Physical activity and sedentary behavior; mechanistic insights and role in disease prevention

Marcel den Hoed (marcel.den_hoed@igp.uu.se)
Uppsala University https://orcid.org/0000-0001-8081-428X

Zhe Wang
The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai
https://orcid.org/0000-0002-8046-4969

Andrew Emmerich
Uppsala University https://orcid.org/0000-0002-0908-9924

Nicolas Pillon
Karolinska Institutet

Timothy Moore
University of California Los Angeles https://orcid.org/0000-0003-4250-3370

Daiane Hemerich
The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai

Marilyn Cornelis
Northwestern Feinberg School of Medicine

Eugenia Mazzaferro
Uppsala University

Siacia Broos
KU Leuven

Adam Ameur
Science for Life Laboratory, Department of Immunology, Genetics and Pathology, Uppsala University
https://orcid.org/0000-0001-6085-6749

Stefania Bandinelli
Azienda Sanitaria Toscana Centro, Florence

Joshua Bis
Cardiovascular Health Research Unit https://orcid.org/0000-0002-3409-1110

Michael Boehnke
University of Michigan https://orcid.org/0000-0002-6442-7754

Claude Bouchard
Pennington Biomedical Research Center https://orcid.org/0000-0002-0048-491X

Daniel Chasman
Brigham and Women's Hospital https://orcid.org/0000-0003-3357-0862

Eco de Geus

VU University https://orcid.org/0000-0001-6022-2666
Louise Deldicque
UCLouvain
Marcus Dörr
Ernst-Moritz-Arndt-University https://orcid.org/0000-0001-7471-475X
Michele Evans
National Institutes of Health
Luigi Ferrucci
National Institute on Aging https://orcid.org/0000-0002-6273-1613
Myriam Fornage
University of Texas Health Science Center at Houston https://orcid.org/0000-0003-0677-8158
Caroline Fox
Merck
Theodore Garland
University of California, Riverside https://orcid.org/0000-0002-7916-3552
Ulf Gyllensten
Uppsala University https://orcid.org/0000-0002-6316-3355
Torben Hansen
University of Copenhagen https://orcid.org/0000-0001-8748-3831
Caroline Hayward
University of Edinburgh https://orcid.org/0000-0002-9405-9550
Bernardo Horta
Universidade Federal de Pelotas https://orcid.org/0000-0001-9843-412X
Elina Hypponen
University of South Australia https://orcid.org/0000-0003-3670-9399
W. Craig Johnson
University of Washington https://orcid.org/0000-0002-3161-3753
Sharon Kardia
University of Michigan
Lambertus Kiemeney
Radboudumc https://orcid.org/0000-0002-2368-1326
Markku Laakso
University of Eastern Finland https://orcid.org/0000-0002-3394-7749
Claudia Langenberg
University of Cambridge https://orcid.org/0000-0002-5017-7344
Terho Lehtimäki
Department of Clinical Chemistry, Fimlab Laboratories and Finnish Cardiovascular Research Center-Tampere, Faculty of Medicine and Health Technology, Tampere University https://orcid.org/0000-0002-2555-4427

Loic Le Marchand
University of Hawaii
Patrik Magnusson
Karolinska Institutet https://orcid.org/0000-0002-7315-7899
Nicholas Martin
QIMR Berghofer Medical Research Institute https://orcid.org/0000-0003-4069-8020
Mads Melbye
Statens Serum Institut
Andres Metspalu
The Estonian Genome Center, University of Tartu https://orcid.org/0000-0002-3718-796X
David Meyre
https://orcid.org/0000-0003-4850-7444
Kari North
UNC Chapel Hill
Claes Ohlsson
University of Gothenburg, Sweden https://orcid.org/0000-0002-9633-2805
Albertine Oldehinkel
University of Groningen, University Medical Center Groningen https://orcid.org/0000-0003-3925-3913
Marju Orho-Melander
Lund University
Guillaume Pare
Population Health Research Institute, McMaster University https://orcid.org/0000-0002-6795-4760
Taesung Park
Seoul National University https://orcid.org/0000-0002-8294-590X
Oluf Pedersen
Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Blegdamsvej 3B, 2200 Copenhagen https://orcid.org/0000-0002-3321-3972
Brenda Penninx
VU University Amsterdam
Tune Pers
University of Copenhagen https://orcid.org/0000-0003-0207-4831
Ozren Polasek
University of Split School of Medicine https://orcid.org/0000-0002-5765-1862
Inga Prokopenko
University of Surrey https://orcid.org/0000-0003-1624-7457
Charles Rotimi
National Institutes of health https://orcid.org/0000-0001-5759-053X
Nilesh Samani
University of Leicester

Xueling Sim
National University of Singapore https://orcid.org/0000-0002-1233-7642
George Davey Smith
University of Bristol https://orcid.org/0000-0002-1407-8314
Harold Snieder
University Medical Center Groningen https://orcid.org/0000-0003-1949-2298
Thorkild Sorensen
Faculty of Health and Medical Sciences, University of Copenhagen https://orcid.org/0000-0003-4821-430X
Tim Spector
King's College London https://orcid.org/0000-0002-9795-0365
Nicholas Timpson
University of Bristol https://orcid.org/0000-0002-7141-9189
Rob van Dam
National University Health System
Nathalie van der Velde
AMC, Amsterdam UMC, Amsterdam Public Health
Peter Vollenweider
Lausanne University Hospital (CHUV)
Henry Völzke
Institute for Community Medicine Ernst-Moritz-Arndt-University Greifswald 17487 Greifswald Germany
Trudy Voortman
Erasmus University Medical Center https://orcid.org/0000-0003-2830-6813
Gerard Waeber
CHUV University Hospital https://orcid.org/0000-0003-4193-788X
Nick Wareham
MRC Epidemiology Unit https://orcid.org/0000-0003-1422-2993
David Weir
University of Michigan https://orcid.org/0000-0002-1661-2402
Erich Wichmann
Institute of Epidemiology
James Wilson
University of Edinburgh https://orcid.org/0000-0001-5751-9178
Andrea Hevener
University of California
Anna Krook
Karolinska Institute https://orcid.org/0000-0002-0891-0258
Juleen Zierath
Karolinska Institute https://orcid.org/0000-0001-6891-7497

Martine Thomis
KU Leuven
Ruth Loos
Ichan School of Medicine at Mount Sinai https://orcid.org/0000-0002-8532-5087

Article

Keywords:

Posted Date: December 16th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1129252/v1

Version of Record: A version of this preprint was published at Nature Genetics on September 7th, 2022.
See the published version at https://doi.org/10.1038/s41588-022-01165-1.

Physical activity and sedentary behavior; mechanistic insights and role in disease prevention

Wang et al., (a preliminary author list is provided in a separate document)

Abstract

Even though physical activity and sedentary behavior are moderately heritable, little is known about the mechanisms that influence these traits. Here, we combine data for up to 674,980 individuals from 51 studies in a trans-ancestry meta-analysis of genome-wide association studies for self-reported moderate-to-vigorous intensity physical activity during leisure time (MVPA); leisure screen time (LST); sedentary commuting; and sedentary behavior at work. We identify 99 loci that associate with at least one trait. Loci associated with LST are enriched for genes whose expression in skeletal muscle is altered by resistance training. Molecular dynamics simulations suggest that the Glu to Ala substitution encoded by rs2229456 (*ACTN3*) – associated with more MVPA – disrupts salt bridge interactions and makes the alpha actinin 3 filaments more flexible. In isolated type II_A muscle fibers, the Ala-encoding allele is associated with lower maximal force and power during an isometric contraction, suggesting protection from exercise-induced muscle damage. Finally, Mendelian Randomization analyses show that the causal effect of LST on BMI is 2-3 times larger than the effect of body mass index (BMI) on LST, and that beneficial effects of LST and MVPA on several risk factors and diseases are mediated or confounded by BMI. Taken together, our results provide mechanistic insights into the regulation of MVPA and into the role of LST and MVPA in disease prevention. These insights may facilitate the development of tailored physical activity interventions.

Introduction

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

Physical activity and sedentary behavior are affected by public policy and social support, as well as by cultural, environmental and individual factors[7]. Factors like socio-economic status, built environment, and media all influence PA at a population level[7]. In parallel, innate biological factors (e.g. age, sex hormones, preexisting medical conditions, epigenetics and genetics) also explain a moderate proportion of the interindividual variability in PA and sedentary behavior. Heritability estimates (h^2) range from 31% to 71% in large twin studies[8,9]. Pinpointing the genetic factors that influence daily PA 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 PA. A mechanistic understanding of PA at a molecular level may even allow its beneficial effects to be attained through pharmacological intervention[10].

Genome-wide association studies (GWAS) have identified thousands of loci associated with cardiometabolic risk factors and diseases[11]. However, similar efforts for PA have been sparse and have had limited success. This likely reflects the comparatively small sample size of these efforts[12], along with heterogeneous assessments of PA across studies. More recently, GWAS using data from UK Biobank identified three loci associated with self-reported PA (n~377,000 individuals) and five with accelerometry-assessed PA (n~91,000)[13,14]. Hence, on the assumption that PA is a highly polygenic trait, many common variants influencing PA undoubtedly remain to be identified.

Here, we combine data from up to 674,980 individuals from 51 studies in a trans-ancestry meta-analysis of GWAS for moderate-to-vigorous intensity PA during leisure time (MVPA); leisure screen time (LST); sedentary commuting; and sedentary behavior at work. This yields 104 independent association signals in 99 loci, implicating brain and muscle, amongst others organs. Follow-up analyses improve our understanding of the molecular basis of leisure time PA and sedentary behavior and their role in disease prevention.

Results

Genome-wide analyses yield 99 loci associated with physical activity and sedentary traits

In our primary meta-analysis of European-ancestry men and women combined (**Supp Tables 1-2**), we identify 91 loci that are associated ($P<5\times10^{-9}$) with at least one of four self-reported traits, one representing MVPA and three reflecting sedentary behaviors (**Supp Table 3, Figure 2, Supp Figure 1**). The non-European ancestry meta-analyses did not provide novel associations themselves and were only used in trans-ancestry meta-analyses. Trans-ancestry and sex-specific meta-analyses yield eight additional loci, resulting in a total of 104 independent association signals in 99 loci (**Supp Tables 3-4**). The vast majority of these – i.e. 89 independent 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 overlap with LST) and four loci for sedentary behavior at work (all previously reported[13,15], **Supp Table 3**). No loci are identified for sedentary commuting.

SNP-heritability estimates range from 8% for MVPA to 16% for LST (**Supp Table 5**, Online 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 (**Figure 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 PA traits in UK Biobank range from 0.14 to 0.44 (**Figure 1b**). Importantly, five of the eight loci previously identified for objectively assessed daily PA in UK Biobank data[13,14] show directionally consistent associations ($P<0.05$) with self-reported LST and/or MVPA (**Supp Table 6**). Vice versa, 39 LST- and four MVPA-associated loci show directionally consistent associations ($P<0.05$) with at least one objectively-assessed PA trait (using accelerometry) in UK Biobank (**Supp Table 7**). In line with this, each additional LST-decreasing and MVPA-increasing allele in unweighted genetic predisposition scores of the 88 LST- and ten MVPA-

associated loci, respectively, are associated with higher objectively assessed daily PA levels in UK Biobank ($P< 5\times10^{-5}$ for LST; $P<5\times10^{-3}$ for MVPA, **Supp Table 7**). Taken together, these results suggest that in spite of their limitations, self-reported, intensity- and domain-specific PA traits can be meaningful proxies for daily PA in large-scale genetic association studies.

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 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 20 mins/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 20 mins per week on MVPA than individuals at deciles 4 to 6 (OR [95% CI] = 0.74 [0.55-0.99]) (**Figure 3**).

Shared genetic architecture between physical activity, adiposity, and other traits

Using LD score regression implemented in the LD-Hub[17], we observe significant ($P<4.6\times10^{-4}$) genetic correlations of MVPA and LST with adiposity-related traits (r -0.41 to -0.20), especially with body fat percentage (r_g -0.3 and 0.4, respectively, **Figure 4, Supp Figure 2, Supp Table 9).** In line with moderate genetic correlations, 11 of the 99 self-reported PA loci have previously been associated with obesity-related traits[18-24]. In addition, PGSs for higher MVPA and lower LST are associated with lower BMI in up to 23,723 participants from the BioMe Biobank (**Supp Table 8**), and a phenome-wide association study in 8,959 BioMe European ancestry samples shows a negative association between the PGS for MVPA and morbid obesity ($P=1.1\times10^{-5}$, **Supp Figure 3**). Strikingly, genetic correlations with body fat percentage are similar for self-reported MVPA, LST, and accelerometer-assessed physical activity traits (**Figure 4**, **Supp Figure 2**).

Besides adiposity, higher PA 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 better mental health outcomes; a lower risk of lung cancer; and with longevity (**Figure 4**, **Supp Figure 2**).

Causal inference between physical activity, adiposity, and disease outcomes

To assess directions of causality between PA and BMI, we next perform two-sample Mendelian Randomization (MR) analyses using multiple MR methods that utilize genome-wide full summary results or genome-wide significant loci (**Supp Table 10**, Online Methods). CAUSE[25] as well as traditional MR methods (MR-PRESSO[26]; inverse variance–weighted test[27]) consistently show that LST and BMI causally affect each other, with the causal effect (per 1 SD unit increase in each trait) of higher LST on higher BMI being 2-3-fold larger than the effect of BMI on LST (**Figure 5A, Table 1a**). Results are similar for bi-directional causal inference tests using body fat percentage instead of BMI (**Table 1b**). However, CAUSE cannot distinguish a model of causality from horizontal pleiotropy for body fat percentage and LST (**Table 1b**). We do not observe evidence for causal effects between MVPA and adiposity in either direction. CAUSE also illustrates a causal effect of higher LST on higher recalled adiposity and height in childhood (**Table 1b**), supporting our hypothesis that a genetic predisposition for LST later in life represents a lifelong predisposition that already influences adiposity through sedentary behavior early in life.

We next investigate the causal effects of LST and MVPA on common diseases and risk factors, with and without adjusting for BMI. In univariate analyses, we observe effects of lower LST on higher HDL cholesterol levels; and on lower odds of type 2 diabetes, attention deficit hyperactivity disorder, and depression. Effects of higher MVPA are similar for the first three outcomes, albeit without surviving Bonferroni correction for multiple testing. A higher MVPA

does positively affect longevity, as indicated by a causal effect on its proxy parental age at death. Importantly, multivariable MR analyses show that all protective causal effects of lower LST or higher MVPA are either mediated or confounded by BMI. Multivariable MR results for MVPA should be interpreted with caution due to weak instrument bias (conditional F statistics <10)[28] (**Figure 5B-C, Supp Table 11**).

Enrichment of altered gene expression in skeletal muscle following resistance training

While behavior is mainly influenced by signals from the brain, in the case of PA, characteristics of skeletal muscle can play a facilitating or restricting role[29]. Therefore, we next examine if genes in LST- and MVPA-associated loci are enriched for altered mRNA expression in skeletal muscle following an acute bout of exercise or a period of training or inactivity[30] (Online 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). (**Supp Figures 4-5, Supp Table 12**). Of the ten genes driving the enrichment, *PDE10A* may play a critical role in regulating cAMP and cGMP levels in the striatum – a brain region that harbors the central reward system and is important for physical activity regulation[31] – and in regulating striatum output[32]; *ILF3* and *NECTIN2* – near *APOE* – influence the host response to viral infections[33,34]; *EXOC4* plays a role in insulin-stimulated glucose uptake in skeletal muscle[35]; and *IMMP2L* influences transport of proteins across the inner mitochondrial membrane[36] (**Supp Box 1**).

Enrichment for genes involved in visual information processing and the reward system

To further improve the understanding of the biological factors that influence PA, we perform a tissue enrichment analysis using DEPICT[37]. LST- and MVPA-associated loci ($P<1\times10^{-5}$) are most significantly enriched for genes expressed in the retina, visual cortex, occipital lobe, and cerebral cortex. This suggests that: 1) possibly subtle differences in the ability to receive, integrate and process visual information influence the likelihood to engage in MVPA; 2) MVPA

improves the expression of genes that play a role in such visual processes in these tissues; and/or 3) MVPA can slow down age-related perceptual and cognitive decline[38]. The LST-associated loci yield similar tissue enrichment results, with retina being the most significantly enriched tissue. Interestingly, enrichment for genes expressed in retina was also observed in the High Runner mouse model[39]. Areas related to the reward system (e.g. hippocampus, limbic system) and to memory and navigation (e.g. entorhinal cortex, parahippocampal gyrus, temporal lobe and limbic system) are also enriched in both LST- and MVPA-associated loci (**Supp Figure 6, Supp Table 13)**.

We next use CELLECT[40] to identify enriched cell types using single cell RNA sequencing data from the Tabula Muris and mouse brain projects[41]. In Tabula Muris data, we observe enrichment in non-myeloid neurons for MVPA and LST, and of non-myeloid oligodendrocyte precursor cells for MVPA, possibly highlighting a role for signal transduction (**Supp Figure 7, Supp Table 14**). In mouse brain data, we identify enrichment for 13 and 45 cell types from three and 12 distinct brain regions for MVPA and LST, respectively, including enrichment in dopaminergic neurons (**Supp Figure 7, Supp Table 14**); a key feature of physical activity regulation in mice[42].

Prioritized candidate genes point to cell signaling, endocytosis, and myopathy

To explore mechanisms by which the identified loci may influence LST and MVPA, we next pinpoint genes in GWAS identified loci: 1) that contribute to enriched tissues or are identified by DEPICT's gene prioritization algorithm (**Supp Tables 13,15**); 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[43] (**Supp Table 16**); 3) harboring credible variants with a high posterior probability of being causal (>0.80)[44] and a presumed deleterious effect on protein function (**Supp Table 17**)[45]; 4) that show chromatin-chromatin interactions with credible variants in central nervous system cell types (such genes may be further from lead SNPs, **Supp Table**

17); 5) associated with PA in GWAS in humans and mice (see below) and located <100 Kb of the lead variant in either humans or mice (**Supp Tables 18-19**); and 6) driving enrichment of altered expression in skeletal muscle following resistance exercise training (**Supp Table 12**). Integrating results across approaches yields 32 candidate genes in eight MVPA-associated loci; and 232 candidate genes in 68 LST-associated loci. Two MVPA and 27 LST candidate genes are prioritized by >1 approach and point to cell signaling (*YWHAB*, *REPS1*, *TESC*, *PDE10A*), endocytosis (*HERC1*, *PACS1*, *REPS1*, *DNM2*) and myopathy (*HERC1*, *SIL1*, *DNM2*) as relevant pathways (**Supp Tables 20-21, Supp Box 1**). *In vivo* perturbation of these and other candidate genes in model systems is required to confirm or refute a role in PA.

Physical activity loci under selection point to signal transduction and wound healing

As much higher PA levels were required to ascertain sufficient nutrition in times of hunting and gathering and pre-mechanical farming as compared with today's Westernized societies[46], a higher capacity to be physically active may have been selected for. To explore this, we examine if MVPA and LST association signals overlap with regions identified in three genome-wide selection screens[47-49]. Here, we show that 22 genes located <100kb of lead SNPs in three MVPA and/or LST-associated loci are located in three of 412 regions under selection in the past 50,000 years[47] (**Supp Table 22**). The protein-coding genes nearest the lead SNPs (<10kb) – *DNM3*, *MST1R* and *FOXP1* – are also prioritized by other approaches (**Supp Tables 20-22**) and play a role in cell signaling and wound healing, amongst others (**Supp Box 1**). We next identify genes located <10kb of 15 loci under selection in the past 10,000 years – based on results from an ancient DNA scan[48] – and <100kb of PA association signals. This yields one additional gene (*GRM5*) that harbors a GWAS lead SNP for LST (rs1391954, **Supp Table 22**). *GRM5* encodes a metabotropic glutamate receptor that activates phospholipase C[50]; another key player in signal transduction[51], inflammation and wound healing[52], amongst other processes. No lead SNPs for LST or MVPA are located within 1Mb of five loci under very recent

selection[49]. In summary, we show that four loci selected for in the past 10-50,000 years are associated with leisure time PA and sedentary behavior today.

Overlap with genetics of voluntary wheel-running behavior in mice unveils a new human transcript

Many of the biological factors influencing PA levels are likely shared across species[53]. Identifying loci that are associated with PA across multiple species may help prioritize candidate genes in such loci, and shed light on the mechanisms by which overlapping loci influence PA. To this end, we compare our findings with loci identified in a GWAS for spontaneous PA in 100 inbred mouse strains, performed using the Hybrid Mouse Diversity Panel (HMDP)[54] (**Supp Table 18**). Nine genes in two LST-associated loci are also located within ±1Mbp of two lead SNPs for distance run and average running speed in mice ($P<4.1\times10^{-6}$) (**Supp Table 19**). Of the eight genes that overlap across humans and mice in one of these two loci, *TESC* – highly expressed in the striatum – harbors the intronic lead SNP rs2173650 in humans (**Supp Box 1**). In the mouse however, a gene without an established orthologue in humans – the lncRNA 4930413E15Rik – is considered likely causal for high voluntary wheel running behavior in mice selectively bred for 61 generations[39]. Using single cell RNA-sequencing data from GTEx[55], we show that a sequence 1.4 Mb from rs2173650 with high conservation to the mouse 4930413E15Rik is expressed in several human reproductive tissues (**Supp Figure 8**).

Enrichment for associations with previously reported candidate genes

Candidate gene studies in humans have aimed to identify and characterize the role of genes in exercise (i.e., PA behavior) and fitness (i.e., PA ability) for decades. We next examine if variants in genes that have been linked to or associated with exercise and fitness show evidence of associations with self-reported LST and MVPA[12,56-60]. Of the 58 previously described candidate genes (13 for exercise; 45 for fitness), 56 (13 and 43) harbor variants with $P<0.05$ for associations with LST and/or MVPA ($P_{binomial}=2.1\times10^{-70}$, **Supp Table 23, Supp Figure 9**).

Associations reach traditional genome-wide significance ($P<5\times10^{-8}$) for variants in three genes: *APOE*[61], *PPARD*[62] and *ACTN3*[63] (see Online Methods).

The C allele in rs1625595 ~300kb upstream of *ACTN3* is significantly associated with higher MVPA ($P=1.9\times10^{-11}$) as well as with higher *ACTN3* expression in skeletal muscle (GTEx, $P=6.6\times10^{-5}$). alpha-actinin-3 (*ACTN3*) forms a structural component of the muscle's Z-disc that is exclusively expressed in type II_A and II_X muscle fibers[64]. A common *ACTN3* variant that introduces a premature stop codon (rs1815739, i.e. R577X, a.k.a. R620X) has been extensively studied in the context of exercise performance[63]. We observe little evidence for a role of rs1815739 in leisure time PA or sedentary behavior ($P_{MVPA}=0.17$, $P_{LST}=0.017$), in contrast with the intronic *ACTN3* variants rs679228 ($P_{LST}=4.3\times10^{-8}$) and rs2275998 ($P_{MVPA}=1.8\times10^{-7}$). Of these, rs2275998 – located 646bp downstream of R577X – is in full LD ($R^2=1.0$) with the missense variant rs2229456 (encoding E635A), which likely affects protein function (CADD score for the derived, minor, 635A-encoding C allele = 28.6). Each 635A-encoding allele in rs2229456 is associated with less LST ($P=1.4\times10^{-4}$) and higher odds of engaging in MVPA ($P=8.3\times10^{-7}$). Of note: given its downstream location, a potentially causal effect of rs2229456 on PA requires absence of the protein truncating 577X-encoding allele in rs1815739 (**Supp Table 24**).

The ACTN3 spectrin repeat region shows greater flexibility with 635A

Given the striking finding of MVPA and LST being associated with the ACTN3 missense variant rs2229456, but not with the ACTN3-truncating variant rs1815739, we next examine if rs2229456 (encoding E635A) has functional consequences for ACTN3's mechanistic properties at the molecular level. The results of computer-based (steered) molecular dynamics simulations and umbrella sampling (see Online Methods for details) show that the ancestral E635-encoding allele product facilitates salt bridge interactions at residue 635 with surrounding residues (e.g. R638 and Q639, **Figure 6A-B**) via its glutamate side chain. Such salt bridge interactions are not

formed in the presence of the 635A product. Moreover, 635A shows a distinctly different behavior, with a greater magnitude of root mean square fluctuations in the middle section of the spectrin repeats under no-load conditions (**Figure 6C**), suggesting a more flexible structural region. When placed under simulated compressive loads that are likely experienced *in vivo*, 635A shows a more linear force versus distance relationship, with greater variance in the potential of mean force (**Figure 6D**). Taken together, these results indicate that the ACTN3 635A dimer – associated with higher MVPA – exhibits greater flexibility than the E635 dimer.

Maximal force and fiber power lower in single II_A muscle fibers with ACTN3 635A

We next examine if a higher predicted ACTN3 dimer flexibility in the presence of 635A 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[65,66]. Results from a 15,000 iteration Markov chain Monte Carlo model show that within four individuals homozygous for the R577-encoding allele, stable maximal force – with fibers submerged in activating solution – and fiber power during isotonic load clamps are lower in 32±7 fibers (mean±SD) from three E635A heterozygous individuals than in 46 fibers from an individual that is homozygous for the E635-encoding allele (**Figure 6E**, Online methods). Strikingly, both outcomes are similar for fibers from R577 homozygous individuals carrying one 635A-encoding allele and 39±6 fibers from four individuals homozygous for the 577X-encoding allele, who are anticipated to have ACTN2 incorporated in their Z-disc, instead of ACTN3. Associations are most striking after an eccentric exercise intervention, and are – as expected – more pronounced in type II_A than in type I fibers (**Supp Figure 11**). We do not observe evidence that differences at a single fiber level extrapolate to whole-body exercise performance in data from 266 *de novo* genotyped healthy young men[67] (data not shown), which may reflect the relatively low statistical power we have to detect such associations when compared with the

discovery GWAS for MVPA. Taken together, these results suggest that a more flexible ACTN3 dimer with lower peak performance may be less susceptible to exercise-induced muscle damage, 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 during leisure time and sedentary behavior. Around half of these also show evidence of directionally consistent associations with objectively assessed PA traits. Genetic correlations and two-sample MR analyses show that lower LST results in lower adiposity. Protective causal effects of higher PA acting through lowering BMI are observed for longevity and odds of type-2 diabetes, attention deficit hyperactivity disorder and depression. 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. *In silico* annotation using a range of approaches helps prioritize 263 unique candidate genes across 68 LST- and seven MVPA-associated loci. Of these, five genes are located in three loci that have been under selection in the last 10-50,000 years and suggest a role for cell signaling and wound healing. Moreover, the 29 candidate genes flagged by >1 prioritization approach point to pathways related to cell signaling, endocytosis, and myopathy. Two LST-associated loci are also associated with voluntary wheel-running behavior in mice. One of these loci harbors a sequence with high conservation to a presumed causal lncRNA in mice for which we identify a previously unknown transcript in humans that is expressed in reproductive tissues. Finally, results from molecular dynamics simulations, umbrella sampling, and single fiber experiments suggest that a missense variant (rs2229456 encoding E635A in *ACTN3*) likely increases MVPA at least in part by reducing susceptibility to exercise-induce muscle damage.

Recent MR studies reported causal protective effects of objectively assessed PA on depression, colorectal cancer, and breast cancer[68,69], but did not examine mediation by BMI. The MR study for cancer concluded that a 1 SD increase in device-measured PA was associated with lower odds of breast (OR=0.51) and colorectal cancer (OR=0.66)[68]. Both the

direction and size of the effect are consistent with our univariable MR results. Furthermore, a causal effect of objectively assessed, but not self-reported PA (i.e., MVPA) on depression has been reported[69]. Our MR results for LST on depression show that while the PA trait matters, the self-reported nature of it is inconsequential. According to an earlier study, TV viewing has an attenuated effect but still causes coronary artery disease when adjusting for BMI[15]. The discrepancy with our results – suggesting mediation or confounding by BMI – highlights the importance of including variants associated with both PA and BMI in multivariable MR analysis, to prevent loss of precision and potentially even biased estimates[28].

Aiming to improve the understanding of the molecular basis of PA, we perform a range of largely complementary approaches to identify candidate genes through which the association signals are anticipated to act. Strikingly, of the 263 unique genes prioritized across 68 LST- and seven MVPA-associated loci, only seven genes are prioritized by two approaches when using traditional cut-offs within each approach. This illustrates the complexity of *in silico* gene prioritization for complex behaviors, especially when proof-of-concept genes are sparse and a gold standard approach for prioritization is nonexistent. When combining results from multiple approaches, applying more lenient criteria in individual approaches is justifiable. First, because the odds that a gene with an FDR of (e.g.) 0.20 in two methodologically independent approaches represents a false positive is low (i.e. 0.04); and secondly because *in silico* gene prioritization should at most guide the order in which genetic perturbation studies in model systems are performed.

To investigate the molecular basis for the association of *ACTN3* with MVPA, we compare the E635 and 635A variants in rs2229456 using molecular dynamics 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 residue 635 – i.e., residue 356 of the spectrin repeat region – lies outside of the linkers between the α helices of the spectrin repeats where most flexibility is expected and

observed[70]. The absence of salt bridge interactions between position 635 and surrounding residues in the presence of the alanine substitution at residue 635 increases the flexibility of the ACTN3 protein under a compressive load, with far less work required to deform the ACTN3 homodimer beyond a compressive distance of 1.4 nm. The E635A substitution may reduce the stiffness of the muscle fiber whilst undergoing elastic deformation during exercise. While 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 streaming[1], alleviating delayed onset muscle soreness[2] and risk of injuries[3], enabling a more active lifestyle.

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. We also provide evidence that the causal effect of sedentary behavior on adiposity is 2-3 times larger than vice versa, and support the important public health message that a physically active lifestyle mitigates the risk of multiple diseases in major part through an effect on BMI.

Online Methods

Samples and study design

We conducted the largest meta-analyses for physical activity (PA) traits to date, including results from up to 674,980 individuals (including nearly half-a-million from the UK Biobank) to identify genetic loci associated with PA 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) moderate-to-vigorous intensity PA during leisure time (MVPA); 2) leisure screen time (LST); 3) sedentary commuting behavior; and/or 4) sedentary behavior at work, using study-specific, tailored analysis plans (**Supp Table 2)**. Next, we performed ancestry-specific, inverse-variance weighted fixed effects meta-analyses of summary statistics for each of the four self-reported traits (**Figure 1a**), including data from up to 674,980 individuals of European (93.9%), African (1.8%), East Asian (0.9%), South Asian (1.4%), and Hispanic (2.0%) ancestry (**Supp Table 1**). Our primary meta-analyses were restricted to 601,735 European ancestry participants. Secondary meta-analyses were also conducted for: 1) all ancestries (European + other ancestries), 2) European ancestry men, 3) European ancestry women, 4) each non-European ancestry. Details of participating studies are described in **Supp Tables 1 and 2**. Although modest genomic inflation[71] was observed (lambda 1.2-1.4) (**Supp Figure 1**), linkage disequilibrium (LD) score regression analyses indicated this reflects true polygenic architecture rather than cryptic population structure[72].

Each study obtained informed consent from participants and approval from the appropriate institutional review boards.

Self-reported physical activity traits

We defined four self-reported PA traits: MVPA (more than 20 mins/week or not, i.e., the median in most studies); LST (hours/day); sedentary commuting behavior (driving a car vs. using other

modes in those employed and commuting[73]) and sedentary behavior at work (mostly sitting and no heavy lifting vs. other in those employed[74]).

The self-reported outcomes are domain- and intensity-specific PA and sedentary traits that, unlike accelerometry-based outcomes, are subject to misclassification and bias by recall and awareness of the beneficial effects of PA, amongst others. Furthermore, different studies used different questionnaires to capture PA, and so we defined cohort-specific traits that make optimal use of the available data, whilst striving for consistency across studies (**Supp 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.

Besides trait definitions, the average age per cohort ranged from early adulthood to old-age (17-74 years old). The power to detect genetic factors that influence PA was thus likely compromised by misclassification of physically active and inactive individuals, and heterogeneity by the inclusion of older age groups in the meta-analysis, as the heritability of PA decreases with increasing age[75]. However, such factors could have resulted in type II – but not type I – errors in the meta-analysis. Despite these limitations, a large sample size helped us identify 42 previously unreported loci for self-reported PA and sedentary behavior. Genetic correlations with objectively assessed PA traits were modest and five of eight previously reported loci for objectively assessed PA show evidence of association with self-reported PA and sedentary behavior. Hence, despite the well-known limitations of self-reported PA, focusing on domain and intensity PA traits in a large study sample helped increase our understanding of the genetic etiology of this complex behavior in a manner that is not currently possible using objectively assessed PA. Descriptive information of these four traits are reported by study in **Supp Table 1**.

Genotyping, imputation and quality control

Detailed information about the genotyping platform and quality control measures for each study are presented in **Supp Table 2**. Quality control following study level analyses was conducted using standard procedures[76].

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^2, principal components (PCs) reflecting population structure and additional study-specific covariates as presented in **Supp Table 2**. For all outcomes, we examined associations with and without adjusting for BMI. To avoid drawing conclusions that are driven by collider bias[77], we did not use the BMI-adjusted associations. Analyses were limited to genotyped and imputed variants (minor allele frequency [MAF] > 0.1% in UK Biobank, and minor allele count > 3 in other studies). Study-, sex- and ancestry-specific GWAS results were meta-analyzed using the fixed-effects, inverse variance weighted method implemented in METAL[78]. In addition, gene-based analyses aggregating rare (MAF<1%) functional variants as annotated by Ensembl's variant effect predictor (VEP)[79] were conducted in UK Biobank European ancestry participants. Gene-based Burden and SKAT tests were performed using a mixed model approach in the GENESIS package[80]. Genes identified using these approaches were also flagged by the single variant analysis and hence the results are not shown. To further explore potential linear associations with MVPA, we used zero-inflated negative binomial regression and modeled MVPA as a continuous outcome (mins/week). Associations were examined using the same covariates as in the main analyses amongst 371,244 unrelated UK Biobank participants of European ancestry, for variants with $P<1\times10^{-5}$ for association with the dichotomous MVPA outcome in the European ancestry meta-analysis. The approach yielded similar results (not shown).

To identify genome-wide significant loci, we defined a distance criterion of ±1 Mb surrounding each genome-wide significant peak ($P<5\times10^{-9}$). We extracted previously reported genome-wide significant associations within 1 Mb of any index variants we identified from the NHGRI-EBI GWAS Catalog[11] and PhenoScanner V2[81]. 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 PA traits previously. To identify PA-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 PA trait, we used BOLT-REML variance components analysis[82], a Monte Carlo average information restricted maximum likelihood algorithm implemented in the BOLT-LMM v2.3.3 software. Like in most GWAS for complex traits, the SNP heritability (up to 16%) was lower than the heritability estimates from twin studies (31 - 71%)[8,9]. A recent study using whole genome sequencing data showed that additionally including rare variants can fully recover the 'missing' heritability[83]. This warrants further, much larger studies in which the association of PA traits with rare variants is examined.

Although we performed a trans-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 PA etiology.

Joint and conditional analyses

To identify additional independent signals in associated loci, we performed approximate joint and conditional SNP association analyses in each locus, using GCTA[84]. Any lead SNPs

identified in known long-range high-LD regions[85] 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.

Phenome-wide association study with physical activity polygenic scores

To assess the out-of-sample predictive power of the variants associated with self-reported PA, we constructed two polygenic scores (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 software[86]. We subsequently assessed the association of MVPA and BMI with the PGSs in individuals of European, Hispanic and African ancestry in BioMe Biobank participants. Among the 2,765 European ancestry individuals with PA measurements and genotypes, the PGSs were calculated on common variants (MAF>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 mins 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 10 PCs 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 ancestry 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 phenome-wide association study (PheWAS) on 1,039 disease outcomes in 8,959 BioMe European ancestry participants. We used logistic regression to separately model each PheWAS trait as a function of the two PGSs, adjusting for age, age^2, sex and the top ten PCs. Interpretation of results was restricted

to outcomes with more than ten cases. Multiple testing thresholds for statistical significance were set to $P<4.8\times10^{-5}$ (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 PA traits assessed in UK Biobank[14] with relevant complex traits and diseases based on established associations at the trait level using LD score regression implemented in the LD-Hub web resource[17]. To define significance, we applied a Bonferroni correction for the 108 selected phenotypes available on LD-Hub ($P<4.6\times10^{-4}$). **Supp Table 9** shows the complete set of pairwise genetic correlations of the four self-reported PA traits with relevant complex traits and diseases. Next, we prioritized traits and diseases showing evidence of genetic overlap (significantly associated with at least one of the PA traits). These can be divided into six categories: lifestyle traits, anthropometric traits, psychiatric diseases, other diseases (cardiometabolic diseases and cancer), biomarkers and others (**Figure 4**). Using objectively assessed PA traits (i.e. using accelerometry) instead of self-reported traits yielded similar results (**Supp Figure 1**).

Two sample Mendelian Randomization

We performed Mendelian Randomization (MR) analyses to disentangle the causality between LST and MVPA on the one hand and BMI on the other hand. Since UK Biobank provided the majority (>75%) of the samples in the meta-analyses, we selected summary statistics of variants for relevant traits and diseases using data from the largest publicly available GWAS without data from UK Biobank participants on the MR-Base platform[87] (**Supp Table 10**), aiming to minimize bias due to sample overlap in the two sample MR analysis[88]. Genetic instrumental variables for each of the traits and diseases consisted of genome-wide significant ($P<5\times10^{-8}$) 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 PA and BMI were used as instrumental variables.

We followed several steps to evaluate potential causality. As MR results can be severely biased if instrumental SNPs show horizontal pleiotropy and violate the instrumental variable assumptions[27], we prioritized methods that are robust to horizontal pleiotropy when calculating causal estimates. We did not use MR Egger regression to identify the presence of potential pleiotropy, because the MR-Egger estimate of causal effect is biased towards the null when the NO Measurement Error (NOME) assumption is violated, as indicated by lower values of I^2_{GX} in our two sample MR setting[89]. Instead, we applied MR-PRESSO (pleiotropy residual sum and outlier)[26], which removes pleiotropy by identifying and discarding influential outlier predictors from the standard inverse variance–weighted test[27]. 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, and conducted Steiger filtering to remove variants likely influenced by reverse causation as well as Cook's distance filtering to remove outlying heterogeneous variants as deemed necessary. To select the most appropriate approach, we implemented a machine learning framework[90]. 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 regarding disease outcomes at 1.9×10^{-3}, i.e. Bonferroni correction for 13 disease outcomes and two types of risk factors: PA and adiposity (0.05/(13×2)).

We also applied the recently published Bayesian-based MR method CAUSE, which accounts for both correlated and uncorrelated pleiotropy[25] in evaluating bi-directional causal effects between PA 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 $P<1\times10^{-3}$, as recommended by the CAUSE authors). Additionally, we took advantage of the robustness of the CAUSE method – which allows overlapping GWAS samples – to test the assumption that a genetic predisposition for LST assessed later in life reflects a lifetime liability. using the summary statistics of SNPs for childhood adiposity (comparative body size at age 10) and height (comparative height at age 10) in UK Biobank[91], we examined bidirectional causal effects between LST and these two recalled childhood traits.

Enrichment for genes with altered expression in skeletal muscle after an intervention

A high degree of physical fitness and a strong adaptive response to exercise interventions facilitate a physically active lifestyle. To identify plausible candidate genes in GWAS-identified loci, we examined enrichment for transcripts whose expression in skeletal muscle was changed after an acute bout of aerobic exercise, aerobic training, an acute bout of resistance exercise, resistance training, and inactivity[30]. 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 such interventions are likely not representative for the general population[92]. Enrichment was examined for genes nearest to, or within 1 Mb of lead variants for LST- and MVPA-associated loci. We used FDR<0.01 as the threshold for altered expression after intervention. A sensitivity analysis with a series of different FDR cut-offs (0.001 to 0.5) showed that results were robust.

Annotation using DEPICT, CELLECT, SMR, FINEMAP and 3D chromatin interactions

We used DEPICT[37] to identify enriched gene sets and tissues, as well as to prioritize candidate genes in the identified loci, using variants with $P<1\times10^{-5}$ in the primary meta-analysis of European ancestry men and women combined as input. We also used CELLECT[40] to identify enriched cell types for PA, 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 Summary-based MR (SMR) and HEIDI tests[43]. Briefly, this approach integrates summary-level data from GWAS and eQTL studies to test if a transcript and phenotype are likely associated because of a shared causal variant (i.e., pleiotropy). We considered genes candidates if they had a Bonferroni corrected $P_{SMR} < 1.02\times10^{-5}$ and showed no evidence of heterogeneity ($P_{HEIDI}>0.05$), as in earlier studies[43]. Based on tissue enrichment results from DEPICT, the SMR analyses were performed using brain eQTL information obtained from GTEx-brain (n=72)[55,93], CommonMind Consortium - CMC (n=467)[94], ROSMAP (n=494)[95], Brain-eMeta (n=1,194)[93]; blood eQTL summary information obtained from the eQTLGen Consortium[96] – which is based on peripheral blood samples from 31,684 individuals – and skeletal muscle eQTL information from the GTEx project (n=803)[97].

To identify variants in GWAS-identified loci with a high posterior probability of being causal, we used LST and MVPA summary statistics as input for FINEMAP[44]. We used default parameters and selected a maximum of 10 putative causal variants per locus. The output variants identified as credible were mapped to genes using tissue-specific HiC chromatin conformation capture data[98]. We integrated all HiC data in the brain (dorsolateral prefrontal cortex, hippocampus, neural progenitor cell, and adult and fetal cortex) available on FUMA v1.3.5, using the same approach. Genes in GWAS-identified loci containing FINEMAP-identified credible coding variants with a CADD score > 12.33 were also prioritized.

Mouse experiments

Females from 100 genetically distinct strains from the Hybrid Mouse Diversity Panel (HMDP)[99] were purchased from Jackson Laboratories (University of Tennessee Health Science Center). They arrived at UCLA at 5 to 8 weeks of age and were housed 1-4 weeks until wheel testing. All mice were ~3 months old at the start of the experimental protocol, and were randomized into two groups: 1) sedentary or no exercise; and 2) exercise trained. Strains used and sample size per group are shown in **Supp Table 18**. Trained animals were housed unaccompanied on a standard 12-hour light dark cycle (6AM to 6PM local time). They were fed on a standard laboratory chow diet (8604, Teklad) with ad libitum access to food and water for the entire duration of the experiment. Mice were given full-time access to a running wheel for ~30 days. Wheel revolutions were tallied every 15 sec using VitalView® Activity Software (Starr Life Sciences Corp, Oakmont, Pennsylvania, United States). Daily running distances were calculated by converting wheel revolutions into distance using wheel circumference (35.9 cm). Average running speed was calculated by averaging all non-zero-wheel revolutions and normalizing on a per sec basis. Percent time running was calculated by dividing all 15 sec bouts a wheel revolution was recorded to the total number of intervals. GWAS performed using the Hybrid Mouse Diversity Panel (HMDP) were completed as described elsewhere[54].

In a selection study for high voluntary wheel-running behavior[39], the mouse lncRNA 4930413E15Rik was considered to show a strong indication of consensus in a locus that was associated with daily distance run and average voluntary running speed in a GWAS in 100 mouse strains, as well as with LST in humans. The mouse gene 4930413E15Rik is located on chr 5, spanning the coordinates 118,952,339 - 118,961,261 (mm10 assembly). To investigate the corresponding region in the human genome, a lift-over to hg19 was performed in the UCSC genome browser, resulting in the coordinates chr 12: 116,087,265 – 116,097,521. The region on chr 12 contains no established gene models, but was further investigated in the GTEx IGV browser[55] to study expression that might be present at low levels in specific human tissues.

Enrichment for previously reported candidate genes

We next conducted a literature review of previously reported genes with evidence of a role in exercise (i.e., PA behavior) and fitness (i.e., PA ability) and identified 58 such candidate genes (13 for exercise; 45 for fitness)[12,56-59]. 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 E635A) may have a causal effect.

Molecular dynamics simulation for E635A

Alpha-actinin is a structural member of vertebrate muscle Z-discs, and primarily functions to cross-link neighboring actin filaments of opposite polarity from adjacent sarcomeres. This binding can occur over a range of angles from 60 to 120°, creating a tetragonal lattice with a lattice spacing of 19 to 25 nm[100-102]. In addition to its interaction with actin, alpha-actinin binds and anchors titin to the Z-disc[103]. The alpha-actinin homodimer is formed from two antiparallel subunits composed of an N-terminal actin-binding domain and a C-terminal calmodulin homology domain (CAM), separated by four spectrin-like repeats. Each repeat consists of a triple α-helix coiled-coil (**Figure 6A**). Alpha-actinin 3 (*ACTN3*) at 901 amino acids in length is one of four isoforms of alpha -actinin and is exclusively found in human type-II (also known as fast-twitch) skeletal muscle fibers. The naturally occurring truncating mutation R557X in *ACTN3* has a potential impact on injury risk during exercise, increased muscle-damage following eccentric training and increased flexibility for 557X homozygotes[63].

As no structure for human ACTN3 has yet been experimentally determined, we constructed a homology model of the E635 variant monomeric filament using the fully annotated protein (Uniprot ID Q08043) using Phyre2[104], with the 635A variant mutated *in silico*. 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. Molecular dynamics (MD) system preparation and simulation was conducted with GROMACS 2020.1[105]. The MD topology was created with GROMACS pdb2gmx using the ACTN3 dimer model and parameterized with the CHARMM36 all-atom force field[106]. The ACTN3 dimer was 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) at 310 K via a Berendsen thermostat, followed by a constant pressure ensemble (NPT) at 1 bar with a Parinello-Rahman barostat. Molecular dynamics 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 actin binding domain of chain B to prevent self-interaction across the periodic boundaries.

Computer-based molecular dynamics (MD) simulations show that the ancestral E635-encoding A (adenine) allele product facilitates salt bridge interactions at residue 635 with surrounding residues (e.g., R638 and Q639, **Figure 6B**) via its glutamate side chain. Such salt bridge interactions are not formed with the 635A substitution variant. Root mean square fluctuations were performed on the residues of spectrin repeats of each monomer chain using gmx rmsf. Root mean square fluctuations of each amino acid residue, i.e., their average displacement over the simulation compared with the starting structure, were calculated with both variants, with variant 635A – but not E635 – showing higher fluctuations in the monomer with the restrained actin-binding domain (**Figure 6C**, Chain B, orange trace). Root mean square

fluctuations peaks in these interface regions are around 0.8 nm for the 635A variant, while the E635 variant fluctuations are approximately 0.6 nm (**Figure 6C**). In the presence of 635A, moderately higher root mean square fluctuations values were observed in the middle section of the spectrin repeats – over the residue range of 410-540 – though E635 showed a more pronounced peak near residue 440 (**Figure 6C**). Overall, 635A showed a distinctly different behavior, with a greater magnitude of Root mean square fluctuations under no-load conditions, suggesting a more flexible structural region in the presence of the 635A-encoding C (cytosine) allele product, associated with higher MVPA.

Steered molecular dynamics and Umbrella sampling for E635A

We next compared the properties of 635A and E635 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 molecular dynamics simulations of both the E635 and 635A *ACTN3* variants using fully relaxed ACTN3 dimers. Steered molecular dynamics simulations were run for 2 ns with a pulling rate of 0.005 nm ps^{-1} and a harmonic potential of 50 kJ mol^{-1} nm^{-2}. Centre of mass pull groups were defined as the actin-binding domain of each respective monomer, with a weak position restraint placed on the Cα atom of threonine-52 (a centrally located residue in the core of the actin-binding domain) on one actin-binding domain, enabling full rotational freedom of each actin-binding domain during the course of the steered molecular dynamics simulations simulation. The pulling vector was oriented along the axis on which the spectrin repeats were initially aligned. Suitable frames from each steered molecular dynamics simulation were selected that differed by no more than 0.2 nm from 0 to -3 nm (a contraction of the dimer by 3 nm) and were used as the starting topology for a series of 10 ns umbrella sampling simulations. Analysis of the umbrella sampling simulations was conducted using g_wham to yield the potential of mean force versus reaction coordinate for each variant.

When a compressive force was applied between the center of mass of the two actin-binding domains, the force required to compress the two actin-binding domains by 0.2 nm was lower for 635A compared with E635 (~28 v. 55 kJ/mol/nm). Furthermore, the force-to-distance relationship to a compressive distance of -1.2 nm – where the two respective forces converge (67 kJ/mol/nm) – was notably more linear for 635A than for E635 (**Supp Figure 10**). Greater variability is also seen for 635A in the force versus distance relationship among triplicate steered molecular dynamics simulations. To explore this further, we used umbrella sampling to examine the change in potential of mean force (free energy surface) over the reaction coordinate corresponding to the compression of the ACTN3 dimer.

Umbrella sampling of the ACTN3 dimer variants showed that 635A reaches an energy minimum at a distance between actin-binding domains that is 0.35 nm shorter than E635. Initial compression of the two ACTN3 variants from a relaxed state to a compression of 1.4 nm was similar, requiring energy input of approximately 4.8 kJ/mol, though beyond this compressive distance the two variants diverged in their response to compression (**Figure 6D**). The variant 635A required 1.8 kJ/mol to compress the dimer from -1.4 to -3.0 nm, while E635 required ~8.8 kJ/mol from -1.4 to -2.5 nm, before relaxing slightly and increasing to 10.3 kJ/mol over the range of -1.4 to -3.0 nm (**Figure 6D**). Interestingly, bootstrap estimation of the error of the potential of mean force showed greater variance for 635A, in line with and strengthening the root mean square fluctuations and steered molecular dynamics simulations results. Taken together, these results indicate that the ACTN3 635A dimer - associated with higher MVPA – exhibits greater flexibility than the E635 dimer.

Single skeletal muscle fiber functional characteristics in relation to E635A

Single muscle fibers from eight non-athletic young men in which contractile and morphological properties were previously characterized in *Vastus Lateralis* biopsies obtained before and after eccentric exercise bout[65,66] were genotyped for rs2229456. A hierarchical linear mixed effects

model was constructed for each fiber type and time point using rstanarm[107] 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 R577X and E635A were clustered into three groups, i.e., 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 5000 sample burn-in). We calculated 90% credible intervals of the posterior density and distribution-free overlapping indices[108] to compare single fiber properties between genotypes.

Acknowledgements

A full list of acknowledgments can be found in the Supplementary Information.

Author Contributions

Conceived the study: MC, RJFL, MdH

Trait definitions: MC, SB, UE, RJFL, MdH

Study-specific analysis plans: MC, MdH

GWAS QC and meta-analysis: ZW, MdH

SNP-based heritability estimation: ZW

Joint and conditional analyses: ZW

PheWAS with physical activity polygenic scores: ZW

Genetic correlation analysis: ZW

Mendelian randomization analysis: ZW

Enrichment of altered gene expression in skeletal muscle following intervention: NP, AK, JZ, ZW

Biological annotation using DEPICT, SMR, FINEMAP: DH

Cell type enrichment using CELLECT: AMJ & TP

Integration of results across gene prioritization approaches: ZW, MdH

GWAS for physical activity in 100 mouse strain: TM & AH

Comparison across mice and humans: MdH

Exploring expression of a 4930413E15Rik orthologue in humans; CM & AA

Physical activity loci and natural selection: MdH

Candidate gene analysis for exercise and fitness: ZW

Construction of a homology model of the E635 ACTN3 variant: EM

(Steered) molecular dynamics and umbrella sampling E635A: AE

Single skeletal muscle fiber experiments E635A: SB, LD & MT

Single skeletal muscle fiber data analysis AE

De novo genotyping for functional characterization E635A: AP, MV & MT

Manuscript writing: ZW, AE & MdH

Cohort study level contribution (concept, design, data acquisition, QC, and analysis): all other coauthors

All authors provided input and feedback to the final manuscript

Competing Interests

A full list of disclosures can be found in the Supplementary Information.

Table 1 - Bi-directional Mendelian Randomization results for leisure screen time and moderate-to-vigorous intensity physical activity during leisure time with body mass index or body fat percentage.

1A using significant loci only

Exposure	Outcome	beta	SE	P-value	Exposure	Outcome	beta	SE	P-value
LST	Body fat %	0.16	0.07	0.016	LST	BMI	0.25	0.05	2.5E-5*
Body fat %	LST	0.12	0.03	0.005*	BMI	LST	0.09	0.01	7.3E-10*
MVPA	Body fat %	-0.21	0.17	0.22	MVPA	BMI	-0.11	0.07	0.14
Body fat %	MVPA	-0.001	0.036	0.97	BMI	MVPA	0.02	0.023	0.32

BMI: body mass index; LST: leisure screen time; MVPA: moderate-to-vigorous intensity physical activity during leisure time; * indicates significant(P<0.0125) results. We use MR-PRESSO with outliers removed for all pairs of traits except for the causal effect estimation between body fat % and MVPA, as no outliers were detected by MR-PRESSO. For body fat % -> MVPA, we reported the causal estimates using an inverse variance–weighted test; for MVPA -> body fat %, we reported the weighted median method, as these two methods were selected by the machine learning framework (Online Methods) to be the most appropriate approaches for each analysis, respectively.

1B 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 - 0.24	1.8E-3	LST	BMI	0.20	0.16 - 0.24	6.7E-7
Body fat %	LST	0.12	0.04 - 0.18	0.14	BMI	LST	0.12	0.09 - 0.15	2.4E-4
MVPA	Body fat %	-0.12	-0.20 - -0.04	0.07	MVPA	BMI	-0.05	-0.12 - 0.01	0.34
Body fat %	MVPA	-0.03	-0.09 - 0.02	0.53	BMI	MVPA	-0.01	-0.06 - 0.03	0.76
LST	Comparative height at age 10	0.02	0.01 - 0.03	0.04	LST	Comparative body size at age 10	0.03	0.01 - 0.04	0.04

* Gamma: posterior median of gamma, which can be taken as a point estimate of the causal effect. This estimate tends to be shrunk slightly towards zero compared with other methods. # shows 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.

Figure legends

Figure 1: Overview of the four self-reported physical activity (PA) traits and correlations with objectively assessed PA traits. **A)** an overview of the four self-reported PA traits; **B)** phenotypic (top left) and genetic (bottom right) correlation coefficients between the four self-reported PA traits studied here and three accelerometer-assessed PA traits quantified in UK Biobank participants. LST: leisure screen time; MVPA: moderate-to-vigorous intensity PA during leisure time; SDC: sedentary commuting behavior; SDW: sedentary behavior at work. AccMod: accelerometer assessed proportion of time spent in moderate intensity PA; AccWalking: accelerometer assessed proportion of time spent walking; AccSed: accelerometer assessed proportion of time spent sedentary.

Figure 2: Main results of GWAS and downstream gene prioritization for leisure screen time (LST) and moderate-to-vigorous intensity physical activity during leisure time (MVPA). **A)** a dendrogram showing the 100 independent association signals in LST- and MVPA-associated loci. Moving outwards from the center are: 1) chr; 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, i.e. DEPICT gene prioritization (Dg) or tissue enrichment (Dt); Summary-based Mendelian randomization of eQTL signals in blood (Sbl), brain (Sbr) or skeletal muscle (Ssm); credible variants identified by Finemap that a) are coding and likely to have a detrimental effect on protein function (Fcadd) or b) show evidence of 3D interactions with the candidate gene in CNS cell types (Fcrt); contribute 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. **B)** 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 ($P<5\times10^{-9}$) are indicated by red dots. SNP density that counts the number of SNPs within a 2mb region is plotted around the circle.

Figure 3: Validation of associations with moderate-to-vigorous intensity physical activity during leisure time (MVPA) and leisure screen time (LST) using polygenic scores (PGSs) in BioMe participants of three ancestries. The best performing PGSs for MVPA (**A**) and LST (**C**) were derived using regression analyses, i.e., 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. Figures **B** and **D** show the association of MVPA with the PGSs for MVPA (**B**) and LST (**D**) in individuals of African (AA), European (EA) and Hispanic (HA) ancestry in data from the BioMe BioBank.

Figure 4. Genetic correlations of four self-reported physical activity traits with complex traits and diseases. Results are based on published GWAS with $P<4.6\times10^{-4}$ for at least one physical activity or sedentary trait. Darker colors reflect higher negative (purple) or positive (red) correlation coefficients.

Figure 5. Mendelian Randomization analyses between leisure screen time (LST), moderate-to-vigorous intensity physical activity during leisure time (MVPA), BMI and complex diseases. A) Median causal estimates for Mendelian randomization analyses using the CAUSE method and causal estimates from the MR-PRESSO after outlier removal and accounting for horizontal pleiotropy; **B)** The causal effects of MVPA and LST on complex risk factors and diseases without (in orange) and with (in blue) adjusting for BMI.

Figure 6. **Allele 635A in *ACTN3* results in a more flexible alpha-actinin 3 homodimer. A)** ACTN3 is a homodimer of two antiparallel filaments, with each filament consisting of an N-terminal actin binding domain (blue), followed by a structural region comprised of four spectrin repeats (gray) with a C-terminal calmodulin homology domain (cyan); **B)** The glutamate residue in position 635 interacts with both the arginine in position 638 and the glutamine in position 639; **C)** Root mean square fluctuations (RMSF) of the spectrin repeat structural region of the ACTN3 dimer for a 150 ns molecular dynamics simulation for variants E635 (blue) and 635A (orange, higher MVPA) (bottom), with the difference in RMSF shown mapped to the spectrin repeat region (top) with ± 0.1 nm difference (red, positive and blue, negative); **D)** Umbrella sampling of *ACTN3* variants E635 and 635A with traces representing the potential of mean force for variant 635A (orange) and E635 (blue) with shaded regions indicating ± 1 standard deviation. The reaction coordinate is the distance between the two actin binding domain (ABD) centers of mass of each dimer, a negative value indicating a decrease in the 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; **E)** Single fiber experiments show a higher maximal force and fiber power during isotonic contractions in type II_A fibers from a homozygous individual for E635 compared with fibers from three E635A heterozygous individuals.

References

1. Lee, I.M. *et al.* Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *Lancet (London, England)* **380**, 219-229 (2012).
2. Organization, W.H. *Global action plan for the prevention and control of noncommunicable diseases 2013-2020*, (World Health Organization, 2013).
3. Guthold, R., Stevens, G.A., Riley, L.M. & Bull, F.C. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. *The Lancet Global Health* **6**, e1077-e1086 (2018).
4. Wang, Y. *et al.* Secular trends in sedentary behaviors and associations with weight indicators among Chinese reproductive-age women from 2004 to 2015: findings from the China Health and Nutrition Survey. *Int J Obes (Lond)* **44**, 2267-2278 (2020).
5. Wijndaele, K. *et al.* Television viewing time independently predicts all-cause and cardiovascular mortality: the EPIC Norfolk study. *Int J Epidemiol* **40**, 150-9 (2011).
6. Wijndaele, K., Sharp, S.J., Wareham, N.J. & Brage, S. Mortality Risk Reductions from Substituting Screen Time by Discretionary Activities. *Med Sci Sports Exerc* **49**, 1111-1119 (2017).
7. Bauman, A.E. *et al.* Correlates of physical activity: why are some people physically active and others not? *The Lancet* **380**, 258-271 (2012).
8. den Hoed, M. *et al.* Heritability of objectively assessed daily physical activity and sedentary behavior. *The American Journal of Clinical Nutrition* **98**, 1317-1325 (2013).
9. Stubbe, J.H. *et al.* Genetic influences on exercise participation in 37,051 twin pairs from seven countries. *PloS one* **1**, e22-e22 (2006).
10. Fan, W. *et al.* PPARδ Promotes Running Endurance by Preserving Glucose. *Cell Metab* **25**, 1186-1193.e4 (2017).

11. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* **47**, D1005-d1012 (2019).
12. Sarzynski, M.A. *et al.* Advances in Exercise, Fitness, and Performance Genomics in 2015. *Med Sci Sports Exerc* **48**, 1906-16 (2016).
13. Klimentidis, Y.C. *et al.* Genome-wide association study of habitual physical activity in over 377,000 UK Biobank participants identifies multiple variants including CADM2 and APOE. *International Journal of Obesity* **42**, 1161-1176 (2018).
14. Doherty, A. *et al.* GWAS identifies 14 loci for device-measured physical activity and sleep duration. *Nature Communications* **9**, 5257 (2018).
15. van de Vegte, Y.J., Said, M.A., Rienstra, M., van der Harst, P. & Verweij, N. Genome-wide association studies and Mendelian randomization analyses for leisure sedentary behaviours. *Nature Communications* **11**, 1770 (2020).
16. Kilpeläinen, T.O. *et al.* Multi-ancestry study of blood lipid levels identifies four loci interacting with physical activity. *Nat Commun* **10**, 376 (2019).
17. 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).
18. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187-196 (2015).
19. Kichaev, G. *et al.* Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *Am J Hum Genet* **104**, 65-75 (2019).
20. 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).

21. 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).

22. 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).

23. Locke, A.E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197-206 (2015).

24. Justice, A.E. *et al.* Genome-wide meta-analysis of 241,258 adults accounting for smoking behaviour identifies novel loci for obesity traits. *Nature Communications* **8**, 14977 (2017).

25. 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).

26. 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. *Nature Genetics* **50**, 693-698 (2018).

27. Burgess, S., Butterworth, A. & Thompson, S.G. Mendelian Randomization Analysis With Multiple Genetic Variants Using Summarized Data. *Genetic Epidemiology* **37**, 658-665 (2013).

28. Sanderson, E., Davey Smith, G., Windmeijer, F. & Bowden, J. An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. *International journal of epidemiology* **48**, 713-727 (2019).

29. Lightfoot, J.T. *et al.* Biological/Genetic Regulation of Physical Activity Level: Consensus from GenBioPAC. *Med Sci Sports Exerc* **50**, 863-873 (2018).

30. Pillon, N.J. *et al.* Transcriptomic profiling of skeletal muscle adaptations to exercise and inactivity. *Nature Communications* **11**, 470 (2020).

31. 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).

32. Threlfell, S., Sammut, S., Menniti, F.S., Schmidt, C.J. & West, A.R. Inhibition of Phosphodiesterase 10A Increases the Responsiveness of Striatal Projection Neurons to Cortical Stimulation. *J Pharmacol Exp Ther* **328**, 785-95 (2009).

33. 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**, 2640-53 (2010).

34. Zhu, Y. *et al.* Identification of CD112R as a novel checkpoint for human T cells. *J Exp Med* **213**, 167-76 (2016).

35. Inoue, M., Chang, L., Hwang, J., Chiang, S.H. & Saltiel, A.R. The exocyst complex is required for targeting of Glut4 to the plasma membrane by insulin. *Nature* **422**, 629-33 (2003).

36. Burri, L. *et al.* Mature DIABLO/Smac is produced by the IMP protease complex on the mitochondrial inner membrane. *Mol Biol Cell* **16**, 2926-33 (2005).

37. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nature Communications* **6**, 5890 (2015).

38. Muiños, M. & Ballesteros, S. Does physical exercise improve perceptual skills and visuospatial attention in older adults? A review. *European Review of Aging and Physical Activity* **15**, 2 (2018).

39. 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).

40. Timshel, P.N., Thompson, J.J. & Pers, T.H. Mapping heritability of obesity by brain cell types. *bioRxiv*, 2020.01.27.920033 (2020).

41. Schaum, N. *et al.* Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367-372 (2018).

42. Roberts, M.D., Ruegsegger, G.N., Brown, J.D. & Booth, F.W. Mechanisms Associated With Physical Activity Behavior: Insights From Rodent Experiments. *Exerc Sport Sci Rev* **45**, 217-222 (2017).

43. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nature Genetics* **48**, 481 (2016).

44. Benner, C. *et al.* FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493-501 (2016).

45. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310-5 (2014).

46. Pontzer, H., Wood, B.M. & Raichlen, D.A. Hunter-gatherers as models in public health. *Obes Rev* **19 Suppl 1**, 24-35 (2018).

47. Grossman, S.R. *et al.* Identifying recent adaptations in large-scale genomic data. *Cell* **152**, 703-13 (2013).

48. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499-503 (2015).

49. Field, Y. *et al.* Detection of human adaptation during the past 2000 years. *Science* **354**, 760-764 (2016).

50. Krueger, D.D., Osterweil, E.K. & Bear, M.F. Activation of mGluR5 induces rapid and long-lasting protein kinase D phosphorylation in hippocampal neurons. *J Mol Neurosci* **42**, 1-8 (2010).

51. Rhee, S.G. Regulation of phosphoinositide-specific phospholipase C. *Annu Rev Biochem* **70**, 281-312 (2001).

52. Zhu, X. *et al.* Phospholipase Cε deficiency delays the early stage of cutaneous wound healing and attenuates scar formation in mice. *Biochem Biophys Res Commun* **484**, 144-151 (2017).

53. Marck, A. *et al.* Age-Related Changes in Locomotor Performance Reveal a Similar Pattern for Caenorhabditis elegans, Mus domesticus, Canis familiaris, Equus caballus, and Homo sapiens. *J Gerontol A Biol Sci Med Sci* **72**, 455-463 (2017).

54. Norheim, F. *et al.* Gene-by-Sex Interactions in Mitochondrial Functions and Cardio-Metabolic Traits. *Cell metabolism* **29**, 932-949.e4 (2019).

55. Battle, A., Brown, C.D., Engelhardt, B.E. & Montgomery, S.B. Genetic effects on gene expression across human tissues. *Nature* **550**, 204-213 (2017).

56. 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).

57. de Geus, E.J., Bartels, M., Kaprio, J., Lightfoot, J.T. & Thomis, M. Genetics of regular exercise and sedentary behaviors. *Twin Res Hum Genet* **17**, 262-71 (2014).

58. Weyerstraß, J., Stewart, K., Wesselius, A. & Zeegers, M. Nine genetic polymorphisms associated with power athlete status – A Meta-Analysis. *Journal of Science and Medicine in Sport* **21**, 213-220 (2018).

59. Moir, H.J. *et al.* Genes and Elite Marathon Running Performance: A Systematic Review. *Journal of sports science & medicine* **18**, 559-568 (2019).

60. Kim, D.S., Wheeler, M.T. & Ashley, E.A. The genetics of human performance. *Nature Reviews Genetics* (2021).

61. 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**, 943-5 (1999).

62. Gielen, M. *et al.* Heritability and genetic etiology of habitual physical activity: a twin study with objective measures. *Genes Nutr* **9**, 415 (2014).

63. Pickering, C. & Kiely, J. ACTN3: More than Just a Gene for Speed. *Frontiers in Physiology* **8**(2017).

64. Vincent, B. *et al.* ACTN3 (R577X) genotype is associated with fiber type distribution. *Physiol Genomics* **32**, 58-63 (2007).

65. Broos, S. *et al.* Evidence for ACTN3 as a Speed Gene in Isolated Human Muscle Fibers. *PLoS One* **11**, e0150594 (2016).

66. Broos, S. *et al.* The stiffness response of type IIa fibres after eccentric exercise-induced muscle damage is dependent on ACTN3 r577X polymorphism. *Eur J Sport Sci* **19**, 480-489 (2019).

67. Broos, S., Van Leemputte, M., Deldicque, L. & Thomis, M.A. History-dependent force, angular velocity and muscular endurance in ACTN3 genotypes. *Eur J Appl Physiol* **115**, 1637-43 (2015).

68. Papadimitriou, N. *et al.* Physical activity and risks of breast and colorectal cancer: a Mendelian randomisation analysis. *Nature Communications* **11**, 597 (2020).

69. 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).

70. 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).

71. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *Eur J Hum Genet* **19**, 807-12 (2011).

72. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics* **47**, 291 (2015).

73. Panter, J. *et al.* Using alternatives to the car and risk of all-cause, cardiovascular and cancer mortality. *Heart* **104**, 1749-1755 (2018).

74. Kikuchi, H. *et al.* Occupational sitting time and risk of all-cause mortality among Japanese workers. *Scand J Work Environ Health* **41**, 519-28 (2015).

75. Vink, J.M. *et al.* Variance components models for physical activity with age as modifier: a comparative twin study in seven countries. *Twin Research and Human Genetics* **14**, 25-34 (2011).

76. Winkler, T.W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nature protocols* **9**, 1192 (2014).

77. Day, F.R., Loh, P.R., Scott, R.A., Ong, K.K. & Perry, J.R. A Robust Example of Collider Bias in a Genetic Association Study. *Am J Hum Genet* **98**, 392-3 (2016).

78. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191 (2010).

79. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biology* **17**, 122 (2016).

80. Gogarten, S.M. *et al.* Genetic association testing using the GENESIS R/Bioconductor package. *Bioinformatics* (2019).

81. Kamat, M.A. *et al.* PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. *Bioinformatics* (2019).

82. Loh, P.R. *et al.* Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat Genet* **47**, 1385-92 (2015).

83. Wainschtein, P. *et al.* Recovery of trait heritability from whole genome sequence data. *bioRxiv*, 588020 (2019).

84. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369-75, s1-3 (2012).

85. Price, A.L. *et al.* Long-range LD can confound genome scans in admixed populations. *American journal of human genetics* **83**, 132-139 (2008).

86. Choi, S.W. & O'Reilly, P.F. PRSice-2: Polygenic Risk Score software for biobank-scale data. *GigaScience* **8**(2019).

87. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **7**, e34408 (2018).

88. Burgess, S., Davies, N.M. & Thompson, S.G. Bias due to participant overlap in two-sample Mendelian randomization. *Genetic epidemiology* **40**, 597-608 (2016).

89. Bowden, J. *et al.* Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic. *International journal of epidemiology* **45**, 1961-1974 (2016).

90. Hemani, G. *et al.* Automating Mendelian randomization through machine learning to construct a putative causal map of the human phenome. *bioRxiv*, 173682 (2017).

91. Lyon, M.S. *et al.* The variant call format provides efficient and robust storage of GWAS summary statistics. *Genome Biology* **22**, 32 (2021).

92. 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**, R466-72 (2005).

93. Qi, T. *et al.* Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nature communications* **9**, 2282-2282 (2018).

94. Fromer, M. *et al.* Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat Neurosci* **19**, 1442-1453 (2016).

95. Ng, B. *et al.* An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nature neuroscience* **20**, 1418-1426 (2017).

96. Võsa, U. *et al.* Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. *bioRxiv*, 447367 (2018).

97. Barbeira, A.N. *et al.* Exploiting the GTEx resources to decipher the mechanisms at GWAS loci. *Genome Biol* **22**, 49 (2021).

98. Belton, J.M. *et al.* Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268-76 (2012).

99. Lusis, A.J. *et al.* The Hybrid Mouse Diversity Panel: a resource for systems genetics analyses of metabolic and cardiovascular traits. *J Lipid Res* **57**, 925-42 (2016).

100. Gautel, M. & Djinović-Carugo, K. The sarcomeric cytoskeleton: from molecules to motion. *J Exp Biol* **219**, 135-45 (2016).

101. Luther, P.K. The vertebrate muscle Z-disc: sarcomere anchor for structure and signalling. *J Muscle Res Cell Motil* **30**, 171-85 (2009).

102. Luther, P.K. Three-dimensional reconstruction of a simple Z-band in fish muscle. *J Cell Biol* **113**, 1043-55 (1991).

103. Grison, M., Merkel, U., Kostan, J., Djinović-Carugo, K. & Rief, M. α-Actinin/titin interaction: A dynamic and mechanically stable cluster of bonds in the muscle Z-disk. *Proc Natl Acad Sci U S A* **114**, 1015-1020 (2017).

104. 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**, 845-58 (2015).

105. Abraham, M.J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19-25 (2015).

106. Huang, J. & MacKerell, A.D., Jr. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J Comput Chem* **34**, 2135-45 (2013).

107. Goodrich B, G.J., Ali I & Brilleman S. rstanarm: Bayesian applied regression modeling via Stan. (R package version 2.14.1 https://mc-stan.org/rstanarm., 2016).

108. Pastore, M. & Calcagnì, A. Measuring Distribution Similarities Between Samples: A Distribution-Free Overlapping Index. *Front Psychol* **10**, 1089 (2019).

Figures

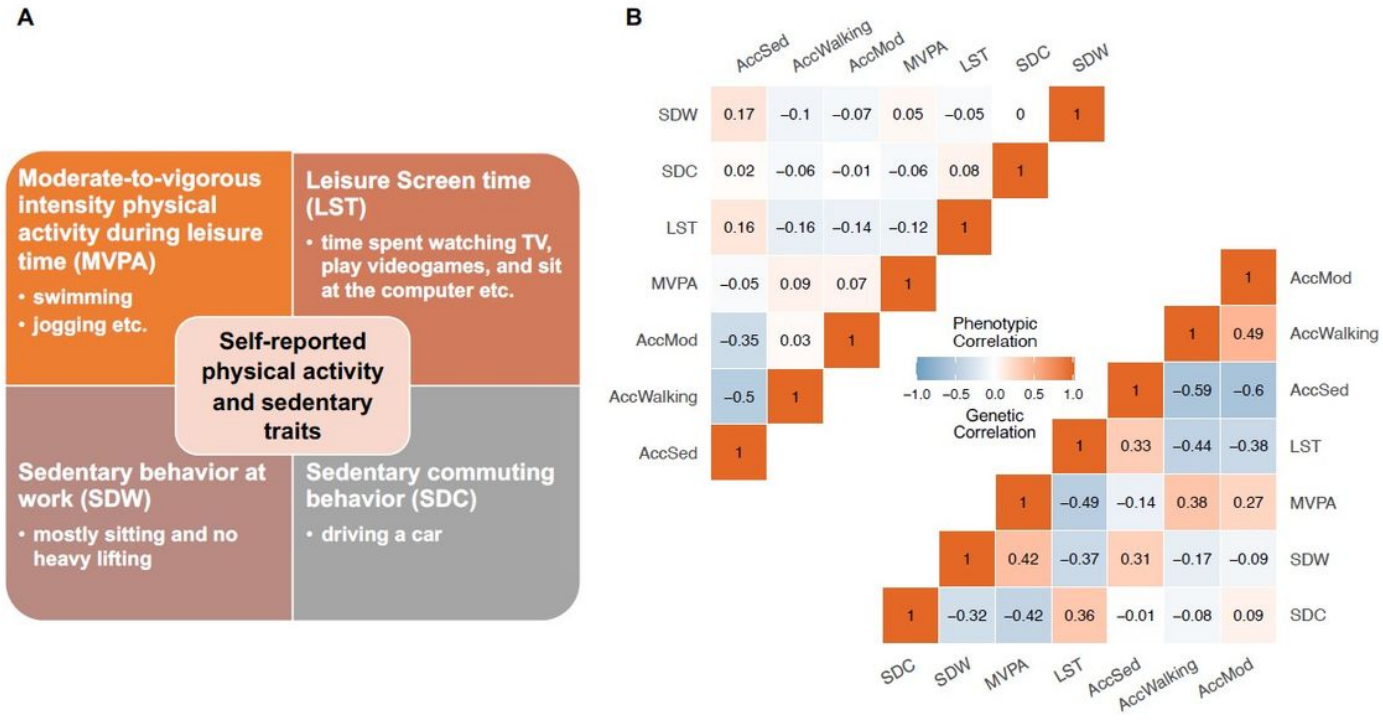

Figure 1

Overview of the four self-reported physical activity (PA) and sedentary traits and correlations with objectively assessed PA and sedentary traits. A) an overview of the four self-reported PA and sedentary traits; B) phenotypic (top left) and genetic (bottom right) correlation coefficients between the four self-reported PA and sedentary traits studied here and three accelerometer-assessed PA and sedentary traits quantified in UK Biobank participants. LST: leisure screen time; MVPA: moderate-tovigorous intensity PA during leisure time; SDC: sedentary commuting behavior; SDW: sedentary behavior at work. AccMod: accelerometer assessed proportion of time spent in moderate intensity PA; AccWalking: accelerometer assessed proportion of time spent walking; AccSed: accelerometer assessed proportion of time spent sedentary.

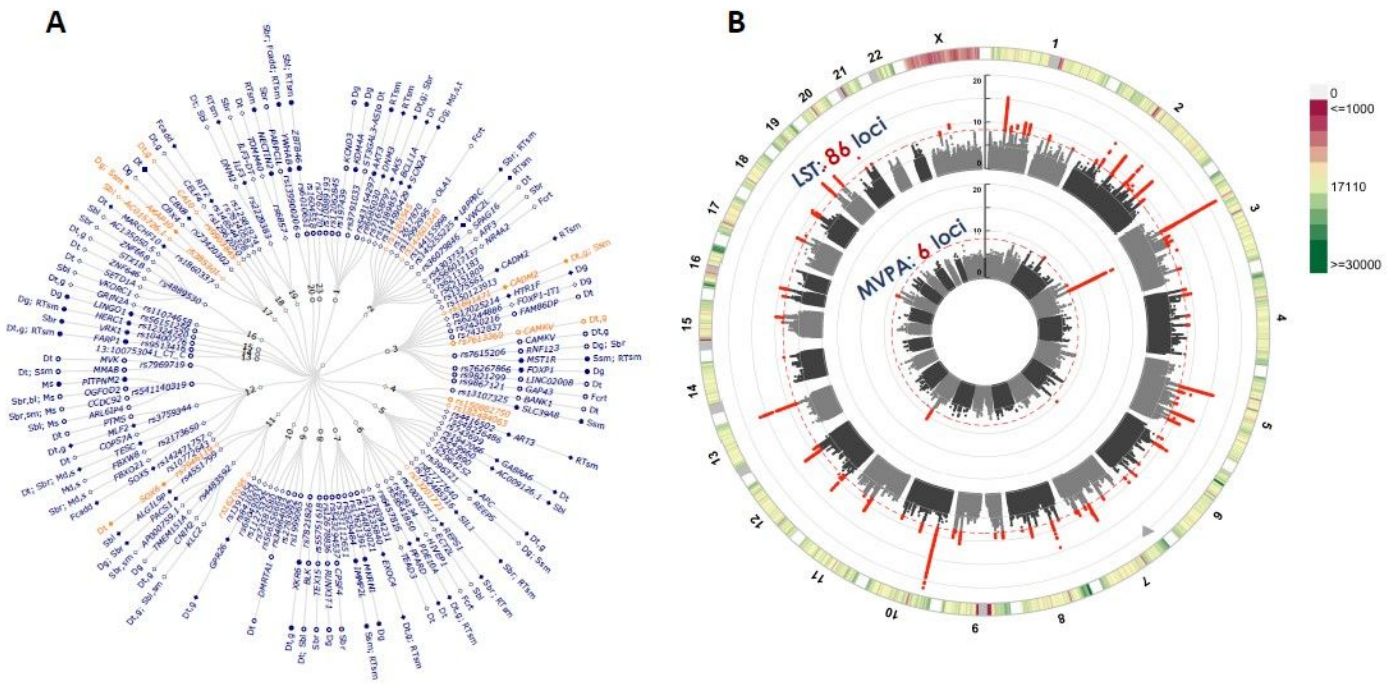

Figure 2

Main results of GWAS and downstream gene prioritization for leisure screen time (LST) and moderate-to-vigorous intensity physical activity during leisure time (MVPA). A) a dendrogram showing the 100 independent association signals in LST- and MVPA-associated loci. Moving outwards from the center are: 1) chr; 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, i.e. DEPICT gene prioritization (Dg) or tissue enrichment (Dt); Summary based Mendelian randomization of eQTL signals in blood (Sbl), brain (Sbr) or skeletal muscle (Ssm); credible variants identified by Finemap that a) are coding and likely to have a detrimental effect on protein function (Fcadd) or b) show evidence of 3D interactions with the candidate gene in CNS cell types (Fcrt); contribute 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. B) 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<5×10−9) are indicated by red dots. SNP density that counts the number of SNP within 2mb region is plotted around the

circle.

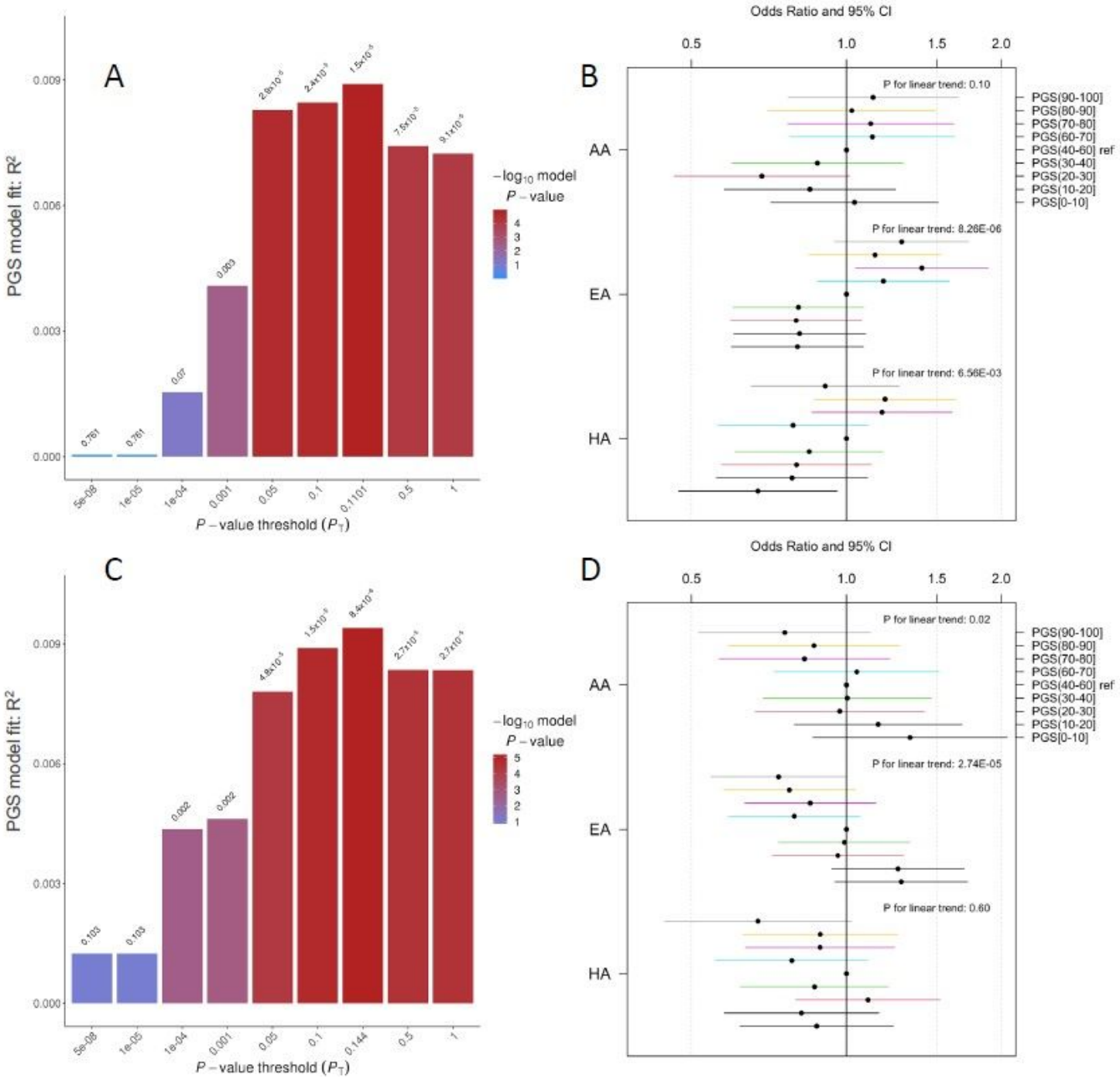

Figure 3

Validation of associations with moderate-to-vigorous intensity physical activity during leisure time (MVPA) and leisure screen time (LST) using polygenic scores (PGSs) in

BioMe participants of three ancestries. The best performing PGSs for MVPA (A) and LST (C) were derived using regression analyses, i.e. those with the highest incremental R2 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. Figures B and D show the association of MVPA with the PGSs

for MVPA (B) and LST (D) in individuals of African (AA), European (EA) and Hispanic (HA) ancestry in data from the BioMe BioBank.

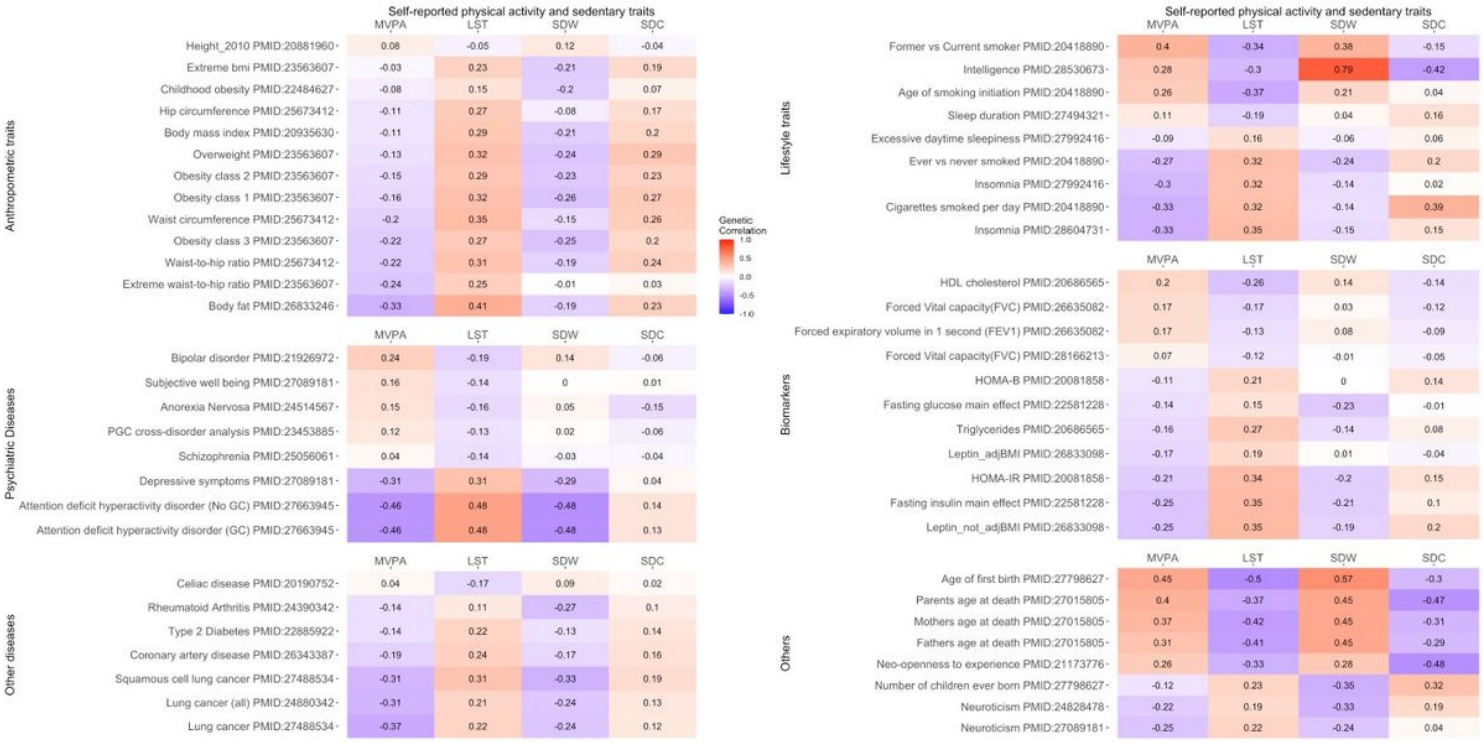

Figure 4

Genetic correlations of four self-reported physical activity and sedentary traits with complex traits and diseases. Results are based on published GWAS with P<4.6×10−4 for at least one physical activity trait. Darker colors reflect higher negative (purple) or positive (red) correlation coefficients. LST: leisure screen time; MVPA: moderate-to-vigorous intensity PA during leisure time; SDC: sedentary commuting behavior; SDW: sedentary behavior at work.

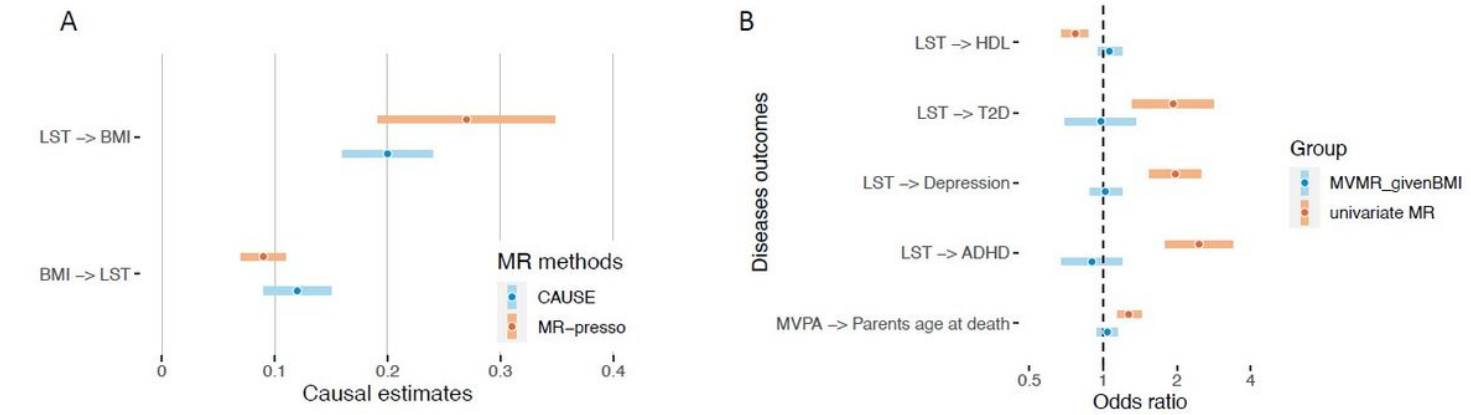

Figure 5

Mendelian Randomization analyses between leisure screen time (LST), moderate-to-vigorous intensity physical activity during leisure time (MVPA), BMI and complex diseases. A) Median causal estimates for Mendelian randomization 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 and MVPA on complex risk factors and diseases without (in orange) and with (in blue) adjusting for BMI

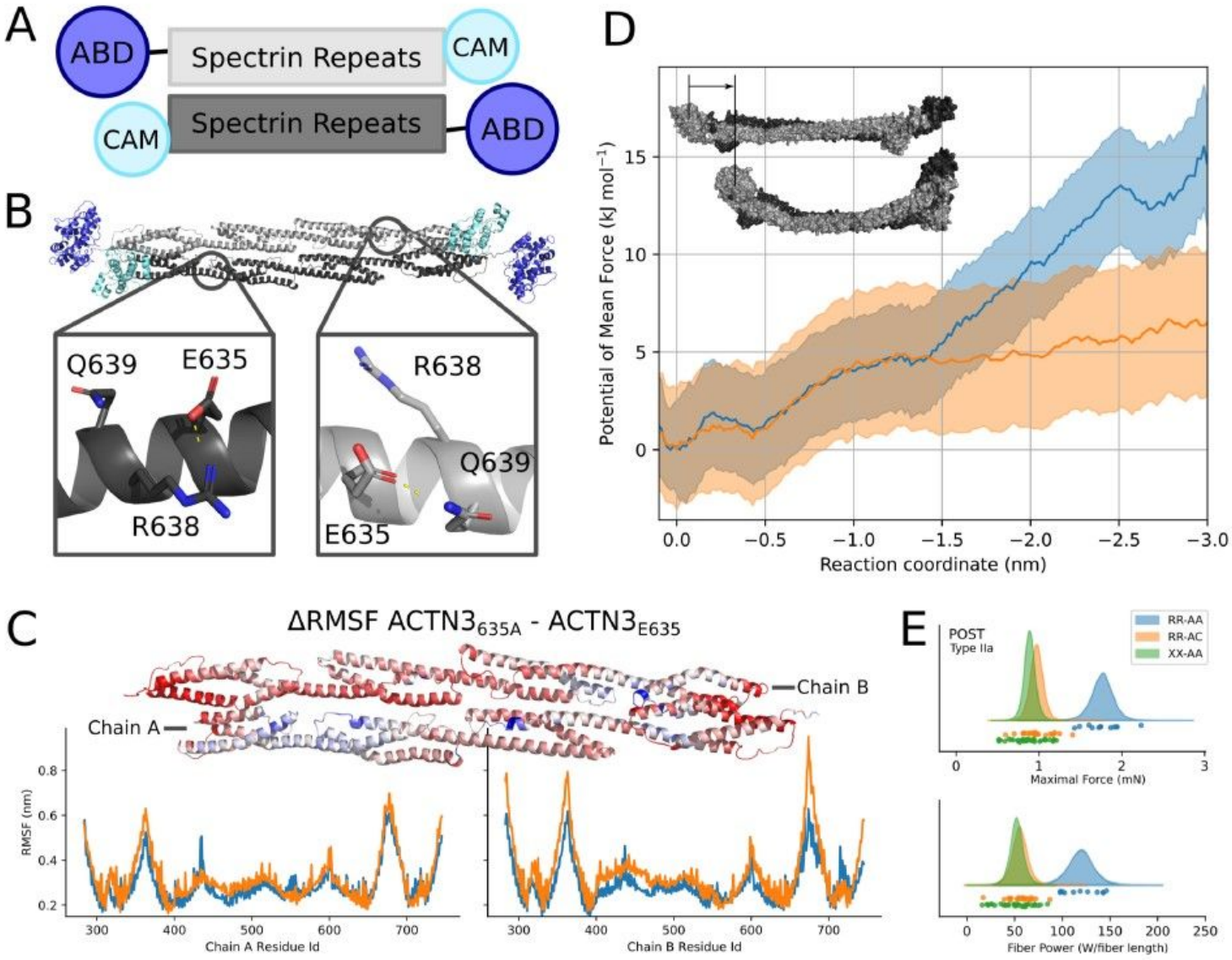

Figure 6

Allele 635A in ACTN3 results in a more flexible alpha actinin 3 homodimer. A) ACTN3 is a homodimer of two antiparallel filaments, with each filament consisting of an N-terminal actin binding domain (blue), followed by a structural region comprised of four spectrin repeats (gray) with a Cterminal calmodulin homology domain (cyan); B) The glutamate residue in position 635 interacts with both the arginine in position 638 and the glutamine in position 639; C) Root mean square fluctuations (RMSF) of the spectrin repeat structural region of the ACTN3 dimer for a 150 ns molecular dynamics simulation for variants E635 (blue) and 635A (orange, higher MVPA) (bottom), with the difference in RMSF shown mapped to the

spectrin repeat region (top) with ± 0.1 nm difference (red, positive and blue, negative); D) Umbrella sampling of ACTN3 variants E635 and 635A with traces representing the potential of mean force for variant 635A (orange) and E635 (blue) with shaded regions indicating ± 1 standard deviation. The reaction coordinate is the distance between the two actin binding domain (ABD) centers of mass of each dimer, a negative value indicating a decrease in the 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; E) Single fiber experiments show a higher maximal force and fiber power during isotonic contractions in type IIA fibers from a homozygous individual for E635 compared with fibers from three E635A heterozygous individuals.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- 520211130SuppInformation.docx
- 620211130SuppTables12.xlsx
- 720211130SuppTables325.xlsx