therefore, we next examine whether genes in LST- and MVPA-associated loci are enriched for altered messenger RNA expression in skeletal muscle following an acute bout of exercise or a period of training or inactivity[33] (Methods). A mild enrichment for transcripts with an altered expression in skeletal muscle after resistance training is observed for genes nearest to lead SNPs in LST-associated loci ($P = 0.02$) (Extended Data Figs. 1 and 2, and Supplementary Table 14). Of the ten genes driving the enrichment, *PDE10A* may play a critical role in regulating cyclic AMP and cyclic GMP levels in the striatum, a brain region that harbors the central reward system and is important for physical activity regulation[34], and in regulating striatum output[35]; *ILF3* and

Table 1 | Bidirectional MR results for LST and MVPA with BMI or body fat percentage using significant loci only

Exposure	Outcome	Beta	s.e.	*P* value	Exposure	Outcome	Beta	s.e.	*P* value
LST	Body fat %	0.16	0.07	0.016	LST	BMI	0.40	0.04	8.4×10^{-14}
Body fat %	LST	0.12	0.03	0.005	BMI	LST	0.16	0.01	1.4×10^{-74}
MVPA	Body fat %	−0.21	0.17	0.22	MVPA	BMI	−0.25	0.04	0.002
Body fat %	MVPA	−0.001	0.036	0.97	BMI	MVPA	−0.10	0.01	5.8×10^{-12}

We use MR-PRESSO with outliers removed for all pairs of traits except for the causal effect estimation between body fat percentage (body fat %) and MVPA because no outliers were detected by MR-PRESSO. For body fat percentage → MVPA, we reported the causal estimates using an inverse variance-weighted test; for MVPA → body fat percentage, we reported the weighted median method because these two methods were selected by the machine learning framework (Methods) to be the most appropriate approaches for each analysis, respectively. P < 0.0125 indicates significant effects.

NECTIN2—near *APOE*—influence the host response to viral infections[36,37]; *EXOC4* plays a role in insulin-stimulated glucose uptake in skeletal muscle[38]; and *IMMP2L* influences the transport of proteins across the inner mitochondrial membrane[39] (Supplementary Note).

Visual information processing and the reward system. To further improve the understanding of the biological factors that influence sedentary behavior and physical activity, we perform a tissue enrichment analysis using DEPICT[40]. LST- and MVPA-associated loci ($P<1\times10^{-5}$) are most significantly enriched for genes expressed in the retina, visual cortex, occipital lobe and cerebral cortex. This suggests that: (1) possibly subtle differences in the ability to receive, integrate and process visual information influence the likelihood to engage in MVPA; (2) MVPA alters the expression of genes that play a role in visual processes in these tissues; and/or (3) MVPA can slow age-related perceptual and cognitive decline[41]. The LST-associated loci yield similar tissue enrichment results, with retina having the lowest *P* value for enrichment. Interestingly, enrichment for genes expressed in retina was also observed in the High Runner mouse model[42]. Areas related to the reward system (for example, the hippocampus and limbic system) and to memory and navigation (for example, the entorhinal cortex, parahippocampal gyrus, temporal lobe and limbic system) are also enriched in both LST- and MVPA-associated loci (Extended Data Fig. 3 and Supplementary Table 15).

We next use CELLECT[43] to identify enriched cell types using single-cell RNA sequencing data from the Tabula Muris and mouse brain projects[44]. In Tabula Muris data, we observe enrichment in nonmyeloid neurons for MVPA and LST, and of nonmyeloid oligodendrocyte precursor cells for MVPA, possibly highlighting a role for signal transduction (Extended Data Fig. 4 and Supplementary Table 16). In mouse brain data, we identify enrichment for 13 and 45 cell types from 3 and 12 distinct brain regions for MVPA and LST, respectively, including enrichment in dopaminergic neurons (Extended Data Fig. 4 and Supplementary Table 16); a key feature of physical activity regulation in mice[45].

Candidate gene prioritization. To explore mechanisms by which the identified loci may influence LST and MVPA, we next pinpoint genes in GWAS-identified loci: (1) contributing to tissue enrichment or identified by DEPICT's gene prioritization algorithm (Supplementary Tables 15 and 17); (2) whose expression in brain, blood and/or skeletal muscle is anticipated to mediate the association between locus and outcome based on Summary-based MR[46] (SMR; Supplementary Table 18); (3) harboring credible variants with a high posterior probability of being causal (>0.80)[47] and a predicted deleterious effect on protein function (Supplementary Table 19)[48]; (4) showing chromatin–chromatin interactions with credible variants in central nervous system cell types (such genes may be further from lead SNPs, Supplementary Table 19); (5) that—across 26 tissues and cell types—are activated by contact with enhancers presumably affected by causal variants flagged by GWAS hits[49] (Supplementary Tables 20–22); (6) associated with physical activity in GWAS in humans and mice and located <100 kb from the lead variant in humans or mice (Supplementary Note, Supplementary Fig. 4 and Supplementary Tables 23 and 24); and (7) driving enrichment of altered expression in skeletal muscle following resistance exercise training (Supplementary Table 14). Twelve (14%) of the LST-associated loci harbor a variant with a high (>80%) posterior probability of being causal, whereas such variants were not identified among the 11 MVPA-associated loci (Supplementary Table 19). Integrating results across approaches yields 268 candidate genes in 70 LST-associated loci and 39 candidate genes in 8 MVPA-associated loci. Forty-six candidate genes are prioritized by multiple approaches (42 for LST and 6 for MVPA; 2 overlap) and point to endocytosis (*CNIH2*, *RAB1B*, *KLC2*, *PACS1*, *REPS1*, *DNM3*, *EXOC4*), locomotion (*CADM2*, *KLC2*) and myopathy (*MLF2*, *HERC1*, *KLC2*, *SIL1*) as relevant pathways (Supplementary Tables 25 and 26, and Supplementary Note). Seven clusters of protein–protein interactions are predicted, involving 17 of the 46 genes (Extended Data Fig. 5). In vivo perturbation in model systems is required to confirm or refute a role in sedentary behavior and physical activity.

Enrichment of previously reported candidate genes. Candidate gene studies in humans have aimed to identify and characterize the role of genes in exercise (physical activity behavior) and fitness (physical activity ability) for decades. We next examine whether variants in genes that have been linked to or associated with exercise and fitness show evidence of associations with self-reported LST and MVPA[12,50–54]. Of the 58 previously described candidate genes (13 for exercise; 45 for fitness), 56 (13 for exercise and 43 for fitness) harbor variants with $P<0.05$ for associations with LST and/or MVPA ($P_{binomial}=2.1\times10^{-70}$; Supplementary Fig. 5 and Supplementary Table 27). Associations reach traditional genome-wide significance ($P<5\times10^{-8}$) for variants in three genes: *APOE*[55], *PPARD*[56] and *ACTN3* (ref. [57]) (Methods).

The SNP in *APOE* with the lowest *P* value for association with LST is rs429358, for which the C allele associated with lower LST was previously associated with higher self-reported MVPA[13] and forms part of the ε4 risk allele for Alzheimer's disease (Discussion). The SNP with the lowest *P* value for association with LST in the locus is rs6857 ($D'=0.90$; $r^2=0.78$ with rs429358), in the 3′ untranslated region of *NECTIN2*. Neither rs429358 ($P=0.16$) nor rs6857 ($P=0.18$) is associated with MVPA in this study.

The C allele in rs1625595, ~300 kb upstream of *ACTN3*, is associated with higher MVPA ($P=1.9\times10^{-11}$) as well as with higher *ACTN3* expression in skeletal muscle (GTEx, $P=6.6\times10^{-5}$). Alpha-actinin-3 (*ACTN3*) forms a structural component of the muscle's Z-disc that is exclusively expressed in type II_A and II_X muscle fibers[58]. rs1815739, a common *ACTN3* variant that introduces a premature stop codon, p.Arg577Ter, also known as p.Arg620Ter, has been extensively studied in the context of exercise performance[57]. Although we observe little evidence for a role

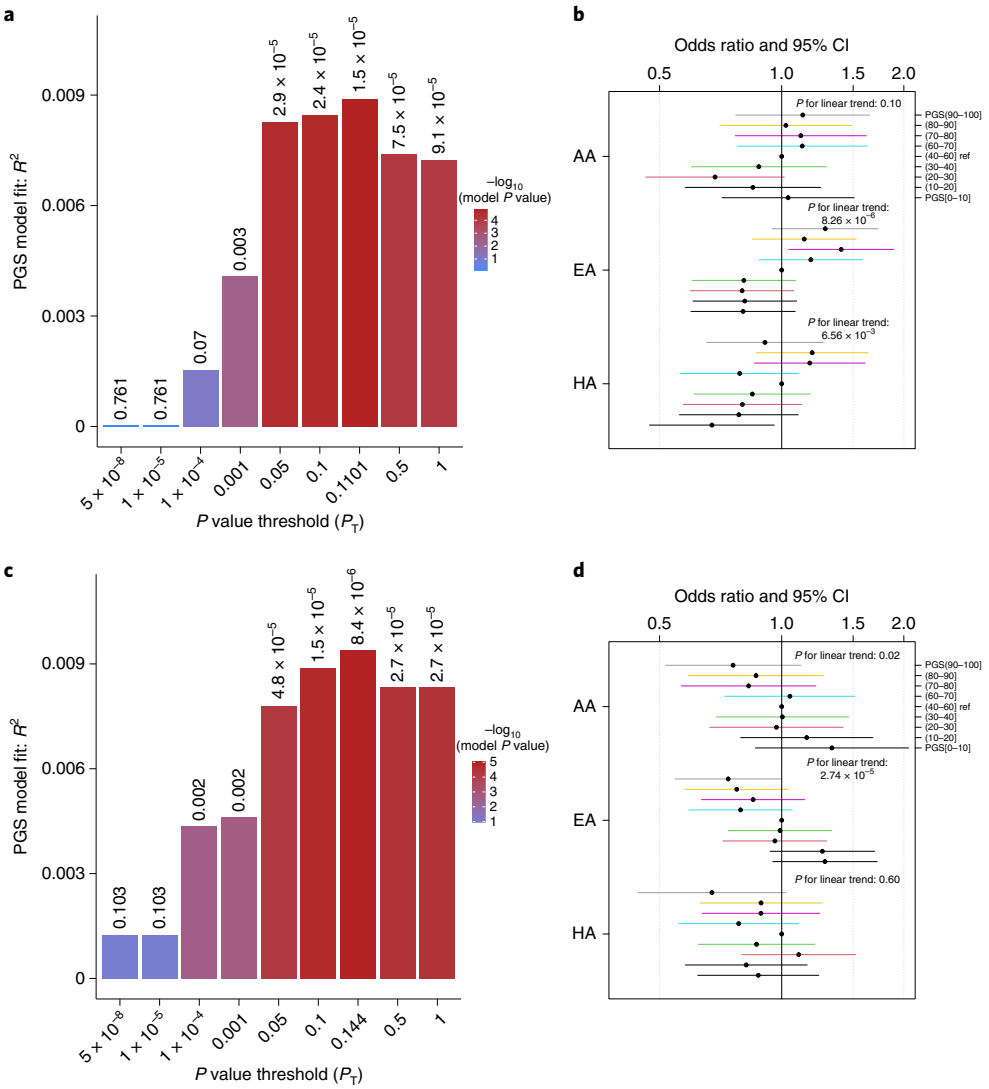


Fig. 3 | Validation of associations with MVPA and LST using PGSs in Bio*Me* participants of three ancestries. a,**c**, The best performing PGSs for MVPA (**a**) and LST (**c**) were derived using logistic/linear regression analyses; that is, those with the highest incremental R^2 above and beyond models with only sex, age and the top ten principal components. This was accomplished using inclusion thresholds of $P < 0.1101$ for MVPA and $P < 0.14$ for LST. **b**,**d**, The association—examined using a logistic regression analysis—of MVPA with the PGSs for MVPA (**b**) and LST (**d**) in individuals of African (AA, $n = 2{,}224$), European (EA, $n = 2{,}765$) and Hispanic (HA, $n = 3{,}206$) ancestry in data from the Bio*Me* BioBank. Dots and error bars show OR and 95% CI.

of rs1815739 in leisure time sedentary behavior or physical activity ($P_{LST} = 0.017$, $P_{MVPA} = 0.17$), the intronic *ACTN3* variants rs679228 ($P_{LST} = 4.3 \times 10^{-8}$) and rs2275998 ($P_{MVPA} = 1.8 \times 10^{-7}$) do show evidence of such associations. Of these, rs2275998—located 646 bp downstream of p.Arg577Ter—is in full LD ($r^2 = 1.0$) with the missense variant rs2229456 (p.Glu635Ala), which likely affects protein function (Combined Annotation Dependent Deletion (CADD) score for the derived, minor, p.635Ala variant =28.6). Each C allele in rs2229456 is associated with less LST ($P = 1.4 \times 10^{-4}$) and higher odds of engaging in MVPA ($P = 8.3 \times 10^{-7}$). Of note, given its downstream location from p.Arg577Ter, a potentially causal effect of rs2229456 on physical activity requires absence of the protein-truncating p.Arg577Ter variant in rs1815739. Haplotype analyses support this (Supplementary Table 28).

Greater ACTN3 flexibility with p.635Ala. Given the striking finding that MVPA and LST are associated with the ACTN3 missense variant rs2229456, but not with the ACTN3-truncating variant rs1815739, we next examine whether rs2229456 (p.Glu635Ala variant) has functional consequences for ACTN3's mechanistic properties at the molecular level. We add ACTN2 to this comparison because it likely compensates for the loss of ACTN3 in the presence of the truncating p.Arg577Ter variant[59]. The results of computer-based (steered) molecular dynamics (MD) simulations and umbrella sampling (see Methods and Supplementary Note for more details) show that the ancestral p.Glu635 variant facilitates salt-bridge and hydrogen-bonding interactions at residue 635 with surrounding residues (for example, R638 and Q639; Fig. 6a,b and Supplementary Fig. 6) via its glutamate side chain. Such interactions are not formed in the presence of the ACTN3 p.635Ala product. They are also less likely to be formed in ACTN2, because of a kink that is present at exactly this location in ACTN2 (Fig. 6c and Supplementary Fig. 6). Moreover, p.635Ala and ACTN2 show distinctly different behavior from p.Glu635, with a greater magnitude of root mean squared fluctuations (r.m.s.f.) in the middle section of the spectrin repeats under no-load conditions (Fig. 6d), suggesting a more flexible structural region. When placed under simulated compressive loads that are likely experienced in vivo, p.635Ala

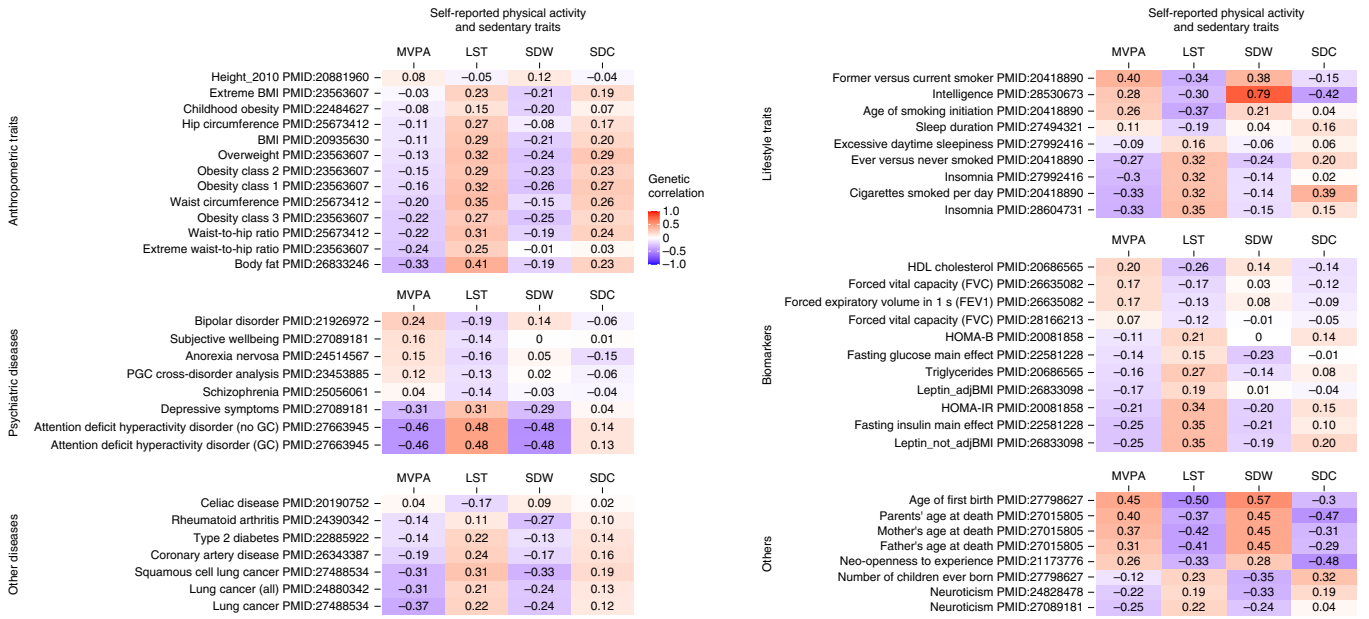


Fig. 4 | Genetic correlations of four self-reported physical activity traits with complex traits and diseases. Results are based on published GWAS with $P < 4.6 \times 10^{-4}$ for at least one physical activity or sedentary trait. Darker colors reflect higher negative (purple) or positive (red) correlation coefficients. GC, genomic control; HDL, high-density lipoprotein; HOMA-B, homeostasis model assessment of beta-cell function; HOMA-IR, homeostatasis model assessment of insulin resistance; PGC, psychiatric genomics consortium.

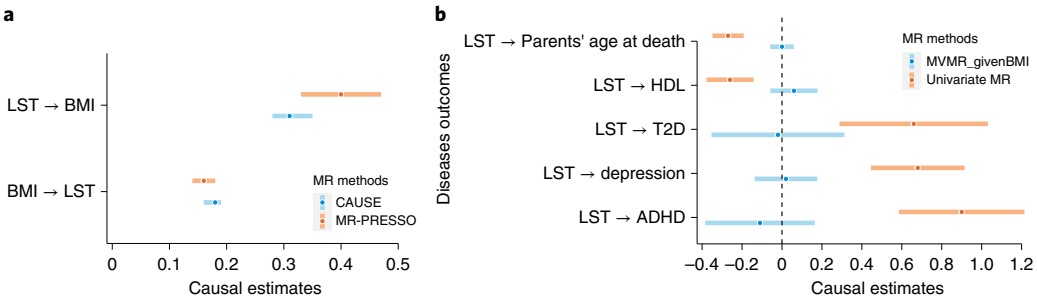


Fig. 5 | MR analyses between LST, MVPA, BMI and complex diseases. a, Median causal estimates for MR analyses using the CAUSE method and causal estimates from the MR-PRESSO method after outlier removal and accounting for horizontal pleiotropy. **b**, The causal effects of LST on complex risk factors and diseases without (in orange) and with (in blue) adjusting for BMI. Dots and error bars show the estimated causal effect sizes and 95% CI. ADHD, attention deficit hyperactivity disorder; T2D, type 2 diabetes.

shows a more linear force versus distance relationship, with greater variance in the potential of mean force (Fig. 6e and Supplementary Fig. 6). Taken together, these results indicate that the ACTN3 p.635Ala dimer—associated with higher MVPA—exhibits similar flexibility to ACTN2 and greater flexibility than the p.Glu635 dimer.

Maximal force and fiber power lower with ACTN3 p.635Ala. We next examine whether a higher predicted ACTN3 dimer flexibility in the presence of p.635Ala has functional consequences in isolated human skeletal muscle fibers. To this end, we compare functional readouts in 298 isolated type I and II_A fibers from vastus lateralis biopsies obtained from eight healthy, young, untrained male participants before and after an eccentric exercise bout[60,61]. Results from a 15,000 iteration Markov chain Monte Carlo model show that stable maximal force—with fibers submerged in activating solution—and fiber power during isotonic load clamps are similar in 32 ± 7 fibers (mean ± s.d.) from three p.Arg577 homozygous, p.Glu635Ala heterozygous individuals compared with 39 ± 6 fibers from four individuals homozygous for the p.577Ter variant; and lower in both groups when compared with 46 fibers from an individual that is homozygous for both the p.Arg577 and p.Glu635 variants (Fig. 6f and Methods). Associations are most striking after an eccentric exercise intervention and are, as expected, more pronounced in type II_A than in type I fibers (Supplementary Fig. 7). Taken together, these results suggest that a more flexible ACTN dimer with lower peak performance (ACTN3 p.635Ala or ACTN2) may be less susceptible to exercise-induced muscle damage than the ancestral ACTN3 p.Glu635, thereby facilitating a more active lifestyle.

Discussion

By doubling the sample size compared with earlier GWAS, we identify 104 independent association signals in 99 loci, including 42 newly identified loci, for self-reported traits reflecting MVPA and sedentary behavior during leisure time. Around half of these also show evidence of directionally consistent associations with objectively assessed physical activity traits. Genetic correlations and two-sample MR analyses show that lower LST results in lower adiposity. Protective causal effects of higher MVPA and lower

Table 2 | Bidirectional MR results for LST and MVPA during leisure time with BMI or body fat percentage using genome-wide summary results (CAUSE method)

Exposure	Outcome	Gamma[a]	95% CI	*P* value[b]	Exposure	Outcome	Gamma[a]	95% CI	*P* value[b]
LST	Body fat %	0.18	0.13 to 0.24	1.8×10^{-3}	LST	BMI	0.31	0.28 to 0.35	6.7×10^{-28}
Body fat %	LST	0.12	0.04 to 0.18	0.14	BMI	LST	0.18	0.16 to 0.19	1.1×10^{-14}
MVPA	Body fat %	−0.12	−0.20 to −0.04	0.07	MVPA	BMI	−0.14	−0.20 to −0.07	6.0×10^{-3}
Body fat %	MVPA	−0.03	−0.09 to 0.02	0.53	BMI	MVPA	−0.09	−0.11 to −0.06	7.4×10^{-3}
LST	Comparative height at age 10	0.03	0.01 to 0.04	0.04	LST	Comparative body size at age 10	0.02	0.01 to 0.03	0.04

[a]Posterior median of gamma, which can be taken as a point estimate of the causal effect. This estimate tends to be shrunk slightly toward zero compared with other methods. [b]The *P* value for comparing the causal model with the sharing model. $P < 0.05$ indicates that posteriors estimated under the causal model predict the data significantly better than posteriors estimated under the sharing model.

LST—acting through or confounded by BMI—are observed for longevity. Tissue and cell-type enrichment analyses suggest a role for visual information processing and the reward system in MVPA and LST, including enrichment for dopaminergic neurons. Loci associated with LST are enriched for genes whose expression in skeletal muscle is altered by resistance training. Forty-six candidate genes are prioritized by more than one approach and point to pathways related to endocytosis, locomotion and myopathy. Finally, results from MD simulations, umbrella sampling and single fiber experiments suggest that a missense variant (rs2229456 encoding ACTN3 p.Glu635Ala) likely increases MVPA, at least in part by reducing susceptibility to exercise-induced muscle damage.

Recent MR studies reported causal protective effects of self-reported and objectively assessed physical activity on breast and colorectal cancer[62,63]. One study concluded that a 1 s.d. increase in self-reported MVPA was associated with lower odds of colorectal cancer (OR = 0.56), with BMI only mediating 2% of the protective effect[63]. Our results—on lung cancer rather than colorectal cancer—show that instrumental variables of MVPA in multivariable MR are weak, and results should be interpreted with caution. Furthermore, a causal effect of objectively assessed, but not self-reported physical activity (MVPA) on depression has been reported[64]. Our MR results for LST on depression show that although the physical activity trait matters, the self-reported nature of it seems inconsequential. According to an earlier study, TV viewing has an attenuated effect but still causes coronary artery disease when adjusting for BMI[15]. The discrepancy with our results—suggesting mediation or confounding by BMI—highlights the importance of including physical activity, as well as BMI-associated variants in multivariable MR analysis, to prevent loss of precision and potentially even biased estimates[31].

It is of interest that a proxy of rs429358, part of the established *APOE* ε4 risk allele for Alzheimer's disease, is associated with lower LST. Klimentidis et al. previously showed that the association of rs429358 with MVPA was stronger in those reporting a family history of Alzheimer's disease, and among older individuals[13]. Based on the direction of the association, it was hypothesized that individuals at higher risk of developing Alzheimer's disease may adopt a healthy lifestyle to mitigate their risk, especially later in life[13]. However, our MR analyses show no evidence of a causal role of MVPA or LST in Alzheimer's disease, and lower average physical activity levels in individuals with a first-degree family history of Alzheimer's disease or dementia[13] suggest other explanations are more likely, although a role for survival bias cannot be ruled out[13]. For example, *APOE* ε4 carriers have a greater increase in aerobic capacity following exercise training[65], which may reinforce a physically active lifestyle independently of Alzheimer's risk. Furthermore, several studies have investigated the moderating role of the *APOE* ε4 allele in the relationship between physical activity and Alzheimer prevention[66]. Although more studies are needed to resolve inconsistencies in the literature, ε4 carriers seem to benefit more from physical activity in terms of reducing the risk of dementia and brain pathology[66].

To investigate the molecular basis for the association of *ACTN3* with MVPA, we compare the ACTN3 p.Glu635 and p.635Ala variants (rs2229456) with each other and with ACTN2—as a functional proxy for ACTN3 p.577Ter—using MD simulations and single fiber experiments. Previous studies using normal mode analysis of alpha-actinin show that several of the natural frequencies have bending flexibility near residue 635. This is interesting because ACTN3's residue 635—the 356th residue of the spectrin repeat region (Fig. 6)—lies outside the linkers between the α-helices of the spectrin repeats, where most flexibility is expected and observed[67]. The absence of salt-bridge and hydrogen-bonding interactions between position 635 (628 in ACTN2) and surrounding residues—due to either the presence of the alanine substitution at ACTN3's residue 635, or a kink in the α-helix at ACTN2's residue 628—increases the flexibility of the dimer under a compressive load, with far less work required to deform the homodimer beyond a compressive distance of 1.2 nm. The p.635Ala substitution may reduce the stiffness of the muscle fiber while undergoing elastic deformation during exercise to a level that is comparable with ACTN2. Although at the expense of the maximal force that single fibers can generate, this may reduce exercise-induced microtrauma caused by Z-disc rupture or streaming[1], alleviating delayed onset muscle soreness[2] and risk of injuries[3], enabling a more active lifestyle. Our results suggest it would be interesting to revisit the plethora of data on p.Arg577Ter, and differentiate between effects of the p.Arg577Ter and p.Glu635Ala variants.

In conclusion, our results shed light on genetic variants and molecular mechanisms that influence physical activity and sedentary behavior in daily life. As would be expected for complex behaviors that involve both motivation and physical ability, these mechanisms occur in multiple organs and organ systems. In addition, our causal inference supports the important public health message that a physically active lifestyle mitigates the risk of multiple diseases, in major part through or confounded by an effect on BMI.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-022-01165-1.

Received: 30 November 2021; Accepted: 18 July 2022;
Published online: 07 September 2022

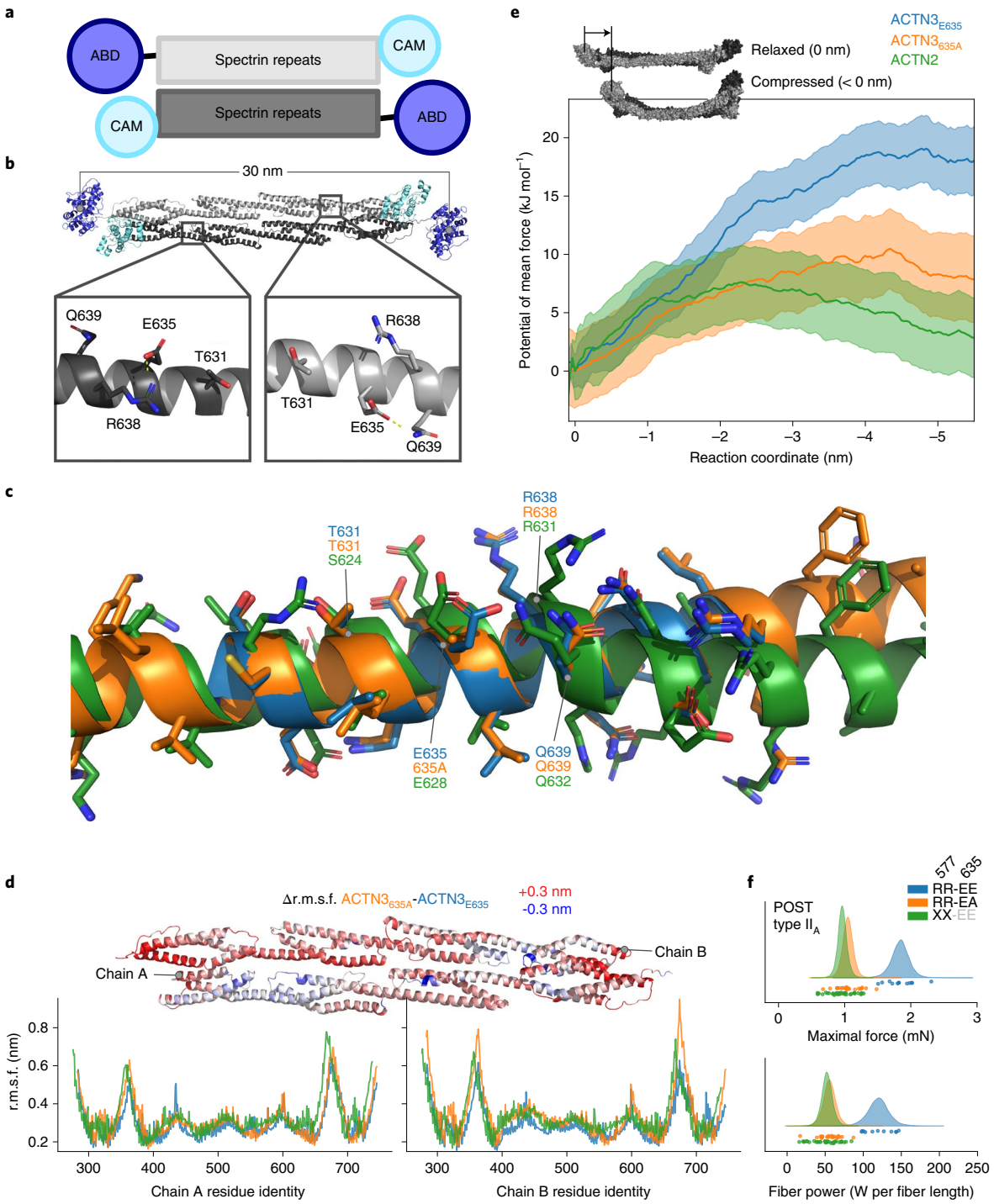


Fig. 6 | Allele p.635Ala in *ACTN3* results in a more flexible ACTN3 homodimer. a, ACTN3 is a homodimer of two antiparallel filaments, with each filament consisting of an N-terminal actin binding domain (ABD, blue), followed by a structural region comprised of four spectrin repeats (gray) with a C-terminal calmodulin (CAM) homology domain (cyan). **b**, The glutamate residue side chain in position 635 of ACTN3 (p.Glu635) interacts primarily with the arginine in position 638 and the glutamine in position 639. **c**, The α-helix comprised of residues adjacent to ACTN3 residue 635 (ACTN2 628) exhibits a pronounced kink in ACTN2 (green) at this α-helical turn compared with ACTN3 p.Glu635 (blue) and p.635Ala (orange), decreasing the likelihood of interactions under load with R631, whereas the alanine substitution of ACTN3 p.635Ala precludes any side chain interactions with neighboring residues p.Arg638 or p.Glu639. **d**, The r.m.s.f. of the spectrin repeat structural region of the ACTN3 dimer for a 150 ns MD simulation for variants p.Glu635 (blue) and p.635Ala (orange, higher MVPA) and ACTN2 (green) (bottom), with the difference in r.m.s.f. between ACTN3 variants shown mapped to the spectrin repeat region (top) with ±0.3 nm difference (red, positive and blue, negative). **e**, Umbrella sampling of ACTN3 variants p.Glu635 and p.635Ala and ACTN2 with orange, blue and green traces representing the potential of mean force for ACTN3 variants p.635Ala (orange) and p.Glu635 (blue) and ACTN2 (green) ±1 s.d. The reaction coordinate is the distance between the two ABD centers of mass of each dimer, a negative value indicating a shorter distance between the two ABDs. Inset shows the relaxed dimer at reaction coordinate of 0 nm (top) and the direction and effect on the compressive force. **f**, Single fiber experiments show a higher maximal force and fiber power during isotonic contractions after an eccentric exercise bout in type II_A fibers from an individual homozygous for p.Arg577 and p.Glu635 (blue) compared with type II_A fibers from three p.Arg577 homozygous, p.Glu635Ala heterozygous individuals (orange); and from four p.577Ter homozygous individuals (green).

References

1. Lee, I. M. et al. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *Lancet* **380**, 219–229 (2012).
2. *Global Action Plan for the Prevention and Control of Noncommunicable Diseases 2013–2020* (World Health Organization, 2013).
3. Guthold, R., Stevens, G. A., Riley, L. M. & Bull, F. C. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1.9 million participants. *Lancet Glob. Health* **6**, e1077–e1086 (2018).
4. Wang, Y. et al. Secular trends in sedentary behaviors and associations with weight indicators among Chinese reproductive-age women from 2004 to 2015: findings from the China Health and Nutrition Survey. *Int J. Obes. (Lond.)* **44**, 2267–2278 (2020).
5. Wijndaele, K. et al. Television viewing time independently predicts all-cause and cardiovascular mortality: the EPIC Norfolk study. *Int J. Epidemiol.* **40**, 150–159 (2011).
6. Wijndaele, K., Sharp, S. J., Wareham, N. J. & Brage, S. Mortality risk reductions from substituting screen time by discretionary activities. *Med Sci. Sports Exerc.* **49**, 1111–1119 (2017).
7. Bauman, A. E. et al. Correlates of physical activity: why are some people physically active and others not? *Lancet* **380**, 258–271 (2012).
8. den Hoed, M. et al. Heritability of objectively assessed daily physical activity and sedentary behavior. *Am. J. Clin. Nutr.* **98**, 1317–1325 (2013).
9. Stubbe, J. H. et al. Genetic influences on exercise participation in 37,051 twin pairs from seven countries. *PLoS ONE* **1**, e22 (2006).
10. Fan, W. et al. PPARδ promotes running endurance by preserving glucose. *Cell Metab.* **25**, 1186–1193.e4 (2017).
11. Buniello, A. et al. The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).
12. Sarzynski, M. A. et al. Advances in exercise, fitness, and performance genomics in 2015. *Med. Sci. Sports Exerc.* **48**, 1906–1916 (2016).
13. Klimentidis, Y. C. et al. Genome-wide association study of habitual physical activity in over 377,000 UK Biobank participants identifies multiple variants including CADM2 and APOE. *Int. J. Obes.* **42**, 1161–1176 (2018).
14. Doherty, A. et al. GWAS identifies 14 loci for device-measured physical activity and sleep duration. *Nat. Commun.* **9**, 5257 (2018).
15. van de Vegte, Y. J., Said, M. A., Rienstra, M., van der Harst, P. & Verweij, N. Genome-wide association studies and Mendelian randomization analyses for leisure sedentary behaviours. *Nat. Commun.* **11**, 1770 (2020).
16. Kilpeläinen, T. O. et al. Multi-ancestry study of blood lipid levels identifies four loci interacting with physical activity. *Nat. Commun.* **10**, 376 (2019).
17. Turley, P. et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* **50**, 229–237 (2018).
18. Zheng, J. et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272–279 (2016).
19. Shungin, D. et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187–196 (2015).
20. Kichaev, G. et al. Leveraging polygenic functional enrichment to improve GWAS power. *Am. J. Hum. Genet.* **104**, 65–75 (2019).
21. Astle, W. J. et al. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* **167**, 1415–1429.e19 (2016).
22. Pulit, S. L. et al. Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum. Mol. Genet.* **28**, 166–174 (2019).
23. Winkler, T. W. et al. The influence of age and sex on genetic associations with adult body size and shape: a large-scale genome-wide interaction study. *PLoS Genet.* **11**, e1005378 (2015).
24. Locke, A. E. et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
25. Justice, A. E. et al. Genome-wide meta-analysis of 241,258 adults accounting for smoking behaviour identifies novel loci for obesity traits. *Nat. Commun.* **8**, 14977 (2017).
26. Morrison, J., Knoblauch, N., Marcus, J. H., Stephens, M. & He, X. Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. *Nat. Genet.* **52**, 740–747 (2020).
27. Verbanck, M., Chen, C.-Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* **50**, 693–698 (2018).
28. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37**, 658–665 (2013).
29. Hartwig, F. P., Davey Smith, G. & Bowden, J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int. J. Epidemiol.* **46**, 1985–1998 (2017).
30. Hemani, G. et al. Automating Mendelian randomization through machine learning to construct a putative causal map of the human phenome. Preprint at *bioRxiv* https://doi.org/10.1101/173682 (2017).
31. Sanderson, E., Smith, G. D., Windmeijer, F. & Bowden, J. An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. *Int. J. Epidemiol.* **48**, 713–727 (2019).
32. Lightfoot, J. T. et al. Biological/genetic regulation of physical activity level: consensus from GenBioPAC. *Med. Sci. Sports Exerc.* **50**, 863–873 (2018).
33. Pillon, N. J. et al. Transcriptomic profiling of skeletal muscle adaptations to exercise and inactivity. *Nat. Commun.* **11**, 470 (2020).
34. Saul, M. C. et al. High motivation for exercise is associated with altered chromatin regulators of monoamine receptor gene expression in the striatum of selectively bred mice. *Genes Brain Behav.* **16**, 328–341 (2017).
35. Threlfell, S., Sammut, S., Menniti, F. S., Schmidt, C. J. & West, A. R. Inhibition of phosphodiesterase 10A increases the responsiveness of striatal projection neurons to cortical stimulation. *J. Pharmacol. Exp. Ther.* **328**, 785–795 (2009).
36. Harashima, A., Guettouche, T. & Barber, G. N. Phosphorylation of the NFAR proteins by the dsRNA-dependent protein kinase PKR constitutes a novel mechanism of translational regulation and cellular defense. *Genes Dev.* **24**, 2640–2653 (2010).
37. Zhu, Y. et al. Identification of CD112R as a novel checkpoint for human T cells. *J. Exp. Med.* **213**, 167–176 (2016).
38. Inoue, M., Chang, L., Hwang, J., Chiang, S. H. & Saltiel, A. R. The exocyst complex is required for targeting of Glut4 to the plasma membrane by insulin. *Nature* **422**, 629–633 (2003).
39. Burri, L. et al. Mature DIABLO/Smac is produced by the IMP protease complex on the mitochondrial inner membrane. *Mol. Biol. Cell* **16**, 2926–2933 (2005).
40. Pers, T. H. et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
41. Muiños, M. & Ballesteros, S. Does physical exercise improve perceptual skills and visuospatial attention in older adults? A review. *Eur. Rev. Aging Phys. Act.* **15**, 2 (2018).
42. Hillis, D. A. et al. Genetic basis of aerobically supported voluntary exercise: results from a selection experiment with house mice. *Genetics* **216**, 781–804 (2020).
43. Timshel, P. N., Thompson, J. J. & Pers, T. H. Genetic mapping of etiologic brain cell types for obesity. *eLife* **9**, e55851 (2020).
44. Schaum, N. et al. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367–372 (2018).
45. Roberts, M. D., Ruegsegger, G. N., Brown, J. D. & Booth, F. W. Mechanisms associated with physical activity behavior: insights from rodent experiments. *Exerc. Sport Sci. Rev.* **45**, 217–222 (2017).
46. Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481 (2016).
47. Benner, C. et al. FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493–1501 (2016).
48. Kircher, M. et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
49. Nasser, J. et al. Genome-wide enhancer maps link risk variants to disease genes. *Nature* **593**, 238–243 (2021).
50. Bray, M. S. et al. The human gene map for performance and health-related fitness phenotypes: the 2006–2007 update. *Med. Sci. Sports Exerc.* **41**, 35–73 (2009).
51. de Geus, E. J., Bartels, M., Kaprio, J., Lightfoot, J. T. & Thomis, M. Genetics of regular exercise and sedentary behaviors. *Twin Res. Hum. Genet* **17**, 262–271 (2014).
52. Weyerstraß, J., Stewart, K., Wesselius, A. & Zeegers, M. Nine genetic polymorphisms associated with power athlete status – a meta-analysis. *J. Sci. Med. Sport* **21**, 213–220 (2018).
53. Moir, H. J. et al. Genes and elite marathon running performance: a systematic review. *J. Sports Sci. Med.* **18**, 559–568 (2019).
54. Kim, D. S., Wheeler, M. T. & Ashley, E. A. The genetics of human performance. *Nat. Rev. Genet.* **23**, 40–54 (2021).
55. Hagberg, J. M. et al. Apolipoprotein E genotype and exercise training-induced increases in plasma high-density lipoprotein (HDL)- and HDL2-cholesterol levels in overweight men. *Metabolism* **48**, 943–945 (1999).
56. Gielen, M. et al. Heritability and genetic etiology of habitual physical activity: a twin study with objective measures. *Genes Nutr.* **9**, 415, 1–12 (2014).
57. Pickering, C. & Kiely, J. ACTN3: more than just a gene for speed. *Front. Physiol.* **8**, 1080 (2017).
58. Vincent, B. et al. ACTN3 (R577X) genotype is associated with fiber type distribution. *Physiol. Genomics* **32**, 58–63 (2007).
59. Norman, B. et al. Strength, power, fiber types, and mRNA expression in trained men and women with different ACTN3 R577X genotypes. *J. Appl. Physiol. (1985)* **106**, 959–965 (2009).
60. Broos, S. et al. Evidence for ACTN3 as a speed gene in isolated human muscle fibers. *PLoS ONE* **11**, e0150594 (2016).

61. Broos, S. et al. The stiffness response of type IIa fibres after eccentric exercise-induced muscle damage is dependent on ACTN3 r577X polymorphism. *Eur. J. Sport Sci.* **19**, 480–489 (2019).
62. Papadimitriou, N. et al. Physical activity and risks of breast and colorectal cancer: a Mendelian randomisation analysis. *Nat. Commun.* **11**, 597 (2020).
63. Zhang, X. et al. Genetically predicted physical activity levels are associated with lower colorectal cancer risk: a Mendelian randomisation study. *Br. J. Cancer* **124**, 1330–1338 (2021).
64. Choi, K. W. et al. Assessment of bidirectional relationships between physical activity and depression among adults: a 2-sample Mendelian randomization study. *JAMA Psychiatry* **76**, 399–408 (2019).
65. Thompson, P. D. et al. Apolipoprotein E genotype and changes in serum lipids and maximal oxygen uptake with exercise training. *Metabolism* **53**, 193–202 (2004).
66. de Frutos-Lucas, J. et al. Does APOE genotype moderate the relationship between physical activity, brain health and dementia risk? A systematic review. *Ageing Res. Rev.* **64**, 101173 (2020).
67. Golji, J., Collins, R. & Mofrad, M. R. Molecular mechanics of the alpha-actinin rod domain: bending, torsional, and extensional behavior. *PLoS Comput. Biol.* **5**, e1000389 (2009).

Zhe Wang[1]✉, Andrew Emmerich[2], Nicolas J. Pillon[3], Tim Moore[4], Daiane Hemerich[1], Marilyn C. Cornelis[5], Eugenia Mazzaferro[6], Siacia Broos[7,8], Tarunveer S. Ahluwalia[9,10,11], Traci M. Bartz[12,13], Amy R. Bentley[14], Lawrence F. Bielak[15], Mike Chong[16], Audrey Y. Chu[17,18], Diane Berry[19], Rajkumar Dorajoo[20,21], Nicole D. Dueker[22,23], Elisa Kasbohm[24,25], Bjarke Feenstra[26], Mary F. Feitosa[27], Christian Gieger[28], Mariaelisa Graff[29], Leanne M. Hall[30,31], Toomas Haller[32], Fernando P. Hartwig[33,34], David A. Hillis[35], Ville Huikari[36], Nancy Heard-Costa[37,38], Christina Holzapfel[28,39], Anne U. Jackson[40], Åsa Johansson[41], Anja Moltke Jørgensen[10], Marika A. Kaakinen[42,43], Robert Karlsson[44], Kathleen F. Kerr[13], Boram Kim[45], Chantal M. Koolhaas[46], Zoltan Kutalik[47,48,49], Vasiliki Lagou[50], Penelope A. Lind[51,52], Mattias Lorentzon[53,54], Leo-Pekka Lyytikäinen[55,56], Massimo Mangino[57,58], Christoph Metzendorf[6], Kristine R. Monroe[59], Alexander Pacolet[7], Louis Pérusse[60,61], Rene Pool[62,63], Rebecca C. Richmond[64], Natalia V. Rivera[65,66,67], Sebastien Robiou-du-Pont[68], Katharina E. Schraut[69], Christina-Alexandra Schulz[70,71], Heather M. Stringham[40], Toshiko Tanaka[72], Alexander Teumer[24,73], Constance Turman[74], Peter J. van der Most[75], Mathias Vanmunster[7], Frank J. A. van Rooij[46], Jana V. van Vliet-Ostaptchouk[76,77], Xiaoshuai Zhang[78,79], Jing-Hua Zhao[80], Wei Zhao[15], Zhanna Balkhiyarova[43,81,82], Marie N. Balslev-Harder[10], Sebastian E. Baumeister[24,83], John Beilby[84], John Blangero[85], Dorret I. Boomsma[62,63], Soren Brage[78], Peter S. Braund[30,31], Jennifer A. Brody[12], Marcel Bruinenberg[86], Ulf Ekelund[87,88], Ching-Ti Liu[89], John W. Cole[90], Francis S. Collins[91], L. Adrienne Cupples[37,89], Tõnu Esko[32], Stefan Enroth[41], Jessica D. Faul[92], Lindsay Fernandez-Rhodes[93], Alison E. Fohner[94], Oscar H. Franco[46,95], Tessel E. Galesloot[96], Scott D. Gordon[51], Niels Grarup[10], Catharina A. Hartman[97], Gerardo Heiss[29], Jennie Hui[84,98,99], Thomas Illig[100,101], Russell Jago[102], Alan James[103], Peter K. Joshi[69,104], Taeyeong Jung[45], Mika Kähönen[56,105], Tuomas O. Kilpeläinen[10], Woon-Puay Koh[106,107], Ivana Kolcic[108], Peter P. Kraft[74], Johanna Kuusisto[109], Lenore J. Launer[110], Aihua Li[68], Allan Linneberg[111,112], Jian'an Luan[78], Pedro Marques Vidal[113], Sarah E. Medland[51,114], Yuri Milaneschi[115], Arden Moscati[1], Bill Musk[99,174], Christopher P. Nelson[30,31], Ilja M. Nolte[75], Nancy L. Pedersen[44], Annette Peters[116], Patricia A. Peyser[15], Christine Power[19], Olli T. Raitakari[117,118,119], Mägi Reedik[32], Alex P. Reiner[120], Paul M. Ridker[17,121], Igor Rudan[69], Kathy Ryan[122], Mark A. Sarzynski[123], Laura J. Scott[40], Robert A. Scott[78], Stephen Sidney[124], Kristin Siggeirsdottir[125], Albert V. Smith[40,125], Jennifer A. Smith[15,92], Emily Sonestedt[70], Marin Strøm[26,126], E. Shyong Tai[127,128,129], Koon K. Teo[68,130], Barbara Thorand[116], Anke Tönjes[131], Angelo Tremblay[60,61], Andre G. Uitterlinden[132], Jagadish Vangipurapu[109], Natasja van Schoor[133], Uwe Völker[73,134], Gonneke Willemsen[62,63],

Kayleen Williams[13], Quenna Wong[13], Huichun Xu[122], Kristin L. Young[29], Jian Min Yuan[135,136], M. Carola Zillikens[132], Alan B. Zonderman[137], Adam Ameur[41], Stefania Bandinelli[138], Joshua C. Bis[12], Michael Boehnke[40], Claude Bouchard[139], Daniel I. Chasman[17,121], George Davey Smith[34,140], Eco J. C. de Geus[62,63], Louise Deldicque[141], Marcus Dörr[73,142], Michele K. Evans[137], Luigi Ferrucci[72], Myriam Fornage[143], Caroline Fox[144], Theodore Garland Jr[145], Vilmundur Gudnason[125,146], Ulf Gyllensten[41], Torben Hansen[10], Caroline Hayward[147], Bernardo L. Horta[33], Elina Hyppönen[148,149,150], Marjo-Riitta Jarvelin[36,151], W. Craig Johnson[13], Sharon L. R. Kardia[15], Lambertus A. Kiemeney[96], Markku Laakso[109], Claudia Langenberg[78,152], Terho Lehtimäki[55,56], Loic Le Marchand[153], Lifelines Cohort Study*, Patrik K. E. Magnusson[44], Nicholas G. Martin[51], Mads Melbye[154,155,156,112], Andres Metspalu[32], David Meyre[16,68], Kari E. North[29], Claes Ohlsson[157,158], Albertine J. Oldehinkel[97], Marju Orho-Melander[70], Guillaume Pare[16], Taesung Park[45,159], Oluf Pedersen[10], Brenda W. J. H. Penninx[115], Tune H. Pers[10], Ozren Polasek[160], Inga Prokopenko[81,82,161], Charles N. Rotimi[14], Nilesh J. Samani[30,31], Xueling Sim[127], Harold Snieder[75], Thorkild I. A. Sørensen[10,162], Tim D. Spector[57], Nicholas J. Timpson[163], Rob M. van Dam[127,164], Nathalie van der Velde[132,165,166], Cornelia M. van Duijn[46,167], Peter Vollenweider[113], Henry Völzke[24,73], Trudy Voortman[46], Gérard Waeber[113], Nicholas J. Wareham[78], David R. Weir[92], Heinz-Erich Wichmann[116], James F. Wilson[69,147], Andrea L. Hevener[168], Anna Krook[3], Juleen R. Zierath[3,10,169], Martine A. I. Thomis[8], Ruth J. F. Loos[1,10,170] and Marcel den Hoed[6]✉

[1]The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [2]Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden. [3]Department of Physiology and Pharmacology, Karolinska Institutet, Stockholm, Sweden. [4]Division of Cardiology, Department of Medicine, University of California, Los Angeles, CA, USA. [5]Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. [6]The Beijer Laboratory and Department of Immunology, Genetics and Pathology, Uppsala University and SciLifeLab, Uppsala, Sweden. [7]Faculty of Movement and Rehabilitation Sciences, Department of Movement Sciences - Exercise Physiology Research Group, KU Leuven, Leuven, Belgium. [8]Faculty of Movement and Rehabilitation Sciences, Department of Movement Sciences - Physical Activity, Sports & Health Research Group, KU Leuven, Leuven, Belgium. [9]Steno Diabetes Center Copenhagen, Herlev, Denmark. [10]Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [11]The Bioinformatics Center, Department of Biology, University of Copenhagen, Copenhagen, Denmark. [12]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA. [13]Department of Biostatistics, University of Washington, Seattle, WA, USA. [14]Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA. [15]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA. [16]Department of Pathology and Molecular Medicine, McMaster University, Hamilton, Ontario, Canada. [17]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, USA. [18]GlaxoSmithKline, Cambridge, MA, USA. [19]Division of Population, Policy and Practice, Great Ormond Street Hospital Institute for Child Health, University College London, London, UK. [20]Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, Singapore. [21]Health Services and Systems Research, Duke-NUS Medical School, Singapore, Singapore. [22]John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL, USA. [23]Department of Epidemiology & Public Health, University of Maryland School of Medicine, Baltimore, MD, USA. [24]Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. [25]Institute of Mathematics and Computer Science, University of Greifswald, Greifswald, Germany. [26]Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark. [27]Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA. [28]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München –Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH), Munich, Germany. [29]Department of Epidemiology, University of North Carolina, Chapel Hill, NC, USA. [30]Department of Cardiovascular Sciences, University of Leicester, Leicester, UK. [31]NIHR Leicester Biomedical Research Centre, Glenfield Hospital, Leicester, UK. [32]Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia. [33]Postgraduate Program in Epidemiology, Federal University of Pelotas, Pelotas, Brazil. [34]MRC Integrative Epidemiology Unit, NIHR Bristol Biomedical Research Center, University of Bristol, Bristol, UK. [35]Genetics, Genomics, and Bioinformatics Graduate Program, University of California, Riverside, CA, USA. [36]Institute of Health Sciences, University of Oulu, Oulu, Finland. [37]Framingham Heart Study, Framingham, MA, USA. [38]Department of Neurology, Boston University School of Medicine, Boston, MA, USA. [39]Institute for Nutritional Medicine, School of Medicine, Technical University of Munich, Munich, Germany. [40]Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA. [41]Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden. [42]Section of Statistical Multi-omics, Department of Clinical and Experimental Medicine, University of Surrey, Guildford, UK. [43]Department of Metabolism, Digestion and Reproduction, Imperial College London, London, UK. [44]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [45]Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul, South Korea. [46]Department of Epidemiology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands. [47]University Center for Primary Care and Public Health, University of Lausanne, Lausanne, Switzerland. [48]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [49]Department of Computational Biology, University of Lausanne, Lausanne, Switzerland. [50]Wellcome Sanger Institute, Cambridge, UK. [51]Mental Health and Neuroscience Research Program, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [52]School of Biomedical Science, Faculty of Medicine, University of Queensland, Brisbane, Queensland,

Australia. [53]Geriatric Medicine, Institute of Medicine, University of Gothenburg and Sahlgrenska University Hospital Mölndal, Gothenburg, Sweden. [54]Mary MacKillop Institute for Health Research, Australian Catholic University, Melbourne, Victoria, Australia. [55]Department of Clinical Chemistry, Fimlab Laboratories, Tampere, Finland. [56]Finnish Cardiovascular Research Center – Tampere, Department of Clinical Chemistry, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland. [57]Department of Twin Research and Genetic Epidemiology, Kings College London, London, UK. [58]NIHR Biomedical Research Centre at Guy's and St Thomas' Foundation Trust, London, UK. [59]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. [60]Department of Kinesiology, Université Laval, Quebec, Quebec, Canada. [61]Centre Nutrition Santé et Société (NUTRISS), Institute of Nutrition and Functional Foods (INAF), Université Laval, Quebec, Quebec, Canada. [62]Department of Biological Psychology, Vrije Universiteit, Amsterdam, the Netherlands. [63]Amsterdam Public Health Research Institute, Amsterdam UMC, Amsterdam, the Netherlands. [64]MRC Integrative Epidemiology Unit and Avon Longitudinal Study of Parents and Children, University of Bristol Medical School, Population Health Sciences and Avon Longitudinal Study of Parents and Children, University of Bristol, Bristol, UK. [65]Respiratory Division, Department of Medicine, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden. [66]Rheumatology Division, Department of Medicine, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden. [67]Center of Molecular Medicine (CMM), Karolinska Institutet, Stockholm, Sweden. [68]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada. [69]Centre for Global Health Research, Usher Institute, University of Edinburgh, Edinburgh, UK. [70]Department of Clinical Sciences Malmö, Lund University, Malmö, Sweden. [71]Department of Nutrition and Food Sciences, Nutritional Epidemiology, University of Bonn, Bonn, Germany. [72]Translational Gerontology Branch, National Institute on Aging, Baltimore, MD, USA. [73]German Centre for Cardiovascular Research (DZHK), partner site Greifswald, Greifswald, Germany. [74]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [75]Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [76]Department of Endocrinology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [77]Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [78]MRC Epidemiology Unit, University of Cambridge, Cambridge, UK. [79]School of Public Health, Department of Biostatistics, Shandong University, Jinan, China. [80]Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [81]Department of Clinical and Experimental Medicine, University of Surrey, Guilford, UK. [82]People-Centred Artificial Intelligence Institute, University of Surrey, Guilford, UK. [83]University of Münster, Münster, Germany. [84]Diagnostic Genomics, PathWest Laboratory Medicine WA, Perth, Western Australia, Australia. [85]South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley, Brownsville, TX, USA. [86]Lifelines Cohort Study, Groningen, the Netherlands. [87]Department of Sports Medicine, Norwegian School of Sport Sciences, Oslo, Norway. [88]Department of Chronic Diseases, Norwegian Institute of Public Health, Oslo, Norway. [89]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA. [90]Vascular Neurology, Department of Neurology, University of Maryland School of Medicine and the Baltimore VAMC, Baltimore, MD, USA. [91]Center for Precision Health Research, National Human Genome Research Institute, NIH, Bethesda, MD, USA. [92]Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, USA. [93]Department of Biobehavioral Health, College of Health and Human Development, Pennsylvania State University, University Park, PA, USA. [94]Department of Epidemiology, Institute of Public Health Genetics, Cardiovascular Health Research Unit, University of Washington, Seattle, WA, USA. [95]Institute of Social and Preventive Medicine (ISPM), University of Bern, Bern, Switzerland. [96]Radboud Institute for Health Sciences, Department for Health Evidence, Radboud University Medical Center, Nijmegen, the Netherlands. [97]Interdisciplinary Center Psychopathology and Emotion Regulation, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [98]School of Population and Global Health, The University of Western Australia, Perth, Western Australia, Australia. [99]Busselton Population Medical Research Institute, Busselton, Western Australia, Australia. [100]Hannover Unified Biobank, Hannover Medical School, Hannover, Germany. [101]Department of Human Genetics, Hannover Medical School, Hannover, Germany. [102]Centre for Exercise Nutrition & Health Sciences, School for Policy Studies, University of Bristol, Bristol, UK. [103]Department of Pulmonary Physiology and Sleep Medicine, Sir Charles Gairdner Hospital, , Western Australia, Perth, Australia. [104]Humanity Inc, Boston, MA, USA. [105]Department of Clinical Physiology, Tampere University Hospital, Tampere, Finland. [106]Healthy Longevity Translational Research Programme, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. [107]Singapore Institute for Clinical Sciences, Agency for Science, Technology, and Research, Singapore, Singapore. [108]Department of Public Health, University of Split School of Medicine, Split, Croatia. [109]Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland. [110]Laboratory of Epidemiology and Population Sciences, National Institutes of Health, Baltimore, MD, USA. [111]Center for Clinical Research and Prevention, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark. [112]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [113]Division of Internal Medicine, Department of Medicine, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland. [114]School of Psychology and Faculty of Medicine, University of Queensland, St Lucia, Queensland, Australia. [115]Department of Psychiatry, Amsterdam UMC, Vrije Universiteit, Amsterdam, the Netherlands. [116]Institute of Epidemiology, Helmholtz Zentrum München –Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH), Munich, Germany. [117]Centre for Population Health Research, University of Turku and Turku University Hospital, Turku, Finland. [118]Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland. [119]Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland. [120]Department of Epidemiology, University of Washington, Seattle, WA, USA. [121]Harvard Medical School, Boston, MA, USA. [122]Division of Endocrinology, Diabetes and Nutrition, Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA. [123]Department of Exercise Science, University of South Carolina, Columbia, SC, USA. [124]Division of Research, Kaiser Permanente Northern California, Oakland, CA, USA. [125]Icelandic Heart Association, Kópavogur, Iceland. [126]Faculty of Health Sciences, University of the Faroe Islands, Tórshavn, Faroe Islands. [127]Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore. [128]Duke-NUS Medical School, Singapore, Singapore. [129]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. [130]Department of Medicine, McMaster University, Hamilton, Ontario, Canada. [131]Department of Medicine, University of Leipzig, Leipzig, Germany. [132]Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, the Netherlands. [133]Department of Epidemiology and Biostatistics, Amsterdam Public Health Research Institute, VU University Medical Center, Amsterdam, the Netherlands. [134]Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany. [135]Division of Cancer Control and Population Sciences, UPMC Hillman Cancer Center, University of Pittsburgh, Pittsburgh, PA, USA. [136]Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA. [137]Laboratory of Epidemiology and Population Science, National Instiute on Aging, National Institutes of Health, Bethesda, MD, USA. [138]Geriatric Unit, Azienda USL Toscana Centro, Florence, Italy. [139]Human Genomics Laboratory, Pennington Biomedical Research Center, Baton Rouge, LA, USA. [140]Population Health Science, Bristol Medical School, NIHR Bristol Biomedical Research Center, University of Bristol, Bristol, UK. [141]Faculty of Movement and Rehabilitation Sciences, Institute of Neuroscience, UC Louvain, Louvain-la-Neuve, Belgium. [142]Department of Internal Medicine B, University Medicine Greifswald, Greifswald, Germany. [143]Brown Foundation Institute of Molecular Medicine, The University of Texas Health Science Center at Houston, Houston, TX, USA. [144]Genetics and Pharmacogenomics (GpGx), Merck Research Labs, Boston, MA, USA. [145]Department of Evolution, Ecology, and Organismal Biology, University of California, Riverside, Riverside, CA, USA. [146]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [147]MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK. [148]Australian Centre for Precision Health, Unit of Clinical and Health Sciences, University of South Australia, Adelaide, South Australia, Australia. [149]South Australian Health and Medical Research Institute, Adelaide, South Australia, Australia. [150]Population, Policy and Practice, Great Ormond Street Hospital Institute for Child Health, University College London, London, UK. [151]Department of Epidemiology and Biostatistics and HPA-MRC Center, School of Public

Health, Imperial College London, London, UK. [152]Computational Medicine, Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Berlin, Germany. [153]Epidemiology Program, University of Hawaii Cancer Center, University of Hawaii at Manoa, Honolulu, HI, USA. [154]K.G.Jebsen Center for Genetic Epidemiology, Norwegian University of Science and Technology, Trondheim, Norway. [155]Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA. [156]Center for Fertility and Health, Norwegian Institute of Public Health, Oslo, Norway. [157]Centre for Bone and Arthritis Research, Department of Internal Medicine and Clinical Nutrition, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden. [158]Department of Drug Treatment, Sahlgrenska University Hospital, Gothenburg, Sweden. [159]Department of Statistics, Seoul National University, Seoul, South Korea. [160]University of Split School of Medicine, Split, Croatia. [161]UMR 8199 – EGID, Institut Pasteur de Lille, CNRS, University of Lille, Lille, France. [162]Department of Public Health, Section of Epidemiology, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [163]MRC Integrative Epidemiology Unit, University of Bristol Medical School, University of Bristol, Bristol, UK. [164]Department of Exercise and Nutrition Sciences, Milken Institute School of Public Health, George Washington University, Washington, DC, USA. [165]Section of Geriatrics, Department of Internal Medicine, Amsterdam UMC, University of Amsterdam, Amsterdam, the Netherlands. [166]Amsterdam Public Health, Aging and Later Life, Amsterdam, the Netherlands. [167]Nuffield Department of Population Health, University of Oxford, Oxford, UK. [168]Division of Endocrinology, Department of Medicine, University of California, Los Angeles, CA, USA. [169]Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden. [170]The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [174]Deceased: Bill Musk. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: zhe.wang@mssm.edu; marcel.den_hoed@igp.uu.se

Lifelines Cohort Study

Behrooz Z. Alizadeh[75], H. Marike Boezen[75], Lude Franke[77], Morris Swertz[77], Cisca Wijmenga[77], Pim van der Harst[171], Gerjan Navis[172], Marianne Rots[173] and Bruce H. R. Wolffenbuttel[76]

[171]Department of Cardiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [172]Department of Internal Medicine, Division of Nephrology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [173]Department of Medical Biology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands.

Methods

Each study (Supplementary Table 2) obtained informed consent from participants and approval from the appropriate institutional review boards or committees.

Samples and study design. We conducted a large meta-analysis for physical activity traits, including results from up to 703,901 individuals (including nearly half-a-million from the UK Biobank) to identify genetic loci associated with physical activity and sedentary behavior across different ancestries. We first examined genome-wide, ancestry- and sex-stratified associations in 51 studies with questionnaire-based data on: (1) MVPA; (2) LST; (3) sedentary commuting behavior; and/or (4) sedentary behavior at work, using study-specific, tailored analysis plans (Supplementary Table 2, see Supplementary Note for rationale). Next, we performed ancestry-specific, inverse variance-weighted fixed-effects meta-analyses of summary statistics for each of the four self-reported traits (Fig. 1a), including data from up to 703,901 individuals consisting of European (94.0%), African (2.1%), East Asian (0.8%) and South Asian (1.3%) ancestries; as well as Hispanics (1.9%) (Supplementary Table 1). Our primary meta-analyses were restricted to 661,399 European ancestry participants. Secondary meta-analyses were also conducted for: (1) all ancestries (European + other ancestries), (2) European ancestry men, (3) European ancestry women, and (4) each non-European ancestry separately. Details of participating studies are described in Supplementary Tables 1 and 2. Although modest genomic inflation[68] was observed (lambda 1.2–1.4) (Supplementary Fig. 1), LD score regression analyses indicated this reflects true polygenic architecture rather than cryptic population structure[69].

Self-reported physical activity and sedentary behavior traits. The self-reported outcomes in this study are domain- and intensity-specific physical activity and sedentary traits that, unlike accelerometry-based outcomes, are subject to misclassification and bias by recall and awareness of the beneficial effects of physical activity, among others. Furthermore, different studies used different questionnaires to capture physical activity, and so we defined cohort-specific traits that make optimal use of the available data, while striving for consistency across studies (Supplementary Table 2). As a result, and based on the zero-inflated negative binomial nature of the distribution of MVPA in most studies, we had to analyze MVPA as a dichotomous outcome, which had a negative impact on statistical power. Descriptive information of these four outcomes is reported by study in Supplementary Table 1.

Genotyping, imputation and quality control. Detailed information about the genotyping platform used, and quality control measures applied within each study are presented in Supplementary Table 2. Quality control following study level analyses was conducted using standard procedures[70].

GWAS and meta-analyses. GWAS were performed within each study in a sex- and ancestry-specific manner. Additive genetic models accounting for family relatedness (where appropriate) were adjusted for age, age-squared, principal components reflecting population structure and additional study-specific covariates as presented in Supplementary Table 2. Analyses were limited to genotyped and imputed variants with minor allele frequency >0.1% in UK Biobank, and minor allele count >3 in other studies. Study-, sex- and ancestry-specific GWAS results were meta-analyzed using the fixed-effects, inverse variance-weighted method implemented in METAL[71], for 19.1 to 22.5 million SNPs per trait. Because we did not include a replication stage and given the high SNP density, we applied a stricter than usual Bonferroni correction and considered associations with $P < 5 \times 10^{-9}$ statistically significant[72].

To identify genome-wide significant loci, we defined a distance criterion of ±1 Mb surrounding each genome-wide significant peak ($P < 5 \times 10^{-9}$). We extracted previously reported genome-wide significant associations within 1 Mb of any index variants we identified from the NHGRI-EBI GWAS Catalog[11] and PhenoScanner V2 (ref. [73]). A locus is considered previously reported if any variant we extracted at that locus was in LD ($r^2 > 0.1$) with a lead variant that has been associated with objectively assessed or self-reported physical activity and sedentary traits previously. To identify physical activity- and sedentary behavior-associated loci that were previously associated with obesity-related traits, we performed a look up for each lead variant (and their proxies with LD $r^2 > 0.2$) in the GWAS catalog and PhenoScanner V2.

SNP-based heritability estimation. To estimate the heritability explained by genotyped SNPs for each physical activity and sedentary trait, we used BOLT-REML variance components analysis[74], a Monte Carlo average information restricted maximum likelihood algorithm implemented in the BOLT-LMM v.2.3.3 software. As in most GWAS for complex traits, the SNP heritability (up to 16%) was lower than the heritability estimates from twin studies (31%–71%)[8,9], likely at least in part due to the absence of rare variants in GWAS[75].

Although we performed a multi-ancestry meta-analysis, data from relatively few individuals of non-European ancestries were available to us, and our functional follow-up analyses were conducted based on the European ancestry results. Studies with data from more individuals of non-European ancestry will no doubt further increase the understanding of physical activity etiology.

Joint and conditional analyses. To identify additional independent signals in associated loci, we performed approximate joint and conditional SNP association analyses in each locus, using GCTA[76]. Any lead SNPs identified in known long-range high-LD regions[77] were treated as a single large locus in the GCTA analysis. We used unrelated European ancestry participants from the UK Biobank as the reference sample to acquire conditional *P* values for association.

MTAG. MTAG results were calculated using the European ancestry meta-analysis results of LST and MVPA, using standard settings[17]. Because MTAG's estimates are biased away from zero when SNPs are null for one trait but non-null for other traits, we applied it to only the two outcomes that were most strongly genetically correlated: MVPA and LST (absolute value of genetic correlation 0.49).

PheWAS with physical activity PGSs. To assess the out-of-sample predictive power of the variants associated with self-reported sedentary behavior and physical activity, we constructed two PGSs—for LST and for MVPA—in up to 23,723 Mount Sinai Bio*Me* BioBank participants, using summary statistics of the primary European ancestry meta-analyses and PRSice software[78]. We subsequently assessed the association of MVPA and BMI with the PGSs in individuals of European and African ancestry, as well as in Hispanic participants, within the Bio*Me* BioBank. Among the 2,765 European ancestry individuals with physical activity measurements and genotypes, the PGSs were calculated on common variants (minor allele frequency >1%) using *P* value thresholds from 5×10^{-8} to 1 (all variants) in the LST and MVPA GWAS, and clumping parameters of $r^2 < 0.5$ over a 250-kb window. Logistic regression models were used to examine the associations between MVPA (defined as at least 30 min per week of MVPA yes/no in Bio*Me*) and the PGSs in European ancestry participants of Bio*Me*. In each analysis, we estimated the variance in MVPA explained by the PGS, adjusting for age, sex and the top ten principal components for population structure. For both LST and MVPA, the *P* value threshold resulting in the best performing PGS was defined based on the highest R^2 increase upon adding the PGS to the regression model. To examine the generalizability of the two PGSs, we next examined their associations with MVPA in 3,206 Hispanic individuals and 2,224 African ancestry participants of Bio*Me*. We then tested each PGS for classification performance and examined whether the generated PGS was associated with any other trait by performing a PheWAS. Briefly, International Classification of Diseases 9 and 10 codes from electronic health records were mapped to phecodes using the PheWAS package[79]. Among 8,959 Bio*Me* European ancestry participants, the 1,039 disease outcomes with at least ten cases were analyzed. We used logistic regression to separately model each phecode as a function of the two PGSs, adjusting for age, age-squared, sex and the top ten principal components. Interpretation of results was restricted to outcomes with more than ten cases. Multiple testing thresholds for statistical significance were set to $P < 4.8 \times 10^{-5}$ (0.05/1,039).

Genetic correlations. To explore a possibly shared genetic architecture, we next estimated genetic correlations of the four self-reported traits examined in this study and five accelerometry-assessed physical activity traits assessed in UK Biobank[14] with relevant complex traits and diseases based on established associations at the trait level using LD score regression implemented in the LD-Hub web resource[18]. To define significance, we applied a Bonferroni correction for the 108 selected phenotypes available on LD-Hub ($P < 4.6 \times 10^{-4}$). Supplementary Table 10 shows the complete set of pairwise genetic correlations of the four self-reported physical activity traits with relevant complex traits and diseases. Next, we prioritized traits and diseases showing evidence of genetic overlap (associated with at least one of the physical activity traits). These can be divided into six categories: lifestyle traits, anthropometric traits, psychiatric diseases, other diseases (cardiometabolic diseases and cancer), biomarkers and others (Fig. 4). Using objectively assessed physical activity traits (accelerometry) instead of self-reported traits yielded similar results (Supplementary Fig. 2).

Two-sample MR. We performed MR analyses to disentangle the causality between LST and MVPA, on the one hand, and BMI, on the other hand. We further investigated the causal effects of LST and MVPA on common diseases and risk factors, while considering BMI through multivariable MR. For multivariable MR, we used BMI (exposure 2) summary statistics based on UK Biobank data, and summary statistics for disease outcomes and other relevant traits based on data from the largest publicly available GWAS without data from UK Biobank participants on the MR-Base platform and OpenGWAS database[80,81]. This way, we aimed to minimize bias due to sample overlap in the two-sample MR analysis[82]. The source of each of the instruments is presented in Supplementary Table 12. Genetic instrumental variables for each of the traits and diseases consisted of genome-wide significant ($P < 5 \times 10^{-8}$) index SNPs. Index SNPs were LD clumped ($r^2 > 0.001$ within a 10-Mb window) to remove any correlated variants. In the multivariable MR that evaluates the independent effects of each risk factor, the genetic instrumental variables from two risk factors were combined. For both LST and MVPA, independent loci associated with physical activity or BMI were used as instrumental variables.

We followed several steps to evaluate potential causality. Because MR results can be severely biased if instrumental SNPs show horizontal pleiotropy and violate the instrumental variable assumptions[28], we prioritized methods that are robust

to horizontal pleiotropy when calculating causal estimates. We did not use the MR-Egger intercept test to identify the presence of potential pleiotropy, because the MR-Egger intercept parameter estimate is positively biased when the NO Measurement Error assumption is violated, as indicated by lower values of I^2_{GX} in our two-sample MR setting[83]. Instead, we applied MR-PRESSO (pleiotropy residual sum and outlier)[27], which removes pleiotropy by identifying and discarding influential outlier predictors from the standard inverse variance-weighted test[28]. For analyses with evidence of no distortion due to pleiotropy (MR-PRESSO Global test $P > 0.05$), we considered other robust methods, for instance fixed- and random-effect inverse variance-weighted, weighted- or simple- median and mode methods. We also conducted Steiger filtering to remove variants likely influenced by reverse causation and used Cook's distance filtering to remove outlying heterogeneous variants as deemed necessary. To select the most appropriate approach, we implemented a machine learning framework[30]. Finally, we performed a leave-one-out analysis to identify potential outliers among the variants included in the instrumental variables tested. We set the multiple testing significance threshold for MR analyses with disease outcomes at 1.9×10^{-3}, that is, Bonferroni correction for 13 disease outcomes and 2 types of risk factors: physical activity or sedentary behavior and adiposity ($0.05/(13 \times 2)$).

We also applied the recently published Bayesian-based MR method CAUSE, which accounts for both correlated and uncorrelated pleiotropy[26], in evaluating bidirectional causal effects between physical activity and adiposity. Compared with the other two-sample MR methods, CAUSE calculates the posterior probabilities of the causal effect and the shared effect, and tests whether the causal model fits the data better than the sharing model. That is, it examines whether the association between the traits is more likely to be explained by causality than horizontal pleiotropy. In addition, CAUSE improves the power of MR analysis by using full genome-wide summary results (LD pruned at $r^2 < 0.1$ with $P < 1 \times 10^{-3}$, as recommended by the CAUSE authors). In addition, we took advantage of the robustness of the CAUSE method—which allows overlapping GWAS samples—to test the assumption that a genetic predisposition for LST assessed later in life reflects a lifetime liability. Using the summary statistics of SNPs for childhood adiposity (comparative body size at age 10) and height (comparative height at age 10) in UK Biobank[84], we examined bidirectional causal effects between LST and these two recalled childhood traits.

Enrichment for genes with altered expression in skeletal muscle after an intervention. A high degree of physical fitness and a strong adaptive response to exercise interventions facilitate a physically active lifestyle. To identify plausible candidate genes in GWAS-identified loci, we examined enrichment for transcripts whose expression in skeletal muscle was changed after an acute bout of aerobic exercise, aerobic training, an acute bout of resistance exercise, resistance training and inactivity[33]. We excluded individuals with pre-existing conditions such as chronic kidney disease, chronic obstructive pulmonary disease, frailty, metabolic syndromes and obesity. We also excluded athletes because in this subgroup, transcripts with differential expression in response to (in)activity interventions are likely not representative for the general population[85]. Enrichment was examined for genes nearest to, or within 1 Mb of lead variants for LST- and MVPA-associated loci. We used false discovery rate <0.01 as the threshold for altered expression after intervention. A sensitivity analysis with a series of different false discovery rate cut-offs (0.001 to 0.5) showed that results were robust.

Gene, tissue and cell-type prioritization. We used DEPICT[40] to identify enriched gene sets and tissues, as well as to prioritize candidate genes in the identified loci, using variants with $P < 1 \times 10^{-5}$ in the primary meta-analysis of European ancestry men and women combined as input. We also used CELLECT[43] to identify enriched cell types for physical activity, by combining MVPA and LST GWAS summary statistics with single-cell RNA sequencing data. We sought to further refine the set of prioritized candidate genes using SMR and HEIDI tests[46]. Briefly, this approach integrates summary-level data from GWAS and expression quantitative trait loci (eQTL) studies to test whether a transcript and phenotype are likely associated because of a shared causal variant (pleiotropy). We considered genes candidates if they had a Bonferroni-corrected $P_{SMR} < 1.02 \times 10^{-5}$ and showed no evidence of heterogeneity ($P_{HEIDI} > 0.05$), as in earlier studies[46]. Based on tissue enrichment results from DEPICT, the SMR analyses were performed using brain eQTL information obtained from GTEx-brain ($n = 72$)[86,87], CommonMind Consortium ($n = 467$)[88], ROSMAP ($n = 494$)[89], and Brain-eMeta ($n = 1{,}194$)[87]; blood eQTL summary information obtained from the eQTLGen Consortium[90], which is based on peripheral blood samples from 31,684 individuals; and skeletal muscle eQTL information from the GTEx project ($n = 803$)[91].

To identify variants in GWAS-identified loci with a high posterior probability of being causal, we used LST and MVPA summary statistics as input for FINEMAP[47]. We used default parameters and selected a maximum of ten putative causal variants per locus. The output variants identified as credible were mapped to genes using tissue-specific HiC chromatin conformation capture data[92]. We integrated all HiC data in the brain (dorsolateral prefrontal cortex, hippocampus, neural progenitor cell, and adult and fetal cortex) available on FUMA v.1.3.5, using the same approach. Genes in GWAS-identified loci containing FINEMAP-identified credible coding variants with a CADD score >12.33 were also prioritized. Finally, we used data from 26 of the 131 available tissues and cell types deemed relevant for sedentary behavior and physical activity (Supplementary Table 20) to identify genes that are contacted by enhancers affected by causal variants flagged by GWAS lead SNPs, using the recently described activity-by-contact model[49].

Enrichment for previously reported candidate genes. We next conducted a literature review of previously reported genes with evidence of a role in exercise (physical activity behavior) and fitness (physical activity ability) and identified 58 such candidate genes (13 for exercise; 45 for fitness)[12,50–53]. For each gene, we identified all variants within the gene, examined their associations with LST and MVPA in our meta-analysis of European ancestry individuals and, for each gene–trait combination, retained the summary statistics for the variant with the lowest P value for association. Variants in three genes reached the traditional threshold for genome-wide significance (*PPARD*, *APOE* and *ACTN3*). Based on LD and predicted effects on protein function, rs2229456 in *ACTN3* (encoding p.Glu635Ala) may have a causal effect.

MD simulation for p.Glu635Ala. Because no structure for human ACTN3 has yet been experimentally determined, we constructed a homology model of the p.Glu635 variant monomeric filament using the fully annotated protein (UniProt ID Q08043) using Phyre2 (ref. [93]), with the p.635Ala variant mutated in silico. Residue 635 of ACTN3 resides in the 356th residue of the spectrin repeat region and corresponds with residue 628 in ACTN2 (see the Supplementary Methods for more information). For each variant, the spectrin repeats of the ACTN3 monomer were aligned with the crystal structure of the rod domain of alpha-actinin (PDB ID 1HCI), to give the dimeric form of ACTN3. MD system preparation and simulation was conducted with GROMACS 2020.1 (ref. [94]) and using mdanalysis v.2.0. The MD topology was created with GROMACS pdb2gmx using the ACTN2 and ACTN3 dimer models and parameterized with the CHARMM36 all-atom force field[95]. The ACTN2 and ACTN3 dimers were placed in a rectangular simulation box with a 1.0-nm buffer between the protein and the box extent, with periodic boundary conditions in all three spatial axes. The system was solvated with TIP3P water molecules and using GROMACS genion, random solvent molecules were replaced with K^+ and Cl^- to a concentration of 150 mM with additional K^+ ions added to provide an electrostatically neutral system. Energy minimization was accomplished using the steepest descent algorithm. To equilibrate the system, two 100-ps simulations were conducted using a constant temperature ensemble (NVT, that is, a constant number of particles [N], volume [V] and temperature [T]) at 310 K via a Berendsen thermostat, followed by a constant pressure ensemble (NPT, that is, a constant number of particles [N], pressure [P] and temperature [T]) at 1 bar with a Parinello–Rahman barostat. MD simulation parameters were set in accordance with the recommendations for the CHARMM36 force field in GROMACS. A short production run of 1 ns without position restraints was followed by a full simulation of 150 ns with weak position restraints on the ABD of chain B to prevent self-interaction across the periodic boundaries.

Steered MD and umbrella sampling for p.Glu635Ala. We next compared the properties of ACTN2 and of ACTN3 p.635Ala and p.Glu635 when placed under the simulated compressive loads that are likely experienced in vivo. The final frame of the 1-ns MD production run was used as the starting topology for steered MD simulations using fully relaxed dimers. Steered MD simulations were run for 2 ns with a pulling rate of 0.005 nm ps^{-1} and a harmonic potential of 50 kJ mol^{-1} nm^{-2}. Center-of-mass pull groups were defined as the ABD of each respective monomer, with a weak position restraint placed on the Cα atom of threonine 52 (ACTN3) or threonine 45 (ACTN2)—a centrally located residue in the core of the ABD—on one ABD, enabling full rotational freedom of each ABD during the course of the steered MD simulations. The pulling vector was oriented along the axis on which the spectrin repeats were initially aligned. Suitable frames from each steered MD simulation were selected that differed by no more than 0.2 nm from 0 to −5.5 nm (a contraction of the dimer by 5.5 nm or ~18%) and were used as the starting topology for a series of 10-ns umbrella sampling simulations. Analysis of the umbrella sampling simulations was conducted using g_wham, to yield the potential of mean force versus reaction coordinate for each variant.

Single skeletal muscle fiber functional characteristics in relation to p.Glu635Ala. Single muscle fibers from eight nonathletic young men in which contractile and morphological properties were previously characterized in vastus lateralis biopsies obtained before and after an eccentric exercise bout[60,61] were genotyped for rs2229456. A hierarchical linear mixed effects model was constructed for each fiber type and time point using rstanarm[96] to test the genotype fixed effect, with muscle fibers nested within each of the eight individuals as random factors for each contractile and morphological variable. Genotypes at p.Arg577Ter and p.Glu635Ala were clustered into three groups: RR-AA ($n = 1$ individual, 46 fibers, reference group); RR-AC ($n = 3$ individuals, 32 ± 5 fibers); and XX-AA ($n = 4$ individuals, 39 ± 6 fibers). Using weakly informative priors, the posterior distribution was estimated with Markov chain Monte Carlo sampling (20,000 samples total with 5,000 sample burn-in). We calculated 90% credible

intervals of the posterior density and distribution-free overlapping indices[97] to compare single fiber properties between genotypes.

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

Data availability

European and multi-ancestry meta-analyses summary statistics for the genome-wide association study are available through the NHGRI-EBI GWAS Catalog (https://www.ebi.ac.uk/gwas/downloads/summary-statistics, GCP ID: GCP000358). UK Biobank individual-level data can be obtained through a data access application available at https://www.ukbiobank.ac.uk/. In this study we made use of data made available by: MetaMex https://www.metamex.eu/; Tabula Muris https://www.czbiohub.org/tabula-muris/; Open GWAS https://gwas.mrcieu.ac.uk/; MR Base https://www.mrbase.org/; GTEx Consortium https://gtexportal.org/home/; eQTLGen Consortium https://www.eqtlgen.org/; CommonMind Consortium https://www.synapse.org/#!Synapse:syn2759792/wiki/69613; Brain zQTLServe http://mostafavilab.stat.ubc.ca/xqtl/; MetaBrain https://www.metabrain.nl/.

Code availability

We made use of publicly available software and tools such as METAL (https://genome.sph.umich.edu/wiki/METAL), GCTA (https://yanglab.westlake.edu.cn/software/gcta/), LD score regression (https://github.com/bulik/ldsc), SMR (https://cnsgenomics.com/software/smr/) and PLINK (www.cog-genomics.org/plink/).

References

68. Yang, J. et al. Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* **19**, 807–812 (2011).
69. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291 (2015).
70. Winkler, T. W. et al. Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9**, 1192 (2014).
71. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
72. Pulit, S. L., de With, S. A. & de Bakker, P. I. Resetting the bar: statistical significance in whole-genome sequencing-based association studies of global populations. *Genet. Epidemiol.* **41**, 145–151 (2017).
73. Kamat, M. A. et al. PhenoScanner V2: an expanded tool for searching human genotype–phenotype associations. *Bioinformatics* **5**, 4851–4853 (2019).
74. Loh, P. R. et al. Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat. Genet.* **47**, 1385–1392 (2015).
75. Wainschtein, P. et al. Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data. *Nat. Genet.* **54**, 263–273 (2022).
76. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).
77. Price, A. L. et al. Long-range LD can confound genome scans in admixed populations. *Am. J. Hum. Genet.* **83**, 132–139 (2008).
78. Choi, S. W. & O'Reilly, P. F. PRSice-2: Polygenic Risk Score software for biobank-scale data. *GigaScience* **8**, giz082 (2019).
79. Carroll, R. J., Bastarache, L. & Denny, J. C. R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. *Bioinformatics* **30**, 2375–2376 (2014).
80. Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **7**, e34408 (2018).
81. Elsworth, B. et al. The MRC IEU OpenGWAS data infrastructure. Preprint at *bioRxiv* https://doi.org/10.1101/2020.08.10.244293 (2020).
82. Burgess, S., Davies, N. M. & Thompson, S. G. Bias due to participant overlap in two-sample Mendelian randomization. *Genet. Epidemiol.* **40**, 597–608 (2016).
83. Bowden, J. et al. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I_2 statistic. *Int. J. Epidemiol.* **45**, 1961–1974 (2016).
84. Lyon, M. S. et al. The variant call format provides efficient and robust storage of GWAS summary statistics. *Genome Biol.* **22**, 32 (2021).
85. Koch, L. G. et al. Test of the principle of initial value in rat genetic models of exercise capacity. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **288**, R466–R472 (2005).
86. Battle, A., Brown, C. D., Engelhardt, B. E. & Montgomery, S. B. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
87. Qi, T. et al. Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat. Commun.* **9**, 2282–2282 (2018).
88. Fromer, M. et al. Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* **19**, 1442–1453 (2016).
89. Ng, B. et al. An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat. Neurosci.* **20**, 1418–1426 (2017).
90. Võsa, U. et al. Large-scale *cis*- and *trans*-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat. Genet.* **53**, 1300–1310 (2021).
91. Barbeira, A. N. et al. Exploiting the GTEx resources to decipher the mechanisms at GWAS loci. *Genome Biol.* **22**, 49 (2021).
92. Belton, J. M. et al. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268–276 (2012).
93. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
94. Abraham, M. J. et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
95. Huang, J. & MacKerell, A. D. Jr. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145 (2013).
96. Goodrich, B., Gabry, J., Ali, I. & Brilleman, S. rstanarm: Bayesian applied regression modeling via Stan. R package version 2.14.1 https://mc-stan.org/rstanarm (2016).
97. Pastore, M. & Calcagnì, A. Measuring distribution similarities between samples: a distribution-free overlapping index. *Front. Psychol.* **10**, 1089 (2019).

Acknowledgements

T.S.A. is supported by the Steno Diabetes Center Copenhagen, Copenhagen, Denmark and the Novo Nordisk Foundation Grant NNF18OC0052457. J.W.C. was supported by grants from the National Institutes of Health (NIH) (R01-NS100178; R01-NS105150), the US Department of Veterans Affairs and the American Heart Association (AHA) (15GPSPG23770000; 17IBDG33700328). B.F. was supported by the Oak Foundation. T.O.K. was supported by the Novo Nordisk Foundation (grant numbers NNF17OC0026848 and NNF18CC0034900). N.G.M. is funded by a National Health and Medical Research Council (NHMRC) Investigator Grant (APP1172990). S.E.M. is funded by NHMRC Investigator Grant (APP1172917). D.M. is supported by a Canada Research Chair in Genetics of Obesity. R.C.R. is a de Pass Vice Chancellor's Research Fellow at the University of Bristol. S.R.-d.-P was supported by the Heart and Stroke Foundation of Ontario (grant number NA 7293). N.J.S. holds a National Institute for Health and Care Research (NIHR) Senior Investigator award. N.J.T. is a Wellcome Trust (WT) Investigator (202802/Z/16/Z), is the principal investigator of the Avon Longitudinal Study of Parents and Children (Medical Research Council (MRC) & WT 217065/Z/19/Z), is supported by the University of Bristol NIHR Biomedical Research Centre (BRC-1215-2001), the MRC Integrative Epidemiology Unit (MC_UU_00011) and works within the Cancer Research UK Integrative Cancer Epidemiology Programme (C18281/A19169). X.Z. is supported by China Scholarship Council 201406220101. T.P. is supported by the Bio-Synergy Research Project (2013M3A9C4078158) of the Ministry of Science, ICT and Future Planning through the National Research Foundation of Korea. S.S. is supported by the Swedish Research Council (grant numbers 2016-06264, 2018-05946 and 2018-05498). J.W.C. was partially supported by an AHA-Bayer Discovery Grant (grant 17IBDG33700328), the AHA Cardiovascular Genome–Phenome Study (grant-15GPSPG23770000), NIH (grants R01-NS114045, R01-NS100178, R01-NS105150), and the US Department of Veterans Affairs. H.X. was supported by AHA grant 19CDA34760258 and NIH grants R01-NS114045, R01-NS100178 and R01-NS105150. K.E.N. is funded by AHA grants 13GRNT16490017 and 15GRNT25880008, and by NIH grants R01DK089256, R01DK101855, R01HD057194, R01DK122503, 01HG010297, R01HL142302, R01HL143885 and R01HG009974. L.F.-R. is supported by an AHA grant (13PRE16100015). C.P.N. is funded by the British Heart Foundation (SP/16/4/32697). L.M.H., C.P.N., P.S.B. and N.J.S. are supported by the NIHR Leicester Cardiovascular Biomedical Research Centre (BRC-1215-20010). T.G. Jr and D.H. were supported by US National Science Foundation grant IOS-2038528. R.J.F.L. is supported by the NIH (R01DK110113, R01DK075787, R01DK107786, R01HL142302, R01HG010297, R01DK124097, R01HL151152). M.d.H. is a fellow of the Swedish Heart–Lung Foundation (20170872, 20200781) and a Kjell and Märta Beijer Foundation researcher. He is further supported by project grants from the Swedish Heart–Lung Foundation (20140543, 20170678, 20180706, 20200602) and the Swedish Research Council (2015-03657, 2019-01417).

Author contributions

M.C., R.J.F.L. and M.d.H. conceived the study. M.C., S. Brage, U.E., R.J.F.L. and M.d.H. conceived trait definitions. M.C. and M.d.H. generated study-specific analysis plans. Z.W. and M.d.H. performed GWAS QC and meta-analyses. Z.W. performed SNP-based heritability estimations. Z.W. performed joint and conditional analyses. Z.W. performed PheWAS with physical activity polygenic scores. Z.W. performed the genetic correlation analysis. Z.W. performed the Mendelian randomization analysis. N.J.P., A.K., J.R.Z. and Z.W. analyzed enrichment of altered gene expression in skeletal muscle following intervention. D.H. performed biological annotation using DEPICT, SMR and FINEMAP.

A.M.J. and T.P. performed cell-type enrichment using CELLECT. Z.W. and M.d.H. integrated results across gene prioritization approaches. T.M. and A.L.H. performed the GWAS for physical activity in 100 mouse strains. M.d.H. compared GWAS results across mice and humans. C.M. and A.A. explored expression of a 4930413E15Rik ortholog in humans. M.d.H. examined overlap between physical activity loci and loci showing evidence of selection. Z.W. performed the candidate gene analysis for exercise and fitness. E.M. constructed a homology model of the ACTN3 p.Glu635 variant. A.E. performed (steered) molecular dynamics simulations and umbrella sampling for p.Glu635Ala. S. Broos, L.D. and M.A.I.T. performed single skeletal muscle fiber experiments. A. Pacolet, M.V. and M.A.I.T. performed de novo genotyping for functional characterization of p.Glu635Ala. A.E. analyzed the isolated skeletal muscle fiber data. Z.W., A.E., R.J.F.L. and M.d.H. wrote the manuscript. T.S.A., T.M.B., A.R.B., L.F.B., M.C., A.Y.C., D.B., R.D., N.D.D., K.E., B.F., M.F.F., C.G., M.G., L.M.H., T. Haller, F.P.H., D.A.H., V.H., N.H.-C., C. Holzapfel, A.U.J., Å.J., M.A.K., R.K., K.F.K., B.K., C.M.K., Z.K., V.L., P.A.L., M. Lorentzon, L.-P.L., M. Mangino, C.M., K.R.M., A. Pacolet, L.P., R.P., R.C.R., N.V.R., S.R.-d.-P., K.E.S., C.-A.S., H.M.S., T.T., A. Tremblay, A. Tönjes, C.T., P.J.v.d.M., M.V., F.J.A.v.R., J.V.v.V.-O., X.Z., J.-H.Z., W.Z., Z.B., M.N.B.-H., S.E.B., J. Beilby, J. Blangero, D.I.B., S. Brage, P.S.B., J.A.B., M. Boehnke, C.L., J.W.C., F.S.C., L.A.C., T.E., S.E., J.D.F., L.F.-R., A.E.F., O.H.F., T.E.G., S.D.G., N.G., C.A.H., G.H., J.H., T.I., R.J., A.J., P.K.J., T.J., M.K., T.O.K., W.-P.K., I.K., P.P.K., J.K., L.J.L., A. Li, A. Linneberg, J.L., P.M.V., S.E.M., Y.M., A. Moscati, B.M., C.P.N., I.M.N., N.L.P., A. Peters, P.A.P., C.P., O.T.R., M.R., A.P.R., P.M.R., I.R., K.R., M.A.S., L.J.S., R.A.S., S.S., K.S., A.V.S., J.A.S., E.S., M.S., E.S.T., K.K.T., B.T., A. Tönjes, A. Tremblay, A.G.U., J.V., N.v.S., U.V., G. Waeber, K.W., Q.W., H.X., K.L.Y., J.M.Y., M.C.Z., A.B.Z., A.A., S. Broos, J.C.B., M. Bruinenberg, C.B., D.I.C., G.D.S., E.J.C.d.G., L.D., M.D., M.K.E., L.F., M.F., C.F., T.G., V.G., U.G., T. Hansen, C. Hayward, B.L.H., E.H., M.-R.J., W.C.J., S.L.R.K., L.A.K., M. Laakso, C.L., T.L., L.L.M., P.K.E.M., N.G.M., M. Melbye, A. Metspalu, D.M., K.E.N., C.O., A.J.O., M.O.-M., G.P., T.P., O. Pedersen, B.W.J.H.P., O. Polasek, I.P., C.N.R., N.J.S., X.S., H.S., T.I.A.S., T.D.S., N.J.T., R.M.v.D., N.v.d.V., C.M.v.D., P.V., H.V., T.V., G. Willemsen, N.J.W., D.R.W., H.-E.W., and J.F.W. contributed cohort study level results (concept, design, data acquisition, QC, and analysis. Z.W., A.E., N.J.P., T.M., D.H., M.C.C., E.M., S. Brage, T.S.A., T.M.B., A.R.B., L.F.B., M.C., A.Y.C., D.B., R.D., N.D.D., K.E., B.F., M.F.F., C.G., M.G., L.M.H., T. Haller, F.P.H., D.A.H., V.H., N.H.-C., C. Holzapfel, A.U.J., Å.J., A.M.J., M.A.K., R.K., K.F.K., B.K., C.M.K., Z.K., V.L., P.A.L., M. Lorentzon, L.-P.L., M. Mangino, C.M., K.R.M., A. Pacolet, L.P., R.P., R.C.R., N.V.R., S.R.-d.-P., K.E.S., C.-A.S., H.M.S., T.T., A. Tremblay, A. Tönjes, C.T., P.J.v.d.M., M.V., F.J.A.v.R., J.V.v.V.-O., X.Z., J.-H.Z., W.Z., Z.B., M.N.B.-H., S.E.B., J. Beilby, J. Blangero, D.I.B., P.S.B., J.A.B., M. Boehnke, U.E., C.-T.L., J.W.C., F.S.C., L.A.C., T.E., S.E., J.D.F., L.F.-R., A.E.F., O.H.F., T.E.G., S.D.G., N.G., C.A.H., G.H., J.H., T.I., R.J., A.J., P.K.J., T.J., M.K., T.O.K., W.-P.K., I.K., P.P.K., J.K., L.J.L., A. Li, A. Linneberg, J.L., P.M.V., S.E.M., Y.M., A. Moscati, B.M., C.P.N., I.M.N., N.L.P., A. Peters, P.A.P., C.P., O.T.R., M.R., A.P.R., P.M.R., I.R., K.R., M.A.S., L.J.S., R.A.S., S.S., K.S., A.V.S., J.A.S., E.S., M.S., E.S.T., K.K.T., B.T., A. Tönjes, A. Tremblay, A.G.U., J.V., N.v.S., U.V., G. Waeber, K.W., Q.W., H.X., K.L.Y., J.M.Y., M.C.Z., A.B.Z., A.A., S. Broos, J.C.B., M. Bruinenberg, C.B., D.I.C., G.D.S., E.J.C.d.G., L.D., M.D., M.K.E., L.F., M.F., C.F., T.G., V.G., U.G., T. Hansen, C. Hayward, B.L.H., E.H., M.-R.J., W.C.J., S.L.R.K., L.A.K., M. Laakso, C.-T.L., T.L., L.L.M., P.K.E.M., N.G.M., M. Melbye, A. Metspalu, D.M., K.E.N., C.O., A.J.O., M.O.-M., G.P., T.P., O. Pedersen, B.W.J.H.P., T.H.P., O. Polasek, I.P., C.N.R., N.J.S., X.S., H.S., T.I.A.S., T.D.S., N.J.T., R.M.v.D., N.v.d.V., C.M.v.D., P.V., H.V., T.V., G. Willemsen, N.J.W., D.R.W., H.-E.W., J.F.W., A.L.H., A.K., J.R.Z., M.A.I.T., R.J.F.L. and M.d.H. provided input and feedback to the final manuscript.

Funding

Open access funding provided by Uppsala University

Competing interests

C.F. is Vice President and Head at Genetics and Pharmacogenomics, Merck labs. M. Lorentzon has received lecture or consulting fees from Amgen, Lilly, UCB Pharma, Radius Health, Meda, GE-Lunar and Santax Medico/Hologic. P.V. received an unrestricted grant from GlaxoSmithKline to build the CoLaus study. These authors played a role in individual studies that contributed to the meta-analysis, but not to the meta-analysis of GWAS studies, downstream experiments and analyses, or interpretation of the data. Hence, it is highly unlikely to have influenced the results of this study. The remaining authors declare no competing interests.

Additional information

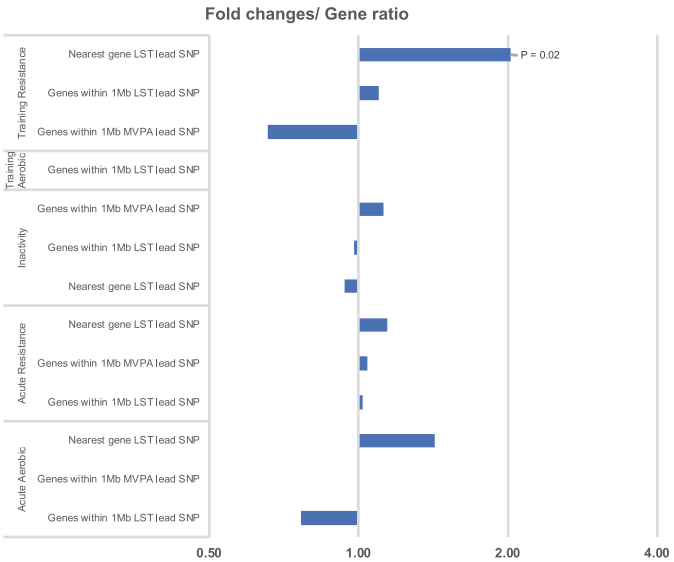
Extended data is available for this paper at https://doi.org/10.1038/s41588-022-01165-1.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-022-01165-1.

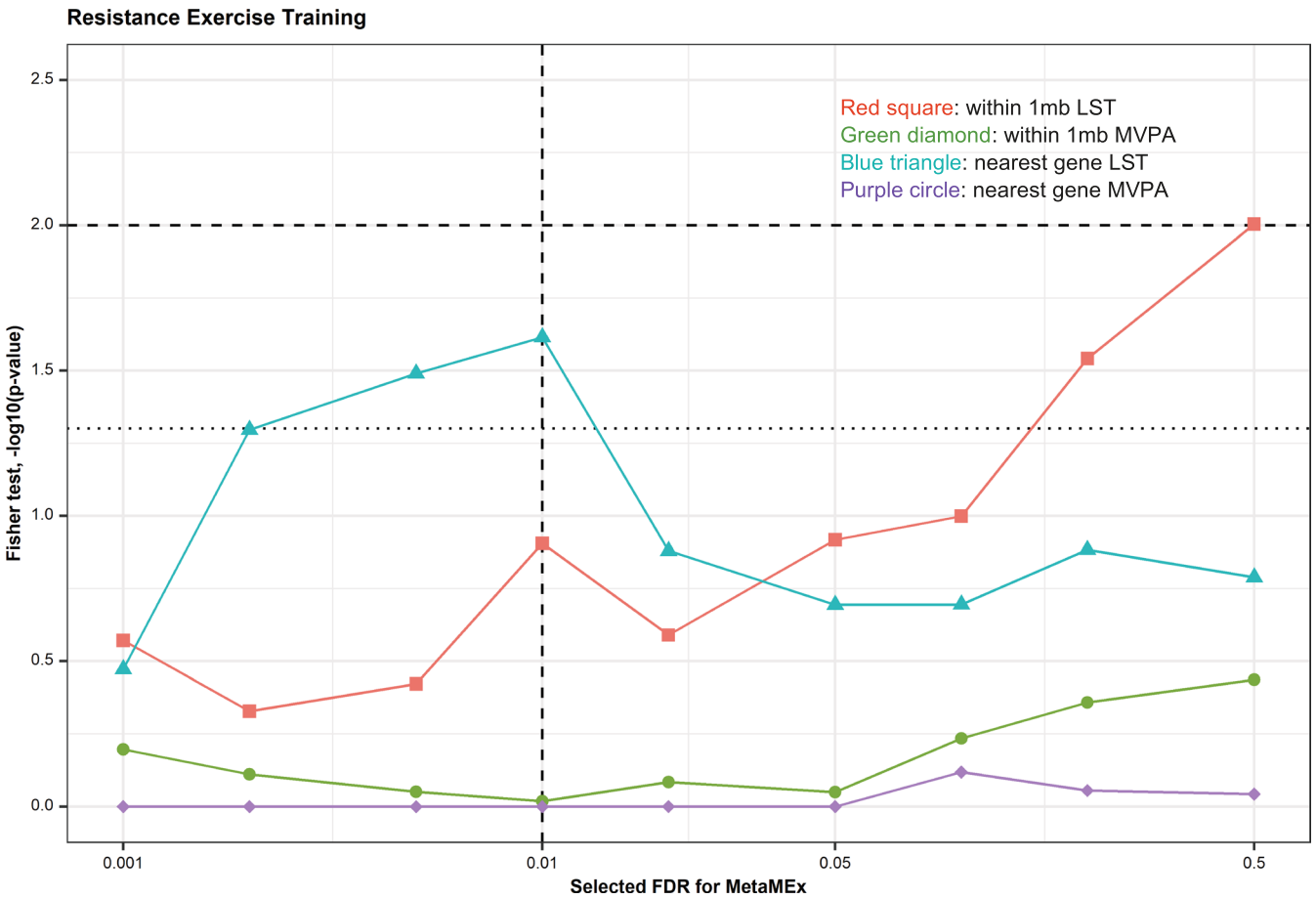
Correspondence and requests for materials should be addressed to Zhe Wang or Marcel den Hoed.

Peer review information *Nature Genetics* thanks Yann Klimentidis and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

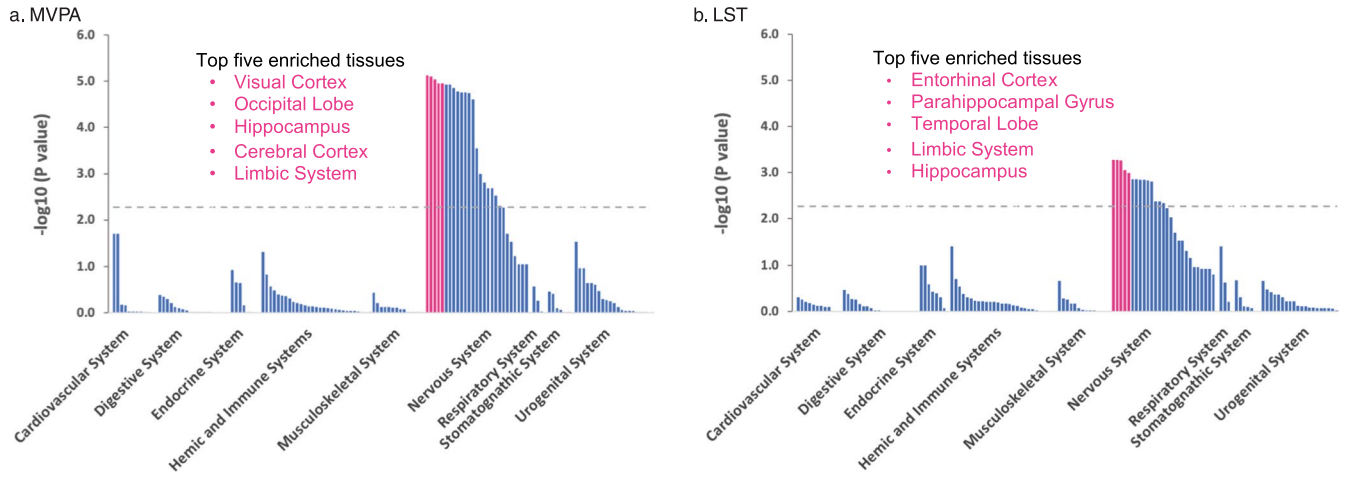
Reprints and permissions information is available at www.nature.com/reprints.



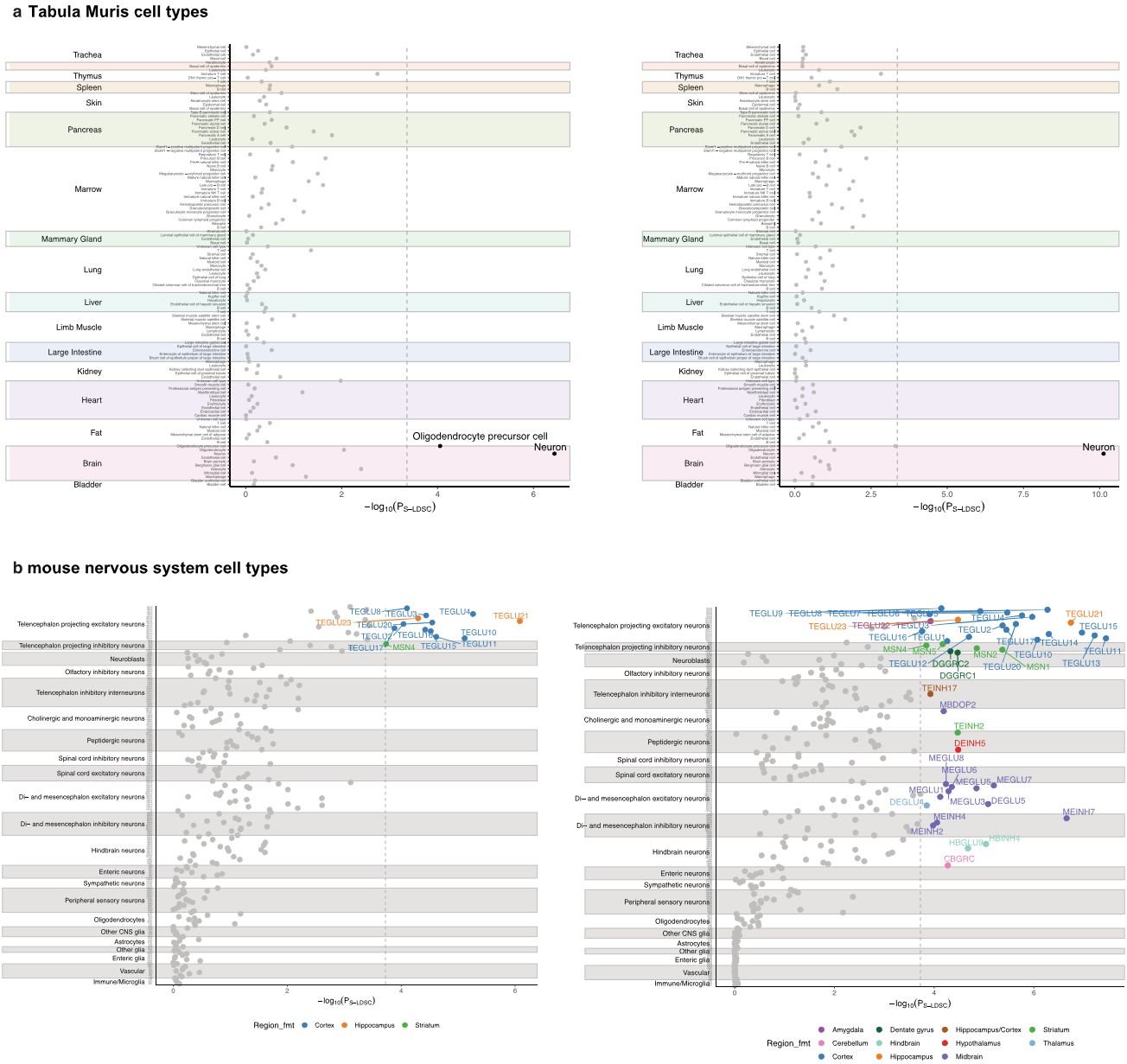
Extended Data Fig. 1 | LST-associated loci are enriched for genes with altered expression in skeletal muscle following resistance training. Fold-change plot in log scale for the ratio between: (1) the proportion of genes in physical activity-associated loci that showed an altered expression in skeletal muscle (FDR < 0.01) across five categories: inactivity, acute bout of resistance exercise, acute bout of aerobic exercise, resistance training, or aerobic training; and (2) the proportion of all genes that showed an altered expression following such (in)activity in the MetaMex database (PMID: 31980607). Tested loci were MVPA or LST-associated loci. In a given set of loci, we either considered only the genes nearest to the lead SNP, or all genes within 1 Mb of the lead SNP. Only loci harboring at least five genes with altered gene expression levels after intervention were included in this figure. A one-sided Fisher exact test was used to calculate the *P*-value for enrichment.



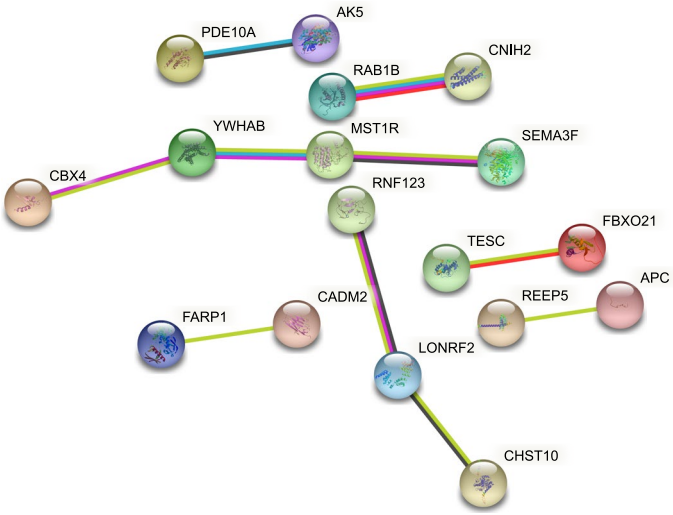
Extended Data Fig. 2 | A sensitivity analysis shows the analysis of altered gene expression following resistance training is robust to FDR threshold. We examined the effect of different FDR thresholds on Fisher's exact test results for the enrichment analysis of alteration in gene expression in skeletal muscle following resistance training. Red square, genes within 1 Mb of the LST lead SNP; green circle, genes within 1 Mb of the MVPA lead SNP; blue triangle, nearest gene LST lead SNP; purple diamond, nearest gene MVPA lead SNP. The horizonal dotted line indicates nominal significance level ($P < 0.05$), and the vertical dashed line indicates the FDR threshold that was used. FDR thresholds explored range from 0.001 to 0.5.



Extended Data Fig. 3 | DEPICT-derived tissue enrichment of MVPA and LST. a, MVPA. **b**, LST. SNPs with $P < 1 \times 10^{-5}$ for association in the European ancestry GWAS of men and women combined were used as input. The dashed line indicates the FDR corrected significance threshold (FDR < 0.05).



Extended Data Fig. 4 | Cell type prioritization using CELLECT for MVPA and LST. a, Prioritization of 115 Tabula Muris cell types across 19 tissues identified two cell types from the brain as significantly associated (stratified linkage disequilibrium score regression) with MVPA (left) and LST (right), namely oligodendrocyte precursor cells and neurons (shown in black; Bonferroni-corrected significance threshold, $P < 0.05/115$). **b**, Prioritization of 265 mouse nervous system cell types identified 13 and 45 cell types from 12 distinct brain regions as significantly associated (stratified linkage disequilibrium score regression) with MVPA and LST, respectively (highlighted; Bonferroni-corrected significance threshold, $P < 0.05/265$.



Extended Data Fig. 5 | Protein-protein interactions involving 17 of the 46 candidate genes in GWAS-identified loci prioritized by at least two approaches. Protein-protein interactions were visualized using String. *LONRF2* and *CHST10* were prioritized in loci associated with MVPA; the remaining genes were prioritized in loci associated with LST.

nature portfolio

Corresponding author(s): Marcel den Hoed, Wang Zhe

Last updated by author(s): June 8, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Genotype calling and imputation were performed by 51 contributing studies using software specified for each study in Supp Table 2. Genetic association analyses were performed by 51 contributing studies using software specified for each study in Supp Table 2.
Data analysis	GWAS in UK Biobank were conducted using BOLT-LMM v2.3.2. GWAS meta-analyses were performed using METAL, version 2017-12-21. Variants' effects on protein function were predicted using Ensembl's Variant Effect Predictor (VEP) Gene-based burden and SKAT tests were performed using the GENESIS package version 3.15. Previously reported summary statistics were extracted from the GWAS catalog, PhenoScanner V2, MRC IEU OpenGWAS. Chip-based heritability was quantified using the BOLT-LMM v2.3.3 software. Joint and conditional SNP association analyses were performed using GCTA version 1.92. Polygenic score analyses were performed using PRSice software v2. Genome-wide genetic correlations were examined using LS score regression implemented in the LD-hub web resource. Mendelian randomisation analyses were performed using R packages TwoSampleMR, MVMR, MR-PRESSO, CAUSE V1.2.0. Candidate genes and enriched tissues were identified using DEPICT version 1 release 194. Enriched cell types were identified using CELLECT. Variants with a high posterior probability of being causal were identified using FINEMAP v 1.4.1. Relevant HiC data in the brain were extracted using FUMA v1.3.5. VitalView Activity Software v 1.5 to calculate wheel running revolutions in mice. The UCSC genome-browser and GTEx IGV browser were used to explore tissue-specific expression of the human orthologue of the mouse gene 4930413E15Rik. A homology model of the E635 variant monomeric filament was generated using Phyre2 v 2.0. Molecular dynamics (MD) system preparation and simulation was conducted with GROMACS 2020 and mdanalysis v 2.0. The MD topology was created with GROMACS pdb2gmx using the ACTN3 dimer model and parameterized with the CHARMM36 all-atom force

field.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

European and multi-ancestry meta-analyses summary statistics for the genome-wide association study are available through the NHGRI-EBI GWAS Catalog (https://www.ebi.ac.uk/gwas/downloads/summary-statistics).
UK Biobank individual-level data can be obtained through a data access application available at https://www.ukbiobank.ac.uk/.
In this study we made use of data made available by: MetaMex: https://www.metamex.eu/; Tabula muris: https://www.czbiohub.org/tabula-muris/; Open GWAS: https://gwas.mrcieu.ac.uk/; MR Base: https://www.mrbase.org/; GTEx Consortium: https://gtexportal.org/home/; eQTLGen Consortium: https://www.eqtlgen.org/; CommonMind Consortium: https://www.synapse.org/#!Synapse:syn2759792/wiki/69613; Brain zQTLServe: http://mostafavilab.stat.ubc.ca/xqtl/; MetaBrain: https://www.metabrain.nl/

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	To ensure the highest possible statistical power to identify robust genetic associations with physical activity and sedentary traits, we aimed to include data from any and all studies with relevant data that were willing and able to participate. This resulted in a sample size of up to 674,980 individuals from 51 studies. For external validation using genetic predisposition scores, we used physical activity data from 8,195 participants of the BioMe Biobank that were available to us. While this sample size is not sufficient to validate associations for individual variants, it is adequate to examine associations of complex traits with genetic predisposition scores. Molecular dynamics simulations were run in triplicate to ensure results are robust. Single fiber experiments were performed in 298 single muscle fibers from four R/R and four X/X carriers at the ACTN3 R577X variant, of which one R/R carrier is an E/E carrier and three are E/A carriers at the ACTN3 E635A variant of interest.
Data exclusions	Standard quality control based on genotype and sample quality was performed at the study level, as well as at the meta-analysis level (as described in Winkler et al., 2014).
Replication	The robustness of findings from the meta-analysis of GWAS for physical activity and sedentary traits was examined in two ways. Firstly, by examining the associations of individual variants identified as being associated with self-reported physical activity and sedentary traits in our study with objectively assessed physical activity traits in UK Biobank data (which makes up a large part of the data in the meta-analysis). Roughly half of the identified loci showed such associations at P<0.05. Secondly, we used data from participants of the BioMe Biobank to examine associations with moderate-to-vigorous intensity leisure time activity for genetic predisposition scores consisting of loci associated with moderate-to-vigorous intensity leisure time activity and leisure screen time identified in our meta-analysis. Both predisposition scores showed significant associations with moderate-to-vigorous intensity leisure time activity in the independent replication data of BioMe.
Randomization	With the exception of the single fiber experiments, there were no experimental groups in this study. In the single fiber experiments, eight healthy young men that volunteered to donate a muscle biopsy were selected from a group of 266 participants based on characteristics including height, weight, physical activity level and maximal knee extension torque at 45 degrees flexion. Participant were selected in such a way that four R/R carriers at the ACTN3 R577X variant were matched to four X/X carriers, as described in more detail in Broos et al., 2016 (PMID 26930663). The physician obtaining the muscle biopsy was unaware of the R577X genotype, the researchers performing the single fiber expriments were aware of the R577X genotype of the donot, but the variant of interest in the current study (the ACTN3 E635A-encoding variant) was not at the time anticipated to be relevant and was only genotyped in 2021 for the current study. Hence, the researchers that performed the single fiber experiments were blinded to the main exposure of interest in the present study.
Blinding	GWAS is a hypothesis-free approach, so in each study contributing to the meta-analysis, researchers assessing physical activity and sedentary traits were blinded to the genotypes that we now know are associated with those outcomes. Similarly, in the GWAS performed in 100 mouse strains, associations analyses were performed hypothesis-free, and researchers were not aware of the loci identified as being associated with physical activity traits in humans.

In the single muscle fiber experiments, the researchers were aware of the genotype at the ACTN3 R577X variant at the time of the study, but they were blinded to the participants' genotype at the variant of interest in the present study, i.e. the ACTN3 E635A-encoding variant, which was de novo genotyped for this study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Mice arrived at UCLA at 5 to 8 weeks of age and were housed 1-4 weeks until wheel testing. All mice were ~3 months old at the start of the experimental protocol, and were randomized into two groups: 1) sedentary or no exercise; and 2) exercise trained. Strains used and sample size per group are shown in Supp Table 23. Trained animals were housed unaccompanied on a standard 12-hour light dark cycle (6AM to 6PM local time). They were fed on a standard laboratory chow diet (8604, Teklad) with ad libitum access to food and water for the entire duration of the experiment. Mice were given full-time access to a running wheel for ~30 days.
Wild animals	The study did not involve wild animals.
Field-collected samples	The study did not involve samples collected from the field.
Ethics oversight	For the GWAS in 100 mouse strains, all studies were approved by the Institutional Animal Care and Use Committee (IACUC) and the Animal Research Committee (ARC # 1992-169-83e) at the University of California, Los Angeles (UCLA).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	Study level genome-wide association studies were run in a sex and ancestry specific manner, and were adjusted for age, age-squared, principal components reflecting population structure and additional study-specific covariates where appropriate. Since we used data from men and women of 51 studies, we kindly refer to supplementary table 1 for descriptive information on age and sex in each study.
Recruitment	Participants were required in the 51 studies contributing to the meta-analysis as described in protocol papers for those studies. To identify studies with data on both genome-wide genotypes and physical activity and sedentary behavior, we contacted all cohort studies we were aware of at the time, benefiting from earlier collaborations and meta-analyses of GWAS performed and published by others.
Ethics oversight	No ethical approval was required for the study since it was a meta-analysis of summary statistics obtained in studies that each had ethical approval provided by local ethics boards.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

nature portfolio

https://doi.org/10.1038/s41588-022-01165-1

Supplementary information

Genome-wide association analyses of physical activity and sedentary behavior provide insights into underlying mechanisms and roles in disease prevention

In the format provided by the authors and unedited

Genome-wide association analyses of physical activity and sedentary behavior provide insights into underlying mechanisms and roles in disease prevention

Wang et al., (the author list is provided in the main manuscript)

SUPPLEMENTARY INFORMATION

Table of Contents

SUPPLEMENTARY RESULTS

Physical activity loci under selection

As much higher physical activity levels were on average required to ascertain sufficient nutrition in times of hunting and gathering and pre-mechanical farming as compared with today's Westernized societies[1], a higher capacity to be physically active may have been selected for. To explore this, we examine if MVPA and LST association signals overlap with regions identified in three genome-wide selection screens[2-4]. Here, we show that 22 genes located <100kb of lead SNPs in three MVPA and/or LST-associated loci are located in three of 412 regions under selection in the past 50,000 years[2] (**Supp Table 29**). The protein-coding genes nearest the lead SNPs (<10kb) – *DNM3*, *MST1R* and *FOXP1* – are also prioritized by other approaches (**Supp Tables 25-26**) and play a role in cell signaling and wound healing, amongst others (**Supp Box 1**). We next identify genes located <10kb of 15 loci under selection in the past 10,000 years – based on results from an ancient DNA scan[3] – and <100kb of physical activity association signals. This yields one additional gene (*GRM5*) that harbors an intronic GWAS lead SNP for LST (rs1391954, **Supp Table 29**). *GRM5* encodes a metabotropic glutamate receptor that activates phospholipase C[5]; another key player in signal transduction[6], inflammation and wound healing[7], amongst other processes. No lead SNPs for LST or MVPA are located within 1Mb of five loci under very recent selection[4]. In summary, we show that a modest number of loci (n=4) selected for in the past 10,000-50,000 years are associated with leisure time physical activity and sedentary behavior today.

Overlap between mouse and human loci

Many of the biological factors influencing sedentary behavior and physical activity are likely shared across species[8]. Identifying such loci may help prioritize candidate genes, and shed light on relevant mechanisms. To this end, we compare our findings with loci identified in a GWAS for spontaneous physical activity in 100 inbred mouse strains, performed using the Hybrid

Mouse Diversity Panel (HMDP)[9] (**Supp Table 23**). Nine genes in two LST-associated loci are also located within ±1 Mbp of two lead SNPs for distance run and average running speed in mice ($P<4.1\times10^{-6}$) (**Supp Table 24**). Of the eight genes that overlap across humans and mice in one of these two loci, *TESC* – highly expressed in the striatum – harbors the intronic lead SNP rs2173650 in humans (**Supp Box 1**). In the mouse however, a gene without an established orthologue in humans – the lncRNA 4930413E15Rik expressed in olfactory and reproductive tissues – ranks 65th out of 16,378 genes (0.4th percentile), indicating it is likely causal for high voluntary wheel running behavior in mice selectively bred for 61 generations[10]. Using single cell RNA-sequencing data from GTEx[11], we show that in humans, a sequence 1.4 Mb from rs2173650 with high conservation to the mouse 4930413E15Rik is expressed in several reproductive tissues (**Supp Figure 4**). Sex hormones are firmly established to influence physical activity regulation in animal models[12], while vomeronasal chemosensory receptors were recently proposed to play a role in voluntary exercise behavior[13]. Unfortunately, no information on gene expression in the olfactory system is available in GTEx.

Molecular dynamics simulation for p.Glu635Ala

Computer-based molecular dynamics (MD) simulations for alpha actinin 3 show that the ancestral p.Glu635 variant (A allele) product facilitates salt bridge and hydrogen bonding interactions at residue 635 with surrounding residues (e.g., p.Arg638 and p.Gln639, **Figure 6b**) via its glutamate side chain. Such interactions are not formed with the p.635Ala substitution variant and have a different pattern of interaction in ACTN2. Root mean square fluctuations were performed on the residues of spectrin repeats of each monomer chain using gmx rmsf. Root mean square fluctuations of each amino acid residue, i.e., their average displacement over the simulation compared with the starting structure, were calculated with both variants and with ACTN2, with ACTN2 as well as ACTN3 variant p.635Ala – but not ACTN3 variant p.Glu635 – showing higher fluctuations in the monomer with the restrained actin-binding domain (**Figure**

6d, Chain B, orange and green traces). Root mean square fluctuations peaks in these interface regions are around 0.8 nm for ACTN2 and the ACTN3 p.635Ala variant, while the ACTN3 p.Glu635 variant fluctuations are approximately 0.6 nm (**Figure 6d**). In the presence of p.635Ala, moderately higher root mean square fluctuations values were observed in the middle section of the spectrin repeats – over the residue range of 410-540 – though p.Glu635 showed a more pronounced peak near residue 440 (**Figure 6d**). Overall, ACTN2 and ACTN3 p.635Ala showed a similar behavior that is distinctly different from p.Glu635, with a greater magnitude of root mean square fluctuations under no-load conditions, suggesting a more flexible structural region in the presence of ACTN2 or the ACTN3 p.635Ala variant product (C allele), associated with higher MVPA.

Steered molecular dynamics and Umbrella sampling for p.Glu635Ala

When a compressive force was applied between the center of mass of the two actin-binding domains, the force required to compress the two actin-binding domains by 0.6 nm was lower for both ACTN3 p.635Ala and ACTN2 compared with ACTN3 p.Glu635 (50, 45 and 74 kJ mol^{-1} nm^{-1}, respectively). Furthermore, the force-to-distance relationship to a compressive distance of -1.2 nm – where the two respective forces converge (67 kJ mol^{-1} nm^{-1}) – was notably more linear for both ACTN3 p.635Ala and ACTN2 than for ACTN3 p.Glu635 (**Supp Figure 6**). Greater variability is also seen for ACTN2 and ACTN3 p.635Ala in the force versus distance relationship among triplicate steered molecular dynamics simulations. To explore these features further, we used umbrella sampling to examine the change in potential of mean force (free energy surface) over the reaction coordinate corresponding to the compression of the ACTN3 dimer.

Umbrella sampling of ACTN2 and the ACTN3 dimer variants showed that the initial compression of the two ACTN3 variants and ACTN2 from a relaxed state to a compression of 1.2 nm was similar, requiring energy input of approximately 4.6 kJ mol^{-1}. Beyond this distance of 1.2 nm ACTN3 p.Glu635Ala diverged from ACTN3 p.635Ala and ACTN2 in its response to

compression (**Figure 6e**). ACTN3 p.635Ala required 2.8 kJ mol^{-1} to compress the dimer from -1.2 to -2.3 nm, while ACTN3 p.Glu635 required ~6.5 kJ mol^{-1} from -1.2 to -2.3 nm, ACTN2 having reached its peak of 7.5 kJ mol^{-1} at a compression of -2.3 nm (**Figure 6e**). Interestingly, bootstrap estimation of the error of the potential of mean force showed greater variance for p.635Ala, in line with and strengthening the root mean square fluctuations and steered molecular dynamics simulations results. Taken together, these results indicate that the ACTN3 p.635Ala dimer - associated with higher MVPA – exhibits greater flexibility than the p.Glu635 dimer.

SUPPLEMENTARY DISCUSSION

Results in post-GWAS analyses were most robust for LST, as a result of a markedly larger number of loci showing associations with LST as compared with the other three outcomes. There are at least five reasons for identifying most loci for LST. First, the SNP heritability was approximately twice as high for LST as for the other outcomes. Secondly, LST is the only outcome that was recorded in a homogenous manner across all studies. For the other three outcomes, the exact questions and response options used to capture the underlying latent variable differed across studies. For those outcomes, we summarized data from all available, relevant questions in each study into the most informative outcome that was still comparable across studies. The more heterogeneous nature in which the other three outcomes were captured negatively affects the statistical power to identify associated loci. Furthermore, while LST was normally distributed and could be analyzed as a continuous outcome, the other three traits were analyzed as dichotomous outcomes, resulting in a lower statistical power. In addition, due to the negative binomial distribution of MVPA – with an inflation of zeros – we used the median of 20 mins/week as a threshold to distinguish between "active" and inactive individuals. While this threshold is statistically sound, it essentially distinguishes between individuals that partake in some MVPA vs. those that do not. This likely negatively affected the likelihood of identifying loci that are truly relevant for participation in MVPA. Finally, the sample size for LST was much larger than for sedentary commuting behavior and sedentary behavior at work.

Aiming to improve the understanding of the molecular basis of physical activity, we perform a range of largely complementary approaches to identify candidate genes through which the association signals are anticipated to act. Strikingly, of the 268 and 39 genes prioritized across 70 LST- and eight MVPA-associated loci, only 22 genes are prioritized by >1 approach when using traditional cut-offs within each approach. This illustrates the complexity of *in silico* gene prioritization for complex behaviors, especially when proof-of-concept genes are sparse and a gold standard approach for prioritization is nonexistent. When combining results

from multiple approaches, applying more lenient criteria in individual approaches is justifiable. First, because the odds that a gene with an FDR of (e.g.) 0.20 in two methodologically independent approaches represents a false positive is low (i.e. 0.04); and secondly because sensitivity is more important than specificity when using *in silico* gene prioritization results to guide the selection of genes for downstream genetic perturbation studies in high throughput model systems. Applying more lenient criteria in the individual gene prioritization approaches increased the number of genes prioritized by >1 approach from 22 to 46 (Supp Tables 25-26).

SUPPLEMENTARY METHODS

Self-reported physical activity and sedentary behavior – rationale for trait definitions

When initiating this effort in 2011 (i.e. pre-UK Biobank), we first explored what physical activity-related questions were available in the first ~20 cohort studies that agreed to participate. This served several purposes: 1) to identify common ground; 2) to explore the distributions of the available outcomes; and 3) to assess how traits could (or should) be analyzed to maximize the statistical power to identify relevant loci. Since we were interested in identifying genetic factors that are relevant for daily physical activity levels, we aimed to capture all moderate-to-vigorous intensity physical activity (MVPA) during leisure time, rather than limiting ourselves to just sports and exercise participation. Since the median time spent on all MVPA during leisure time in the first ~20 cohort studies was merely ~20 mins per week (with a zero-inflated negative binomial distribution), we decided against further dissecting this trait into MPA and VPA separately. Leisure screen time was the only outcome that we could analyze in a continuous manner, thanks to the uniform and continuous manner in which it had been acquired in all studies, and thanks to its normal distribution. For commuting behavior and physical activity at work, the distributions were such that contrasting the most active individuals with the remainder of the population would have resulted in a much lower statistical power than instead contrasting the least active individuals with the remainder of the population. Since then, it has been shown that: 1) among regular commuters, those exclusively commuting by car had an 11% higher risk of cardiovascular events, and a 30% higher risk of fatal cardiovascular events compared with individuals with a more active mode of commuting, during a mean follow-up of seven years[14]; and 2) more sitting at work is associated with higher mortality during follow-up in primary industry, i.e. non-office professions[15]. Taken together, these results suggest that identifying genetic factors associated with the outcomes as defined in our study had the potential to yield clinically meaningful and possibly actionable insights.

Besides trait definitions varying across studies, the average age per cohort ranged from early adulthood to old-age (17-74 years old). The power to detect genetic factors that influence physical activity was thus likely compromised by misclassification of physically active and inactive individuals, and heterogeneity by the inclusion of older age groups in the meta-analysis, as the heritability of physical activity decreases with increasing age[16]. However, such factors could have resulted in type II – but not type I – errors in the meta-analysis. Despite these limitations, a large sample size helped us identify 42 previously unreported loci for self-reported physical activity and sedentary behavior. Genetic correlations with objectively assessed physical activity traits were modest and five of eight previously reported loci for objectively assessed physical activity show evidence of association with self-reported physical activity and sedentary behavior. Hence, despite the well-known limitations of self-reported physical activity, focusing on domain and intensity specific physical activity traits in a large study sample helped increase the understanding of the genetic etiology of this complex behavior in a manner that is not currently possible using objectively assessed physical activity.

GWAS and meta-analyses – additional analyses

Several analyses were performed at the discovery stage for which we decided not to report associations. First, for all outcomes, we examined genome-wide associations with and without adjusting for BMI. To avoid drawing conclusions that are driven by collider bias[17], we did not use the BMI-adjusted associations. Secondly, to further explore potential linear associations with MVPA, we used zero-inflated negative binomial regression and modeled MVPA as a continuous outcome (mins/week). Associations were examined using the same covariates as in the main genome-wide analyses amongst 371,244 unrelated UK Biobank participants of European ancestry, for variants with $P<1\times10^{-5}$ for association with the dichotomous MVPA outcome in the European ancestry meta-analysis. The approach yielded similar results and are therefore not shown. Finally, gene-based analyses aggregating rare (MAF<1%) functional variants as

annotated by Ensembl's variant effect predictor (VEP)[18] were conducted in UK Biobank European ancestry participants. Gene-based Burden and SKAT tests were performed using a mixed model approach in the GENESIS package[19]. Genes identified using these approaches were also flagged by the single variant analysis and hence, the results are not shown.

Mouse experiments

Females from 100 genetically distinct strains from the Hybrid Mouse Diversity Panel (HMDP)[20] were purchased from Jackson Laboratories (University of Tennessee Health Science Center). They arrived at UCLA at 5 to 8 weeks of age and were housed 1-4 weeks until wheel testing. All mice were ~3 months old at the start of the experimental protocol, and were randomized into two groups: 1) sedentary or no exercise; and 2) exercise trained. Strains used and sample size per group are shown in **Supp Table 23**. Trained animals were housed unaccompanied on a standard 12-hour light dark cycle (6AM to 6PM local time). They were fed on a standard laboratory chow diet (8604, Teklad) with ad libitum access to food and water for the entire duration of the experiment. Mice were given full-time access to a running wheel for ~30 days. Wheel revolutions were tallied every 15 sec using VitalView® Activity Software (Starr Life Sciences Corp, Oakmont, Pennsylvania, United States). Daily running distances were calculated by converting wheel revolutions into distance using wheel circumference (35.9 cm). Average running speed was calculated by averaging all non-zero-wheel revolutions and normalizing on a per sec basis. Percent time running was calculated by dividing all 15 sec bouts a wheel revolution was recorded to the total number of intervals. Additional information on GWAS performed using the Hybrid Mouse Diversity Panel (HMDP) is described elsewhere[9]. All studies were approved by the Institutional Animal Care and Use Committee (IACUC) and the Animal Research Committee (ARC # 1992-169-83e) at the University of California, Los Angeles (UCLA).

In a selection study for high voluntary wheel-running behavior[10], the mouse lncRNA 4930413E15Rik was considered to show a strong indication of consensus in a locus that was associated with daily distance run and average voluntary running speed in a GWAS in 100 mouse strains, as well as with LST in humans. The mouse gene 4930413E15Rik is located on chr 5, spanning the coordinates 118,952,339 - 118,961,261 (mm10 assembly). To investigate the corresponding region in the human genome, a lift-over to hg19 was performed in the UCSC genome browser, resulting in the coordinates chr 12: 116,087,265 – 116,097,521. The region on chr 12 contains no established gene models, but was further investigated in the GTEx IGV browser[11] to study expression that might be present at low levels in specific human tissues.

Molecular dynamics simulations for the ACTN3 p.Glu635Ala variant

Alpha-actinin is a structural member of vertebrate muscle Z-discs, and primarily functions to cross-link neighboring actin filaments of opposite polarity from adjacent sarcomeres. This binding can occur over a range of angles from 60 to 120°, creating a tetragonal lattice with a lattice spacing of 19 to 25 nm[21-23]. In addition to its interaction with actin, alpha-actinin binds and anchors titin to the Z-disc[24]. The alpha-actinin homodimer is formed from two antiparallel subunits composed of an N-terminal actin-binding domain and a C-terminal calmodulin homology domain (CAM), separated by four spectrin-like repeats. Each repeat consists of a triple α-helix coiled-coil (**Figure 6A**). Alpha-actinin 3 (*ACTN3*) at 901 amino acids in length is one of four isoforms of alpha -actinin and is exclusively found in human type-II (also known as fast-twitch) skeletal muscle fibers. The naturally occurring truncating mutation R557X in *ACTN3* has a potential impact on injury risk during exercise, increased muscle-damage following eccentric training and increased flexibility for 557X homozygotes, who are generally presumed to have ACTN2 incorporated in their type II muscle fibers to compensate for the functional loss of ACTN3[25,26].

Steered molecular dynamics and Umbrella sampling for p.Glu635Ala

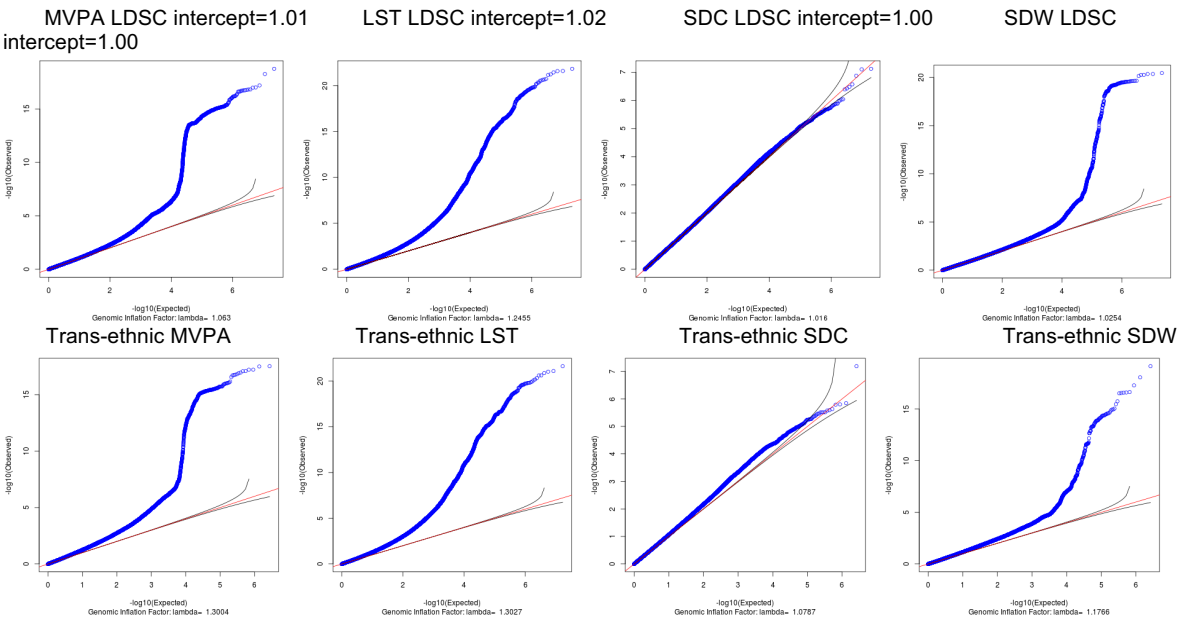
We compared the properties of ACTN2 and of the ACTN3 p.635Ala and p.Glu635 variants when placed under the simulated compressive loads that are likely experienced *in vivo*. The final frame of the 1 ns molecular dynamics production run was used as the starting topology for steered molecular dynamics simulations using fully relaxed dimers. Steered molecular dynamics simulations were run for 2 ns with a pulling rate of 0.005 nm ps^{-1} and a harmonic potential of 50 kJ mol^{-1} nm^{-2}. Centre of mass pull groups were defined as the actin-binding domain of each respective monomer, with a weak position restraint placed on the Cα atom of threonine-52 (ACTN3) or threonine-45 (ACTN2) – a centrally located residue in the core of the actin-binding domain – on one actin-binding domain, enabling full rotational freedom of each actin-binding domain during the course of the steered molecular dynamics simulations. The pulling vector was oriented along the axis on which the spectrin repeats were initially aligned. Suitable frames from each steered molecular dynamics simulation were selected that differed by no more than 0.2 nm from 0 to -5.5 nm (a contraction of the dimer by 5.5 nm or ~18%) and were used as the starting topology for a series of 10 ns umbrella sampling simulations. Analysis of the umbrella sampling simulations was conducted using g_wham, to yield the potential of mean force versus reaction coordinate for each variant.

When a compressive force was applied between the center of mass of the two actin-binding domains, the force required to compress the two actin-binding domains by 0.6 nm was lower for both ACTN3 p.635Ala and ACTN2 compared with ACTN3 p.Glu635 (50, 45 and 74 kJ mol^{-1} nm^{-1}, respectively). Furthermore, the force-to-distance relationship to a compressive distance of -1.2 nm – where the two respective forces converge (67 kJ mol^{-1} nm^{-1}) – was notably more linear for both ACTN3 p.635Ala and ACTN2 than for ACTN3 p.Glu635 (**Supp Figure 6**). Greater variability is also seen for ACTN2 and ACTN3 p.635Ala in the force versus distance relationship among triplicate steered molecular dynamics simulations. To explore these features further, we used umbrella sampling to examine the change in potential of mean force (free

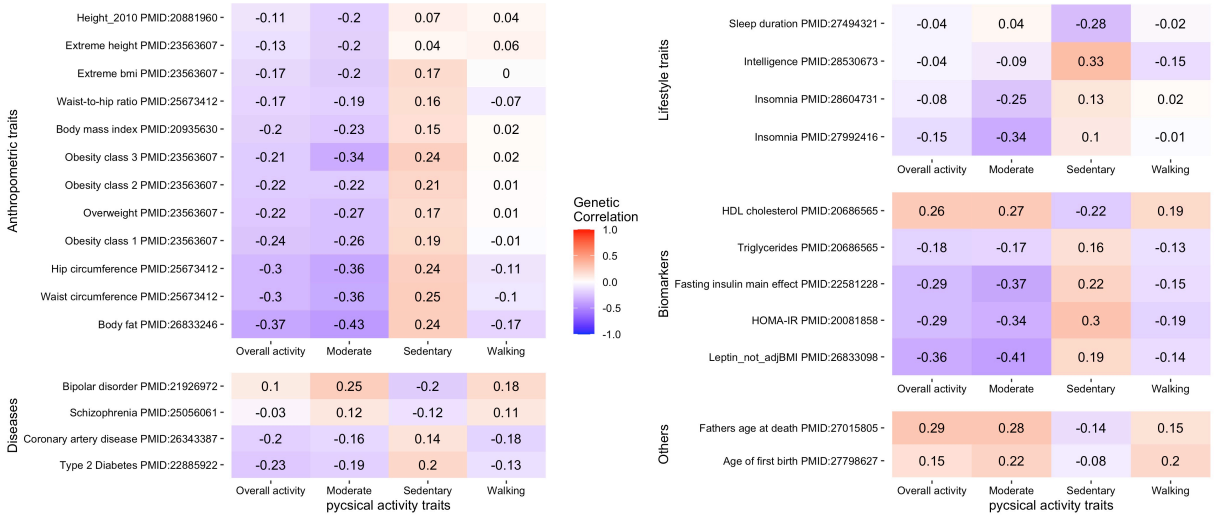
energy surface) over the reaction coordinate corresponding to the compression of the ACTN3 dimer.

Umbrella sampling of ACTN2 and the ACTN3 dimer variants showed that the initial compression of the two ACTN3 variants and ACTN2 from a relaxed state to a compression of 1.2 nm was similar, requiring energy input of approximately 4.6 kJ mol^{-1}. Beyond this distance of 1.2 nm, ACTN3 p.Glu635 diverged from ACTN3 p.635Ala and ACTN2 in its response to compression (**Figure 6E**). ACTN3 p.635Ala required 2.8 kJ mol^{-1} to compress the dimer from -1.2 to -2.3 nm, while ACTN3 p.Glu635 required ~6.5 kJ mol^{-1} from -1.2 to -2.3 nm, ACTN2 having reached its peak of 7.5 kJ mol^{-1} at a compression of -2.3 nm (**Figure 6E**). Interestingly, bootstrap estimation of the error of the potential of mean force showed greater variance for p.635Ala, in line with and strengthening the root mean square fluctuations and steered molecular dynamics simulations results. Taken together, these results indicate that the ACTN3 p.635Ala dimer - associated with higher MVPA – exhibits greater flexibility than the p.Glu635 dimer.

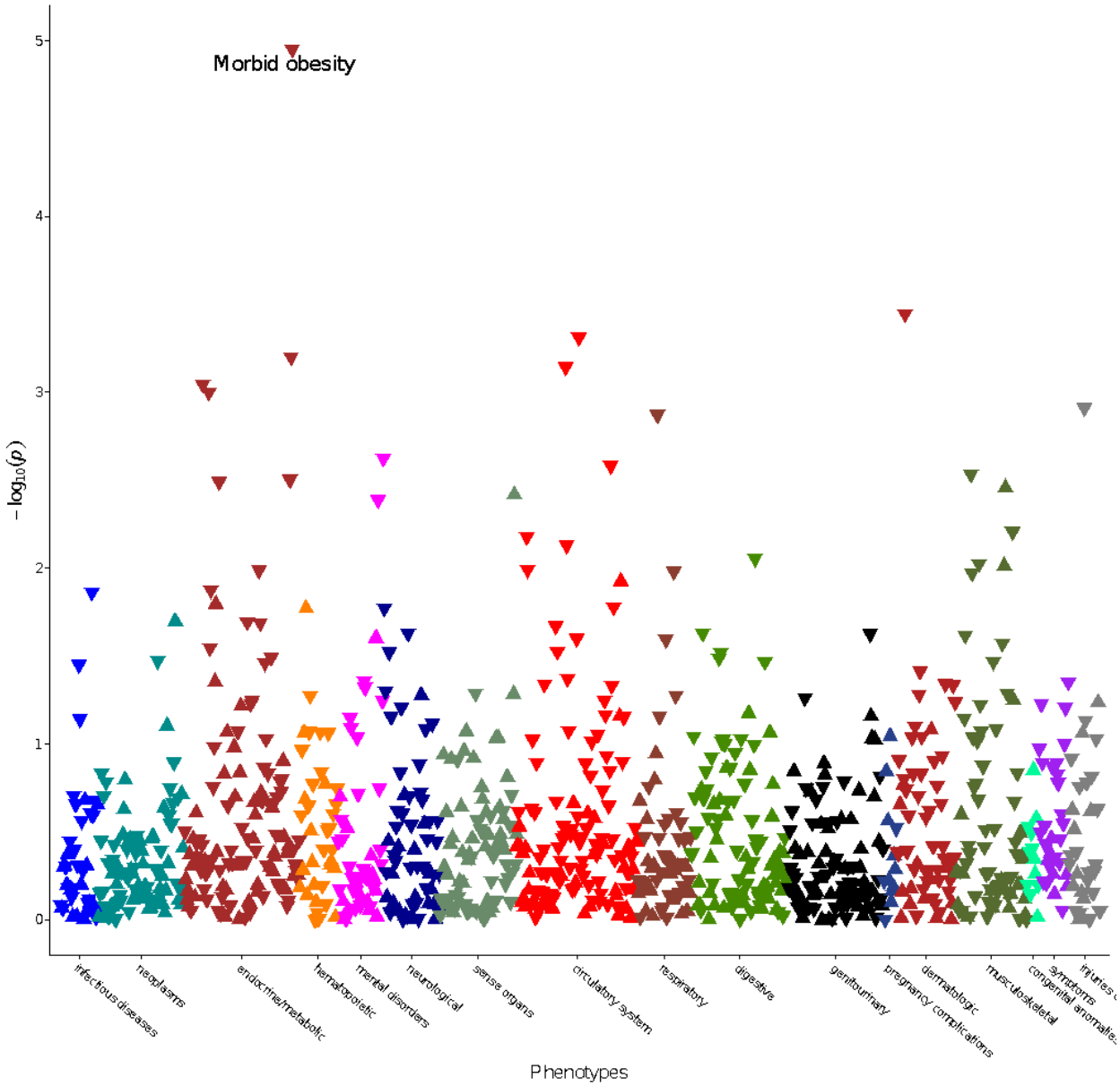
SUPPLEMENTARY FIGURES



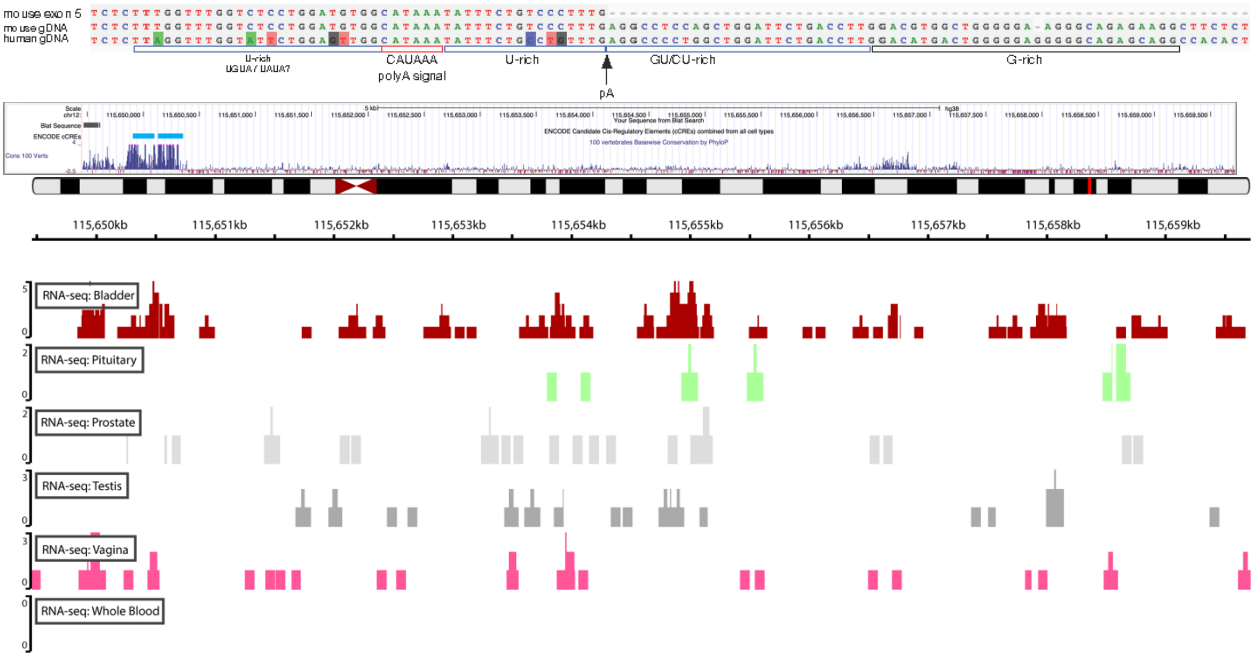
Supplementary Figure 1: **Quantile-quantile plots for the primary GWAS of self-reported physical activity and sedentary traits.** Moderate-to-vigorous intensity physical activity during leisure time (MVPA); leisure screen time (LST); sedentary behavior at work (SDW); and sedentary commuting behavior (SDC) in individuals of European ancestry only (top) as well as from the multi-ancestry meta-analysis (bottom). The estimated LD Score intercept for the primary GWAS is indicated for the former.



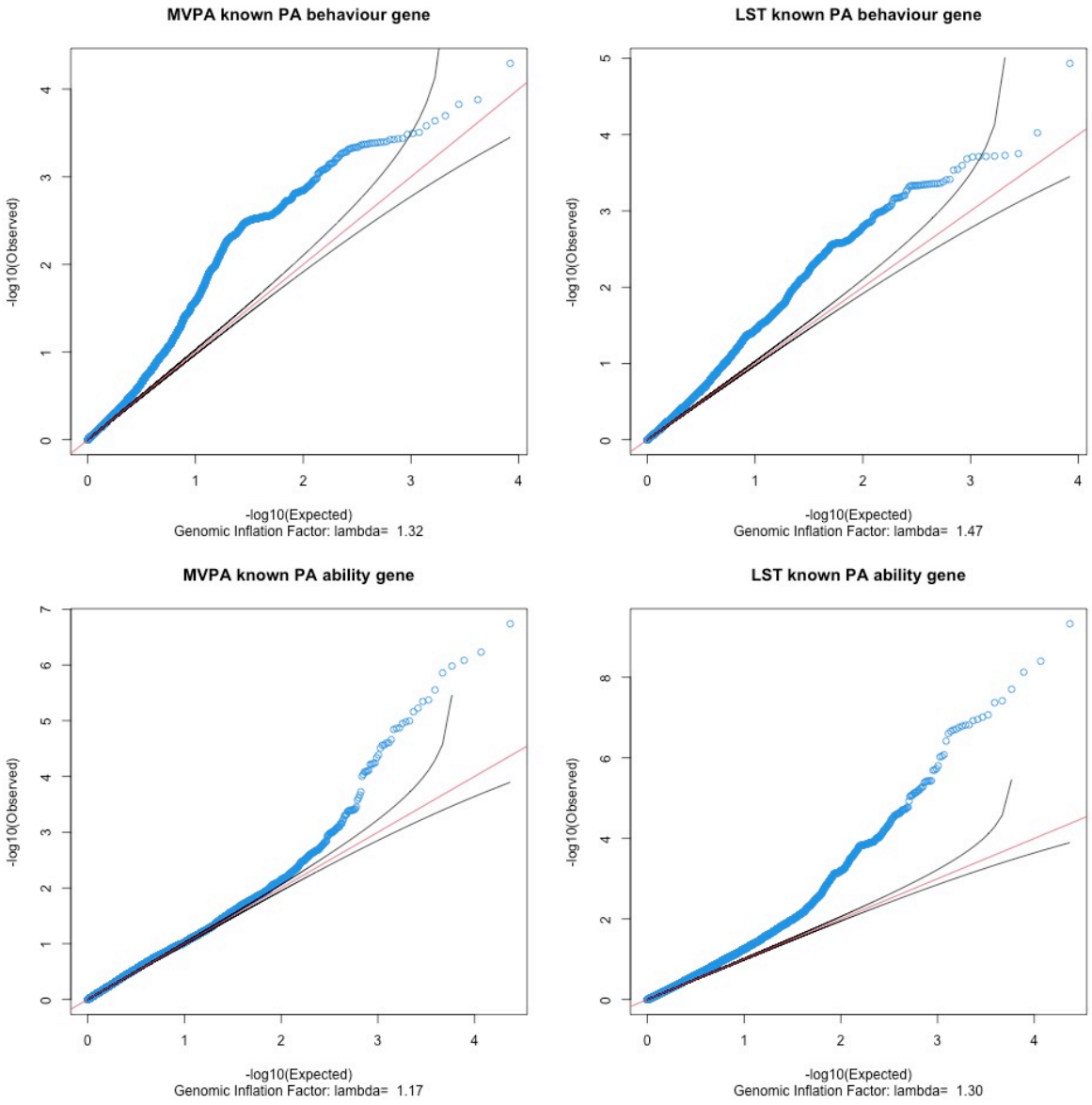
Supplementary Figure 2: **Genetic correlations of accelerometer-assessed physical activity with other traits and diseases in 91,105 UK Biobank participants**. We computed genetic correlations of four objectively assessed physical activity traits with 108 other traits and diseases using LD score regression, and show results for traits and diseases with at least one genetic correlation of $P < 4.6 \times 10^{-4}$ with an objectively assessed physical activity trait.



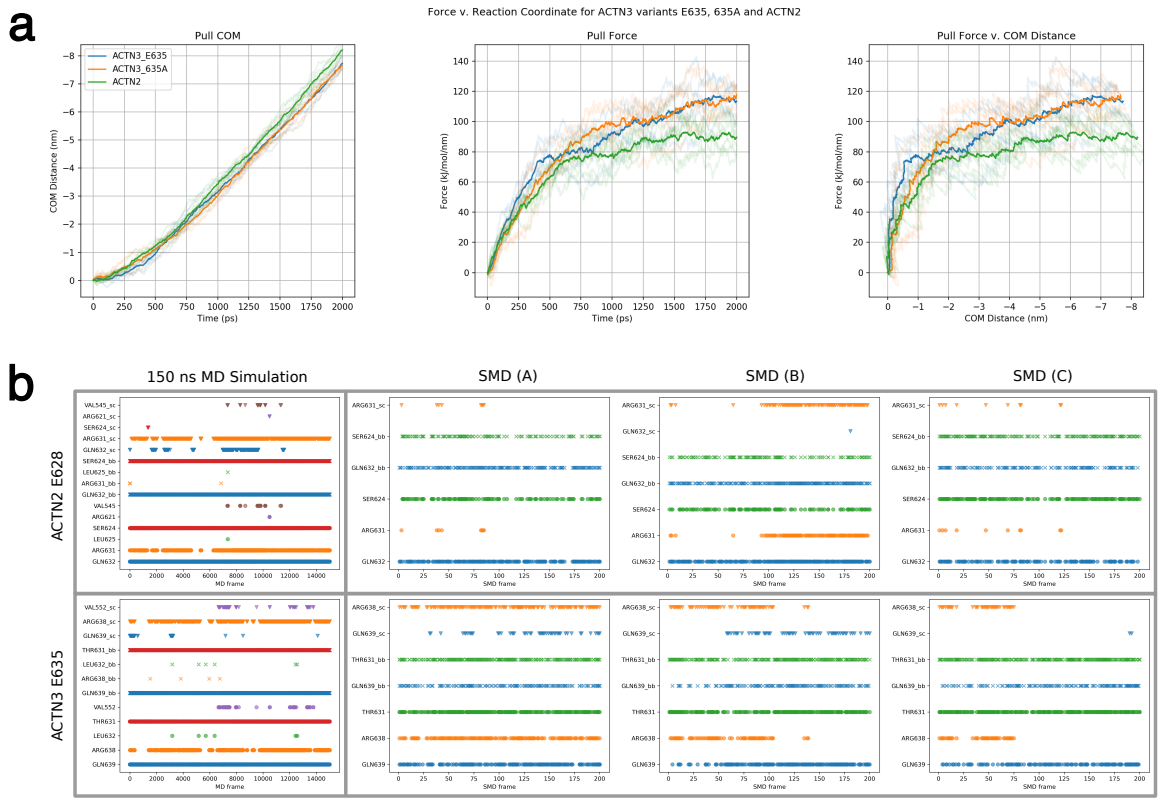
Supplementary Figure 3: Manhattan plot of PheWAS for polygenic score of MVPA shows association with morbid obesity in European ancestry individuals (N=8,959) in the BioMe Biobank. MVPA: moderate-to-vigorous intensity physical activity during leisure time, PheWAS: phenome-wide association study.



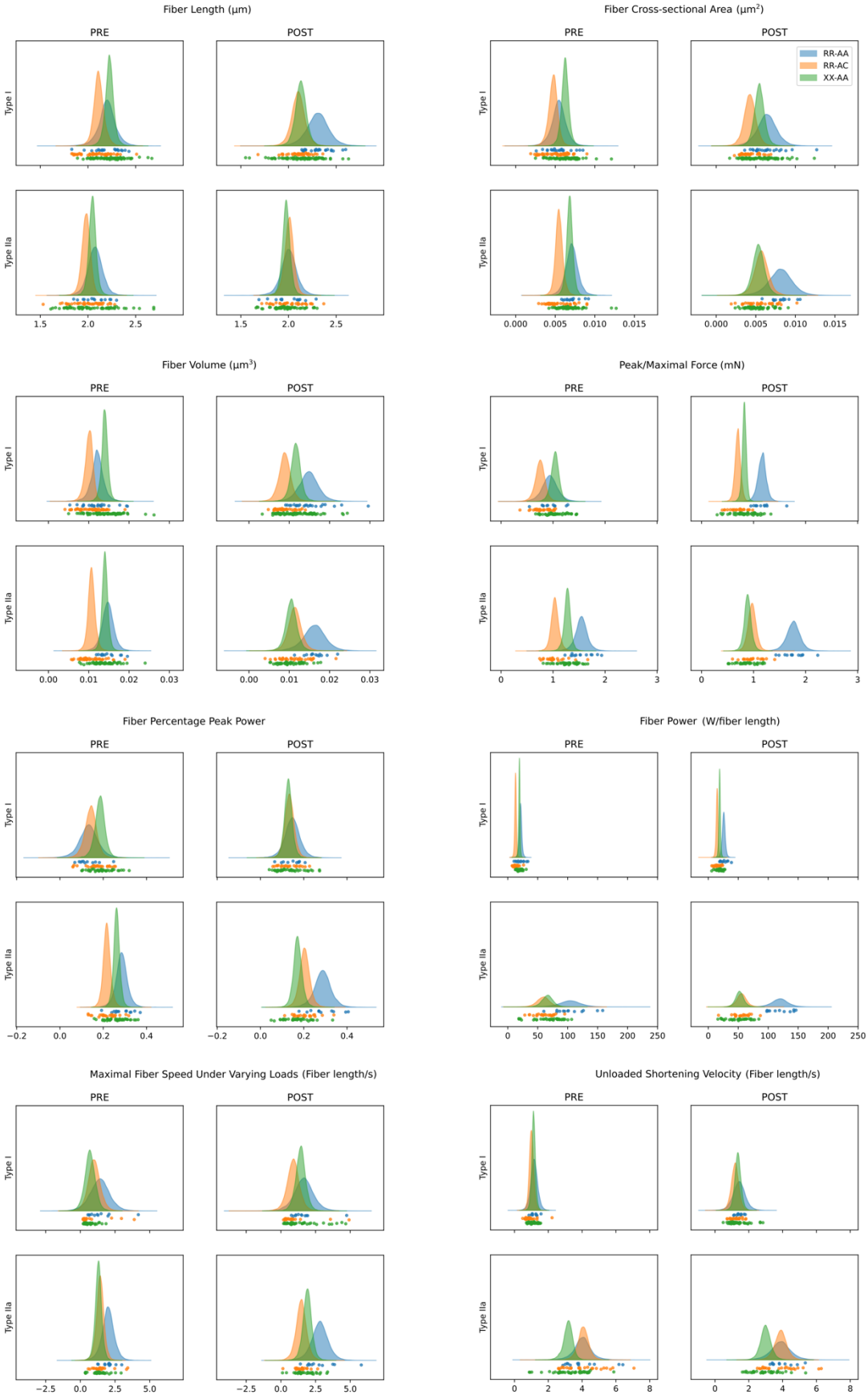
Supplementary Figure 4: RNA-seq data from GTEx displaying expression levels in the region chr12: 116,087,265 - 116,097,521 across several human tissues. Top: the region surrounding and including a poly-A signal is conserved across the mouse lncRNA 4930413E15Rik-encoding sequence and a locus on the human Chr 12 that is associated with leisure screen time; **Middle**: the human region on Chr 12 corresponding with an exonic sequence from the mouse 4930413E15Rik contains regulatory elements; **Bottom**: the human transcript is most highly expressed across the five highlighted tissues. In other tissues in the GTEx data collection, including whole blood (bottom), the region shows little or no evidence of expression. Chromosomal coordinates shown are from human genome built GRCh38.



Supplementary Figure 5: QQ plots of 31,673 variants show enrichment for association with MVPA and LST in 58 previously reported exercise (i.e. physical activity (PA) behavior) and fitness (i.e. physical activity ability) genes. 8,345 variants within 13 physical activity behavior genes, 23,328 variants within 45 PA ability genes. MVPA: moderate-to-vigorous intensity physical activity during leisure time; LST: leisure screen time.



Supplementary Figure 6. Steered Molecular Dynamics (SMD) and Hydrogen Bond Analysis (HBA) of ACTN3 p.Glu635, ACTN3 p.635Ala and ACTN2 from a homology structure shows a divergence in behavior under compressive force. a) SMD (replicates shown in transparency) of ACTN3 p.Glu635 (blue, 3 replicates), ACTN3 p.635Ala (orange, 3 replicates) and ACTN2 (green, 6 replicates), showing, left to right: the pulling center-of-mass (COM) distance between actin binding domains versus time; the pull force versus time; and the pull force versus pulling COM. **b)** HBA of interactions between the glutamate of: ACTN2 residue p.Glu628 (top row); ACTN3 residue p.Glu635 (bottom row) and neighboring residues within the relaxed 150 ns MD simulation (left column) and three SMD replicates (rightmost three columns). Interacting residues are indicated on the y-axis with the suffix '_sc' denoting side chain interactions, '_bb' backbone interactions and without suffix are total interactions. ACTN2 p.Glu628's side chain interacts with p.Arg631 in the relaxed dimer (top left), but this interaction tends to break upon application of force in SMD simulations (top right). ACTN3 p.Glu635's side chain interacts with p.Arg638 in the relaxed dimer and to a lesser extent with p.Gln639 (bottom left), though as increased force is applied in SMD simulations, the interaction with p.Arg638 is generally sustained for a longer period than that seen in ACTN2, and the interaction with p.Gln639 tends to become more extensive under compression, which is not seen in ACTN2 (and p.Gln632).


Fiber Length (μm)
Fiber Cross-sectional Area (μm²)
Fiber Volume (μm³)
Peak/Maximal Force (mN)
Fiber Percentage Peak Power
Fiber Power (W/fiber length)
Maximal Fiber Speed Under Varying Loads (Fiber length/s)
Unloaded Shortening Velocity (Fiber length/s)
PRE
POST
Type I
Type IIa
RR-AA
RR-AC
XX-AA

Supplementary Figure 7. Single muscle fiber experiments show a higher maximal stable force and fiber power in p.Glu635 compared with p.635Ala. Muscle biopsies from eight healthy young men (four Arg/Arg and four Ter/Ter at p.Arg577Ter) obtained before (pre) and after (post) an eccentric exercise bout were used to isolate single fibers, which were then functionally characterized. Of the four Arg/Arg carriers, one was heterozygous at p.Glu635Ala (46 fibers) and three were homozygous for the p.Glu635 variant (32±5 fibers). All four Ter/Ter carriers were homozygous for the p.Glu635 variant (39±6 fibers). Posterior distributions from 15,000 iteration Markov chain Monte Carlo models are shown separately for type I and type II_A fibers pre and post eccentric intervention.

SUPPLEMENTARY BOX 1

A Brief description of candidate genes prioritized by at least two approaches using information from Genecards, NCBI and Uniprot.

For MVPA and LST

CHST10 (chr 2, flagged by the intronic *LINC01104* variants rs4303732 (LST) and rs1160545 (MVPA)
Prioritized by: Activity by contact in adipose tissue, coronary artery, ovary; SMR brain (lenient)
Carbohydrate sulfotransferase 10 participates in the biosynthesis of human natural killer-1 (HNK-1) carbohydrate structure, which is involved in neurodevelopment and synaptic plasticity of the hippocampus.

MST1R (chr 3, flagged by the intronic *MST1R* variant rs7615206 for MVPA and LST)
Prioritized by: SMR skeletal muscle; activity by contact in skeletal muscle and fetal thymus; altered expression in skeletal muscle following resistance training (lenient)
Also: in a locus under selection in the past 50,000 years[2]
Macrophage Stimulating 1 Receptor is a cell surface receptor for macrophage-stimulating protein (MSP) that has tyrosine kinase activity. It is expressed at the protein level on the ciliated epithelia of the mucociliary transport apparatus of the lung[27], and together with MSP plays a role in host defense. MST1R regulates physiological processes that include cell survival, migration and differentiation. Ligand binding at the cell surface induces autophosphorylation of RON[28], which activates the wound healing response by promoting epithelial cell migration, proliferation and survival at the wound site[29,30]. Following activation, MST1R interacts with PIK3R1, PLCG1 or the adapter GAB1. Recruitment of these downstream effectors by RON leads to the activation of several signaling cascades including the RAS-ERK, PI3 kinase-AKT, or PLCgamma-PKC. MST1R also plays a role in the innate immune response by regulating the migration and phagocytic activity of macrophages.

For MVPA

LONRF2 (chr 2, flagged by the intronic *LINC01104* variants rs4303732 (LST) and rs1160545 (MVPA)
Prioritized by: Activity by contact in coronary artery and ovary; SMR blood (lenient) for MVPA. Only prioritized by Activity by contact in coronary artery and ovary for LST
LON peptidase N-terminal domain and ring finger 2 is predicted to enable metal ion binding activity.

AKAP10 (chr 17, flagged by rs385301 downstream of *AKAP10*
Prioritized by: Activity by contact in pancreas, adipose tissue, coronary artery, cardiac ventricle, skeletal muscle, ovary, derived neuronal progenitor cultured cells; SMR skeletal muscle (lenient); DEPICT gene prioritization (lenient).

Encoded by a gene in a locus previously identified for physical activity[31], alpha-kinase anchoring protein 10 is known to confine regulatory subunits of protein kinase A to discrete regions of mitochondria[32]. Animal studies have shown evidence for a role of the protein encoded by *AKAP10* in heart rhythm regulation[33], but skeletal muscle phenotypes were not reported previously in knockout models. However, A-kinase anchoring proteins (AKAPs) partially restrict cAMP-PKA signaling, especially in highly structured cell types like skeletal myofibers[34]. cAMP signaling participates in muscle development and regeneration mediated by muscle precursor cells[35].

SPECC1 (chr 17, flagged by rs385301 downstream of *AKAP10*
Prioritized by: Activity by contact in coronary artery and cardiac ventricle; DEPICT gene prioritization (lenient).
Sperm antigen with calponin homology and coiled-coil domains 1 belongs to the cytospin-A family and is highly expressed in testis.

For LST

KDM4A (chr 1, flagged by the intronic *KDM4A* variant rs71658797)
Prioritized by: Activity by contact in pancreas, skeletal muscle, adipose tissue, thymus, ovary, derived neuronal progenitor cultured cells, bipolar neuron from iPSC; and DEPICT gene prioritization (lenient)
Lysine Demethylase 4A is a nuclear protein that functions as a trimethylation-specific histone demethylase and as a transcriptional repressor. It is crucial for muscle differentiation and promotes transcriptional activation of MYOG[36], which in turn is essential for the development of functional embryonic skeletal muscle.

AK5 (chr 1, flagged by the intronic *AK5* variant rs3791033)
Prioritized by: DEPICT tissue and gene prioritization; SMR brain (lenient)
Adenylate Kinase 5 is a cytosolic protein that is exclusively expressed in the brain. It plays a role in regulating the adenine nucleotide composition in a cell by catalyzing the reversible transfer of the terminal phosphate group between nucleoside triphosphates and monophosphates[37].

DNM3 (chr 1, flagged by the intronic DNM3 variant rs6685030)
Prioritized by: Activity by contact in neuronal progenitors, adrenal gland and skeletal muscle; SMR blood (lenient)
Dynamins represent a subfamily of GTP-binding proteins, which are associated with microtubules and bind actin and other cytoskeletal proteins. DNM3 plays a role in the development of megakaryocytes and vesicle-mediated transport and endocytosis.

LRPPRC (chr 2, flagged by the indel rs145255225 (a.k.a. rs34908368))
Prioritized by: SMR brain (lenient); altered expression in skeletal muscle following resistance training (lenient).

This gene encodes a Leucine-rich pentatricopeptide repeat containing protein. Mutations in *LRPPRC* cause a monogenic mitochondrial disease (Leigh syndrome French Canadian Type) that involves severe muscle and movement problems[38]. In addition to the altered expression in skeletal muscle following resistance training we observed, LRPPRC is also up-regulated by exposure to environmental enrichment that is a complex combination of physical, cognitive, and social stimuli in striatum, which may improve locomotor performance[39].

SCN2A (chr 2, flagged by the *SCN2A* missense variant rs114590429)
Prioritized by: DEPICT gene prioritization; running speed, distance and time run in mice (lenient)
SCN2A encodes Sodium Voltage-Gated Channel Alpha Subunit 2, which mediates the voltage-dependent sodium ion permeability of excitable membranes. Mutations in this gene have been associated with seizure disorders, autism spectrum disorder and general cognitive ability[40]. *Scn2a* haploinsufficient mice display a spectrum of phenotypes affecting anxiety, sociability, memory flexibility that commonly observed in models of schizophrenia and autism spectrum disorder[41].

RNF123 (chr 3, flagged by the intronic *MST1R* variant rs7615206)
Prioritized by: SMR brain; DEPICT gene prioritization (lenient) for LST. Only prioritized by SMR brain (lenient) for MVPA
The *RNF123* gene encodes E3 ubiquitin-protein ligase, a motif present in a variety of functionally distinct proteins and known to be involved in protein-protein and protein-DNA interactions. It promotes the ubiquitination and proteasome-mediated degradation of CDKN1B, which is the cyclin-dependent kinase inhibitor at the G0-G1 transition of the cell cycle by the ubiquitin-proteasome pathway [42,43]. It also functions as a novel inhibitor of innate antiviral signaling, independently of its E3 ligase activity[44]. This gene is more highly expressed in skeletal muscle than in other tissues. Recent studies involving UK Biobank samples also associated the locus with musculoskeletal pain[45,46].

SEMA3F *(chr 3, flagged by* the intronic *MST1R* variant rs7615206; group 2 LST locus)
Prioritized by: SMR brain; Activity by contact in pancreas, adipose tissue, hepatocytes, bipolar neuron from iPSC. Not prioritized for MVPA
Semaforin 3F – encoded by *SEMA3F* - is involved in axon guidance during neuronal development. This gene is expressed in endothelial cells where it induces apoptosis, inhibits cell proliferation and survival, and acts as an anti-tumorigenic agent.

FOXP1 *(chr 3, flagged by the intronic SAMMSON variant rs76267866)*
Prioritized by: DEPICT gene prioritization; Activity by contact in adipose tissue, adrenal gland, skeletal muscle, myotube, astrocyte
Forkhead Box P1 acts as a tumor suppressor and is involved in regulation of cardiac muscle cell proliferation and columnar organization of spinal motor neurons. It also plays a role in B-cell development and promotes the formation of the lateral motor neuron column (LMC) and the preganglionic motor column (PGC) and is required for appropriate motor axon projections.

CADM2 (chr 3; flagged by the intronic CADM2 variants rs1691471 (MVPA) and rs1375561 (LST)
Prioritized by: DEPICT tissue enrichment; DEPICT gene prioritization (nominal); SMR skeletal muscle. Only prioritized by altered expression in skeletal muscle following resistance training (lenient) for LST.
CADM2 (also known as SynCAM2, Igsf4d, and Nectin-like molecule 3) encodes the synaptic cell adhesion molecule 2. SNPs in the locus have been associated with a series of psychological traits, such as educational attainment[47]; self-reported physical activity[31]; risk-taking behaviour[48]; alcohol consumption[49]; substance use and risk-taking[50]; and obesity[51]. In addition to lower adiposity, lower systemic glucose levels, and better insulin sensitivity, *Cadm2*-knockout mice exhibited more locomotor activity, higher energy expenditure, and higher core body temperature, suggesting *cadm2* is a potent regulator of systemic energy homeostasis[52]. While C*ADM2* is predominantly expressed in the brain, the top SMR SNP for MVPA in skeletal muscle (rs382210) is in LD with the lead MVPA GWAS SNP rs1691471 (r2=0.29, D'=0.95), it is independent of the previously identified BMI-associated SNP rs13078960 (r2=0.03, D'=0.48)[51]). This suggests that while *CADM2* likely influences other complex traits through the brain, it possibly influences PA locally through skeletal muscle.

HTR1F *(chr 3, flagged by the intronic HTR1F variant rs17025214)*
Prioritized by: DEPICT gene prioritization; Activity by contact in thyroid
5-Hydroxytryptamine receptor 1F is primarily located in the hippocampus, cortex and dorsal raphe nucleus and enables G protein-coupled serotonin receptor activity and serotonin binding activity. It also functions as a receptor for various alkaloids and psychoactive substances.

APC *(chr 5, flagged by the intronic APC variant rs396321)*
Prioritized by: DEPICT gene prioritization and tissue enrichment; Activity by contact in skeletal muscle
APC regulator of Wnt signaling pathway acts as an antagonist of the Wnt signaling pathway. It is also involved in cell migration and adhesion, transcriptional activation and apoptosis. It is a tumor suppressor.

REEP5 (chr 5, flagged by the intronic *APC* variant rs396321)
Prioritized by: DEPICT gene prioritization; SMR skeletal muscle (lenient)
Receptor Accessory Protein 5 (*REEP5*) expression is muscle-specific, with the highest protein expression in the mouse ventricles and skeletal muscle. *In vitro* and *in vivo* experiments have demonstrated that the protein encoded by *REEP5* plays a critical role in sarco/endoplasmic reticulum organization and function, as well as in normal heart function and development[53].

SIL1 (chr 5, flagged by the intronic *SIL1* indel rs752485316)
Prioritized by: SMR brain; Activity by contact in adrenal gland, cardiac ventricle and ovary; altered expression in skeletal muscle following resistance training (lenient)
SIL1 (SIL1 Nucleotide Exchange Factor) encodes a resident endoplasmic reticulum (ER), N-linked glycoprotein with an N-terminal ER targeting sequence, 2 putative N-glycosylation sites, and a C-terminal ER retention signal. Mutations in this gene have been associated with

Marinesco-Sjogren syndrome, which is clinically characterized by progressive myopathy and other tissue pathologies. Experimental characterization in mice reveals a disruption in ER homeostasis upon SIL1 loss, leading to loss of skeletal muscle mass, strength and function[54].

SOBP (chr 6, flagged by the intronic *PDSS2* variant rs78394231)
Prioritized by: DEPICT gene prioritization (lenient); Activity by contact in adrenal gland, bipolar neuron from iPSC, skeletal muscle, cardiac ventricle, ovary, thryoid
Sine oculis binding protein homolog is involved in development of the cochlea and has been linked to intellectual disability.

REPS1 (chr 6, flagged by the intronic *REPS1* deletion rs200307517)
Prioritized by: SMR brain (lenient); altered expression in skeletal muscle following resistance training (lenient)
REPS1 (RALBP1 Associated Eps Domain Containing 1) encodes a signaling adaptor protein with two EH domains that interacts with proteins that participate in signaling, endocytosis and cytoskeletal changes.

PDE10A (chr 6; flagged by the intronic *PDE10A* SNP rs58541850)
Prioritized by: altered expression in skeletal muscle following resistance training; DEPICT gene prioritization
Phosphodiesterase 10A plays a role in signal transduction by regulating the intracellular concentration of cyclic nucleotides. The protein can hydrolyze cAMP and cGMP, and may play a critical role in regulating cAMP and cGMP levels in the striatum[55], a region of the brain contributing to the control of movement and cognition. cAMP and cGMP both mediate the effects of dopamine D1 and D2 receptors on the activity of medium-sized spiny neurons[56]. Pharmacological inhibition of PDE10A increases cAMP and cGMP levels; and increases striatal output[57].

IMMP2L (chr 7; flagged by the intronic *IMMPL2* SNP rs2529484)
Prioritized by: altered expression in skeletal muscle following resistance training; SMR skeletal muscle (lenient); Activity by contact in coronary artery and liver;
Inner Mitochondrial Membrane Peptidase Subunit 2 resides in the mitochondria and is required for the catalytic activity of the mitochondrial inner membrane peptidase (IMP) complex. It catalyzes the removal of transit peptides required to transport proteins from the mitochondrial matrix, across the inner membrane, to the intermembrane space[58].

EXOC4 (chr 7; flagged by the intronic lead SNP rs13235840 in *EXOC4* and *LOC101928861*)
Prioritized by: altered expression in skeletal muscle following resistance training; DEPICT gene prioritization;
Exocyst Complex Component 4 is part of the highly conserved exocyst complex that is essential for targeting exocytic vesicles to specific docking sites on the plasma membrane. Exocyst Complex Component 4 participates in GLUT4 translocation and docking to the plasma membrane[59], and is essential for insulin-stimulated glucose uptake in skeletal muscle[59].

MKRN1 (chr 7; flagged by the intronic *MKRN1* lead SNP rs17621391)
Prioritized by: DEPICT gene prioritization (lenient); Activity by contact in pancreas, adrenal gland, bipolar neuron from iPSC, skeletal muscle, hepatocyte, ovary, thymus
Makorin ring finger protein 1 is thought to regulate RNA polymerase II-catalyzed transcription. It keeps cells alive by suppressing p53/TP53 under normal conditions, but stimulates apoptosis by repressing CDKN1A under stress.

BLK (chr 8, flagged by the intronic *XKR6* variant rs7821826)
Prioritized by: DEPICT tissue enrichment; SMR blood (lenient)
BLK Proto-Oncogene, Src Family Tyrosine Kinase is a nonreceptor tyrosine-kinase of the Src family of proto-oncogenes that are typically involved in cell proliferation and differentiation. Mutations at the *BLK* locus have been linked to Maturity-onset diabetes of the young (MODY) and β-cell dysfunction[60]. In pancreatic islets, it acts as a modulator of beta-cells function through the up-regulation of PDX1 and NKX6-1 and consequent stimulation of insulin secretion in response to glucose [60].

PACS1 (chr 11, flagged by the intronic *PACS1* variant rs4483592)
Prioritized by: SMR brain; DEPICT gene prioritization
Phosphofurin Acidic Cluster Sorting Protein 1 is involved in the trans-Golgi-membrane traffic[61]. A *de novo* mutation in *PACS1* was recently shown to cause defective migration of cranial-neural-crest cells and resulted in an intellectual disability syndrome and global developmental delay[62].

KLC2 (chr 11, flagged by the intronic *PACS1* variant rs4483592)
Prioritized by: SMR blood and skeletal muscle; DEPICT gene prioritization and tissue enrichment; Activity by contact in derived neural progenitors, adipose tissue, cardiac ventricle, hepatocytes and fetal thymus
KCL2 encodes Kinesin Light Chain 2, a light chain of kinesin and molecular motor responsible for moving vesicles and organelles along microtubules. Defects in this gene cause the rare, autosomal recessive mendelian disorder Spastic Paraplegia, Optic Atropy, and Neuropathy (SPOAN) Syndrome[63]. This syndrome is characterized by an early-onset, progressive weakness and spasticity of the legs. Zebrafish embryos with morpholino-mediated downregulation of *klc2* had a dose-dependent, shortened, twisted tail and were unable to swim. A similar motor phenotype was observed in zebrafish embryos upon upregulation of *klc2*[64].

RAB1B (chr11, flagged by the intronic *PACS1* variant rs4483592)
Prioritized by: DEPICT gene prioritization (lenient); Activity by contact in skeletal muscle, mytube, thymus
RAB1B member RAS oncogene Family functions in the early secretory pathway and is essential for vesicle transport between the endoplasmic reticulum and Golgi[65].

CNIH2 (chr11, flagged by the intronic *PACS1* variant rs4483592)
Prioritized by: DEPICT gene prioritization; Activity by contact in bipolar neuron from iPSC, thymus, hepatocyte

Cornichon family AMPA receptor auxiliary protein 2 mediates fast synaptic neurotransmission in the CNS and plays a role in assembly of hippocampal AMPA receptor complexes, thus modulating receptor gating and pharmacology.

TMEM151A (chr11, flagged by the intronic *PACS1* variant rs4483592)
Prioritized by: DEPICT gene prioritization; Activity by contact in bipolar neuron from iPSC, pancreas, hepatocyte
Transmembrane protein 151A is predicted to be an integral component of membranes.

MLF2 (chr12, flagged by rs3759344 upstream of *MLF2*)
Prioritized by: DEPICT gene prioritization (lenient); Activity by contact in derived neuronal progenitor cultured cells, pancreas, adipose tissue, adrenal gland, astrocyte, bipolar neuron from iPSC, cardiac muscle, coronary artery, skeletal muscle, cardiac ventricle, hepatocyte, myotube, ovary, thymus
Myeloid leukemia factor 2 is a membrane protein that is predicted to be involved in the regulation of transcription. Diseases associated with this gene include fatal infantile hypertonic myofibrillar myopathy

PTMS (chr12, flagged by rs3759344 upstream of *MLF2*)
Prioritized by: DEPICT tissue enrichment; Activity by contact in adrenal gland, astrocyte, pancreas, adipose tissue, cardiac muscle, skeletal muscle, hepatocyte, liver, myotube, thyroid
Parathymosin is predicted to be a nuclear protein that is involved in DNA replication and may mediate immune function.

COPS7A (chr12, flagged by rs3759344 upstream of *MLF2*)
Prioritized by: DEPICT gene prioritization (lenient); Activity by contact in cardiac muscle, liver
COP9 signalosome subunit 7A is a component of the COP9 signalosome complex that plays a role in various cellular and developmental processes.

MMAB (chr 12, flagged by the intronic *MYO1H* variant rs7969719)
Prioritized by: DEPICT tissue prioritization; SMR skeletal muscle (lenient)
MMAB (Metabolism Of Cobalamin Associated B) encodes a protein that catalyzes the final step in the conversion of vitamin B(12) into adenosylcobalamin (AdoCbl), a vitamin B12-containing coenzyme for methylmalonyl-CoA mutase. GWAS has reported variants in this gene to be associated with Apolipoprotein A1, HDL, and BMI, amongst others[66-68].

TESC (chr 12; flagged by the intronic *TESC* lead SNP rs2173650)
Prioritized by: SMR brain; association with daily running distance and average voluntary running speed in mice
Tescalcin, also known as Calcineurin B Homologous Protein 3 is highly expressed in the striatum[69], which harbours the central reward system and which represents a major site of physical activity regulation[70-72]. *TESC* encodes a Ca^{2+}- and Mg^{2+} -binding protein that is essential for extracellular signal-regulated kinase (ERK) cascade activation, which in turn is critical for normal cell differentiation[73], as well as for the motivating effects of reward-associated

stimuli along with other important roles related to learning, reinforcing and addiction in the striatum[74].

FBXO21 (chr 12, flagged by the intronic *TESC* lead SNP rs2173650)
Prioritized by: SMR brain (lenient); spontaneous running speed and distance in mice
FBXO21 encodes F-Box Protein 21, a member of the F-box protein family, fbxs, containing either different protein-protein interaction modules or no recognizable motifs. It is anticipated to play a role in the innate immune function.

ARL6IP4 (chr 12, flagged by the intronic *PITPNM2* indel rs541140319 (aka rs59131741))
Prioritized by: SMR blood; spontaneous running speed in mice (lenient)
ADP Ribosylation Factor Like GTPase 6 Interacting Protein 4 is involved in modulating alternative pre-mRNA splicing with either 5' distal site activation or preferential use of 3' proximal site.

OGFOD2 (chr 12, flagged by the intronic *PITPNM2* indel rs541140319 (aka rs59131741))
Prioritized by: SMR brain and blood; spontaneous running speed in mice (lenient)
Gene Ontology annotations related to *OGFOD2* (2-Oxoglutarate And Iron Dependent Oxygenase Domain Containing 2) include iron ion binding and oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, using 2-oxoglutarate as a donor, and incorporation of one atom each of oxygen into both donors.

CCDC92 (chr12, flagged by the intronic *PITPNM2* indel rs541140319 (aka rs59131741))
Prioritized by: SMR brain (lenient); SMR skeletal muscle (lenient); spontaneous running speed in mice (lenient)
CCDC92 is a coiled coil domain protein which interacts with proteins in the centriole/ciliary interface[75]. The *CCDC92* locus has been associated with higher insulin, higher triglyceride, and lower HDL-cholesterol levels. Further experimental studies showed that knockout of *CCDC92* resulted in less lipid accumulation in a mouse model. These results suggested a role for *CCDC92* in adipocyte differentiation[76].

FARP1 (chr 13; flagged by the intronic *FARP1* lead SNP rs9513416)
Prioritized by: altered expression in skeletal muscle following resistance training; DEPICT gene prioritization
FARP1 encodes FERM, ARH/RhoGEF And Pleckstrin Domain Protein 1, which promotes dendritic growth in neurons.

HERC1 (chr 15, flagged by the intronic *HERC1* variant rs12324720)
Prioritized by: DEPICT gene prioritization; altered expression in skeletal muscle following resistance training (lenient)
HECT And RLD Domain Containing E3 Ubiquitin Protein Ligase Family Member 1 encodes a protein that stimulates guanine nucleotide exchange on ARF1 and Rab proteins and may be involved in membrane transport processes via some guanine nucleotide exchange factor (GEF) activity and its ability to bind clathrin. HERC1 is involved in the ubiquitin-proteasome system, for

which the role in cachexia and sarcopenia is well-described[77]. An Intronic *HERC1* variant was associated with heel bone mineral density in UK BioBank data[78].

CBX4 (chr 17, flagged by the *CBX8* 3' UTR variant rs73420302)
Prioritized by: DEPICT gene prioritization; Activity by contact in pancreas
Chromobox 4 is involved in the negative regulation of transcription by RNA polymerase II.

CELF4 (chr 18, flagged by the intronic SNP rs12962050)
Prioritized by: DEPICT tissue enrichment and gene prioritization; Activity by contact in skeletal muscle, pancreas, adrenal gland, bipolar neuron from iPSC, coronary artery, hepatocyte
CUGBP Elav-like family member 4 belongs to a protein family that regulates pre-mRNA alternative splicing. It specifically activates exon 5 inclusion of cardiac isoforms of troponin 2 (TNNT2) during heart remodeling and the juvenile to adult transition[79].

YWHAB (chr 20, flagged by the intronic *YWHAB* indel rs139900206 (a.k.a. rs3838037))
Prioritized by: Finemapp & CADD score; Activity by contact in white adipose tissue; SMR brain (lenient); altered expression in skeletal muscle following resistance training (lenient)
YWHAB (Tyrosine 3-Monooxygenase/Tryptophan 5-Monooxygenase Activation Protein Beta, also known as 14-3-3 protein) encodes an adapter protein that is implicated in the regulation of a large spectrum of both general and specialized signaling pathways and plays a role in the cell cycle[80]. Previous proteomic analyses showed expression of YWHAB is upregulated in rat dorsal hippocampus following consumption of a diet high in fat and refined sugar[81], as well as in plasma after exercise[82].

STK4 (chr 20, flagged by the intronic *YWHAB* indel rs139900206 (a.k.a. rs3838037))
Prioritized by: SMR brain (lenient); Activity by contact in white adipose tissue
Serine/Threonine kinase 4 is a cytoplasmic, stress-activated kinase that can phosphorylate myelin basic protein and undergoes autophosphorylation. Following caspase-cleavage it enters the nucleus and induces chromatin condensation, followed by internucleosomal DNA fragmentation. The phosphorylation that is catalyzed by this protein has been associated with apoptosis.

ZBTB46 (chr 20, flagged by the intronic *ZBTB46* variant rs6010651)
Prioritized by: altered expression in skeletal muscle following resistance training; Activity by contact in bipolar neuron from iPSC; SMR blood (lenient)
Gene Ontology (GO) annotations for *ZBTB46* (Zinc Finger And BTB Domain Containing 46) include nucleic acid binding. *ZBTB46* functions as a transcriptional repressor for PRDM1 that mediates a transcriptional program in various immune tissue-resident lymphocyte T cell types.

STUDY SPECIFIC ACKNOWLEDGEMENTS AND FUNDING

AGES

Age, Gene/Environment Susceptibility-Reykjavik Study (AGES) was funded by NIH contract N01-AG-1-2100 and HHSN271201200022C, the NIA Intramural Research Program, Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament)

ALSPAC

The UK Medical Research Council and Wellcome (Grant ref: 217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors and Marcel den Hoed and Zhe Wang will serve as guarantors for the contents of this paper.
We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses.

ARIC

The Atherosclerosis Risk in Communities study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services (contract numbers HHSN268201700001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700004I and HHSN268201700005I), R01HL087641, R01HL059367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research.

B58C

The management of the 1958 Birth Cohort is funded by the Economic and Social Research Council (grant number ES/M001660/1). Access to these resources was enabled via the 58READIE Project funded by Wellcome Trust and Medical Research Council (grant numbers WT095219MA and G1001799). DNA collection was funded by MRC grant G0000934 and cell-line creation by Wellcome Trust grant 068545/Z/02. This study makes use of data generated by the Wellcome Trust Case-Control Consortium. A full list of investigators who contributed to generation of the data is available from the Wellcome Trust Case-Control Consortium website. Funding for the project was provided by the Wellcome Trust under the award 076113. This research used resources provided by the Type 1 Diabetes Genetics Consortium, a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases, National Human Genome Research Institute, National Institute of Child Health and Human Development, and Juvenile Diabetes Research Foundation International (JDRF) and supported by U01 DK062418.

BioMe

The Mount Sinai BioMe Biobank is supported by The Andrea and Charles Bronfman Philanthropies. We thank all participants and all our recruiters who have assisted and continue to assist in data collection and management. We are grateful for the computational resources and staff expertise provided by Scientific Computing at the Icahn School of Medicine at Mount Sinai.

BPROOF

We thank the participants of the B-PROOF study for their enthusiasm and cooperation. Furthermore, we thank the dedicated research team that conducted the study. This study is supported and funded by The Netherlands Organization for Health Research and Development (ZonMw, Grant 6130.0031), the Hague; unrestricted grant from NZO (Dutch Dairy Association), Zoetermeer; Orthica, Almere; NCHA (Netherlands Consortium Healthy Ageing) Leiden/Rotterdam; Ministry of Economic Affairs, Agriculture and Innovation (project KB-15-004-003), the Hague; Wageningen University, Wageningen; VUmc, Amsterdam; Erasmus Medical Center, Rotterdam; Unilever, Colworth, UK. All organisations, except Unilever, are based in the Netherlands. The sponsors have no role in the design or implementation of the study, data collection, data management, data analysis, data interpretation, or in the preparation, review, or approval of the manuscript.

Busselton

The Busselton Health Study acknowledges the generous support for the 1994/5 follow-up study from Healthway, Western Australia and the numerous Busselton community volunteers who assisted with data collection and the study participants from the Shire of Busselton. The Busselton Health Study is supported by The Great Wine Estates of the Margaret River region of Western Australia.

CARDIA

The Coronary Artery Risk Development in Young Adults Study (CARDIA) is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with the University of Alabama at Birmingham (HHSN268201800005I & HHSN268201800007I), Northwestern University (HHSN268201800003I), University of Minnesota (HHSN268201800006I), and Kaiser Foundation Research Institute (HHSN268201800004I). CARDIA was also partially supported by the Intramural Research Program of the National Institute on Aging (NIA) and an intra-agency agreement between NIA and NHLBI (AG0005). Genotyping was funded as part of the NHLBI Candidate-gene Association Resource (N01-HC-65226) and the NHGRI Gene Environment Association Studies (GENEVA) (U01-HG004729, U01-HG04424, and U01-HG004446). This manuscript has been reviewed and approved by CARDIA for scientific content.

CHS

Cardiovascular Health Study: This CHS research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086,

75N92021D00006; and NHLBI grants U01HL080295, R01HL085251, R01HL087652, R01HL105756, R01HL103612, R01HL120393, and U01HL130114 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through R01AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org.
The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.
The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

CoLaus

The CoLaus study was and is supported by research grants from GlaxoSmithKline, the Faculty of Biology and Medicine of Lausanne, and the Swiss National Science Foundation (grants 33CSCO-122661, 33CS30-139468, 33CS30-148401 and 33CS30_177535/1).

DNBC

The Danish National Birth Cohort (DNBC) is a result of major grants from the Danish National Research Foundation, the Danish Pharmacists' Fund, the Egmont Foundation, the March of Dimes Birth Defects Foundation, the Augustinus Foundation, and the Health Fund of the Danish Health Insurance Societies. The DNBC biobank is a part of the Danish National Biobank resource, which is supported by the Novo Nordisk Foundation. The generation of GWAS genotype data for the DNBC samples was carried out within the Gene Environment Association Studies (GENEVA) consortium with funding provided through the National Institutes of Health's Genes, Environment, and Health Initiative (U01HG004423; U01HG004446; U01HG004438). We are very grateful to all DNBC families who took part in the study. We would also like to thank everyone involved in data collection and biological material handling.

EGCUT

This research was supported by the European Union through the European Regional Development Fund (Project No. 2014-2020.4.01.15-0012), by the Estonian Research Council grant PUT (PRG687), by the Estonian Research Council grant PUT (PRG1291). We acknowledge the Estonian Biobank research team. Data analyses were carried out in part in the High-Performance Computing Center of University of Tartu.

EPIC-Norfolk

We are grateful to all the participants who have been part of the EPIC-Norfolk project and to the many members of the study teams at the University of Cambridge who have enabled this research. The study is funded by Medical Research Council (MR/N003284/1, MC-UU_12015/1, MC_PC_13048) Cancer Research UK (C864/A14136).

FamHS

The FamHS is funded by R01HL118305 and R01HL117078 NHLBI grants, and 5R01DK07568102 and 5R01DK089256 NIDDK grant.

FAMILY

We would like to thank all the participants of the FAMILY study. We acknowledge the internal support from the Population Health Research Institute for centralizing the data collect. The FAMILY genetic study has been funded by the Heart and Stroke Foundation of Ontario (grant # NA 7293 “Early genetic origins of cardiovascular risk factors”). The FAMILY study was funded by the Hamilton Health Science Foundation, the Canadian Institutes of Health Research and by Heart & Stroke Foundation of Ontario as well as additional grants from the Population Health Research Institute internal funds.

Fenland

The Fenland Study (10.22025/2017.10.101.00001) is funded by the Medical Research Council (MC_UU_12015/1). We are grateful to all the volunteers and to the General Practitioners and practice staff for assistance with recruitment. We thank the Fenland Study Investigators, Fenland Study Co-ordination team and the Epidemiology Field, Data and Laboratory teams. We further acknowledge support for genomics from the Medical Research Council (MC_PC_13046).

FHS

The Framingham Heart Study is funded by 75N92019D00031

FUSION

Support for FUSION was provided by NIH grants R01-DK062370 (to M.B.) and intramural project number 1Z01-HG000024 (to F.S.C.). Genome-wide genotyping was conducted by the Johns Hopkins University Genetic Resources Core Facility SNP Center at the Center for Inherited Disease Research (CIDR), with support from CIDR NIH contract no. N01-HG-65403.

GENOA

Genetic Epidemiology Network of Arteriosclerosis (GENOA) was supported by the National Institutes of Health grant numbers HL054457, HL054464, HL054481, HL087660, and HL119443 from the National Heart, Lung, and Blood Institute. Genotyping was performed at the Mayo Clinic by Stephan T. Turner, MD, Mariza de Andrade PhD, Julie Cunningham, PhD. We thank Eric Boerwinkle, PhD and Megan L. Grove from the Human Genetics Center and Institute of Molecular Medicine and Division of Epidemiology, University of Texas Health Science Center, Houston, Texas, USA for their help with genotyping. We would also like to thank the families that participated in the GENOA study.

GEOS/BWYSS

GEOS/BWYSS was supported by the Department of Veterans Affairs Biomedical Laboratory Research and Development Service, the Centers for Disease Control and Prevention, and the National Institutes of Health (Grants: R01 NS45012 and R01 NS105150).

GOOD

The GOOD study was funded by the Swedish Research Council, the Swedish state under the agreement between the Swedish government and the county councils, the ALF-agreement, the Lundberg Foundation, the Torsten Söderberg Foundation, the Novo Nordisk Foundation and the Knut and Alice Wallenberg Foundation.

GOYA

The GOYA study was conducted as part of the activities of the Danish Obesity Research Center(DanORC,www.danorc.dk) and the MRC centerfor Causal Analyses in Translational Epidemiology (MRC CAiTE).

GRAPHIC

The GRAPHIC Study was funded by the British Heart Foundation (RG/200004).

HANDLS

The Healthy Aging in Neighborhoods of Diversity across the Life Span study is supported by the National Institute on Aging Intramural Research Program, NIH Project number AG000513. We thank the HANDLS participants for agreeing to donate samples for the study. We also recognize the HANDLS medical staff for their careful evaluation of study participants.

HCHS/SOL

The authors thank the staff and participants of HCHS/SOL for their important contributions. Investigators website - http://www.cscc.unc.edu/hchs/ The Hispanic Community Health Study/Study of Latinos is a collaborative study supported by contracts from the National Heart, Lung, and Blood Institute (NHLBI) to the University of North Carolina (HHSN268201300001I / N01-HC-65233), University of Miami (HHSN268201300004I / N01-HC-65234), Albert Einstein College of Medicine (HHSN268201300002I / N01-HC-65235), University of Illinois at Chicago – HHSN268201300003I / N01-HC-65236 Northwestern Univ), and San Diego State University (HHSN268201300005I / N01-HC-65237).

The following Institutes/Centers/Offices have contributed to the HCHS/SOL through a transfer of funds to the NHLBI: National Institute on Minority Health and Health Disparities, National Institute on Deafness and Other Communication Disorders, National Institute of Dental and Craniofacial Research, National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of Neurological Disorders and Stroke, NIH Institution-Office of Dietary Supplements. The Genetic Analysis Center at the University of Washington was supported by NHLBI and NIDCR contracts (HHSN268201300005C AM03 and MOD03)

Health 06

The Health2006 study was financially supported by grants from the Velux Foundation; the Danish Medical Research Council, Danish Agency for Science, Technology and Innovation; the Aase and Ejner Danielsens Foundation; ALK-Abello´ A/S (Hørsholm, Denmark), Timber Merchant Vilhelm Bangs Foundation, MEKOS Laboratories (Denmark) and Research Centre for Prevention and Health, the Capital Region of Denmark. This project was also supported by the Lundbeck Foundation and produced by The Lundbeck Foundation Centre for Applied Medical

Genomics in Personalized Disease Prediction, Prevention and Care (LuCamp, www.lucamp.org). The Novo Nordisk Foundation Center for Basic Metabolic Research is an independent Research Center at the University of Copenhagen partially funded by an unrestricted donation from the Novo Nordisk Foundation (NNF18CC0034900). Further funding came from the Danish Council for Independent Research (Medical Sciences).

HPFS (Health Professional Follow-up Study)

The HPFS and NHS are funded by the National Institute of Deafness and Other Communication Disorders (R03 DC013373 to M.C.C). The NHS (UM1 CA186107) and HPFS (U01 CA167552) are additionally supported by the National Cancer Institute. We thank all participants of the NHS and HPFS for their continued contributions to research.

HRS (Health and retirement study)

HRS is supported by the National Institute on Aging (NIA U01AG009740). The genotyping was funded separately by the National Institute on Aging (RC2 AG036495, RC4 AG039029). Our genotyping was conducted by the NIH Center for Inherited Disease Research (CIDR) at Johns Hopkins University. Genotyping quality control and final preparation of the data was performed by the Genetics Coordinating Center at the University of Washington.

HUFS

The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Human Genome Research Institute; the National Institutes of Health; or the U.S. Department of Health and Human Services. This project was supported in part by the Intramural Research Program of the National Human Genome Research Institute of the National Institutes of Health through the Center for Research on Genomics and Global Health (CRGGH). The CRGGH is supported by the National Human Genome Research Institute, the National Institute of Diabetes and Digestive and Kidney Diseases, the Center for Information Technology, and the Office of the Director at the National Institutes of Health (Z01HG200362).

Inchianti / BLSA

The InCHIANTI study baseline (1998-2000) was supported as a "targeted project" (ICS110.1/RF97.71) by the Italian Ministry of Health and in part by the U.S. National Institute on Aging (Contracts: 263 MD 9164 and 263 MD 821336).
The study protocol for both studies were reviewed and approved by the Internal Review Board of the National Institute for Environmental Health Sciences (NIEHS) and all participants provided written informed consent.

Inter99

This project was funded by the Lundbeck Foundation and produced by The Lundbeck Foundation Centre for Applied Medical Genomics in Personalised Disease Prediction, Prevention and Care (LuCamp, www.lucamp.org). The Novo Nordisk Foundation Center for Basic Metabolic Research is an independent Research Center at the University of Copenhagen

partially funded by an unrestricted donation from the Novo Nordisk Foundation (NNF18CC0034900). Further funding came from the Danish Council for Independent Research (Medical Sciences). The Inter99 study was initiated by Torben Jørgensen (PI), Knut Borch-Johnsen (co-PI), Hans Ibsen, and Troels F. Thomsen. The steering committee comprises the former two and Charlotta Pisinger. The study was financially supported by research grants from the Danish Research Council, the Danish Centre for Health Technology Assessment, Novo Nordisk Inc., Research Foundation of Copenhagen County, Ministry of Internal Affairs and Health, the Danish Heart Foundation, the Danish Pharmaceutical Association, the Augustinus Foundation, the Ib Henriksen Foundation, the Becket Foundation, and the Danish Diabetes Association. We are indebted to the staff and participants of the Inter99, cohort for their important contributions. The authors wish to thank the staff at Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Denmark: A. Forman, T. Lorentzen, B. Andreasen and G. J. Klavsen for technical assistance and A. L. Nielsen, P. Sandbeck and G. Lademann for management assistance.

INTERSTROKE

This study was supported by Canadian Stroke Network, Canadian Institutes of Health Research, and Heart & Stroke Foundation

KARE

This work was supported by the Bio-Synergy Research Project (2013M3A9C4078158) of the Ministry of Science, ICT and Future Planning through the National Research Foundation

KORA 3/4

The KORA study was initiated and financed by the Helmholtz Zentrum München – German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria.

Lifelines Cohort Study

The LifeLines Cohort Study, and generation and management of GWAS genotype data for the LifeLines Cohort Study is supported by the Netherlands Organization of Scientific Research NWO (grant 175.010.2007.006), the Economic Structure Enhancing Fund (FES) of the Dutch government, the Ministry of Economic Affairs, the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the Northern Netherlands Collaboration of Provinces (SNN), the Province of Groningen, University Medical Center Groningen, the University of Groningen, Dutch Kidney Foundation and Dutch Diabetes Research Foundation. The authors wish to acknowledge the services of the Lifelines Cohort Study, the contributing research centers delivering data to Lifelines, and all the study participants.

LifeLines Cohort Study authors at the *University of Groningen, University Medical Center Groningen, The Netherlands* are:
Behrooz Z Alizadeh, H Marike Boezen and Harold Snieder (*Department of Epidemiology*); Lude Franke, Morris Swertz and Cisca Wijmenga (*Department of Genetics*); Pim van der Harst (*Department of Cardiology*); Gerjan Navis (*Department of Internal Medicine, Division of*

Nephrology); Marianne Rots (*Department of Medical Biology*); and Bruce HR Wolffenbuttel (*Department of Endocrinology*)

Malmö DC (MDC)

This study was supported by the European Research Council (Consolidator grant nr 649021, Orho-Melander), the Swedish Research Council, the Swedish Heart and Lung Foundation, the Novo Nordic Foundation, the Swedish Diabetes Foundation, and the Påhlsson Foundation, and by equipment grants from the Knut and Alice Wallenberg Foundation, the Region Skåne, Skåne University Hospital, the Linneus Foundation for the Lund University Diabetes Center and Swedish Foundation for Strategic Research for IRC15-0067.

MEC

The Multiethnic Cohort Study was supported by the National Cancer Institute (U01CA164973)

MESA

MESA and the MESA SHARe project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420, UL1-TR-001881, and DK063491. This publication was developed under Science to Achieve Results (STAR) research assistance agreements, No. RD831697 (MESA Air), and RD-83830001 (MESA Air Next Stage), awarded by the U.S Environmental protection Agency. It has not been formally reviewed by the EPA. The views expressed in this document are solely those of the authors and the EPA does not endorse any products or commercial services mentioned in this publication.

METSIM

The METSIM study was funded by the Academy of Finland (grants no. 77299 and 124243).

NBS1

The Nijmegen Biomedical Study was funded by the Radboud university medical center

NESDA

The infrastructure for the NESDA study (www.nesda.nl) is funded through the Geestkracht program of the Netherlands Organisation for Health Research and Development (ZonMw, grant number 10-000-1002) and financial contributions by participating universities and mental health care organizations (VU University Medical Center, GGZ inGeest, Leiden University Medical Center, Leiden University, GGZ Rivierduinen, University Medical Center Groningen, University of Groningen, Lentis, GGZ Friesland, GGZ Drenthe, Rob Giel Onderzoekscentrum).

NFBC66

We thank all cohort members and researchers who participated in the 31 yrs study. We also wish acknowledge the work of the NFBC project center. NFBC1966 received financial support

from University of Oulu Grant no. 65354, Oulu University Hospital Grant no. 2/97, 8/97, Ministry of Health and Social Affairs Grant no. 23/251/97, 160/97, 190/97, National Institute for Health and Welfare, Helsinki Grant no. 54121, Regional Institute of Occupational Health, Oulu, Finland Grant no. 50621, 54231.

NHS (Nurse Health Study)

The HPFS and NHS are funded by the National Institute of Deafness and Other Communication Disorders (R03 DC013373 to M.C.C). The NHS (UM1 CA186107) and HPFS (U01 CA167552) are additionally supported by the National Cancer Institute. We thank all participants of the NHS and HPFS for their continued contributions to research.

NSPHS

We acknowledge all the participants and staff involved in NSPHS for their valuable contribution. The NSPHS was funded by the Foundation for Strategic Research and the European Commission FP6. The computations and data handling were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX). This work was also funded by the Swedish Research Council 2019-01497, and the Swedish Heart Lung Foundation nr. 20200687.

NTR

Netherland Twin Register: Funding was obtained from the Netherlands Organization for Scientific Research (NWO) and The Netherlands Organisation for Health Research and Development (ZonMW) grants 904-61-090, 985-10-002, 912-10-020, 904-61-193,480-04-004, 463-06-001, 451-04-034, 400-05-717, Addiction-31160008, 016-115-035, 481-08-011, 400-07-080, 056-32-010, Middelgroot-911-09-032, OCW_NWO Gravity program –024.001.003, NWO-Groot 480-15-001/674, Center for Medical Systems Biology (CSMB, NWO Genomics), NBIC/BioAssist/RK(2008.024), Biobanking and Biomolecular Resources Research Infrastructure (BBMRI –NL, 184.021.007 and 184.033.111), X-Omics 184-034-019; Spinozapremie (NWO- 56-464-14192) and KNAW Academy Professor Award (PAH/6635); Amsterdam Public Health research institute (former EMGO+) , Neuroscience Amsterdam research institute (former NCA) ; the European Community's Fifth and Seventh Framework Program (FP5- LIFE QUALITY-CT-2002-2006, FP7- HEALTH-F4-2007-2013, grant 01254: GenomEUtwin, grant 01413: ENGAGE and grant 602768: ACTION); the European Research Council (ERC Starting 284167, ERC Consolidator 771057, ERC Advanced 230374), Rutgers University Cell and DNA Repository (NIMH U24 MH068457-06), the National Institutes of Health (NIH, R01D0042157-01A1, R01MH58799-03, MH081802, DA018673, R01 DK092127-04, Grand Opportunity grants 1RC2 MH089951, and 1RC2 MH089995); the Avera Institute for Human Genetics, Sioux Falls, South Dakota (USA). Part of the genotyping and analyses were funded by the Genetic Association Information Network (GAIN) of the Foundation for the National Institutes of Health.

ORCADES

The Orkney Complex Disease Study (ORCADES) was supported by the Chief Scientist Office of the Scottish Government (CZB/4/276, CZB/4/710), a Royal Society URF to J.F.W., the MRC Human Genetics Unit quinquennial programme “QTL in Health and Disease”, Arthritis Research UK and the European Union framework program 6 EUROSPAN project (contract no. LSHG-CT-2006-018947). DNA extractions were performed at the Wellcome Trust Clinical Research Facility in Edinburgh. We would like to acknowledge the invaluable contributions of the research nurses in Orkney, the administrative team in Edinburgh and the people of Orkney.

PELOTAS1982

The 1982 Pelotas Birth Cohort Study is conducted by the Postgraduate Program in Epidemiology at Universidade Federal de Pelotas with the collaboration of the Brazilian Public Health Association (ABRASCO). From 2004 to 2013, the Wellcome Trust supported the study. The International Development Research Center, World Health Organization, Overseas Development Administration, European Union, National Support Program for Centers of Excellence (PRONEX), the Brazilian National Research Council (CNPq), and the Brazilian Ministry of Health supported previous phases of the study.
Genotyping of 1982 Pelotas Birth Cohort Study participants was supported by the Department of Science and Technology (DECIT, Ministry of Health) and National Fund for Scientific and Technological Development (FNDCT, Ministry of Science and Technology), Funding of Studies and Projects (FINEP, Ministry of Science and Technology, Brazil), Coordination of Improvement of Higher Education Personnel (CAPES, Ministry of Education, Brazil).

QFS

The Quebec Family Study (QFS) was funded by multiple grants from the Medical Research Council of Canada and the Canadian Institutes for Health Research. This work was supported by a team grant from the Canadian Institutes for Health Research (FRCN-CCT-83028)

QIMR

The QIMR cohort consists of twins recruited to two studies conducted at QIMR Berghofer Medical Research Institute: the Over 50’s (Aged) and Alcohol Cohort 1 (AL1) Studies. Twins from the Aged Study were drawn from the Australian National Health and Medical Research Council (NHMRC) Twin Registry. This work was partly supported by a donation from Mr George Landers, and benefited from funding from NHMRC to Ian B. Hickie (931215-Project Grant, and 953208-Program Grant) and Nicholas G. Martin (941177). We thank Fran Boyle and Len Roberts for their work in constructing the questionnaire, Olivia Zheng for administering the mail-out, John Pearson for data management and Nirmala Pandeya for data cleaning. The AL1 study was carried out in co-operation with the Australian Twin Registry, and was supported in part by grants from NIAA (USA) AA07535, AA013320, AA013326, and NHMRC 941177, 951023, 950998, 981339, 241916 and 941944. We are extremely grateful to all the twins who took part in these studies.

SHIP & SHIP-TREND

SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants no. 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania, and the network 'Greifswald Approach to Individualized Medicine (GANI_MED)' funded by the Federal Ministry of Education and Research (grant 03IS2061A). Genome-wide data have been supported by the Federal Ministry of Education and Research (grant no. 03ZIK012) and a joint grant from Siemens Healthineers, Erlangen, Germany and the Federal State of Mecklenburg- West Pomerania. The University of Greifswald is a member of the Caché Campus program of the InterSystems GmbH.

Singapore Chinese Health Study (SCHS)

The Singapore Chinese Health Study is supported by the National Institutes of Health, USA (RO1 CA144034 and UM1 CA182876), the nested case-control study of myocardial infarction by the Singapore National Medical Research Council (NMRC 1270/2010) and genotyping by the HUJ-CREATE Programme of the National Research Foundation, Singapore (Project Number 370062002).

SP2

The Singapore Prospective Study Program (SP2) was funded through grants from the Biomedical Research Council of Singapore (BMRC) and the National Medical Research Council of Singapore (NMRC). Genome Institute of Singapore provided services for genotyping.

STR / Twingene

The Swedish Twin Registry is managed by Karolinska Institutet and receives funding through the Swedish Research Council under the grant no 2017-00641.

TRAILS

TRAILS (TRacking Adolescents' Individual Lives Survey) is a collaborative project involving various departments of the University Medical Center and University of Groningen, the University of Utrecht, the Radboud Medical Center Nijmegen, and the Parnassia Group, all in the Netherlands. TRAILS has been financially supported by various grants from the Netherlands Organization for Scientific Research NWO (Medical Research Council program grant GB-MW 940-38-011; ZonMW Brainpower grant 100-001-004; ZonMw Risk Behavior and Dependence grants 60-60600-97-118; ZonMw Culture and Health grant 261-98-710; Social Sciences Council medium-sized investment grants GB-MaGW 480-01-006 and GB-MaGW 480-07-001; Social Sciences Council project grants GB-MaGW 452-04-314 and GB-MaGW 452-06-004; NWO large-sized investment grant 175.010.2003.005; NWO Longitudinal Survey and Panel Funding 481-08-013 and 481-11-001; NWO Vici 016.130.002 and 453-16-007/2735; NWO Gravitation 024.001.003), the Dutch Ministry of Justice (WODC), the European Science Foundation (EuroSTRESS project FP-006), the European Research Council (ERC-2017-STG-757364 en ERC-CoG-2015-681466), Biobanking and Biomolecular Resources Research Infrastructure BBMRI-NL (CP 32), the Gratama foundation, the Jan Dekker foundation, the participating universities, and Accare Centre for Child and Adolescent Psychiatry. Statistical analyses were

carried out on the Genetic Cluster Computer (http://www.geneticcluster.org), which is financially supported by the Netherlands Scientific Organization (NWO 480-05-003) along with a supplement from the Dutch Brain Foundation. We are grateful to everyone who participated in this research or worked on this project to make it possible.

TwinsUK

TwinsUK is funded by the Wellcome Trust, Medical Research Council, European Union, the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London.

VIS

We would like to acknowledge the staff of several institutions in Croatia that supported the field work, including but not limited to The University of Split and Zagreb Medical Schools, the Institute for Antroplogical Research in Zagreb and the Croatian Institute for Public Health. We would like to acknowledge the invaluable contributions of the recruitment team in Korcula, the administrative teams in Croatia and Edinburgh and the participants. The SNP genotyping was performed in the core genotyping laboratory of the Wellcome Trust Clinical Research Facility at the Western General Hospital, Edinburgh, Scotland. The study was funded by the Medical Research Council UK, the European Union framework program 6 EUROSPAN project (contract no. LSHG-CT-2006-018947), the Croatian National Centre of Research Excellence in Personalized Healthcare grant (number KK.01.1.1.01.0010), and the Centre of Competence in Molecular Diagnostics (KK.01.2.2.03.0006).

WGHS

The WGHS is funded by grants from the NHLBI (HL043851 and HL080467) and NCI (CA047988 and UM1CA182913), with funding for genotyping provided by Amgen.

WHI

WHI: The WHI program 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, 75N92021D00005. A listing of WHI investigators can be found at: https://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Long%20List.pdf.

YFS

The Young Finns Study has been financially supported by the Academy of Finland: grants 322098, 286284, 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), and 41071 (Skidi); the Social Insurance Institution of Finland; Competitive State Research Financing of the Expert Responsibility area of Kuopio, Tampere and Turku University Hospitals (grant X51001); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research ; Finnish Cultural Foundation; The Sigrid Juselius Foundation; Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; Diabetes Research Foundation of Finnish Diabetes

Association; EU Horizon 2020 (grant 755320 for TAXINOMISIS; grant 848146 for To_Aition); European Research Council (grant 742927 for MULTIEPIGEN project); and Tampere University Hospital Supporting Foundation.
We thank the teams that collected data at all measurement time points; the persons who participated as both children and adults in these longitudinal studies; and biostatisticians Irina Lisinen, Johanna Ikonen, Noora Kartiosuo, Ville Aalto, and Jarno Kankaanranta for data management and statistical advice.

SUPPLEMENTARY NOTE REFERENCES

1. Pontzer, H., Wood, B.M. & Raichlen, D.A. Hunter-gatherers as models in public health. *Obes Rev* **19 Suppl 1**, 24-35 (2018).
2. Grossman, S.R. *et al.* Identifying recent adaptations in large-scale genomic data. *Cell* **152**, 703-13 (2013).
3. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499-503 (2015).
4. Field, Y. *et al.* Detection of human adaptation during the past 2000 years. *Science* **354**, 760-764 (2016).
5. Krueger, D.D., Osterweil, E.K. & Bear, M.F. Activation of mGluR5 induces rapid and long-lasting protein kinase D phosphorylation in hippocampal neurons. *J Mol Neurosci* **42**, 1-8 (2010).
6. Rhee, S.G. Regulation of phosphoinositide-specific phospholipase C. *Annu Rev Biochem* **70**, 281-312 (2001).
7. Zhu, X. *et al.* Phospholipase Cε deficiency delays the early stage of cutaneous wound healing and attenuates scar formation in mice. *Biochem Biophys Res Commun* **484**, 144-151 (2017).
8. Marck, A. *et al.* Age-Related Changes in Locomotor Performance Reveal a Similar Pattern for Caenorhabditis elegans, Mus domesticus, Canis familiaris, Equus caballus, and Homo sapiens. *J Gerontol A Biol Sci Med Sci* **72**, 455-463 (2017).
9. Norheim, F. *et al.* Gene-by-Sex Interactions in Mitochondrial Functions and Cardio-Metabolic Traits. *Cell metabolism* **29**, 932-949.e4 (2019).
10. Hillis, D.A. *et al.* Genetic Basis of Aerobically Supported Voluntary Exercise: Results from a Selection Experiment with House Mice. *Genetics* **216**, 781-804 (2020).
11. Battle, A., Brown, C.D., Engelhardt, B.E. & Montgomery, S.B. Genetic effects on gene expression across human tissues. *Nature* **550**, 204-213 (2017).
12. Lightfoot, J.T. Sex hormones' regulation of rodent physical activity: a review. *International journal of biological sciences* **4**, 126-132 (2008).
13. Nguyen, Q.A.T. *et al.* Coadaptation of the chemosensory system with voluntary exercise behavior in mice. *PLoS One* **15**, e0241758 (2020).
14. Panter, J. *et al.* Using alternatives to the car and risk of all-cause, cardiovascular and cancer mortality. *Heart* **104**, 1749-1755 (2018).
15. Kikuchi, H. *et al.* Occupational sitting time and risk of all-cause mortality among Japanese workers. *Scand J Work Environ Health* **41**, 519-28 (2015).
16. Vink, J.M. *et al.* Variance components models for physical activity with age as modifier: a comparative twin study in seven countries. *Twin Research and Human Genetics* **14**, 25-34 (2011).
17. Day, F.R., Loh, P.R., Scott, R.A., Ong, K.K. & Perry, J.R. A Robust Example of Collider Bias in a Genetic Association Study. *Am J Hum Genet* **98**, 392-3 (2016).
18. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biology* **17**, 122 (2016).
19. Gogarten, S.M. *et al.* Genetic association testing using the GENESIS R/Bioconductor package. *Bioinformatics* (2019).

20. Lusis, A.J. *et al.* The Hybrid Mouse Diversity Panel: a resource for systems genetics analyses of metabolic and cardiovascular traits. *J Lipid Res* **57**, 925-42 (2016).
21. Gautel, M. & Djinović-Carugo, K. The sarcomeric cytoskeleton: from molecules to motion. *J Exp Biol* **219**, 135-45 (2016).
22. Luther, P.K. The vertebrate muscle Z-disc: sarcomere anchor for structure and signalling. *J Muscle Res Cell Motil* **30**, 171-85 (2009).
23. Luther, P.K. Three-dimensional reconstruction of a simple Z-band in fish muscle. *J Cell Biol* **113**, 1043-55 (1991).
24. Grison, M., Merkel, U., Kostan, J., Djinović-Carugo, K. & Rief, M. α-Actinin/titin interaction: A dynamic and mechanically stable cluster of bonds in the muscle Z-disk. *Proc Natl Acad Sci U S A* **114**, 1015-1020 (2017).
25. Pickering, C. & Kiely, J. ACTN3: More than Just a Gene for Speed. *Frontiers in Physiology* **8**(2017).
26. Norman, B. *et al.* Strength, power, fiber types, and mRNA expression in trained men and women with different ACTN3 R577X genotypes. *J Appl Physiol (1985)* **106**, 959-65 (2009).
27. Mallakin, A. *et al.* Gene expression profiles of Mst1r-deficient mice during nickel-induced acute lung injury. *Am J Respir Cell Mol Biol* **34**, 15-27 (2006).
28. Wang, M.H. *et al.* Identification of the ron gene product as the receptor for the human macrophage stimulating protein. *Science* **266**, 117-9 (1994).
29. Nanney, L.B. *et al.* Proteolytic cleavage and activation of pro-macrophage-stimulating protein and upregulation of its receptor in tissue injury. *J Invest Dermatol* **111**, 573-81 (1998).
30. Feres, K.J., Ischenko, I. & Hayman, M.J. The RON receptor tyrosine kinase promotes MSP-independent cell spreading and survival in breast epithelial cells. *Oncogene* **28**, 279-88 (2009).
31. Klimentidis, Y.C. *et al.* Genome-wide association study of habitual physical activity in over 377,000 UK Biobank participants identifies multiple variants including CADM2 and APOE. *International Journal of Obesity* **42**, 1161-1176 (2018).
32. Scott, J.D., Dessauer, C.W. & Taskén, K. Creating order from chaos: cellular regulation by kinase anchoring. *Annu Rev Pharmacol Toxicol* **53**, 187-210 (2013).
33. Tingley, W.G. *et al.* Gene-trapped mouse embryonic stem cell-derived cardiac myocytes and human genetics implicate AKAP10 in heart rhythm regulation. *Proc Natl Acad Sci U S A* **104**, 8461-6 (2007).
34. Dessauer, C.W. Adenylyl cyclase--A-kinase anchoring protein complexes: the next dimension in cAMP signaling. *Mol Pharmacol* **76**, 935-41 (2009).
35. Berdeaux, R. & Stewart, R. cAMP signaling in skeletal muscle adaptation: hypertrophy, metabolism, and regeneration. *American journal of physiology. Endocrinology and metabolism* **303**, E1-E17 (2012).
36. Zhu, Q. *et al.* KDM4A regulates myogenesis by demethylating H3K9me3 of myogenic regulatory factors. *Cell Death Dis* **12**, 514 (2021).
37. Van Rompay, A.R., Johansson, M. & Karlsson, A. Identification of a novel human adenylate kinase. cDNA cloning, expression analysis, chromosome localization and characterization of the recombinant protein. *Eur J Biochem* **261**, 509-17 (1999).

38. Debray, F.G. *et al.* LRPPRC mutations cause a phenotypically distinct form of Leigh syndrome with cytochrome c oxidase deficiency. *J Med Genet* **48**, 183-9 (2011).
39. Song, S.-Y. *et al.* Environmental Enrichment Upregulates Striatal Synaptic Vesicle-Associated Proteins and Improves Motor Function. *Frontiers in neurology* **9**, 465-465 (2018).
40. Dickinson, D. *et al.* Differential Effects of Common Variants in SCN2A on General Cognitive Ability, Brain Physiology, and messenger RNA Expression in Schizophrenia Cases and Control Individuals. *JAMA Psychiatry* **71**, 647-656 (2014).
41. Tatsukawa, T. *et al.* Scn2a haploinsufficient mice display a spectrum of phenotypes affecting anxiety, sociability, memory flexibility and ampakine CX516 rescues their hyperactivity. *Molecular Autism* **10**, 15 (2019).
42. Kamura, T. *et al.* Cytoplasmic ubiquitin ligase KPC regulates proteolysis of p27(Kip1) at G1 phase. *Nat Cell Biol* **6**, 1229-35 (2004).
43. Hara, T. *et al.* Role of the UBL-UBA protein KPC2 in degradation of p27 at G1 phase of the cell cycle. *Mol Cell Biol* **25**, 9292-303 (2005).
44. Wang, S. *et al.* RNF123 has an E3 ligase-independent function in RIG-I-like receptor-mediated antiviral signaling. *EMBO Rep* **17**, 1155-68 (2016).
45. Johnston, K.J.A. *et al.* Genome-wide association study of multisite chronic pain in UK Biobank. *PLOS Genetics* **15**, e1008164 (2019).
46. Rahman, M.S. *et al.* Genome-wide association study identifies RNF123 locus as associated with chronic widespread musculoskeletal pain. *Ann Rheum Dis* (2021).
47. Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539-542 (2016).
48. Strawbridge, R.J. *et al.* Genome-wide analysis of self-reported risk-taking behaviour and cross-disorder genetic correlations in the UK Biobank cohort. *Transl Psychiatry* **8**, 39 (2018).
49. Clarke, T.K. *et al.* Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N=112 117). *Mol Psychiatry* **22**, 1376-1384 (2017).
50. Pasman, J.A. *et al.* GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. *Nat Neurosci* **21**, 1161-1170 (2018).
51. Locke, A.E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197-206 (2015).
52. Yan, X. *et al.* Cadm2 regulates body weight and energy homeostasis in mice. *Molecular metabolism* **8**, 180-188 (2018).
53. Lee, S.-H. *et al.* REEP5 depletion causes sarco-endoplasmic reticulum vacuolization and cardiac functional defects. *Nature Communications* **11**, 965 (2020).
54. Ichhaporia, V.P. *et al.* SIL1, the endoplasmic-reticulum-localized BiP co-chaperone, plays a crucial role in maintaining skeletal muscle proteostasis and physiology. *Disease Models & Mechanisms* **11**(2018).
55. Chen, J., Zook, D., Crickard, L. & Tabatabaei, A. Effect of phosphodiesterase (1B, 2A, 9A and 10A) inhibitors on central nervous system cyclic nucleotide levels in rats and mice. *Neurochem Int* **129**, 104471 (2019).

56. Greengard, P. The neurobiology of slow synaptic transmission. *Science* **294**, 1024-30 (2001).
57. Threlfell, S., Sammut, S., Menniti, F.S., Schmidt, C.J. & West, A.R. Inhibition of Phosphodiesterase 10A Increases the Responsiveness of Striatal Projection Neurons to Cortical Stimulation. *J Pharmacol Exp Ther* **328**, 785-95 (2009).
58. Burri, L. *et al.* Mature DIABLO/Smac is produced by the IMP protease complex on the mitochondrial inner membrane. *Mol Biol Cell* **16**, 2926-33 (2005).
59. Inoue, M., Chang, L., Hwang, J., Chiang, S.H. & Saltiel, A.R. The exocyst complex is required for targeting of Glut4 to the plasma membrane by insulin. *Nature* **422**, 629-33 (2003).
60. Borowiec, M. *et al.* Mutations at the BLK locus linked to maturity onset diabetes of the young and β-cell dysfunction. *Proceedings of the National Academy of Sciences* **106**, 14460-14465 (2009).
61. Wan, L. *et al.* PACS-1 defines a novel gene family of cytosolic sorting proteins required for trans-Golgi network localization. *Cell* **94**, 205-216 (1998).
62. Schuurs-Hoeijmakers, J.H.M. *et al.* Recurrent de novo mutations in PACS1 cause defective cranial-neural-crest migration and define a recognizable intellectual-disability syndrome. *American journal of human genetics* **91**, 1122-1127 (2012).
63. Macedo-Souza, L.I. *et al.* Spastic paraplegia, optic atrophy, and neuropathy is linked to chromosome 11q13. *Ann Neurol* **57**, 730-7 (2005).
64. Melo, U.S. *et al.* Overexpression of KLC2 due to a homozygous deletion in the non-coding region causes SPOAN syndrome. *Hum Mol Genet* **24**, 6877-85 (2015).
65. Alvarez, C., Garcia-Mata, R., Brandon, E. & Sztul, E. COPI recruitment is modulated by a Rab1b-dependent mechanism. *Mol Biol Cell* **14**, 2116-27 (2003).
66. Richardson, T.G. *et al.* Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease: A multivariable Mendelian randomisation analysis. *PLoS Med* **17**, e1003062 (2020).
67. Nielsen, J.B. *et al.* Loss-of-function genomic variants highlight potential therapeutic targets for cardiovascular disease. *Nature Communications* **11**, 6417 (2020).
68. Pulit, S.L. *et al.* Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum Mol Genet* **28**, 166-174 (2019).
69. Girard, F., Venail, J., Schwaller, B. & Celio, M.R. The EF-hand Ca2+-binding protein super-family: A genome-wide analysis of gene expression patterns in the adult mouse brain. *Neuroscience* **294**, 116-155 (2015).
70. Lightfoot, J.T. *et al.* Biological/Genetic Regulation of Physical Activity Level: Consensus from GenBioPAC. *Medicine and science in sports and exercise* **50**, 863-873 (2018).
71. Rhodes, J.S., Gammie, S.C. & Garland, T., Jr. Neurobiology of Mice Selected for High Voluntary Wheel-running Activity1. *Integrative and Comparative Biology* **45**, 438-455 (2005).
72. Roberts, M.D. *et al.* Nucleus accumbens neuronal maturation differences in young rats bred for low versus high voluntary running behaviour. *The Journal of physiology* **592**, 2119-2135 (2014).

73. Levay, K. & Slepak, V.Z. Tescalcin is an essential factor in megakaryocytic differentiation associated with Ets family gene expression. *J Clin Invest* **117**, 2672-83 (2007).
74. Shiflett, M.W. & Balleine, B.W. Contributions of ERK signaling in the striatum to instrumental learning and performance. *Behavioural brain research* **218**, 240-247 (2011).
75. Gupta, G.D. *et al.* A Dynamic Protein Interaction Landscape of the Human Centrosome-Cilium Interface. *Cell* **163**, 1484-99 (2015).
76. Lotta, L.A. *et al.* Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance. *Nature genetics* **49**, 17-26 (2017).
77. Bowen, T.S., Schuler, G. & Adams, V. Skeletal muscle wasting in cachexia and sarcopenia: molecular pathophysiology and impact of exercise training. *Journal of cachexia, sarcopenia and muscle* **6**, 197-207 (2015).
78. Kim, S.K. Identification of 613 new loci associated with heel bone mineral density and a polygenic risk score for bone mineral density, osteoporosis and fracture. *PLOS ONE* **13**, e0200785 (2018).
79. Wang, X. *et al.* CELF4 Variant and Anthracycline-Related Cardiomyopathy: A Children's Oncology Group Genome-Wide Association Study. *J Clin Oncol* **34**, 863-70 (2016).
80. Wilker, E. & Yaffe, M.B. 14-3-3 Proteins--a focus on cancer and human disease. *J Mol Cell Cardiol* **37**, 633-42 (2004).
81. Francis, H.M., Mirzaei, M., Pardey, M.C., Haynes, P.A. & Cornish, J.L. Proteomic analysis of the dorsal and ventral hippocampus of rats maintained on a high fat and refined sugar diet. *PROTEOMICS* **13**, 3076-3091 (2013).
82. Wei, Y. *et al.* Long-term moderate exercise enhances specific proteins that constitute neurotrophin signaling pathway: A TMT-based quantitative proteomic analysis of rat plasma. *Journal of Proteomics* **185**, 39-50 (2018).