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 IIA muscle fibers, and possibly protection from exercise-induced muscle damage. Finally, Mendelian randomization analyses show that beneficial effects of lower LST and higher MVPA on several risk factors and diseases are mediated or confounded by body mass index (BMI). Our results provide insights into physical activity mechanisms and its role in disease prevention.

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

Physical activity and sedentary behavior are affected by public policy and social support, as well as by cultural, environmental and individual factors. Factors like socioeconomic status, built environment and media all influence physical activity at a population level. In parallel, innate biological factors (for example, age, sex hormones, pre-existing medical conditions, epigenetics and genetics) also explain a moderate proportion of the individual variability in physical activity and sedentary behavior. Heritability estimates \( h^2 \) range from 31% to 71% in large twin studies\(^4\). Identifying the genetic factors that influence daily physical activity will improve our understanding of this complex behavior, and may (1) facilitate unbiased causal inference; (2) help identify vulnerable subpopulations; and (3) fuel the design of tailored interventions to effectively promote physical activity. A mechanistic understanding of physical activity at a molecular level may even allow its beneficial effects to be attained through pharmacological intervention\(^5\).

Genome-wide association studies (GWAS) have identified thousands of loci associated with cardiometabolic risk factors and diseases\(^6\). However, similar efforts for physical activity have been sparse and initially had limited success. This likely reflects the comparatively small sample size of these efforts\(^7\), 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 \( n \approx 91,000 \)\(^13\). 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 ancestors, 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\(^13\))—are associated with LST, explaining 2.75% of its variance. We also identify 11 loci for MVPA (six not previously reported\(^13\), four that overlap with LST) and four loci for sedentary behavior at work (all previously reported\(^13\); 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
13 additional loci: eight loci for MVPA and eight for LST, with three loci overlapping (Supplementary Table 5)\(^1\)\(^2\).

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 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\(^1\)\(^3\)\(^4\) show directionally consistent associations (\(P<0.05\)) with self-reported LST and/or MVPA in our study (Supplementary Table 7). By contrast, 39 LST- and 4 MVPA-associated loci observed here show directionally consistent associations (\(P<0.05\)) with at least one objectively assessed physical activity and/or sedentary trait (using accelerometry) in UK Biobank (Supplementary Table 8). In line with this, each additional LST-decreasing and MVPA-increasing allele in unweighted genetic predisposition scores of the 88 LST- and 11 MVPA-associated loci, respectively, are associated with higher objectively assessed daily physical activity levels in UK Biobank (\(P=5\times10^{-23}\) for LST; \(P=2\times10^{-3}\) for MVPA, Supplementary Table 8).

As external validation, we use the European ancestry summary statistics of LST and MVPA to construct polygenic scores (PGSs), and examine their associations with MVPA in 8,195 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-Hub\(^18\), we observe significant (\(P<4.6\times10^{-5}\)) genetic correlations of LST and MVPA with adiposity-related traits (\(r=−0.41\) to −0.20), especially with body fat percentage (\(r=0.4\) 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 PGSs for MVPA and morbid obesity (\(P=1.1\times10^{-5}\), Supplementary Fig. 3). Strikingly, genetic correlations with body fat percentage are similar for self-reported LST, MVPA (Fig. 4) and accelerometer-assessed physical activity traits\(^19\)–\(^24\) (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...
Results are similar for bidi-rectangular causal inference tests using body fat percentage instead of BMI (Table 2). CAUSE also illustrates a causal effect of higher LST on 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.

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.

Gene expression in skeletal muscle following training. Although behavior is mainly influenced by signals from the brain, in the case of physical activity, characteristics of skeletal muscle can play a facilitating or restricting role\(^4\). Therefore, we next examine whether genes in LST- and MVPA-associated loci are enriched for altered messenger RNA expression in skeletal muscle following an acute bout of exercise or a period of training or inactivity\(^5\) (Methods).

A mild enrichment for transcripts with an altered expression in skeletal muscle after resistance training is observed for genes near-est to lead SNPs in LST-associated loci (\(P = 0.02\)) (Extended Data Figs. 1 and 2, and Supplementary Table 14). Of the ten genes driving the enrichment, \(PDE10A\) may play a critical role in regulating cyclic AMP and cyclic GMP levels in the striatum, a brain region that harbors the central reward system and is important for physical activity regulation\(^6\), and in regulating striatum output\(^7\). \(ILF3\) and...
NECTIN2—near APOE—influence the host response to viral infections\(^{43,44}\); EXOC4 plays a role in insulin-stimulated glucose uptake in skeletal muscle\(^{45}\); and IMMQL2 influences the transport of proteins across the inner mitochondrial membrane\(^{46}\) (Supplementary Note).

**Visual information processing and the reward system.** To further improve the understanding of the biological factors that influence sedentary behavior and physical activity, we perform a tissue enrichment analysis using DEPICT\(^ {40} \). LST- and MVPA-associated loci (\(P < 1 \times 10^{-5}\)) 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 decline\(^ {47} \). The LST-associated loci yield similar tissue enrichment results, with retina having the lowest \(P\) value for enrichment. Interestingly, enrichment for genes expressed in retina was also observed in the High Runner mouse model\(^ {42} \). Areas related to the reward system (for example, the hippocampus and limbic system) and to memory and navigation (for example, the entorhinal cortex, parahippocampal gyrus, temporal lobe and limbic system) are also enriched in both LST- and MVPA-associated loci (Extended Data Fig. 3 and Supplementary Table 15).

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

**Candidate gene prioritization.** To explore mechanisms by which the identified loci may influence LST and MVPA, we next pinpoint genes in GWAS-identified loci: (1) contributing to tissue enrichment or identified by DEPICT’s gene prioritization algorithm (Supplementary Tables 15 and 17); (2) whose expression in brain, blood and/or skeletal muscle is anticipated to mediate the association between locus and outcome based on Summary-based MR\(^ {46} \) (SMR; Supplementary Table 18); (3) harboring credible variants with a high posterior probability of being causal (\(>0.80\))\(^ {47} \) and a predicted effect on protein function (Supplementary Table 19); (4) showing chromatin–chromatin interactions with credible variants in central nervous system cell types (such genes may be further from lead SNPs, Supplementary Table 19); (5) that—across 26 tissues and cell types—are activated by contact with enhancers presumably affected by causal variants flagged by GWAS hits\(^ {49} \) (Supplementary Tables 20–22); (6) associated with physical activity in GWAS in humans and mice and located <100kb 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, RAB18, KL2, PACS1, REPS1, DNLM, EXOC4), locomotion (CADM2, KLC2) and myopathy (MLF2, HERC1, KLC2, SILI) 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 (\(P_{\text{untrans}}=2.1 \times 10^{-7}\)); 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 MVPA\(^ {55} \) and forms part of the E4 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 (\(P=1.9 \times 10^{-14}\)) as well as with higher ACTN3 expression in skeletal muscle (GTEx, \(P=6.6 \times 10^{-7}\)). Alpha-actinin-3 (ACTN3) forms a structural component of the muscle’s Z-disc that is exclusively expressed in type II\(a\) and II\(x\) muscle fibers\(^ {48} \), rs1815739, a common ACTN3 variant that introduces a premature stop codon, p.Arg577Ter, also known as p.Arg620Ter, has been extensively studied in the context of exercise performance\(^ {58} \). Although we observe little evidence for a role

| Exposure | Outcome | Beta  | s.e.  | \(P\) value |
|----------|---------|-------|-------|-------------|
| LST      | Body fat % | 0.16  | 0.07  | 0.016       |
| Body fat % | LST | 0.12  | 0.03  | 0.005       |
| MVPA | Body fat % | \(-0.21\) | 0.17  | 0.22        |
| Body fat % | MVPA | \(-0.001\) | 0.036 | 0.97        |

**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 |
|----------|---------|-------|-------|-------------|
| LST      | BMI     | 0.40  | 0.04  | \(8.4 \times 10^{-14}\) |
| Body fat % | LST | 0.16  | 0.01  | \(1.4 \times 10^{-7}\) |
| MVPA | BMI     | \(-0.25\) | 0.04  | 0.002       |
| Body fat % | MVPA | \(-0.010\) | 0.01  | \(5.8 \times 10^{-12}\) |

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.

---

### Supplementary Information

1. **Supplementary Table 18**: Additional information about the candidates identified in the previous step.
2. **Supplementary Table 19**: Additional information about the candidates identified in the previous step.
3. **Supplementary Table 20**: Additional information about the candidates identified in the previous step.
4. **Supplementary Table 21**: Additional information about the candidates identified in the previous step.
5. **Supplementary Table 22**: Additional information about the candidates identified in the previous step.
6. **Supplementary Table 23**: Additional information about the candidates identified in the previous step.
7. **Supplementary Table 24**: Additional information about the candidates identified in the previous step.
8. **Supplementary Table 25**: Additional information about the candidates identified in the previous step.
9. **Supplementary Table 26**: Additional information about the candidates identified in the previous step.
10. **Supplementary Table 27**: Additional information about the candidates identified in the previous step.

---

**References**

1. NATURE GENETICS
2. Articles
3. September 2022
4. 1332-1344
5. www.nature.com/naturegenetics
6. 1335
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 $P<0.0101$ for MVPA and $P<0.014$ 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 ($P_{AA}=0.017$, $P_{MVPA}=0.17$), the intronic $ACTN3$ variants rs679228 ($P_{AA}=4.3\times10^{-8}$) and rs2275998 ($P_{MVPA}=1.8\times10^{-8}$) do show evidence of such associations. Of these, rs2275998—located 646bp downstream of p.Arg577Ter—is in full LD ($r^2=1.0$) with the nonsense 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 ($P=1.4\times10^{-8}$) and higher odds of engaging in MVPA ($P=8.3\times10^{-7}$). 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 $ACTN3$ nonsense variant rs2229456, but not with the $ACTN3$-truncating variant rs1815739, we next examine whether rs2229456 (p.Glu635Ala variant) has functional consequences for $ACTN3$'s mechanistic properties at the molecular level. We add $ACTN2$ to this comparison because it likely compensates for the loss of $ACTN3$ in the presence of the truncating p.Arg577Ter variant. 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
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.577Ter variant—associated with higher MVPAs—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 fibers from vastus lateralis biopsies obtained from eight healthy, young, untrained male participants before and after an eccentric exercise bout. Results from a 15,000 iteration Markov chain Monte Carlo model show that stable maximal force—with fibers submerged in activating solution—and fiber power during isotonic load clamps are similar in 32 ± 7 fibers (mean ± s.d.) from three p.Arg577 homozygous, p.Glu635Ala heterozygous individuals compared with 39 ± 6 fibers from four individual 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 than in type I fibers (Supplementary Fig. 7). Taken together, these results suggest that a more flexible ACTN3 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 MVPAs 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 self-reported physical activity and sedentary traits in women. However, the magnitude of these effects is small, with correlations ranging from 0.15 to 0.34.

In conclusion, we provide support for the genetic basis of physical activity and sedentary traits, with evidence for pleiotropic effects of these traits on a wide range of other diseases. The findings provide a framework for understanding the genetic basis of physical activity and sedentary traits and highlight potential targets for future intervention strategies.
LST—a proxy for physical activity and sedentary behavior in daily life. As would be expected for complex behaviors that involve both motivation and physical ability, these behaviors are prioritized by more than one approach and point to pathways related to endocytosis, locomotion and myopathy. Results from MD simulations, umbrella sampling and single fiber experiments suggest that a missense variant (rs2229456 encoding ACTN3 p.Glu635Ala) 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 susceptibility to exercise-induced muscle damage. Forty-six candidate genes were 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.

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* | 95% CI     | P value* | Exposure | Outcome                  | Gamma* | 95% CI     | P value* |
|----------|--------------------------|--------|------------|----------|----------|--------------------------|--------|------------|----------|
| 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^{-3} |
| Body fat % | LST                      | 0.12   | 0.04 to 0.18 | 0.14     | BMI      | LST                      | 0.18   | 0.16 to 0.19 | 1.1 × 10^{-3} |
| 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     |

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

To investigate the molecular basis for the association of ACTN3 with MVPA, we compared the ACTN3 p.Glu635 and p.635Ala variants (rs2229456) with each other and with ACTN2—as a functional proxy for ACTN3 p.577Ter—using MD simulations and single fiber experiments. Previous studies using normal mode analysis of alpha-actinin show that several of the natural frequencies have bending flexibility near residue 635. This is interesting because ACTN3’s residue 635—the 356th residue of the spectrin repeat region (Fig. 6)—lies outside the linkers between the α-helices of the spectrin repeats, where most flexibility is expected and observed. The absence of salt-bridge and hydrogen-bonding interactions between position 635 (628 in ACTN2) and surrounding residues—due to either the presence of the alanine substitution at ACTN3’s residue 635, or a kink in the α-helix at ACTN2’s residue 628—increases the flexibility of the dimer under a compressive load, with far less work required to deform the homodimer beyond a compressive distance of 1.2 nm. The p.635Ala substitution may reduce the stiffness of the muscle fiber while undergoing elastic deformation during exercise to a level that is comparable with ACTN2. Although at the expense of the maximal force that single fibers can generate, this may reduce exercise-induced microtrauma caused by Z-disc rupture or streaming, alleviating delayed onset muscle soreness and risk of injuries, 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.
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 mDpA) 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 fibers from an individual homozygous for p.Arg577 and p.Glu635 (blue) compared with type II fibers from three p.Arg577 homozygous, p.Glu635Ala heterozygous individuals (orange); and from four p.577Ter homozygous individuals (green).
Zhe Wang, Andrew Emmerich, Nicolas J. Pilon, Tim Moore, Daiane Hemerich, Marilyn C. Cornelis, Eugenia Mazzafarrello, Siacia Broos, Tarunveer S. Ahiuluwalia, Traci M. Bartz, Amy R. Bentley, Lawrence F. Bielak, Mike Chong, Audrey Y. Chu, Diane Berry, Rajkumar Dorajoo, Nicole D. Dueker, Elisa Kasbohm, Bjarke Feenstra, Zhanna Balkhiyarova, Mariea Elisabeth Graaff, Leanne M. Hall, Toomas Hailer, Fernando P. Hartwig, David A. Hillis, Ville Huikari, Nancy Heard-costa, Fernando P. Hartwig, David A. Hillis, Ville Huikari, Nancy Heard-costa, Christina Holzpfel, Anne U. Jackson, Åsa Johansson, Anja Moltke Jørgensen, Marika A. Kaakinen, Robert Karlsson, Kathleen F. Kerr, Boram Kim, Chantal M. Koolhaas, Zoltan Kutalik, Vasiliki Lagou, Penelope A. Lind, Mattias Lorentzon, Laura J. Scott, Robert A. Scott, Stephen Sidney, Kristin Siggeirsdottir, Albert V. Smith, Jennifer A. Smith, Emily Sonestedt, Marin Strøm, Koon K. Teo, Aihua Li, Allan Linneberg, Jian'an Luan, Pedro Marques Vidal, Sarah E. Medland, Nancy L. Pedersen, Annette Peters, Patricia A. Peyser, Christine Power, Olli T. Raitakari, Mägi Reedik, Alex P. Reiner, Paul M. Ridker, Igor Rudan, Kathy Ryan, Mark A. Sarzynski, Laura J. Scott, Jennifer A. Smith, Emily Sonestedt, Marin Strøm, E. Shyong Tai, Koon K. Teo, Barbara Thorand, Anke Tönjes, Angelo Tremblay, Andre G. Utterlinden, Jagadish Vangipurapu, Natasja van Schoor, Uwe Völker, Gonneke Willemsen
Health, Imperial College London, London, UK. 132Computational Medicine, Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Berlin, Germany. 133Epidemiology Program, University of Hawaii Cancer Center, University of Hawaii at Manoa, Honolulu, HI, USA. 134K.G. Jebsen Center for Genetic Epidemiology, Norwegian University of Science and Technology, Trondheim, Norway. 135Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA. 136Center for Fertility and Health, Norwegian Institute of Public Health, Oslo, Norway. 137Centre for Bone and Arthritis Research, Department of Internal Medicine and Clinical Nutrition, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden. 138Department of Drug Treatment, Sahlgrenska University Hospital, Gothenburg, Sweden. 139Department of Statistics, Seoul National University, Seoul, South Korea. 140University of Split School of Medicine, Split, Croatia. 141UMR 8199 – EGIS, Institut Pasteur de Lille, CNRS, University of Lille, Lille, France. 142Department of Public Health, Section of Epidemiology, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. 143MRC Integrative Epidemiology Unit, University of Bristol Medical School, University of Bristol, Bristol, UK. 144Department of Exercise and Nutrition Sciences, Milken Institute School of Public Health, George Washington University, Washington, DC, USA. 145Section of Geriatrics, Department of Internal Medicine, Amsterdam UMC, University of Amsterdam, Amsterdam, the Netherlands. 146Amsterdam Public Health, Aging and Later Life, Amsterdam, the Netherlands. 147Nuffield Department of Population Health, University of Oxford, Oxford, UK. 148Division of Endocrinology, Department of Medicine, University of California, Los Angeles, CA, USA. 149Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden. 150The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. 151Deceased: 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. Marike Boezen75, Lude Franke77, Morris Swertz77, Cisca Wijmenga77, Pim van der Harst71, Gerjan Navis172, Marianne Rots173 and Bruce H. R. Wolffenbuttel76

70Department of Cardiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. 71Department of Internal Medicine, Division of Nephrology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. 72Department of Molecular 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 Bonferroni 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 data from European ancestry separately. Details of participating studies are described in Supplementary Tables 1 and 2. Although modest genomic inflation was observed (lambda 1.2–1.4) (Supplementary Fig. 1), LD score regression analyses indicated this reflects true polygenic architecture rather than cryptic population structure.

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 accelerometer-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 genotyping platform used, and quality control measures applied within each study was conducted using standard procedures.

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- and ancestry-specific GWAS results were meta-analyzed using the fixed effects, inverse variance-weighted method implemented in METAL1, for 19.1 to 22.5 million SNPs per trait. Because we did not include a replication stage and given the high power of the variants associated with self-reported sedentary behavior and physical activity 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).

To identify genome-wide significant loci, we defined a distance criterion of ≤1 Mb surrounding each genome-wide significant peak (P < 5 × 10^-8). We extracted previously reported genome-wide significant associations within 1 Mb of any index variants we identified from the NHGRI-EBI GWAS Catalog and PhenOScanner V2 (ref. 7). 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 analysis8, 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%)9, likely at least in part due to the absence of rare variants in GWAS.

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 studies were conducted based on 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 GCTA10. Any lead SNPs identified in known loci only tagged LD regions (r^2 > 0.2) within a 5-Mb window) to remove any correlated variants. In the multivariable MR that evaluates the complete set of pairwise genetic correlations of the four self-reported physical activity 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).

PhenWAS with physical activity PGs. To assess the out-of-sample predictive power of the variants associated with self-reported sedentary behavior and physical activity traits, we constructed two PGSs per trait. For LST and MVPA, we constructed two PGSs to cover European and non-European ancestry. Using study-specific variance components reflecting population structure and additional study-specific covariates as presented in Supplementary Table 2. Analyses were limited to outcomes with more than ten cases. Multiple testing thresholds for statistical significance were set to P < 4.8 × 10^-10 (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 accelerometer-assessed physical activity traits assessed in UK Biobank with relevant complex traits and diseases based on established associations at the trait level using LD score regression implemented in the LD-Hub web resource11. To estimate the genetic correlation, we applied a Bonferroni correction for the 108 selected phenotypes available on LD-Hub (P < 4.6 × 10^-8), 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 database12,13. This way, we aimed to minimize bias due to sample overlap in the two-sample MR analysis14. The source of each of the instruments is presented in Supplementary Table 12. Genetic instrumental variables were selected on the basis of the traits they correlated with the traits of interest, using a Bonferroni correction for genome-wide significance (P < 5 × 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 factor combinations were used. For both LST and MVPA, independent loci associated with physical activity or BMI were used as instrumental variables. Because the potential causal effects of LST and MVPA results can be severely biased if instrumental SNPs show horizontal pleiotropy and violate the instrumental variable assumptions, 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 $R^2_{MR}$ in our two-sample MR setting. Instead, we applied MR-PRESSO (pleiotropy residual sum and outlier) test, which removes pleiotropy by identifying and discarding influential outlier predictors from the standard inverse variance-weighted test. For analyses with evidence of no distortion due to pleiotropy (MR-PRESSO Global $p$-value > 0.05) we considered other methods for 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 reversal causation and used Cook’s distance filtering to remove outlying heterogeneous variants as deemed necessary. To select the most appropriate approach, we implemented a machine learning framework. 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 outcome at $1 \times 10^{-4}$, that is, Bonferroni correction for 13 disease outcomes and 2 types of risk factors: physical activity or sedentary behavior and adiposity (0.05/(13 × 2)).

We applied the recently published Bayesian-based MR method CAUSE, which accounts for both correlated and uncorrelated pleiotropy, 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 two confounding model. That is, to examine whether 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 $P < 10^{-8}$, 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 cumulative effect. We have found that CAUSE assessed last trait reflects a lifetime liability. Using the summary statistics of SNPs for childhood adiposity (comparative body size at age 10) and height (comparative height at age 10) in UK Biobank, 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 exercise interventions facilitate a physically active lifestyle. To identify plausible intervention.

For genes $n$ = 467)88, ROSMAP (n = 494)89, and Brain-eQTL (n = 1,194)90; blood eQTL information obtained from the eQTLGen Consortium, which is based on peripheral blood samples from 31,684 individuals; and skeletal muscle eQTL information from the GTEx project (n = 803). To identify prioritized candidate genes, we used the following criteria: (i) a significant association for each phenotype with the simulated compressive loads that are likely experienced in vivo. The final frame of the 1-ns MD production run was used as the starting topology for steered MD simulations using fully relaxed dimers. Steered MD simulations were run for 2 ns with a pulling rate of 0.005 nm ps$^{-1}$ and a harmonic potential of 50 kJ mol$^{-1}$ nm$^{-2}$. Center-of-mass pull groups were defined as the ABD of each respective monomer, with a weak position restraint placed on the Co atom of threonine 52 (ACTN3) or threonine 45 (ACTN2)—a centrally located residue in the core of the ABD—on one ABD, enabling full rotational freedom of each ABD during the course of the steered MD simulations. The pulling vector was oriented along the axis on which the spectrin repeats were initially aligned. Suitable frames from each steered MD simulation were selected that differed by no more than 0.2 nm from 0 to 5.5 nm (a contraction of the dimer by 5.5 nm or ~18%) and were used as the starting topology for a series of 10-ns umbrella sampling simulations. Analysis of the umbrella sampling simulations was conducted using g. wham, to yield the potential of mean force versus reaction coordinate for each variant.

Single skeletal muscle fiber functional characteristics in relation to p.Glu635Ala. Single muscle fibers from eight nonathletic young men in which contractile and morphological properties were previously characterized in vastus lateralis biopsies obtained before and after an eccentric exercise bout were genotyped for rs2229456. A hierarchical linear mixed model was constructed for each fiber type and time point using rstanarm 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 ± 4 fibers); and AC-XX (n = 4 individuals, 44 ± 2 fibers). Using weakly informative priors, the posterior distribution was estimated with Markov chain Monte Carlo sampling (20,000 samples total with 5,000 sample burn-in). We calculated 90% credible
intervals of the posterior density and distribution-free overlapping indices 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/download summary-statistics, GCP-ID: GP000388). 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: MetaMedx (https://www.metamex.eu/); Tabula Muris https://www.czbiohub.org/tabula-muris/; Open GWAS https://www.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: syn7759792/wiki/69613; Brain eQTLserve 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:// genomewideassociationstats.wellcome.ac.uk/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. Pultit, 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, 4851–4859 (2013).

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. & Deney, J. C. G. 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 from polygenic architecture. *Nature* 597, 588–589 (2021).
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 490413E15Rik 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.Glu635 Ala. 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.Glu635 Ala. 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., I.F.B., M.C., A.Y.C., D.B., R.D., N.D.D., K.E., B.F., M.E.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., A.J., M.A.K., R.K., K.E.K., R.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., 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. wrote the manuscript. T.S.A., T.M.B., A.R.B., L.P., R.P., R.C.R., V.H., N.H.-C., C. Holzapfel, A.U.J., A.J., M.A.K., R.K., K.E.K., R.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., 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. wrote the manuscript. T.S.A., T.M.B., A.R.B., L.P., R.P., R.C.R., V.H., N.H.-C., C. Holzapfel, A.U.J., A.J., M.A.K., R.K., K.E.K., R.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., 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. wrote the manuscript. T.S.A., T.M.B., A.R.B., L.P., R.P., R.C.R., V.H., N.H.-C., C. Holzapfel, A.U.J., A.J., M.A.K., R.K., K.E.K., R.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., 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. wrote the manuscript. T.S.A., T.M.B., A.R.B., L.P., R.P., R.C.R., V.H., N.H.-C., C. Holzapfel, A.U.J., A.J., M.A.K., R.K., K.E.K., R.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., 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. wrote the manuscript. T.S.A., T.M.B., A.R.B., L.P., R.P., R.C.R., V.H., N.H.-C., C. Holzapfel, A.U.J., A.J., M.A.K., R.K., K.E.K., R.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., 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. wrote the manuscript. T.S.A., T.M.B., A.R.B., L.P., R.P., R.C.R., V.H., N.H.-C., C. Holzapfel, A.U.J., A.J., M.A.K., R.K., K.E.K., R.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., N.v.d.V., C.M.v.D., P.V., H.V., T.V., G. Willemsen, N.J.W., D.R.W., H.-E.W., J.F.W., A.L.H., A.K., J.R.Z., M.A.I.T., R.J.F.L. and M.d.H. provided input and feedback to the final manuscript.

Funding
Open access funding provided by Uppsala University.

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

Additional information
Extended data is available for this paper at https://doi.org/10.1383/s41588-022-01165-1.
Supplementary information The online version contains supplementary material available at https://doi.org/10.1383/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 MetaMx 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.
An extended data figure showing a sensitivity analysis of altered gene expression following resistance training is robust to FDR threshold. The analysis examines 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 squares represent genes within 1 Mb of the LST lead SNP; green circles represent genes within 1 Mb of the MVPA lead SNP; blue triangles indicate the nearest gene LST lead SNP; purple diamonds represent the nearest gene MVPA lead SNP. The horizontal dotted line indicates the nominal significance level (P < 0.05), and the vertical dashed line indicates the FDR threshold used. FDR thresholds explored range from 0.001 to 0.5.
Extended Data Fig. 3 | DEPICT-derived tissue enrichment of MVPA and LST. a. MVPA. b. LST. SNPs with $P < 1 \times 10^{-5}$ for association in the European ancestry GWAS of men and women combined were used as input. The dashed line indicates the FDR corrected significance threshold (FDR < 0.05).
Extended Data Fig. 4 | Cell type prioritization using CELLECT for MVPA and LST. a, Prioritization of 115 Tabula Muris cell types across 19 tissues identified two cell types from the brain as significantly associated (stratified linkage disequilibrium score regression) with MVPA (left) and LST (right), namely oligodendrocyte precursor cells and neurons (shown in black; Bonferroni-corrected significance threshold, \( P < 0.05/115 \)). b, Prioritization of 265 mouse nervous system cell types identified 13 and 45 cell types from 12 distinct brain regions as significantly associated (stratified linkage disequilibrium score regression) with MVPA and LST, respectively (highlighted; Bonferroni-corrected significance threshold, \( P < 0.05/265 \)).
Extended Data Fig. 5 | Protein-protein interactions involving 17 of the 46 candidate genes in GWAS-identified loci prioritized by at least two approaches. Protein-protein interactions were visualized using String. LONRF2 and CHST10 were prioritized in loci associated with MVPA; the remaining genes were prioritized in loci associated with LST.
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.
Vita/View 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 4930413E15Rk.
A homology model of the F635 variant monomeric filament was generated using Phyre2 v 2.0.
Molecular dynamics (MD) system preparation and simulation was conducted with GROMACS 2020 and msanalysis 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: MetaMx: https://www.metamx.eu/; Tabula muris: https://www.czbiohub.org/tabula-muris/; Open GWAS: https://gwas.mrcieu.ac.uk/; MR Base: https://www.mrbase.org/; GTex Consortium: https://gtexportal.org/home/; eQTLGen Consortium: https://www.eqtlgen.org/; CommonMind Consortium: https://www.synapse.org/_modal/Synapse.syn2759792/wiki/69613; Brain zQTLServe: http://mostafavifar.stat.ubc.ca/zqtl/; MetaBrain: https://www.metabraini.m/.

**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.

- [x] 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-first.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 | n/a |
| ☒ Antibodies | ☒ Involved in the study |
| ☒ Eukaryotic cell lines | ☒ ChiP-seq |
| ☒ Palaeontology and archaeology | ☒ Flow cytometry |
| ☒ Animals and other organisms | ☒ MRI-based neuroimaging |
| ☒ Human research participants | |
| ☒ Clinical data | |
| ☒ Dual use research of concern | |

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.