

UC Riverside

UC Riverside Previously Published Works

Title

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

Permalink

<https://escholarship.org/uc/item/1xq2h3nb>

Journal

Nature Genetics, 54(9)

Authors

Wang, Zhe

Emmerich, Andrew

Pillon, Nicolas

et al.

Publication Date

2022-09-01

DOI

10.1038/s41588-022-01165-1

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed



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 levels¹. Despite efforts to increase physical activity levels², 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 sedentary⁴, 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 factors⁷. Factors like socioeconomic status, built environment and media all influence physical activity at a population level⁷. In parallel, innate biological factors (for example, age, sex hormones, pre-existing medical conditions, epigenetics and genetics) also explain a moderate proportion of the inter-individual 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 intervention¹⁰.

Genome-wide association studies (GWAS) have identified thousands of loci associated with cardiometabolic risk factors and diseases¹¹. However, similar efforts for physical activity have been sparse and initially had limited success. This likely reflects the comparatively small sample size of these efforts¹², 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.

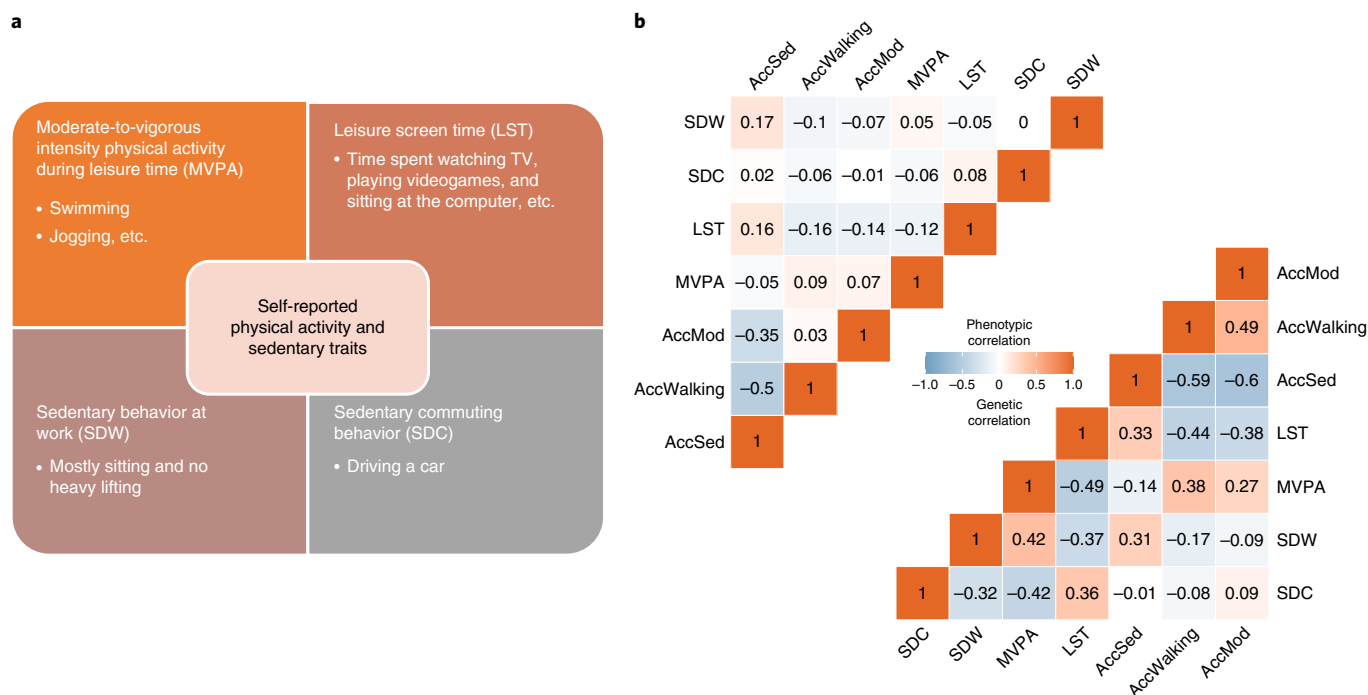


Fig. 1 | Overview of the four self-reported physical activity and sedentary traits and correlations with objectively assessed traits. a, An overview of the four self-reported physical activity and sedentary traits. **b**, Phenotypic (upper left) and genetic (lower right) correlation coefficients between 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 ($P = 5 \times 10^{-23}$ for LST; $P = 2 \times 10^{-3}$ 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 BioMe 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-Hub¹⁸, we observe significant ($P < 4.6 \times 10^{-4}$) 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 BioMe BioBank (Supplementary Table 9), and a phenome-wide association study (PheWAS) in 8,959 BioMe European ancestry samples shows a negative association between the PGS for MVPA and morbid obesity ($P = 1.1 \times 10^{-5}$, 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

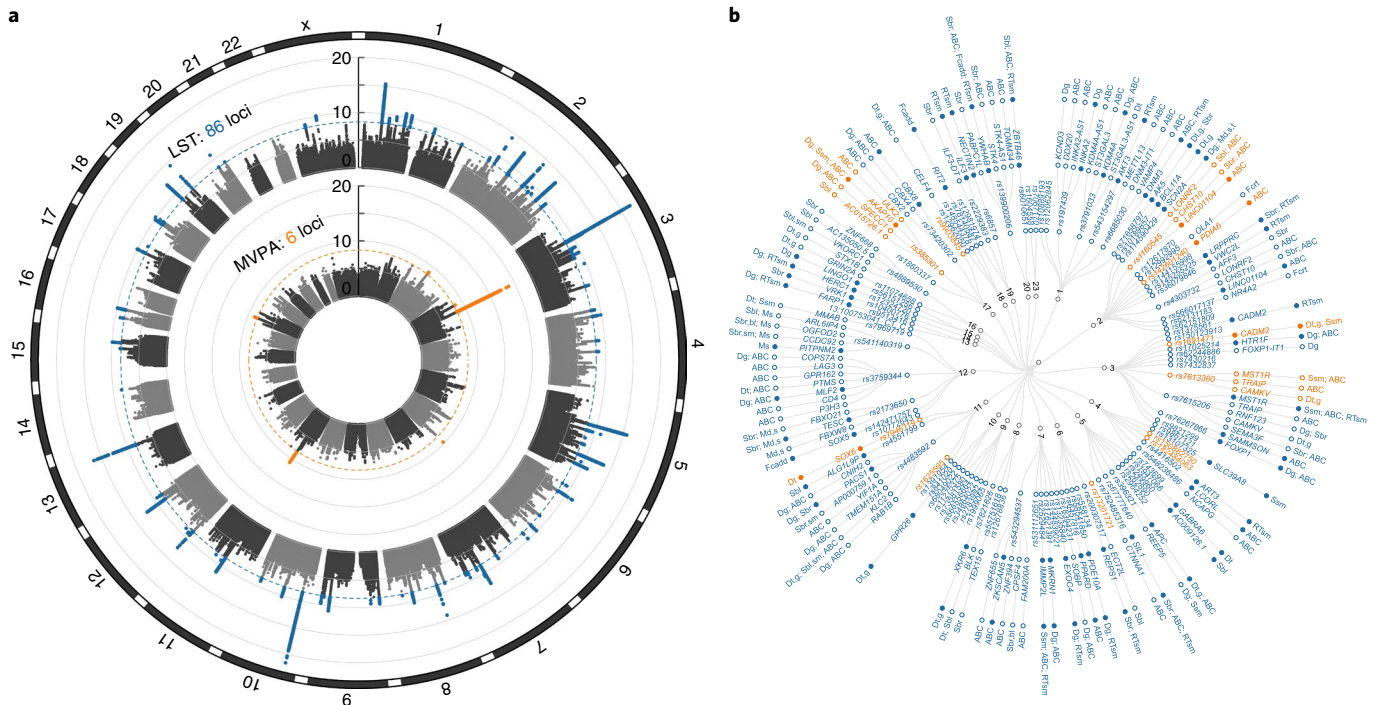


Fig. 2 | Main results of GWAS and downstream gene prioritization for LST and MVPA. **a**, Circular Manhattan plot summarizing the results from European ancestry meta-analyses for LST and MVPA. Outer track, LST; inner track, MVPA. Genome-wide significant variants ($P < 5 \times 10^{-9}$) are highlighted 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); 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 role³². 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 inactivity³³ (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 regulation³⁴, and in regulating striatum output³⁵; *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 muscle³⁸; and *IMMP2L* influences the transport of proteins across the inner mitochondrial membrane³⁹ (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 DEPICT⁴⁰. LST- and MVPA-associated loci ($P < 1 \times 10^{-5}$) 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 decline⁴¹. 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 model⁴². 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 CELLECT⁴³ to identify enriched cell types using single-cell RNA sequencing data from the Tabula Muris and mouse brain projects⁴⁴. 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 mice⁴⁵.

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 MR⁴⁶ (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 hits⁴⁹

(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 ($P_{\text{binomial}} = 2.1 \times 10^{-70}$; Supplementary Fig. 5 and Supplementary Table 27). Associations reach traditional genome-wide significance ($P < 5 \times 10^{-8}$) 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 MVPA¹³ and forms part of the $\epsilon 4$ 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 ($P = 1.9 \times 10^{-11}$) as well as with higher *ACTN3* expression in skeletal muscle (GTEx, $P = 6.6 \times 10^{-5}$). 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 fibers⁵⁸. 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 performance⁵⁷. Although we observe little evidence for a role

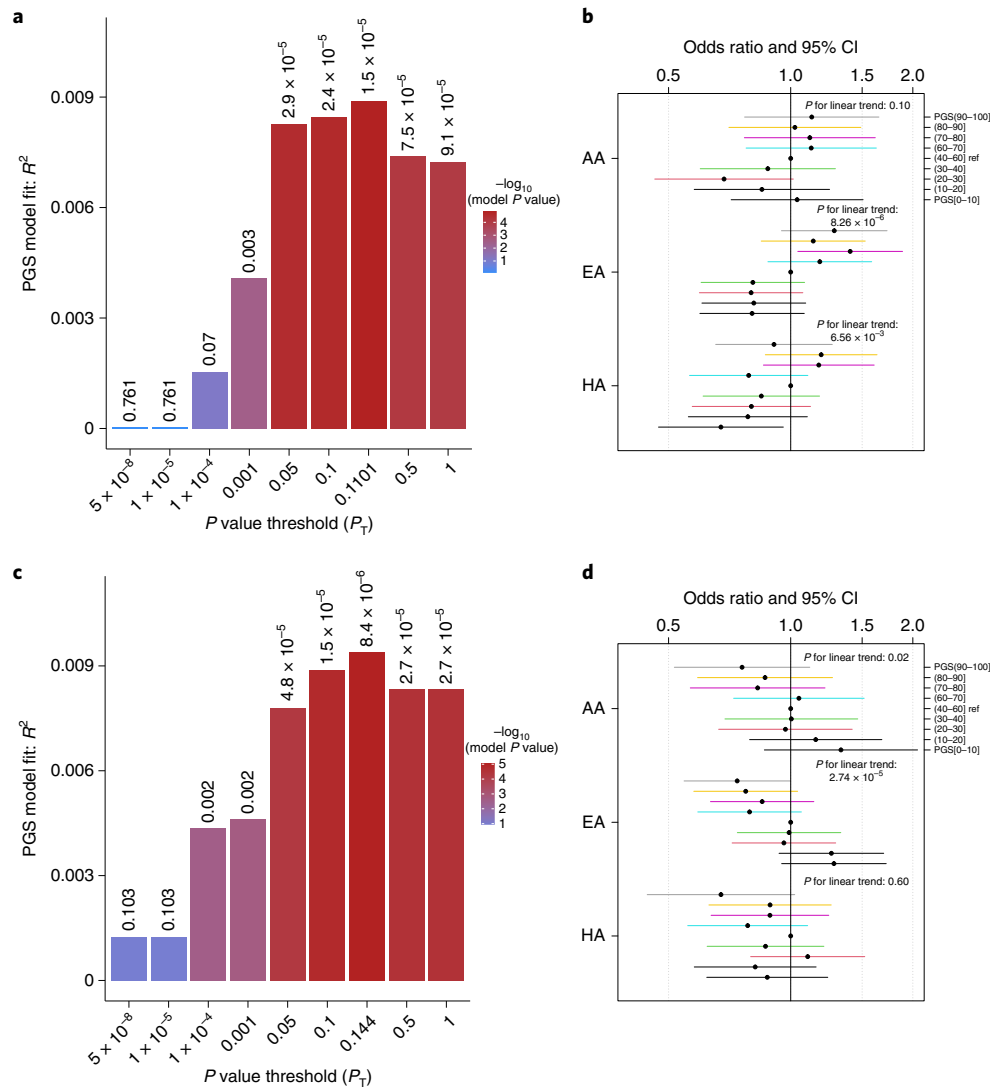


Fig. 3 | Validation of associations with MVPA and LST using PGSs in BioMe participants of three ancestries. a,c, The best performing PGSs for MVPA (a) and LST (c) were derived using logistic/linear regression analyses; that is, those with the highest incremental R^2 above and beyond models with only 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 BioMe 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 variant⁵⁹. 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

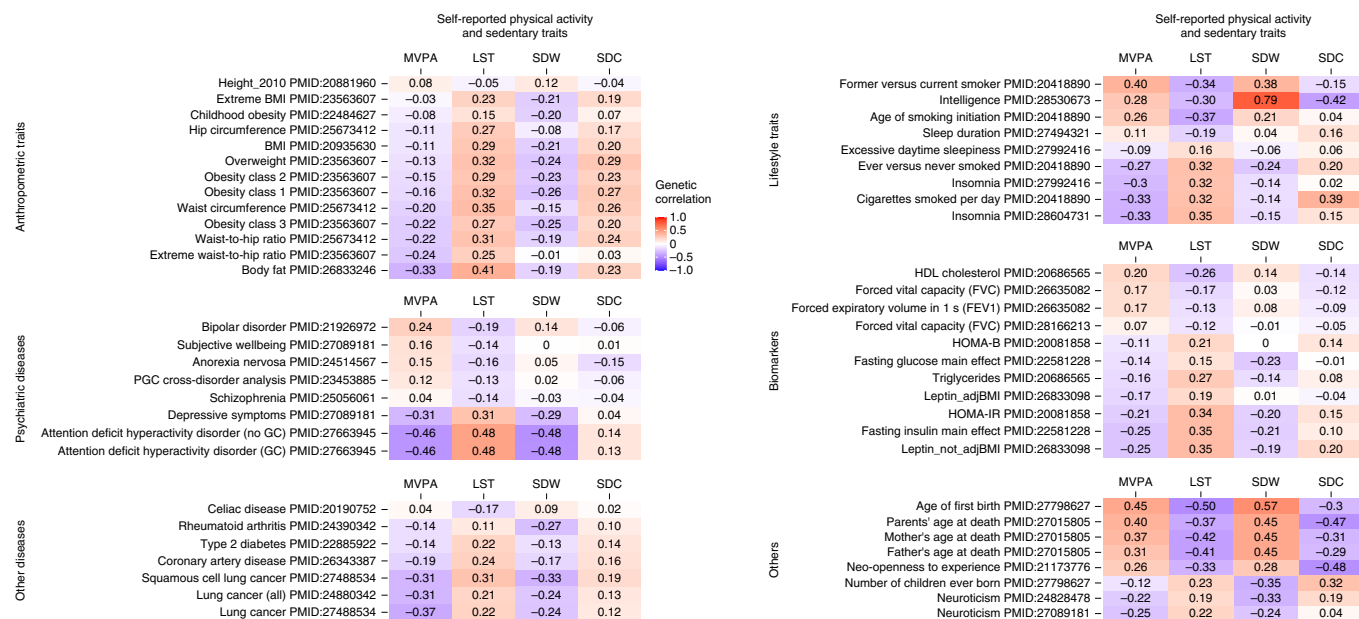


Fig. 4 | Genetic correlations of four self-reported physical activity traits with complex traits and diseases. Results are based on published GWAS with $P < 4.6 \times 10^{-4}$ for at least one physical activity or sedentary trait. Darker colors reflect higher negative (purple) or positive (red) correlation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

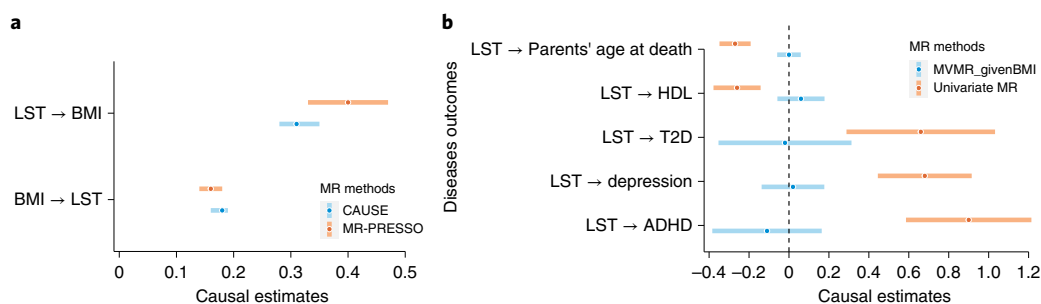


Fig. 5 | MR analyses between LST, MVPA, BMI and complex diseases. **a**, Median causal estimates for MR analyses using the CAUSE method and causal estimates from the MR-PRESSO method after outlier removal and accounting for horizontal pleiotropy. **b**, The causal effects of LST on complex risk factors 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 \pm 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 |

^aPosterior 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. ^bThe 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 effect⁶³. 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 reported⁶⁴. 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 BMI¹⁵. 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 estimates³¹.

It is of interest that a proxy of *rs429358*, part of the established *APOE* $\epsilon 4$ 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 individuals¹³. 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 life¹³. 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 dementia¹³ suggest other explanations are more likely, although a role for survival bias cannot be ruled out¹³. For example, *APOE* $\epsilon 4$ carriers have a greater increase in aerobic capacity following exercise training⁶⁵, which may reinforce a physically active lifestyle independently of Alzheimer's risk. Furthermore, several studies have investigated the moderating role of the *APOE* $\epsilon 4$ allele in the relationship between physical activity

and Alzheimer prevention⁶⁶. Although more studies are needed to resolve inconsistencies in the literature, $\epsilon 4$ carriers seem to benefit more from physical activity in terms of reducing the risk of dementia and brain pathology⁶⁶.

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 observed⁶⁷. 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 micro-trauma caused by Z-disc rupture or streaming³, alleviating delayed onset muscle soreness³ and risk of injuries³, 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: 7 September 2022

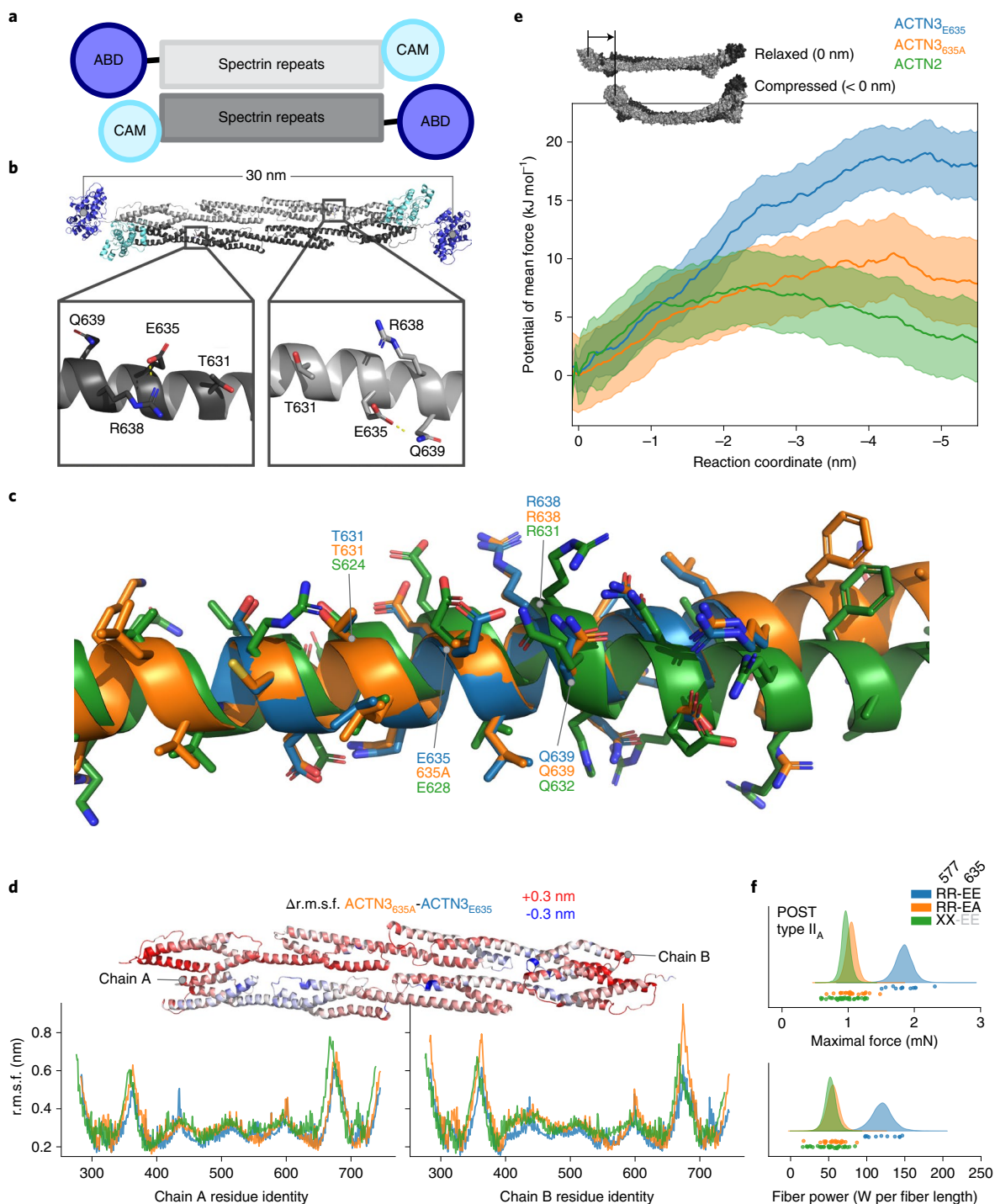


Fig. 6 | Allele p.635Ala in *ACTN3* results in a more flexible *ACTN3* homodimer. **a**, *ACTN3* is a homodimer of two antiparallel filaments, with each filament consisting of an N-terminal actin binding domain (ABD, blue), followed by a structural region comprised of four spectrin repeats (gray) with 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

- 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).
- Global Action Plan for the Prevention and Control of Noncommunicable Diseases 2013–2020* (World Health Organization, 2013).
- 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).
- 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).
- 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).
- 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).
- Bauman, A. E. et al. Correlates of physical activity: why are some people physically active and others not? *Lancet* **380**, 258–271 (2012).
- den Hoed, M. et al. Heritability of objectively assessed daily physical activity and sedentary behavior. *Am. J. Clin. Nutr.* **98**, 1317–1325 (2013).
- Stubbe, J. H. et al. Genetic influences on exercise participation in 37,051 twin pairs from seven countries. *PLoS ONE* **1**, e22 (2006).
- Fan, W. et al. PPAR δ promotes running endurance by preserving glucose. *Cell Metab.* **25**, 1186–1193.e4 (2017).
- 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).
- Sarzynski, M. A. et al. Advances in exercise, fitness, and performance genomics in 2015. *Med. Sci. Sports Exerc.* **48**, 1906–1916 (2016).
- 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).
- Doherty, A. et al. GWAS identifies 14 loci for device-measured physical activity and sleep duration. *Nat. Commun.* **9**, 5257 (2018).
- 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).
- Kilpeläinen, T. O. et al. Multi-ancestry study of blood lipid levels identifies four loci interacting with physical activity. *Nat. Commun.* **10**, 376 (2019).
- Turley, P. et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* **50**, 229–237 (2018).
- 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).
- Shungin, D. et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187–196 (2015).
- Kichaev, G. et al. Leveraging polygenic functional enrichment to improve GWAS power. *Am. J. Hum. Genet.* **104**, 65–75 (2019).
- 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).
- 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).
- 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).
- Locke, A. E. et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
- 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).
- 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).
- 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).
- Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37**, 658–665 (2013).
- 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).
- 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).
- 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).
- Lightfoot, J. T. et al. Biological/genetic regulation of physical activity level: consensus from GenBioPAC. *Med. Sci. Sports Exerc.* **50**, 863–873 (2018).
- Pillon, N. J. et al. Transcriptomic profiling of skeletal muscle adaptations to exercise and inactivity. *Nat. Commun.* **11**, 470 (2020).
- 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).
- 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).
- 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).
- Zhu, Y. et al. Identification of CD112R as a novel checkpoint for human T cells. *J. Exp. Med.* **213**, 167–176 (2016).
- 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).
- 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).
- Pers, T. H. et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
- Muñoz, 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).
- 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).
- Timshel, P. N., Thompson, J. J. & Pers, T. H. Genetic mapping of etiologic brain cell types for obesity. *eLife* **9**, e55851 (2020).
- Schaum, N. et al. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367–372 (2018).
- Roberts, M. D., Rueggsegger, 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).
- Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481 (2016).
- Benner, C. et al. FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493–1501 (2016).
- Kircher, M. et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
- Nasser, J. et al. Genome-wide enhancer maps link risk variants to disease genes. *Nature* **593**, 238–243 (2021).
- 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).
- de Geus, E. J., Bartels, M., Kaprio, J., Lightfoot, J. T. & Thoms, M. Genetics of regular exercise and sedentary behaviors. *Twin Res. Hum. Genet* **17**, 262–271 (2014).
- 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).
- Moir, H. J. et al. Genes and elite marathon running performance: a systematic review. *J. Sports Sci. Med.* **18**, 559–568 (2019).
- Kim, D. S., Wheeler, M. T. & Ashley, E. A. The genetics of human performance. *Nat. Rev. Genet.* **23**, 40–54 (2021).
- 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).
- 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).
- Pickering, C. & Kiely, J. ACTN3: more than just a gene for speed. *Front. Physiol.* **8**, 1080 (2017).
- Vincent, B. et al. ACTN3 (R577X) genotype is associated with fiber type distribution. *Physiol. Genomics* **32**, 58–63 (2007).
- 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).
- 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 Ila 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).

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



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

© The Author(s) 2022

Zhe Wang¹✉, Andrew Emmerich², Nicolas J. Pillon³, Tim Moore⁴, Daiane Hemerich¹, Marilyn C. Cornelis⁵, Eugenia Mazzaferro⁶, Siacia Broos^{7,8}, Tarunveer S. Ahluwalia^{9,10,11}, Traci M. Bartz^{12,13}, Amy R. Bentley¹⁴, Lawrence F. Bielak¹⁵, Mike Chong¹⁶, Audrey Y. Chu^{17,18}, Diane Berry¹⁹, Rajkumar Dorajoo^{20,21}, Nicole D. Dueker^{22,23}, Elisa Kasbohm^{24,25}, Bjarke Feenstra²⁶, Mary F. Feitosa²⁷, Christian Gieger²⁸, Mariaelisa Graff²⁹, Leanne M. Hall^{30,31}, Toomas Haller³², Fernando P. Hartwig^{33,34}, David A. Hillis³⁵, Ville Huikari³⁶, Nancy Heard-Costa^{37,38}, Christina Holzappel^{28,39}, Anne U. Jackson⁴⁰, Åsa Johansson⁴¹, Anja Moltke Jørgensen¹⁰, Marika A. Kaakinen^{42,43}, Robert Karlsson⁴⁴, Kathleen F. Kerr¹³, Boram Kim⁴⁵, Chantal M. Koolhaas⁴⁶, Zoltan Kutalik^{47,48,49}, Vasiliki Lagou⁵⁰, Penelope A. Lind^{51,52}, Mattias Lorentzon^{53,54}, Leo-Pekka Lyytikäinen^{55,56}, Massimo Mangino^{57,58}, Christoph Metzendorf⁶, Kristine R. Monroe⁵⁹, Alexander Pacolet⁷, Louis Pérusse^{60,61}, Rene Pool^{62,63}, Rebecca C. Richmond⁶⁴, Natalia V. Rivera^{65,66,67}, Sebastien Robiou-du-Pont⁶⁸, Katharina E. Schraut⁶⁹, Christina-Alexandra Schulz^{70,71}, Heather M. Stringham⁴⁰, Toshiko Tanaka⁷², Alexander Teumer^{24,73}, Constance Turman⁷⁴, Peter J. van der Most⁷⁵, Mathias Vanmunster⁷, Frank J. A. van Rooij⁴⁶, Jana V. van Vliet-Ostaptchouk^{76,77}, Xiaoshuai Zhang^{78,79}, Jing-Hua Zhao⁸⁰, Wei Zhao¹⁵, Zhanna Balkhiyarova^{43,81,82}, Marie N. Balslev-Harder¹⁰, Sebastian E. Baumeister^{24,83}, John Beilby⁸⁴, John Blangero⁸⁵, Dorret I. Boomsma^{62,63}, Soren Brage⁷⁸, Peter S. Braund^{30,31}, Jennifer A. Brody¹², Marcel Bruinenberg⁸⁶, Ulf Ekelund^{87,88}, Ching-Ti Liu⁸⁹, John W. Cole⁹⁰, Francis S. Collins⁹¹, L. Adrienne Cupples^{37,89}, Tõnu Esko³², Stefan Enroth⁴¹, Jessica D. Faul⁹², Lindsay Fernandez-Rhodes⁹³, Alison E. Fohner⁹⁴, Oscar H. Franco^{46,95}, Tessel E. Galesloot⁹⁶, Scott D. Gordon⁵¹, Niels Grarup¹⁰, Catharina A. Hartman⁹⁷, Gerardo Heiss²⁹, Jennie Hui^{84,98,99}, Thomas Illig^{100,101}, Russell Jago¹⁰², Alan James¹⁰³, Peter K. Joshi^{69,104}, Taeyeong Jung⁴⁵, Mika Kähönen^{56,105}, Tuomas O. Kilpeläinen¹⁰, Woon-Puay Koh^{106,107}, Ivana Kolcic¹⁰⁸, Peter P. Kraft⁷⁴, Johanna Kuusisto¹⁰⁹, Lenore J. Launer¹¹⁰, Aihua Li⁶⁸, Allan Linneberg^{111,112}, Jian'an Luan⁷⁸, Pedro Marques Vidal¹¹³, Sarah E. Medland^{51,114}, Yuri Milaneschi¹¹⁵, Arden Moscati¹, Bill Musk^{99,174}, Christopher P. Nelson^{30,31}, Ilja M. Nolte⁷⁵, Nancy L. Pedersen⁴⁴, Annette Peters¹¹⁶, Patricia A. Peyser¹⁵, Christine Power¹⁹, Olli T. Raitakari^{117,118,119}, Mägi Reedik³², Alex P. Reiner¹²⁰, Paul M. Ridker^{17,121}, Igor Rudan⁶⁹, Kathy Ryan¹²², Mark A. Sarzynski¹²³, Laura J. Scott⁴⁰, Robert A. Scott⁷⁸, Stephen Sidney¹²⁴, Kristin Siggeirsdottir¹²⁵, Albert V. Smith^{40,125}, Jennifer A. Smith^{15,92}, Emily Sonestedt⁷⁰, Marin Strøm^{26,126}, E. Shyong Tai^{127,128,129}, Koon K. Teo^{68,130}, Barbara Thorand¹¹⁶, Anke Tönjes¹³¹, Angelo Tremblay^{60,61}, Andre G. Uitterlinden¹³², Jagadish Vangipurapu¹⁰⁹, Natasja van Schoor¹³³, Uwe Völker^{73,134}, Gonneke Willemsen^{62,63}

Kayleen Williams¹³, Quenna Wong¹³, Huichun Xu¹²², Kristin L. Young²⁹, Jian Min Yuan^{135,136}, M. Carola Zillikens¹³², Alan B. Zonderman¹³⁷, Adam Ameer¹⁴¹, Stefania Bandinelli¹³⁸, Joshua C. Bis¹², Michael Boehnke⁴⁰, Claude Bouchard¹³⁹, Daniel I. Chasman^{17,121}, George Davey Smith^{34,140}, Eco J. C. de Geus^{62,63}, Louise Deldicque¹⁴¹, Marcus Dörr^{73,142}, Michele K. Evans¹³⁷, Luigi Ferrucci⁷², Myriam Fornage¹⁴³, Caroline Fox¹⁴⁴, Theodore Garland Jr¹⁴⁵, Vilmundur Gudnason^{125,146}, Ulf Gyllensten⁴¹, Torben Hansen¹⁰, Caroline Hayward¹⁴⁷, Bernardo L. Horta³³, Elina Hyppönen^{148,149,150}, Marjo-Riitta Jarvelin^{36,151}, W. Craig Johnson¹³, Sharon L. R. Kardia¹⁵, Lambertus A. Kiemeny⁹⁶, Markku Laakso¹⁰⁹, Claudia Langenberg^{78,152}, Terho Lehtimäki^{55,56}, Loic Le Marchand¹⁵³, Lifelines Cohort Study^{*}, Patrik K. E. Magnusson⁴⁴, Nicholas G. Martin⁵¹, Mads Melbye^{154,155,156,112}, Andres Metspalu³², David Meyre^{16,68}, Kari E. North²⁹, Claes Ohlsson^{157,158}, Albertine J. Oldehinkel⁹⁷, Marju Orho-Melander⁷⁰, Guillaume Pare¹⁶, Taesung Park^{45,159}, Oluf Pedersen¹⁰, Brenda W. J. H. Penninx¹¹⁵, Tune H. Pers¹⁰, Ozren Polasek¹⁶⁰, Inga Prokopenko^{81,82,161}, Charles N. Rotimi¹⁴, Nilesh J. Samani^{30,31}, Xueling Sim¹²⁷, Harold Snieder⁷⁵, Thorkild I. A. Sørensen^{10,162}, Tim D. Spector⁵⁷, Nicholas J. Timpson¹⁶³, Rob M. van Dam^{127,164}, Nathalie van der Velde^{132,165,166}, Cornelia M. van Duijn^{46,167}, Peter Vollenweider¹¹³, Henry Völzke^{24,73}, Trudy Voortman⁴⁶, Gérard Waeber¹¹³, Nicholas J. Wareham⁷⁸, David R. Weir⁹², Heinz-Erich Wichmann¹¹⁶, James F. Wilson^{69,147}, Andrea L. Hevener¹⁶⁸, Anna Krook³, Juleen R. Zierath^{3,10,169}, Martine A. I. Thomis⁸, Ruth J. F. Loos^{1,10,170} and Marcel den Hoed⁶ ✉

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

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

Health, Imperial College London, London, UK. ¹⁵²Computational Medicine, Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Berlin, Germany. ¹⁵³Epidemiology Program, University of Hawaii Cancer Center, University of Hawaii at Manoa, Honolulu, HI, USA. ¹⁵⁴K.G.Jebsen Center for Genetic Epidemiology, Norwegian University of Science and Technology, Trondheim, Norway. ¹⁵⁵Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA. ¹⁵⁶Center for Fertility and Health, Norwegian Institute of Public Health, Oslo, Norway. ¹⁵⁷Centre for Bone and Arthritis Research, Department of Internal Medicine and Clinical Nutrition, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden. ¹⁵⁸Department of Drug Treatment, Sahlgrenska University Hospital, Gothenburg, Sweden. ¹⁵⁹Department of Statistics, Seoul National University, Seoul, South Korea. ¹⁶⁰University of Split School of Medicine, Split, Croatia. ¹⁶¹UMR 8199 – EGID, Institut Pasteur de Lille, CNRS, University of Lille, Lille, France. ¹⁶²Department of Public Health, Section of Epidemiology, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. ¹⁶³MRC Integrative Epidemiology Unit, University of Bristol Medical School, University of Bristol, Bristol, UK. ¹⁶⁴Department of Exercise and Nutrition Sciences, Milken Institute School of Public Health, George Washington University, Washington, DC, USA. ¹⁶⁵Section of Geriatrics, Department of Internal Medicine, Amsterdam UMC, University of Amsterdam, Amsterdam, the Netherlands. ¹⁶⁶Amsterdam Public Health, Aging and Later Life, Amsterdam, the Netherlands. ¹⁶⁷Nuffield Department of Population Health, University of Oxford, Oxford, UK. ¹⁶⁸Division of Endocrinology, Department of Medicine, University of California, Los Angeles, CA, USA. ¹⁶⁹Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden. ¹⁷⁰The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ¹⁷⁴Deceased: 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. Alizadeh⁷⁵, H. Marika Boezen⁷⁵, Lude Franke⁷⁷, Morris Swertz⁷⁷, Cisca Wijmenga⁷⁷, Pim van der Harst¹⁷¹, Gerjan Navis¹⁷², Marianne Rots¹⁷³ and Bruce H. R. Wolffenbuttel⁷⁶

¹⁷¹Department of Cardiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. ¹⁷²Department of Internal Medicine, Division of Nephrology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. ¹⁷³Department 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 inflation⁶⁸ was observed (λ 1.2–1.4) (Supplementary Fig. 1), LD score regression analyses indicated this reflects true polygenic architecture rather than cryptic population structure⁶⁹.

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 procedures⁷⁰.

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 METAL⁷¹, 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 significant⁷².

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 Catalog¹¹ 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 analysis⁷⁴, a Monte Carlo average information restricted maximum likelihood algorithm implemented in the BOLT-LMM v2.3.3 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%)⁸⁹, likely at least in part due to the absence of rare variants in GWAS⁷⁵.

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 GCTA⁷⁶. Any lead SNPs identified in known long-range high-LD regions⁷⁷ 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 settings¹⁷. 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 BioMe BioBank participants, using summary statistics of the primary European ancestry meta-analyses and PRSice software⁷⁸. 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 BioMe 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 BioMe) and the PGSs in European ancestry participants of BioMe. 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 BioMe. 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 package⁷⁹. Among 8,959 BioMe 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 Biobank¹⁴ with relevant complex traits and diseases based on established associations at the trait level using LD score regression implemented in the LD-Hub web resource¹⁸. 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 analysis⁸². 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 assumptions²⁸, 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 P_{GX}^2 in our two-sample MR setting⁸³. Instead, we applied MR-PRESSO (pleiotropy residual sum and outlier)⁸⁷, which removes pleiotropy by identifying and discarding influential outlier predictors from the standard inverse variance-weighted test²⁸. 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 framework⁴⁰. 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 pleiotropy²⁶, 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 Biobank⁸⁴, 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 inactivity³³. 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 population⁸⁵. 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 DEPICT⁴⁰ 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 CELLECT³³ 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 tests⁴⁶. 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 studies⁴⁶. 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 Consortium⁹⁰, 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 FINEMAP⁴⁷. 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 data⁹². 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 model⁴⁹.

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 field⁹⁵. 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 $50 \text{ kJ mol}^{-1} \text{ nm}^{-2}$. Center-of-mass pull groups were defined as the ABD of each respective monomer, with a weak position restraint placed on the $C\alpha$ 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 $\sim 18\%$) 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 indices⁹⁷ 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: MetaMix <https://www.metamex.org/>; 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. Vösa, 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. & Calcagni, 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) (15GSPSG23770000; 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 15GSPSG23770000), 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., 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.

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 of 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

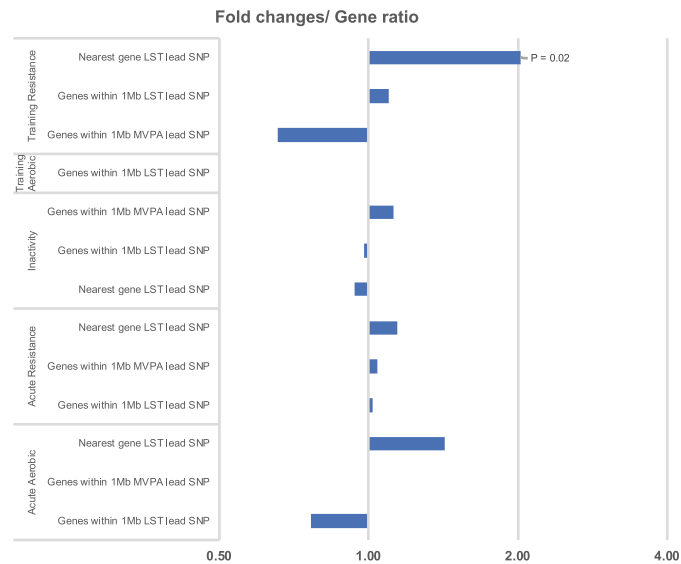
Extended data is available for this paper at <https://doi.org/10.1038/s41588-022-01165-1>.

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

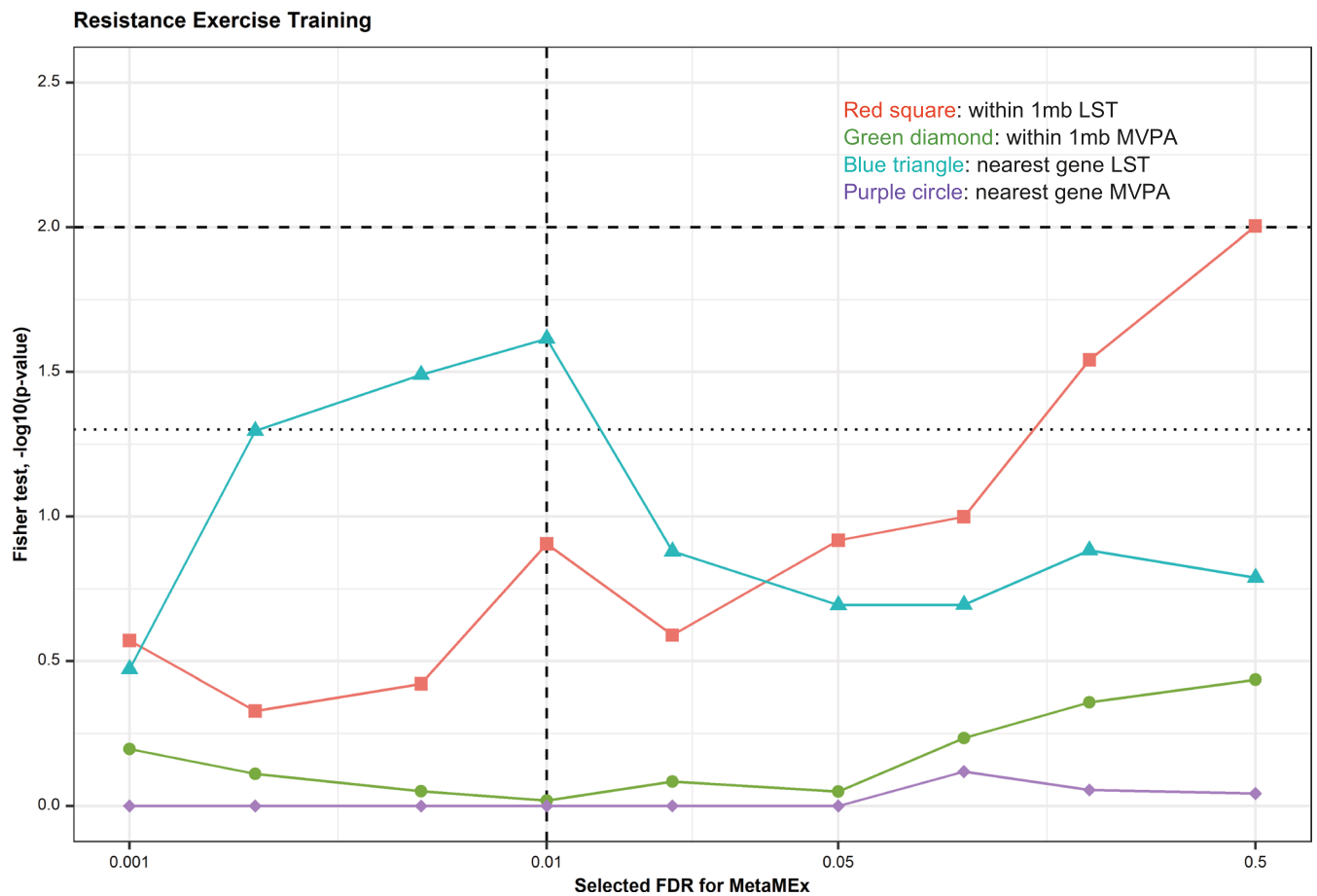
Correspondence and requests for materials should be addressed to Zhe Wang or Marcel den Hoed.

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

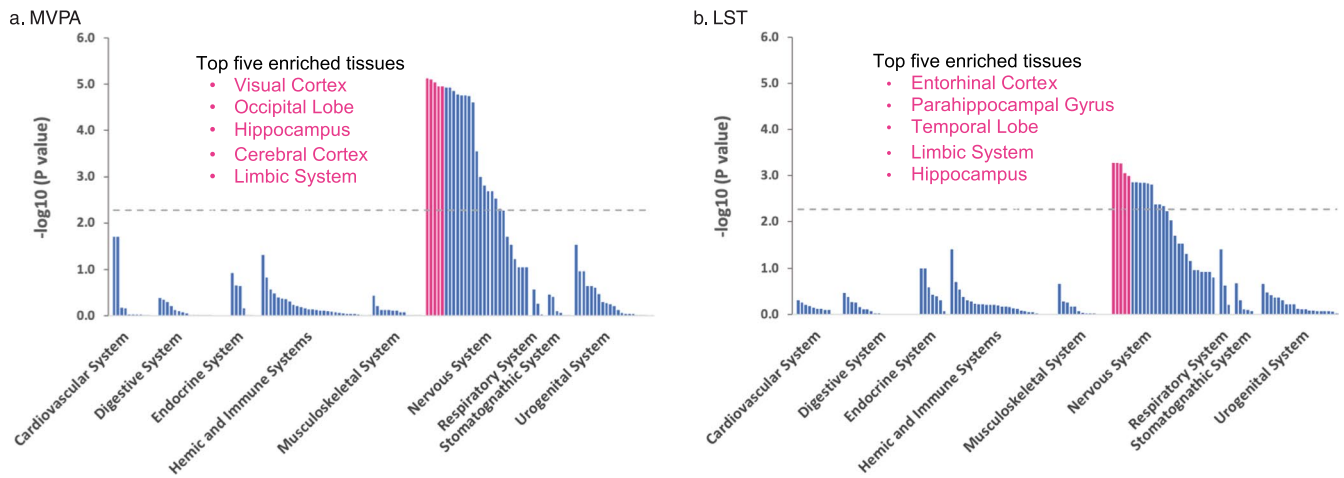
Reprints and permissions information is available at www.nature.com/reprints.



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.

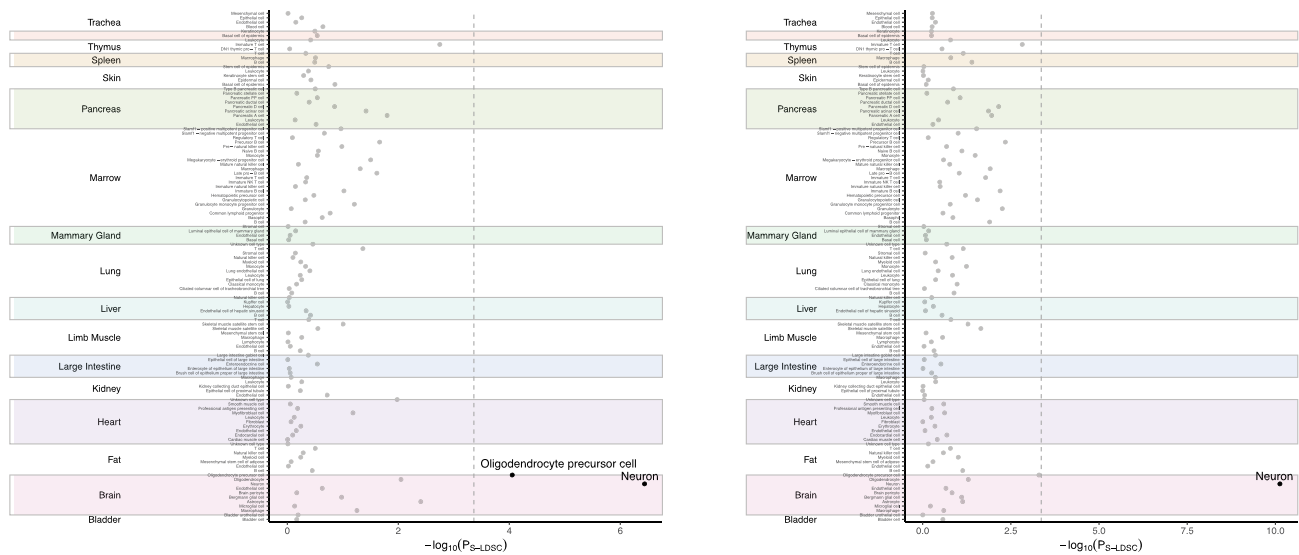


Extended Data Fig. 2 | A sensitivity analysis shows the analysis of altered gene expression following resistance training is robust to FDR threshold. We examined the effect of different FDR thresholds on Fisher's exact test results for the enrichment analysis of alteration in gene expression in skeletal 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 horizontal 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).

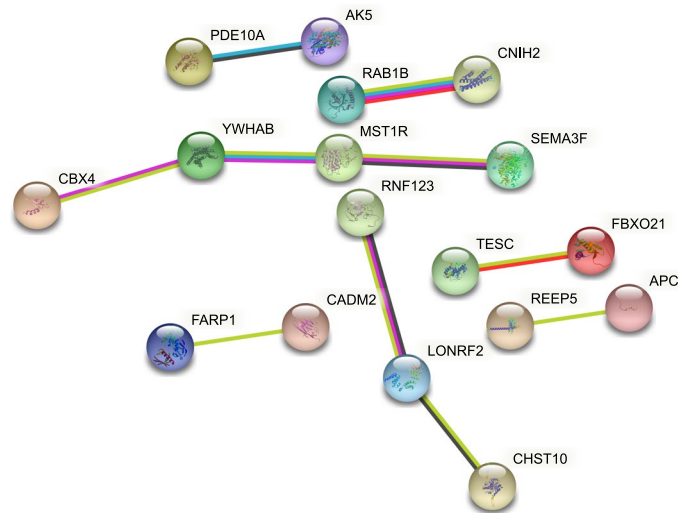
a Tabula Muris cell types



b mouse nervous system cell types



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.

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 mdatanalysis v 2.0.
 The MD topology was created with GROMACS pdb2gmx using the ACTN3 dimer model and parameterized with the CHARMM36 all-atom force

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.eqtngen.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](https://www.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 experiments were aware of the R577X genotype of the donor, 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

Methods

- | n/a | Included in the study |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Animals and other organisms |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Human research participants |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern |

- | n/a | Included in the study |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> 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.