Genome-wide association study of habitual physical activity in over 377,000 UK Biobank participants identifies multiple variants including CADM2 and APOE

Genome-wide association study of habitual physical activity in over 377,000 UK Biobank participants identifies multiple variants including *CADM2* and *APOE*

Short Title: Genome-wide association study of habitual physical activity

Yann C. Klimentidis[1*], David A. Raichlen[2], Jennifer Bea[3,4], David O. Garcia[5], Nathan E. Wineinger, Lawrence J. Mandarino[7], Gene E. Alexander[8], Zhao Chen[1], Scott B. Going[4]

[1] Department of Epidemiology and Biostatistics, Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, Arizona, USA

[2] School of Anthropology, University of Arizona, Tucson, Arizona, USA

[3] Department of Medicine, University of Arizona, Tucson, Arizona, USA

[4] Department of Nutritional Sciences, University of Arizona, Tucson, Arizona, USA

[5] Department of Health Promotion Sciences, Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, Arizona, USA

[6] Scripps Translational Science Institute, La Jolla, CA, USA

[7] Center for Disparities in Diabetes, Obesity and Metabolism, Division of Endocrinology, Diabetes and Metabolism, Department of Medicine, University of Arizona, Tucson, Arizona, USA

[8] Departments of Psychology and Psychiatry, Neuroscience and Physiological Sciences Interdisciplinary Programs, BIO5 Institute, and Evelyn F. McKnight Brain Institute, University of Arizona, Tucson, Arizona, USA; Arizona Alzheimer's Consortium, Phoenix, Arizona, USA

*** Corresponding Author:** Department of Epidemiology and Biostatistics, Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, Arizona, 85724, 520-621-0147, yann@email.arizona.edu

Conflict of interest: The authors declare no conflict of interest.

Abstract

Background/Objectives: Physical activity (PA) protects against a wide range of diseases. Habitual PA appears to be heritable, motivating the search for specific genetic variants that may inform efforts to promote PA and target the best type of PA for each individual.

Subjects/Methods: We used data from the UK Biobank to perform the largest genome-wide association study of PA to date, using three measures based on self-report (n_{max}=377,234) and two measures based on wrist-worn accelerometry data (n_{max}=91,084). We examined genetic correlations of PA with other traits and diseases, as well as tissue-specific gene expression patterns. With data from the Atherosclerosis Risk in Communities (ARIC; n=8,556) study, we performed a meta-analysis of our top hits for moderate-to-vigorous PA (MVPA).

Results: We identified ten loci across all PA measures that were significant in both a basic and a fully adjusted model ($p<5 \times 10^{-9}$). Upon meta-analysis of the nine top hits for MVPA with results from ARIC, eight were genome-wide significant. Interestingly, among these, the rs429358 variant in the *APOE* gene was the most strongly associated with MVPA, whereby the allele associated with higher Alzheimer's risk was associated with greater MVPA. However, we were not able to rule out possible selection bias underlying this result. Variants in *CADM2*, a gene previously implicated in obesity, risk-taking behavior and other traits, were found to be associated with habitual PA. We also identified three loci consistently associated ($p<5 \times 10^{-5}$) with PA across both self-report and accelerometry, including *CADM2*. We find genetic correlations of PA with educational attainment, chronotype, psychiatric traits, and obesity-related traits. Tissue enrichment analyses implicate the brain and pituitary gland as locations where PA-associated loci may exert their actions.

Conclusions: These results provide new insight into the genetic basis of habitual PA, and the genetic links connecting PA with other traits and diseases.

Introduction

A physically active lifestyle has been shown to protect against a wide range of diseases, including cardiovascular disease, cancer, type-2 diabetes, osteoporosis, and Alzheimer's disease [1–4]. Levels of engagement in physical activity (PA) vary across individuals, and most people do not meet recommended levels to achieve health benefits. Although cultural, economic, and other environmental factors influence PA engagement [5,6], genetic factors also likely play a role. Understanding the genetic factors underlying inter-individual variation will better inform efforts to promote PA and potentially allow targeting the best type of PA for each person, what might be called "Precision Exercise Prescription".

Evidence of genetic factors underlying the propensity to exercise in humans has been demonstrated in a number of studies [7–13]. Several studies have utilized a candidate gene approach to identify specific genetic variants associated with a proclivity towards PA [8,14–18]. This work generally focused on genes related to the serotonin and dopamine systems, energy metabolism, and neurotrophic factors. However, to our knowledge there have been only two previous reports of genome-wide association studies (GWAS) of PA [19,20], neither of which identified a locus at genome-wide significance, likely due to relatively small sample sizes. Thus, while previous work strongly suggests a genetic basis for engagement in PA, the genes that contribute to this healthy lifestyle behavior remain unknown.

In this study, we conduct the largest GWAS of PA to date, aiming to identify genetic variants associated with self-reported and accelerometry-based levels of habitual, leisure-time PA. We sought to identify variants in the UK Biobank, a large cohort study of 500,000 adults measured across a wide range of characteristics including genome-wide markers. We then examined the genetic correlation of PA with other traits, examined putative tissues where PA genes may exert their effects, and meta-analyzed the identified loci for MVPA with data on self-reported PA in an independent cohort from the Atherosclerosis Risk in Communities (ARIC) study.

Methods

Studies

Data from the UK Biobank study were used for discovery of variants. Briefly, the UK Biobank is a large prospective cohort study of approximately a half-million adults (ages 40-69) living in the United Kingdom (UK), recruited from 22 centers across the UK [21]. All participants provided written informed consent. Ethical approval of the UK Biobank study was given by the North West Multicentre Research Ethics Committee, the National Information Governance Board for Health & Social Care, and the Community Health Index Advisory Group. We also used data from the ARIC study (n=8,556), which is a prospective cohort study of over 15,000 adults aged 45-64 years that took place in four United States communities. The selection of this cohort for replication was based on 1) the quality of the PA phenotype which incorporates multiple questions assessing types, intensities, and frequency of PA (see below), 2) the focus on habitual, leisure-time PA, and 3) the relatively large sample size. In the absence of previous effect size estimates for genetic variants on PA, the sample size in ARIC, although comparatively much smaller than the UK Biobank, was deemed, on an a-priori basis, to serve as a suitable replication cohort. Details of the ARIC study can be found elsewhere [22]. All participants in ARIC provided written informed consent. Institutional review board approval was obtained by each participating field center, and this study was approved by the University of Arizona Human Subjects Protection Program (Protocol number: 1300000659R001). To reduce the potential for confounding by population stratification, we included only individuals of white race/ethnicity in both studies.

Physical activity

In the UK Biobank, self-reported levels of physical activity during work and leisure time were measured via a touchscreen questionnaire, in a fashion similar to the International Physical Activity Questionnaire [23]. For moderate PA (MPA), participants were asked: “In a typical WEEK, on how many

days did you do 10 minutes or more of moderate physical activities like carrying light loads, cycling at normal pace? (Do not include walking)". For vigorous PA (VPA), participants were asked: "In a typical WEEK, how many days did you do 10 minutes or more of vigorous physical activity? (These are activities that make you sweat or breathe hard such as fast cycling, aerobics, heavy lifting)". For each of these questions, those who indicated 1 or more such days were then asked "How many minutes did you usually spend doing moderate/vigorous activities on a typical DAY". Participants were asked to include activities performed for work, leisure, travel and around the house. We excluded individuals who selected "prefer not to answer" or "do not know" on the above questions, those reporting not being able to walk, and individuals reporting more than 16 hours of either MPA or VPA per day. Those reporting >3hr/day of VPA or MPA were recoded to 3 hours, as recommended [24]. Moderate-to-vigorous PA (MVPA) was calculated by taking the sum of total minutes/week of MPA multiplied by four and the total number of VPA minutes/week multiplied by eight, corresponding to their metabolic equivalents, as previously described [23,25].

Since heritability has previously been shown to be higher for intense/vigorous physical activity [12], we also considered VPA on its own. Because the distribution of minutes/week of VPA was highly skewed and zero-inflated, we chose to dichotomize minutes/week of VPA into those who reported 0 days of VPA, and those reporting 3 or more days of VPA and also reporting a typical duration of VPA that is 25 minutes or greater, corresponding to common physical activity guidelines [26]. Individuals that did not fall into either of these two groups were excluded. We decided to pick extremes because of the heavily skewed and zero-inflated nature of vigorous activity duration, and in order to increase our power to detect associations. We also performed a sensitivity analysis in which we included individuals who did not fall into either of the two groups described above, and placed these individuals in the group that did not meet 3 days of VPA/week of 25 minutes or greater per day (i.e. those meeting the 3 days/week of VPA at 25/mins per day vs. not meeting this amount).

We used responses to the question "In the last 4 weeks did you spend any time doing the following?" and follow-up questions assessing the frequency and typical duration of "strenuous sports" and of "other exercises". The possible responses to the initial question were: 'walking for pleasure', 'other exercises', 'strenuous sports', 'light DIY', 'heavy DIY', 'none of the above', and 'prefer not to answer'. We identified individuals spending 2-3 days/week or more doing strenuous sports or other exercises (SSOE), for a duration of 15-30 minutes or greater. Controls were those individuals who did not indicate spending any time in the last 4 weeks doing either strenuous sports or other exercises. Individuals that did not fall into either of these two groups were excluded. Extremes were chosen because of the heavily skewed and zero-inflated distributions of these variables.

Also, in the UK Biobank, approximately 100,000 participants wore an Axivity AX3 wrist-worn accelerometer, as previously described [27]. We examined two measures derived from up to seven days of accelerometer wear: overall acceleration average, and fraction of accelerations > 425 milli-gravities (mg) [27]. Since the variable that is available in the UK Biobank is the fraction < 425 mg, we subtracted 1 from this variable. The 425 mg cutoff was chosen because this corresponds to an equivalent of vigorous physical activity (6 METs), as previously reported [28]. For both accelerometry variables, individuals with less than three days (72 hours) of data, or those not having data in each one-hour period of the 24-hour cycle were excluded. Based on missing data simulations by Doherty et al, 72 hours of wear was determined to be needed to be within 10% of a complete seven-day measure [27]. Device non-wear time, defined as consecutive stationary episodes >=60 minutes where all three axes had a standard deviation <13 mg, was imputed using the average of similar time-of-day vector magnitude and intensity distribution data points on different days [27]. This accounts for wear-time diurnal bias that may occur if the device was less worn during sleep in some individuals [27]. Finally, we also excluded outliers with values more than 4 standard deviations above the mean.

In ARIC, self-reported PA was assessed for sports/exercise, within the previous year, based on a modification of the Baecke questionnaire [29,30]. The sport/exercise index is based on up to four

sports/exercises (including modalities of mild, moderate, and strenuous energy exertion) that participants reported in the past year, and was calculated with responses to 4 items: frequency of participation in sports/exercise; frequency of sweating during sports/exercise; a subjective rating of the frequency of participation in sports/exercise compared to others in the same age group; the sum of frequency, duration, and intensity of up to 4 reported sports/exercises. This derived index is described in greater detail elsewhere, along with an assessment of its reliability and accuracy [31].

Genotypes

The majority of UK Biobank participants were genotyped with the Affymetrix UK Biobank Axiom Array (Santa Clara, CA, USA), while 10% of participants were genotyped with the Affymetrix UK BiLEVE Axiom Array. Detailed quality control and imputation procedures are described elsewhere [32]. Briefly, phasing was performed by the UK Biobank team in chunks of 15,000 markers, using SHAPEIT3 [33] software and 1,000 Genomes phase 3[34] dataset as a reference panel. Imputation was performed using a combined panel of the Haplotype Reference Consortium [35] and the UK10K haplotype resource [36] after appropriate marker and sample QC in chunks of 50,000 imputed markers. Principal Components Analysis was also performed by the UK Biobank team, using fastPCA [37] software on a set of 147,604 high-quality directly genotyped markers (pruned to minimize LD), and a set of 407,219 unrelated high-quality samples. All other samples were then projected onto the principal components [32]. Since corrections for potential problems with the position assignment of the SNPs from the UK10K haplotype resource were not available at the time of analysis, we only included SNPs imputed from the Haplotype Reference Consortium. To minimize the possibility of confounding due to population stratification, only participants who self-identified as European were included. Individuals were excluded based on unusually high heterozygosity or >5% missing rate, a mismatch between self-reported and genetically-inferred sex. These criteria resulted in a total available sample size of 458,969 individuals with genotype data. SNP

exclusions were made based on Hardy-Weinberg equilibrium ($p<1x10^{-6}$), high missingness (>1.5%), low minor allele frequency (<0.1%), and low imputation quality (info<0.4). A total of approximately 11.8 million SNPs were used in analyses. Of these, 4.1 million have a minor allele frequency < 1%.

In ARIC, participants were genotyped with the Affymetrix Genome-Wide Human SNP Array 6.0 (Affymetrix, Santa Clara, CA, USA). Standard quality control procedures were implemented prior to imputation with IMPUTE2 [38], using all individuals in the 1,000 Genomes phase 1 integrated v3 release. Quality-control procedures consisted of excluding SNPs with minor allele frequency < 1%, with missingness > 10%, and SNPs out of Hardy-Weinberg equilibrium ($p<1 \times 10^{-6}$), and excluding individuals with SNP missingness > 10%. We used principal components for the European-ancestry group as provided by ARIC in dbGaP. Briefly, LD pruning resulted in 71,702 SNPs that were used to derive principal components. A total sample size of 8,556 participants was used in the analysis.

Statistical analyses

For the continuous variables in the UK Biobank (MVPA and accelerometry variables) we created an adjusted phenotype corresponding to the residual of the regression of the following independent variables on the respective dependent PA variable: age, sex, genotyping chip, first ten genomic principal components, center, season (month) at center visit or wearing accelerometer (coded 0 for Winter, 1 for Fall or Spring, and 2 for Summer). In another model (Model 2), we considered the additional inclusion of the following covariates: levels of physical activity at work (coded as 0 by default, 1 for 'sometimes', 2 for 'usually', and 3 for 'always'), extent of walking or standing at work (coded similarly as previous variable), and the Townsend Deprivation Index (TDI; a composite measure of deprivation as previously described [39,40]). We also considered a third model (Model 3) in which body mass index (BMI) was included as an additional covariate. These covariates were considered since both self-reported and accelerometer-based measures of PA could include PA done as part of one's employment, as opposed to

PA during leisure-time. Additionally, both SES and BMI may affect participation in leisure-time PA [5]. Since the MVPA and fraction of accelerations > 425 mg variables exhibited skewed distributions, we inverse-normalized these variables prior to inclusion in the models. Model residuals conformed to the assumptions of normality and homoscedasticity. GWAS were performed with BOLT-LMM software [41,42], which implements a mixed-model linear regression that includes a random effect consisting of the SNPs other than the one being tested, and thus takes into account relatedness among subjects. Since BOLT-LMM implements a linear regression, effect size estimates for case-control outcomes are unreliable. Therefore, as previously done elsewhere [43,44], we derived effect size estimates for the genome-wide significant SNPs for the binary outcomes (VPA and SSOE) using logistic regression in R with the same set of fixed-effect covariates. Given the number of low-frequency SNPs [45] and phenotypes tested, we used a more stringent genome-wide significance threshold: $p<5 \times 10^{-9}$. To examine the relationship of PA-associated SNPs with BMI, we tested the association of identified SNPs with BMI, which was first inverse-normalized, then adjusted via a linear regression with age, sex, genotyping chip, first ten genomic principal components, and center as independent variables. We also sought to identify variants consistently associated with PA across self-report and accelerometry PA measures, for overall PA and for high-intensity PA. We thus searched for variants associated in the same direction, with $p<5 \times 10^{-5}$ for: 1) MVPA and average acceleration, and 2) VPA, SSOE, and fraction of accelerations >425 mg.

To determine the extent to which the loci identified in Model 3 may have been subject to collider bias on account of including BMI as a covariate, we derived an approach to estimate the unbiased effect of each SNP of interest on each metric of PA. Our approach was an extension of the methodology employed in Yaghootkar et al. [46] – the primary difference being the collider in our method (i.e., BMI) is a quantitative trait as opposed to a categorical/disease trait. The unbiased coefficient can be expressed as:

$$\beta^{*}_{SNP \rightarrow PA} = \frac{\beta_{SNP,\mathrm{I}} + \beta_{SNP,\mathrm{II}}\beta_{BMI,\mathrm{I}}}{1 - \beta_{BMI,\mathrm{I}}\beta_{PA,\mathrm{II}}}$$

where each $\beta_{i,j}$ is the corresponding coefficient of the i[th] variable in the j[th] model:

$$\mathrm{I}: PA \sim SNP + BMI$$
$$\mathrm{II}: BMI \sim SNP + PA.$$

We confirmed this approach via simulation under a variety of conditions, including the inclusion of additional covariates and different relationships between BMI, PA, and a genetic marker (results not shown). For simplicity we modeled PA as a quantitative trait in all cases.

Given the association that we identified with the rs428358 variant in *APOE* (see Results), we performed several additional analyses. First, we examined the associations with the *APOE* ε4 haplotype, using this SNP along with the rs7412 SNP. Different protein isoforms of APOE, which is a component of various lipoproteins, are produced by the different haplotypes defined by these two SNPs, and these haplotypes are well-established risk factors for Alzheimer's disease [47] and coronary artery disease [48,49]. Individuals with homozygous CC genotypes at both of these SNPs were classified as homozygous for the APOE ε4 allele. Individuals with homozygous CC genotypes at either SNP and heterozygous at the other SNP were classified as being heterozygous for the ε4 allele. We excluded a relatively small number of individuals heterozygous at both SNPs (n≈10,000), because it is not possible to assign a haplotype status when both loci are heterozygous. We assumed an additive model in association testing. Second, to examine whether this association may be driven by individuals with a known family history of Alzheimer's disease increasing their levels of PA, we examined the association of a binary variable indicating any self-reported first-degree family history (mother, father, or siblings) of Alzheimer's disease or dementia with MVPA. Third, we examined the interaction of family history with the rs429358 SNP on MVPA. Fourth, we examined whether the association of rs429358 with MVPA was modified by age, by testing the interaction of this SNP with age, and testing the association of rs429358 with MVPA among individuals in their 40s, 50s, and 60s. Finally, given prior evidence of an association of *APOE* variants with BMI and the slightly attenuated associations upon our adjustment for BMI, we tested whether BMI

mediated the association of rs429358 with MVPA. For this analysis, we used the mediation package [50] in R statistical software [51].

All genome-wide significant loci were examined in ARIC, where we modeled PA as a continuous variable (as described above). We used multiple linear regression to model PA as a function of age, sex, first ten genomic principal components, center, season (coded in the same way as described above). Residuals from this model conformed to the assumptions of normality and homoscedasticity. They were standardized to have a mean of 0 and standard deviation of 1, and were used as the outcome in the genome-wide SNP association analysis. We performed meta-analysis of the top hits for MVPA in the UK Biobank with the corresponding SNP association results in ARIC, using fixed-effects inverse-variance weighted meta-analysis. We also used a method that uses only the p-values [52] to perform meta-analyses of the top hits for the other UK Biobank PA measures. Additional analyses were performed with R statistical software.

To examine the association of genes identified in the UK Biobank with gene expression patterns in different tissues, we used the web-based platform, Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA GWAS) [53], which uses data from GTEx [54] and the MAGMA gene-based analysis [55]. Also from this platform, we examined results of gene-set analyses performed for curated gene sets and Gene Ontology terms. We also used the summary statistics from our UK Biobank GWAS to examine heritability of PA traits and their genetic correlation with over 200 traits and diseases using LD score regression [56–58], implemented in an online interface (http://ldsc.broadinstitute.org/). Briefly, this method tests the correlation between the LD score of each SNP (reflecting how correlated it is with nearby SNPs) and its test statistic, where the slope corresponds to the SNP/chip heritability, and the intercept is an estimate of inflation. It uses only SNPs that are likely well imputed (1,000 Genomes [34] EUR MAF>5%), while removing strand-ambiguous SNPs, insertions/deletions, those that do not match those in the 1000 Genomes data phase 3 , those in the MHC region, and those with extremely large effect sizes [57,58]. The genetics of other traits and diseases are inferred from previously published GWAS. A

significant genetic correlation was considered if $p < 2.5 \times 10^{-4}$, assuming a correction for 200 different tests, which is conservative given that many of the traits/diseases tested are correlated with each other. Finally, we queried our top hits in the Oxford Brain Imaging Genetics Server to examine associations with other traits, including brain imaging phenotypes, in the UK Biobank (http://big.stats.ox.ac.uk/) [59], and in the GTEx Portal for expression quantitative trait locus (eQTL) analysis.

Results

Self-reported PA in UK Biobank

There were 377,234 individuals with non-missing MVPA data. 80,721 individuals were excluded due to insufficient data on either moderate or vigorous PA days and/or duration. There were 261,055 individuals with non-missing VPA data. 175,965 individuals were excluded from the VPA analysis because they belonged to neither of the two defined groups. 21,946 individuals were excluded from VPA analyses because of insufficient data on VPA days and/or duration. For both measures, individuals excluded because of insufficient data were more likely to be female, older, and have a higher BMI ($p < 2 \times 10^{-16}$). A summary of self-report PA variables can be found in Table 1. BMI and TDI were consistently negatively associated with these variables, whereas warmer season and male gender were consistently positively associated with them (see Supplementary Table 1). Physical activity at work was positively associated with MVPA and VPA, and negatively associated with SSOE. Self-report PA measures were weakly correlated with accelerometry-based measures (see Supplementary Table 2). 'Chip heritability' estimates for self-report PA measures were approximately 5% (Supplementary Table 3). Although Q-Q plots show some evidence of inflation (see Supplementary Figure 1), LD score regression intercepts (<1.03) suggest no significant systematic inflation of test statistics.

We found nine loci significantly associated ($p < 5 \times 10^{-9}$) with MVPA (see Figure 1 and Table 2). Among these, four were significantly associated with MVPA in both Models 1 and 3: *APOE*, *EXOC4*,

CADM2, and PAX5. The *RPP21* and *ZNF165* were significant in Models 1 and 2, but not in Model 3. However, effect estimates were generally similar for all nine loci across Models 1 and 3. The *CALU* locus was only significant in Model 1. The *PBX2* and *ARHGEF26-AS1* loci were only significant in Models 2 and 3. Two loci, *C11orf80* and *CCDC188*, were only significant in Model 3 (see Supplementary Tables 4 and 5, Supplementary Figures 2 and 3). Among the nine identified SNPs, six exhibited trends with BMI in the expected direction, based on the negative phenotypic correlation between PA and BMI. Among the other three, the PA-increasing alleles at the *CADM2* and *PAX5* loci are associated with higher BMI ($p=7.9 \times 10^{-13}$ and 5.2×10^{-8}, respectively; see Table 2).

Most notably among the MVPA associations, the C allele at SNP rs429358 in *APOE* was associated with higher self-reported MVPA. This MVPA-increasing allele is the one associated with higher Alzheimer's disease risk (see Discussion). We found it to also be associated with higher levels of the four other PA measures, though not always reaching nominal significance (VPA: $p=5.3 \times 10^{-7}$; SSOE: $p=0.097$; average acceleration: $p=8.2 \times 10^{-3}$; fraction of accelerations>425mg: $p=0.24$). Testing the association of the Alzheimer's disease-related *APOE* ε4 allele with MVPA resulted in nearly identical findings. In models adjusted for other covariates, including BMI, this *APOE* variant remained genome-wide significant (see Supplementary Tables 4 and 5 and Supplementary Figures 2 and 3). Mediation analysis suggests partial mediation (~14%, $p<2 \times 10^{-16}$) by BMI on the association of rs429358 with MVPA. There were 45,440 individuals reporting any family history of Alzheimer's disease or dementia among parents and siblings. These individuals reported lower levels of MVPA ($p=1.2 \times 10^{-4}$). We found a significant interaction of rs429358 with family history ($p=0.012$), whereby the association of rs429358 with MVPA was stronger among those reporting a family history ($\beta=0.042$, $p=5.13 \times 10^{-8}$) compared to among those without a reported family history ($\beta=0.019$, $p=6.38 \times 10^{-9}$). We also observed a significant interaction of age with this variant ($p=0.005$). Specifically, among individuals in their forties, the association was weaker ($\beta=0.011$, $se=0.006$, $p=0.067$) than among individuals in their fifties ($\beta=0.017$, $se=0.005$, $p=0.0013$) and sixties ($\beta=0.030$, $se=0.005$, $p=1.28 \times 10^{-10}$). In addition, the frequency of the C

allele decreased slightly across these age groups (r=-0.006, $p<5 \times 10^{-5}$), at 0.314, 0.309, and 0.305 among individuals in their forties, fifties and sixties, respectively.

Five loci were significantly associated with VPA using Model 1. The strongest among these were variants in *CADM2*. Four of these were significantly associated with VPA in all three models tested: *CADM2*, *EXOC4*, *CTBP2*, and *DPY19L1*. The *FOXO3* locus was significant in Models 1 and 2, but not Model 3, and loci *NEGR1* and *MYOM3* were significant only in Model 3 (see Supplementary Tables 4 and 5, Supplementary Figures 2 and 3). The VPA-increasing G allele at the *NEGR1* SNP (rs3101340) is in LD (r^2=0.60) with a previously identified variant (rs3101336-C) associated with increased BMI [60]. Although different individual *CADM2* SNPs were identified across models and with MVPA, these SNPs were in strong to moderate LD, suggesting allelic heterogeneity at this locus (see Supplementary Figures 4 and 5). Among all five VPA-associated SNPs, only two had consistent trends of association with BMI in the direction expected based on the negative phenotypic correlation. Notably, the PA-increasing allele at the *CADM2* SNP was associated with higher BMI ($p=6.8 \times 10^{-7}$; see Table 2). In a sensitivity analysis in which all individuals who did not self-report at least 3 days/week at 25 minutes/day of VPA were included as controls (as opposed to only individuals reporting 0 days of VPA), results were similar but generally weaker despite a larger sample size, with a chip heritability of approximately one half of that for the analysis using more extreme controls (h^2=0.026 (0.002) vs. 0.054 (0.003); see Supplementary Figure 6).

Six loci were significantly associated with SSOE using Model 1(see Table 2 and Figure 1). *CADM2* was the most strongly associated locus. Four loci (*CADM2*, *AKAP10*, *CTC-436P18.1*, and *SIPA1L1*) were consistently significantly associated in both Models 1 and 3. Locus *GATAD2A* was associated with SSOE only in Model 1, and *HIST1H1D* was associated with SSOE only in Models 1 and 2 (see Supplementary Tables 4 and 5, Supplementary Figures 2 and 3). The C allele at a variant in *FTO* (rs55872725) was associated with lower odds of SSOE only upon adjustment for BMI (Model 3). This allele is in complete LD (r^2=1) with the T allele of the well-established variant (rs1558902) associated

with lower BMI. Correction for collider bias, shown in Supplementary Table 5, reduces the strength of this *FTO* SNP association with SSOE from $p=7.0 \times 10^{-13}$ to $p=3.0 \times 10^{-3}$ in the unbiased model, with the same direction of association. Among all five SSOE-associated SNPs, three showed consistent trends of association with BMI in the direction expected based on the negative phenotypic correlation. Notably, the PA-increasing allele at the *CADM2* SNP was associated with higher BMI ($p=1.2 \times 10^{-7}$; see Table 2).

Upon meta-analysis of the 9 top hits for MVPA with the results in ARIC, 8 were genome-wide significant ($p<5 \times 10^{-9}$), including the *APOE*, *EXOC4*, and *CADM2* variants (see Supplementary Table 6). The direction of effect was consistent across ARIC and the UK Biobank for all 9 loci. For both VPA and SSOE, we observed consistent directions of effect for 3 out of the 5 top loci (see Supplementary Table 7).

Accelerometer-based PA in UK Biobank

There were approximately 91,000 individuals with non-missing accelerometry data. Approximately 6,500 individuals were excluded because of insufficient wear-time. These excluded individuals were slightly more likely to be male (p=0.03), younger ($p<2 \times 10^{-16}$) and have a higher BMI ($p<2 \times 10^{-16}$). 'Chip heritability' estimates for the accelerometry-based measures were higher (14% for average acceleration, and 11% for fraction of accelerations >425 mg) than for self-report PA measures (Supplementary Table 3). Although Q-Q plots show some evidence of inflation (see Supplementary Figure 1), LD score regression intercepts (<1.008) suggest no significant systematic inflation of test statistics.

Using Model 1, two loci were found to be significantly associated with average acceleration and one locus with fraction of accelerations >425 mg (see Table 2 and Figure 1). Only the *CRHR1* locus remained genome-wide significant in Model 3. The *RIT2/SYT4* locus was only associated with average acceleration in Model 1, and the *PML* locus was only significant with fraction of accelerations >425 mg in Models 1 and 2 (see Table 2, Supplementary Tables 4 and 5, and Supplementary Figures 1 and 2).

In general, with the exception of the *FTO* locus as mentioned above, we observed minimal evidence of collider bias because of adjustment for BMI (see Supplementary Table 5).

Consistent loci across self-report and accelerometry

We found a total of seven loci associated ($p < 5 \times 10^{-5}$) with both self-report and accelerometry measures using Model 1 (see Supplementary Table 8). For MVPA and average acceleration, we identified four loci (*MEF2C*, *RCOR1*, *STOML1* and *CRHR1*). For VPA, SSOE, and fraction of acceleration >425 mg, we identified three loci (*CADM2*, *PML*, and *CCNE1*). However, among these, only *RCOR1*, *CRHR1*, and *CADM2* remained significant in Models 2 and 3 (see Supplementary Table 9).

Follow-up analyses

We found highly significant negative genetic correlations of both MVPA and VPA with intelligence (see Figure 2). We also found significant positive genetic correlations of MVPA and VPA with early-morning chronotype and psychiatric diseases, and negative correlations with body fat and waist circumference. In contrast to the genetic correlations with MVPA and VPA, we found a positive correlation of SSOE with years of schooling and intelligence. We also found positive genetic correlations with age at first birth and negative correlations with neuroticism, depressive symptoms, insomnia, body fat, and waist circumference (see Figure 2). Among the accelerometry-based measures, we found highly significant negative genetic correlations of PA with waist and hip circumference, body fat, obesity, BMI, and other cardiometabolic traits (see Figure 3). Genetic correlation results remained very similar with GWAS models including activity at work and TDI as covariates, except for generally attenuated correlations with intelligence in the model with all covariates except BMI (Model 2, see Figure 2). However, upon the addition of BMI as a covariate (Model 3), the direction of genetic correlation between

PA and obesity traits was reversed (see Figures 2 and 3). As we note below, caution may be warranted in interpreting results from these adjusted models, especially since we observed a reversal of direction of correlations with obesity-related traits upon BMI adjustment.

Gene-based tissue enrichment analysis using data from GTEx generally implicate the brain and pituitary gland as primary tissues through which the PA-associated loci may exert their effects (see Figure 4). Examination of more specific tissues reveals several different parts of the brain. The cerebellum and the frontal cortex appear most consistently implicated across the five PA phenotypes (see Supplementary Figures 7 and 8). Results remained similar when using Models 2 and 3. Gene-set analyses reveal several nervous system gene sets across the PA phenotypes, but the only significant gene set after correction for multiple testing was for enrichment of genes involved in the synapse, for SSOE (see Supplementary Table 10).

Look-up of top SNPs in the Oxford Brain Imaging server suggests associations with mental health, body composition, educational attainment, sleep and psychiatric traits, in addition to physical activity traits. The rs62253088-T PA-increasing allele in *CADM2* was also associated with decreased neuroticism, and decreased self-reported nervous and anxious feelings. The rs7804463-C allele (*EXOC4*) associated with less PA is also associated with higher self-reported time spent using computer, fewer mood swings, and greater daytime dozing. The rs55657917-G allele (*CRHR1*) associated with greater PA was also associated with greater neuroticism, lower pulmonary function, greater sense of hurt feelings, and fewer naps during the day (see Supplementary Figure 9). Gene expression analyses imply several different tissues including the brain, adrenal and thyroid gland, skeletal muscle and adipose tissue, among others (see Supplementary Tables 11 and 12). Genes that we identified have previously been implicated in a range of other traits and diseases, including behavioral, cardiometabolic, psychiatric, educational attainment, and pulmonary function traits (see Supplementary Table 11).

Discussion

Given the importance of PA for many dimensions of health, and its' reported heritability, we sought to identify genetic variants that are associated with engagement in habitual physical activity, while considering important covariates such as season, physical activity at work, socio-economic status, and BMI. In the UK Biobank, with a very large sample size and multiple measures of PA, we identified ten loci that were genome-wide significant for at least one of the PA measures and were consistently associated with the respective PA measure in both the basic (Model 1) and the fully adjusted model (Model 3). We also identified three loci that exhibit consistent associations across both self-report and accelerometry measures.

Although most of the identified loci were novel, the genes that they were in or in proximity to have previous links to various diseases and traits (see Supplementary Table 11). Among these, variants in *CADM2*, a gene which encodes cell adhesion molecule 2, and is primarily expressed in the brain, has been linked to BMI variation [60,61], risk-taking behavior and other personality and behavioral traits [62–65], as well as with information processing speed [66]. The previously identified BMI-associated variant (rs13078960) [60,61] is not in LD (r^2<0.07) with the PA–associated variants that we identified, except for the SSOE-increasing allele at rs62253088 being positively, but weakly, correlated with the BMI-increasing allele at rs13078960 (r^2=0.2). The previously identified G alleles at both rs13084531 [64] and rs57401290 [63] associated with risk taking are weakly to moderately correlated (r^2=0.52 and 0.23, respectively) with the SSOE-increasing allele that we identified at rs62253088 (see Supplementary Figure 5). It thus appears that this locus may be important for several personality, cognitive, and behavioral traits, and may potentially be involved in reward systems. We found that the association of *CADM2* variants with PA in Model 1 was unaffected by the inclusion of BMI as a covariate. Furthermore, the PA-increasing alleles at this locus are associated with higher BMI, in the opposite direction of the phenotypic correlation. Along these lines, but with slightly deviating results, a recent study in mice found that *Cadm2*-deficient mice exhibit increased locomotor activity along with reduced adiposity [67]. Finally, it is important to note that

this locus appeared to be more strongly associated with VPA and SSOE as compared to MVPA. It may thus be specifically implicated in the proclivity to engage in intentional high-intensity exercise and sport, as opposed to more general and/or lower intensity PA.

Interestingly, a well-established variant in *APOE* (part of *APOE* ε4 allele), strongly implicated in Alzheimer's disease [47,68], exhibited one of the strongest associations with PA, and remained significant upon meta-analysis. How the *APOE* risk allele is associated with greater PA is not clear. An exercise training study found that *APOE* ε4 carriers had a greater increase in aerobic capacity [69]. This increased responsivity to PA could reinforce engagement in PA or be related to other factors that influence the tendency to engage in PA. Although another potential explanation for our finding is that individuals with a known family history of dementia or Alzheimer's disease purposefully increase their levels of PA in the hope of reducing risk for developing the disease, our findings do not suggest that individuals with a first-degree family history of Alzheimer's disease or dementia engage in higher levels of PA. However, we could not rule out the possibility of selection bias. Since the association was markedly stronger among older participants and the frequency of the risk allele decreased slightly with age from 40 to 69 years, it may be that the older *APOE* risk allele carriers are particularly enriched for healthy lifestyles. It is important to note that an association between *APOE* and PA may lead to spurious gene-environment interactions [70], and thus further work is needed to confirm and clarify this observed association.

Among the other specific loci that we identified, we did not find any of the loci that have previously been linked to PA [15,16]. The pattern of tissue-specific expression of the identified genes (or nearby genes) varied quite widely, although we observe an overall enrichment of genes expressed in the brain and pituitary gland, and more specifically in the cerebellum and frontal cortex. The cerebellum is involved in the precise coordination of motor activity, and the frontal cortex is involved in decision making, personality expression, and executive function. We also observed an enrichment of genes involved in the nervous system, including in the synapse. Other than *CADM2* and *APOE*, the other identified genes have been previously associated with a wide variety of traits, including intelligence,

cognitive decline, blood cell traits, schizophrenia, and obesity among others. They are also expressed in a wide variety of tissues. We suspect that there are many potential paths leading to differences in PA. These could include response to exercise, personality, hormonal levels, body composition. Future research is needed to help elucidate the genetic underpinnings of these proximate mechanisms, and to provide insight into how each of the identified loci contribute to habitual PA behavior.

Previous studies have shown that BMI-associated genetic variants are also associated with PA [71,72]. Similarly, we found an overall shared genetic basis for PA (especially accelerometer-based measures) with several obesity-related traits (in the expected negative direction of association), suggesting that genetic risk for obesity coincides with genetic propensity for lower PA. There is likely a complex set of genetic, environmental, and phenotypic factors that connect PA and obesity across the lifespan, that involve many pleiotropic genetic factors. Although we identified previously identified BMI-associated genes (*FTO* and *NEGR1*) in Model 3, these results appear to be at least partly attributed to collider bias. Similarly, for all five PA traits, we observed that the direction of the genetic correlation between PA and obesity-related traits is reversed when BMI is included as a covariate, despite a strong negative phenotypic correlation between PA and BMI. In addition to the caution warranted by potential collider bias which occurs when one controls for a variable (i.e. BMI) that is caused by both another covariate (i.e. gene) and the outcome variable in the model (i.e. PA) [46,73], caution is also warranted in interpreting results of genetic associations in which heritable covariates are included in the association model [74]. On the other hand, however, adjustment for the covariates may help identify/confirm loci that may or may not be spuriously associated with PA because of confounding via correlated factors.

Our study is strengthened by the large sample size, the availability of both self-reported and objective accelerometer-based measures of PA, and the availability of a replication cohort from a different country. However, we note several limitations. Given the relatively small genetic effect sizes observed for these PA phenotypes, we were insufficiently powered to formally replicate associations in the much smaller sample size in ARIC. Our inability to firmly replicate these findings does detract from

our confidence in the generalizability of the UK Biobank results. It could be, for example, that the genetic architecture and implicated genes for habitual PA differ widely by country, as well as by age group (see below), and by PA measure. Additional and larger replication studies are thus needed to more robustly identify PA-associated loci. Furthermore, the self-report measures of PA used in ARIC differed from the one used in the UK Biobank. The ARIC measure focuses more explicitly on leisure-time PA and incorporates more detailed information about PA, such as the frequency of sweating and a comparison of PA frequency with others of the same age. Both self-reported and accelerometer-based measures of PA are subject to various biases. Since both the UK Biobank and ARIC cohorts are comprised of middle- to late-middle-aged adults, the extent to which these results generalize to other age groups is not known. For example, it has been shown that the heritability of PA changes with age, with a decreased heritability in older ages [75]. Thus our power to detect strong effects may have been compromised by the older age range in both cohorts that we examined. Furthermore, our results may not generalize to other ethnic/racial groups.

In conclusion, our study revealed several important new findings. Effect sizes were generally very small, given the very large sample size, the common variants identified, and the modest p-values. We identified over 20 variants, most of which were novel, and thus need further study. We identified a variant in *CADM2*, a gene previously been found to be associated with obesity, as well as several personality traits. We also identified a well-established major risk variant for Alzheimer's disease in *APOE*, which was associated with higher levels of PA, suggesting the need for follow up studies to help clarify the nature of this observed association and its implication for understanding gene-environment interactions related to PA. We found genetic correlation of PA with obesity [60,76], psychiatric [77,78], educational [79], chronotype [80], and other traits. Genetic correlations with obesity may indicate extensive pleiotropy involving genes associated with both PA and obesity. The identification of genetic factors that predispose to high or low levels of PA will lead to a better understanding of the biological mechanisms underlying these proclivities. It may also lead to the identification of individuals less likely to engage in

and/or adhere to PA, and consequently to the development of tailored behavioral strategies. Finally, the integration of genetic characteristics with lifestyle and environmental information may point to how lifestyle/environmental factors interact with genetic factors to influence levels of PA.

Acknowledgments

This research was conducted using the UK Biobank Resource under Application Number 15678. We thank the participants and organizers of the UK Biobank. We also thank the participants and organizers of the ARIC study. Data from ARIC was obtained from dbGaP through accession number phs000280.v2.p1. The authors would like to acknowledge support from the National Institute of Diabetes and Digestive and Kidney Diseases grant (K01DK095032), the National Institute on Aging (AG019610), the State of Arizona and Arizona Department of Health Services (ADHS), and the McKnight Brain Research Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Atherosclerosis Risk in Communities

The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C). Funding for GENEVA was provided by National Human Genome Research Institute grant U01HG004402 (E. Boerwinkle). The authors thank the staff and participants of the ARIC study for their important contributions.

LDHUB Acknowledgements

We gratefully acknowledge all the studies and databases that made GWAS summary data available: ADIPOGen (Adiponectin genetics consortium), C4D (Coronary Artery Disease Genetics Consortium), CARDIoGRAM (Coronary ARtery DIsease Genome wide Replication and Meta-analysis), CKDGen (Chronic Kidney Disease Genetics consortium), dbGAP (database of Genotypes and Phenotypes), DIAGRAM (DIAbetes Genetics Replication And Meta-analysis), ENIGMA (Enhancing Neuro Imaging Genetics through Meta Analysis), EAGLE (EArly Genetics & Lifecourse Epidemiology Eczema Consortium, excluding 23andMe), EGG (Early Growth Genetics Consortium), GABRIEL (A Multidisciplinary Study to Identify the Genetic and Environmental Causes of Asthma in the European Community), GCAN (Genetic Consortium for Anorexia Nervosa), GEFOS (GEnetic Factors for OSteoporosis Consortium), GIANT (Genetic Investigation of ANthropometric Traits), GIS (Genetics of Iron Status consortium), GLGC (Global Lipids Genetics Consortium), GPC (Genetics of Personality Consortium), GUGC (Global Urate and Gout consortium), HaemGen (haemotological and platelet traits genetics consortium), HRgene (Heart Rate consortium), IIBDGC (International Inflammatory Bowel Disease Genetics Consortium), ILCCO (International Lung Cancer Consortium), IMSGC (International Multiple Sclerosis Genetic Consortium), MAGIC (Meta-Analyses of Glucose and Insulin-related traits Consortium), MESA (Multi-Ethnic Study of Atherosclerosis), PGC (Psychiatric Genomics Consortium), Project MinE consortium, ReproGen (Reproductive Genetics Consortium), SSGAC (Social Science Genetics Association Consortium) and TAG (Tobacco and Genetics Consortium), TRICL (Transdisciplinary Research in Cancer of the Lung consortium), UK Biobank. We gratefully acknowledge the contributions of Alkes Price (the systemic lupus erythematosus GWAS and primary biliary cirrhosis GWAS) and Johannes Kettunen (lipids metabolites GWAS).

Supplementary information is available at IJO's website

552 **References**

553 1. Fiuza-Luces C, Garatachea N, Berger NA, Lucia A. Exercise is the real polypill. Physiol. 2013;

554 2. US Surgeon General. Physical Activity and Health: A Report of the Surgeon General. S/N 017-
555 023-00196-5. 1996.

556 3. Blair SN. Physical inactivity: the biggest public health problem of the 21st century. Br J Sports
557 Med [Internet]. 2009 Jan 9;43(1):1 LP-2.

558 4. Kaplan GA, Strawbridge WJ, Cohen RD, Hungerford LR. Natural history of leisure-time physical
559 activity and its correlates: associations with mortality from all causes and cardiovascular disease
560 over 28 years. Am J Epidemiol. 1996;

561 5. Bauman AE, Reis RS, Sallis JF, Wells JC, Loos RJF, Martin BW. Correlates of physical activity:
562 Why are some people physically active and others not? Lancet. 2012. p. 258–71.

563 6. Trost SG, Owen N, Bauman AE, Sallis JF, Brown W. Correlates of adults' participation in
564 physical activity: review and update. Med Sci Sport Exerc. 2002;

565 7. den Hoed M, Brage S, Zhao JH, Westgate K, Nessa A, Ekelund U, et al. Heritability of objectively
566 assessed daily physical activity and sedentary behavior. Am. J. Clin. Nutr. 2013. p. 1317–25.

567 8. Gielen M, Westerterp-Plantenga MS, Bouwman FG, Joosen AMCP, Vlietinck R, Derom C, et al.
568 Heritability and genetic etiology of habitual physical activity: a twin study with objective
569 measures. [Internet]. Genes Nutr. 2014. p. 415.

570 9. Stubbe JH, Boomsma DI, Vink JM, Cornes BK, Martin NG, Skytthe A, et al. Genetic influences
571 on exercise participation in 37.051 twin pairs from seven countries. PLoS One. 2006.

572 10. Joosen AMCP, Gielen M, Vlietinck R, Westerterp KR. Genetic analysis of physical activity in
573 twins. Am J Clin Nutr. United States; 2005 Dec;82(6):1253–9.

11. Pérusse L, Tremblay A, Leblanc C, Bouchard C. Genetic and environmental influences on level of habitual physical activity and exercise participation. Am J Epidemiol. 1989;

12. Lauderdale DS, Fabsitz R, Meyer JM, Sholinsky P, Ramakrishnan V, Goldberg J. Familial determinants of moderate and intense physical activity: a twin study. Med Sci Sports Exerc. United States; 1997 Aug;29(8):1062–8.

13. Kaprio J, Koskenvuo M, Sarna S. Cigarette smoking, use of alcohol, and leisure-time physical activity among same-sexed adult male twins. Prog Clin Biol Res. United States; 1981;69 Pt C:37–46.

14. Thompson PD, Tsongalis GJ, Ordovas JM, Seip RL, Bilbie C, Miles M, et al. Angiotensin-converting enzyme genotype and adherence to aerobic exercise training. Prev Cardiol. 2006;

15. Herring MP, Sailors MH, Bray MS. Genetic factors in exercise adoption, adherence and obesity. Obes Rev. England; 2014 Jan;15(1):29–39.

16. Wilkinson A V., Gabriel KP, Wang J, Bondy ML, Dong Q, Wu X, et al. Sensation-seeking genes and physical activity in youth. Genes, Brain Behav. 2013;

17. Caldwell Hooper AE, Bryan AD, Hagger MS. What keeps a body moving? The brain-derived neurotrophic factor val66met polymorphism and intrinsic motivation to exercise in humans. J Behav Med. United States; 2014 Dec;37(6):1180–92.

18. Lightfoot JT. Current understanding of the genetic basis for physical activity. J Nutr. United States; 2011 Mar;141(3):526–30.

19. De Moor MHM, Liu Y-J, Boomsma DI, Li J, Hamilton JJ, Hottenga J-J, et al. Genome-wide association study of exercise behavior in Dutch and American adults. Med Sci Sports Exerc. United States; 2009 Oct;41(10):1887–95.

20. Kim J, Min H, Oh S, Kim Y, Lee AH, Park T. Joint identification of genetic variants for physical activity in Korean population. Int J Mol Sci. 2014;15(7):12407–21.

21. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLoS Med. Public Library of Science; 2015;12(3).

22. The Aric Investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. Am J Epidemiol [Internet]. 1989;129(4):687–702.

23. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc [Internet]. 2003 Aug [cited 2013 Aug 6];35(8):1381–95.

24. Ipaq, Sci M, Med S, Hardcastle S, Blake N, Hagger MS, et al. Guidelines for Data Processing and Analysis of the International Physical Activity Questionnaire (IPAQ) - Short Form. J Am Diet Assoc. 2002;

25. Ekelund U, Sepp H, Brage S, Becker W, Jakes R, Hennings M, et al. Criterion-related validity of the last 7-day, short form of the International Physical Activity Questionnaire in Swedish adults. Public Health Nutr. 2006;

26. American Heart Association. American Heart Association Recommendations for Physical Activity in Adults [Internet]. 2016 [cited 2003 Aug 20]. Available from: http://www.heart.org/HEARTORG/HealthyLiving/PhysicalActivity/FitnessBasics/American-Heart-Association-Recommendations-for-Physical-Activity-in-Adults_UCM_307976_Article.jsp#.WqF2eOjwbIV

27. Doherty A, Jackson D, Hammerla N, Plötz T, Olivier P, Granat MH, et al. Large Scale Population Assessment of Physical Activity Using Wrist Worn Accelerometers: The UK Biobank Study.

619 Buchowski M, editor. PLoS One [Internet]. Public Library of Science; 2017 Feb 1 [cited 2017 Apr
620 17];12(2):e0169649.

621 28. Hildebrand M, Van Hees VT, Hansen BH, Ekelund U. Age group comparability of raw
622 accelerometer output from wrist-and hip-worn monitors. Med Sci Sports Exerc. 2014;

623 29. Baecke JA, Burema J, Frijters JE. A short questionnaire for the measurement of habitual physical
624 activity in epidemiological studies. Am J Clin Nutr. 1982;36(5):936–42.

625 30. Folsom AR, Arnett DK, Hutchinson RG, Liao F, Clegg LX, Cooper LS. Physical activity and
626 incidence of coronary heart disease in middle-aged women and men. Med Sci Sport Exerc.
627 1997;29(7):901–9.

628 31. Richardson MT, Ainsworth BE, Wu HC, Jacobs DRJ, Leon AS. Ability of the Atherosclerosis
629 Risk in Communities (ARIC)/Baecke Questionnaire to assess leisure-time physical activity. Int J
630 Epidemiol. England; 1995 Aug;24(4):685–93.

631 32. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. Genome-wide genetic data
632 on ~500,000 UK Biobank participants. bioRxiv [Internet]. 2017 Jul 20;

633 33. O'Connell J, Sharp K, Shrine N, Wain L, Hall I, Tobin M, et al. Haplotype estimation for biobank-
634 scale data sets. Nat Genet. 2016;

635 34. Consortium GP, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An
636 integrated map of genetic variation from 1,092 human genomes. Nature [Internet]. Nature
637 Publishing Group; 2012;491(7422):56–65.

638 35. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel
639 of 64,976 haplotypes for genotype imputation. Nat Genet. 2016;

640 36. UK10K Consortium. The UK10K project identifies rare variants in health and disease. Nature.

2015.

37. Galinsky KJ, Bhatia G, Loh PR, Georgiev S, Mukherjee S, Patterson NJ, et al. Fast Principal-Component Analysis Reveals Convergent Evolution of ADH1B in Europe and East Asia. Am J Hum Genet. 2016;

38. Howie B, Marchini J, Stephens M, Chakravarti A. Genotype Imputation with Thousands of Genomes. G3 GenesGenomesGenetics [Internet]. Genetics Society of America; 2011;1(6):457–70.

39. Townsend P. Deprivation. J Soc Policy [Internet]. 2009/01/01. Cambridge University Press; 1987;16(2):125–46.

40. Smith GD, Whitley E, Dorling D, Gunnell D. Area based measures of social and economic circumstances: cause specific mortality patterns depend on the choice of index. J Epidemiol Community Health. 2001;

41. Loh P-R, Tucker G, Bulik-Sullivan BK, Vilhjalmsson BJ, Finucane HK, Salem RM, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat Genet. United States; 2015 Mar;47(3):284–90.

42. Loh P-R, Kichaev G, Gazal S, Schoech AP, Price AL. Mixed model association for biobank-scale data sets. bioRxiv [Internet]. Cold Spring Harbor Laboratory; 2017;

43. Klarin D, Zhu QM, Emdin CA, Chaffin M, Horner S, McMillan BJ, et al. Genetic analysis in UK Biobank links insulin resistance and transendothelial migration pathways to coronary artery disease. Nat Genet. United States; 2017 Sep;49(9):1392–7.

44. Klarin D, Emdin CA, Natarajan P, Conrad MF, Kathiresan S. Genetic Analysis of Venous Thromboembolism in UK Biobank Identifies the ZFPM2 Locus and Implicates Obesity as a Causal Risk FactorCLINICAL PERSPECTIVE. Circ Cardiovasc Genet [Internet]. American Heart

663 Association, Inc.; 2017;10(2).

664 45. Fadista J, Manning AK, Florez JC, Groop L. The (in)famous GWAS P-value threshold revisited
665 and updated for low-frequency variants. Eur J Hum Genet. 2016;24:1202–5.

666 46. Yaghootkar H, Bancks MP, Jones SE, McDaid A, Beaumont R, Donnelly L, et al. Quantifying the
667 extent to which index event biases influence large genetic association studies. Hum Mol Genet.
668 England; 2017 Mar;26(5):1018–30.

669 47. Saunders AM, Strittmatter WJ, Schmechel D, St. George-Hyslop PH, Pericak-Vance MA, Joo SH,
670 et al. Association of apolipoprotein E allele ϵ4 with late-onset familial and sporadic alzheimer's
671 disease. Neurology. 1993;

672 48. Lahoz C, Schaefer EJ, Cupples LA, Wilson PW, Levy D, Osgood D, et al. Apolipoprotein E
673 genotype and cardiovascular disease in the Framingham Heart Study. Atherosclerosis. Ireland;
674 2001 Feb;154(3):529–37.

675 49. Eichner JE, Dunn ST, Perveen G, Thompson DM, Stewart KE, Stroehla BC. Apolipoprotein E
676 polymorphism and cardiovascular disease: a HuGE review. Am J Epidemiol. United States; 2002
677 Mar;155(6):487–95.

678 50. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. mediation: R Package for Causal Mediation
679 Analysis. J Stat Softw [Internet]. 2014;59(5):1–38.

680 51. Team RDC. R: A language and environment for statistical computing. 2011.

681 52. Chyou PH. A simple and robust way of concluding meta-analysis results using reported P values,
682 standardized effect sizes, or other statistics. Clin. Med. Res. 2012. p. 219–23.

683 53. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. FUMA: Functional mapping and
684 annotation of genetic associations. bioRxiv. 2017 Feb;

685 54. Consortium TGte. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene
686 regulation in humans. Sci [Internet]. 2015 May 8;348(6235):648–60.

687 55. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of
688 GWAS Data. Tang H, editor. PLOS Comput Biol [Internet]. Public Library of Science; 2015 Apr
689 17 [cited 2018 Mar 5];11(4):e1004219.

690 56. Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, et al. LD Hub: a
691 centralized database and web interface to perform LD score regression that maximizes the
692 potential of summary level GWAS data for SNP heritability and genetic correlation analysis.
693 Bioinformatics [Internet]. 2017;33(2):272.

694 57. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic
695 correlations across human diseases and traits. Nat Genet. 2015;47(11):1236–41.

696 58. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Consortium SWG of the PG, et al.
697 LD Score regression distinguishes confounding from polygenicity in genome-wide association
698 studies. Nat Genet [Internet]. Nature Publishing Group, a division of Macmillan Publishers
699 Limited. All Rights Reserved.; 2015 Mar;47(3):291–5.

700 59. Elliott L, Sharp K, Alfaro-Almagro F, Douaud G, Miller K, Marchini J, et al. The genetic basis of
701 human brain structure and function: 1,262 genome-wide associations found from 3,144 GWAS of
702 multimodal brain imaging phenotypes from 9,707 UK Biobank participants. bioRxiv [Internet].
703 Cold Spring Harbor Laboratory; 2017;

704 60. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass
705 index yield new insights for obesity biology. Nature [Internet]. Nature Publishing Group, a
706 division of Macmillan Publishers Limited. All Rights Reserved.; 2015 Feb 12;518(7538):197–206.

707 61. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association

analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet [Internet]. Metabolism Initiative and Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA; 2010 Nov;42(1546–1718 (Electronic)):937–48.

62. Boutwell B, Hinds D, Tielbeek J, Ong KK, Day FR, Perry JRB. Replication and characterization of CADM2 and MSRA genes on human behavior. Heliyon. England; 2017 Jul;3(7):e00349.

63. Day FR, Helgason H, Chasman DI, Rose LM, Loh P-R, Scott RA, et al. Physical and neurobehavioral determinants of reproductive onset and success. Nat Genet [Internet]. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2016 Jun;48(6):617–23.

64. Strawbridge RJ, Ward J, Cullen B, Tunbridge EM, Hartz S, Bierut L, et al. Genome-wide analysis of risk-taking behaviour and cross-disorder genetic correlations in 116,255 individuals from the UK Biobank cohort. bioRxiv [Internet]. 2017 Aug 16;

65. Clarke T-K, Adams MJ, Davies G, Howard DM, Hall LS, Padmanabhan S, et al. Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N=112 117). Mol Psychiatry. England; 2017 Oct;22(10):1376–84.

66. Ibrahim-Verbaas CA, Bressler J, Debette S, Schuur M, Smith A V, Bis JC, et al. GWAS for executive function and processing speed suggests involvement of the CADM2 gene [Internet]. Mol Psychiatry. Macmillan Publishers Limited; 2016. p. 189–97.

67. Yan X, Wang Z, Schmidt V, Gauert A, Willnow TE, Heinig M, et al. Cadm2 regulates body weight and energy homeostasis in mice. Mol Metab [Internet]. Elsevier GmbH; 2017;1–9.

68. Pericak-Vance MA, Bebout JL, Gaskell Jr. PC, Yamaoka LH, Hung WY, Alberts MJ, et al. Linkage studies in familial Alzheimer disease: evidence for chromosome 19 linkage. AmJHumGenet. 1991;

731 69. Thompson PD, Tsongalis GJ, Seip RL, Bilbie C, Miles M, Zoeller R, et al. Apolipoprotein e
732 genotype and changes in serum lipids and maximal oxygen uptake with exercise training.
733 Metabolism [Internet]. 2004 Feb [cited 2017 Aug 16];53(2):193–202.

734 70. Dudbridge F, Fletcher O. Gene-environment dependence creates spurious gene-environment
735 interaction. Am J Hum Genet. United States; 2014 Sep;95(3):301–7.

736 71. Lee H, Ash GI, Angelopoulos TJ, Gordon PM, Moyna NM, Visich PS, et al. Obesity-Related
737 Genetic Variants and their Associations with Physical Activity [Internet]. Sport. Med Open. 2015.
738 p. 34.

739 72. Richmond RC, Davey Smith G, Ness AR, den Hoed M, McMahon G, Timpson NJ. Assessing
740 Causality in the Association between Child Adiposity and Physical Activity Levels: A Mendelian
741 Randomization Analysis. PLoS Med. 2014.

742 73. Day FR, Loh P-R, Scott RA, Ong KK, Perry JR. A Robust Example of Collider Bias in a Genetic
743 Association Study. Am J Hum Genet [Internet]. Elsevier; 2016 Feb 4;98(2):392–3.

744 74. Aschard H, Vilhjalmsson BJ, Joshi AD, Price AL, Kraft P. Adjusting for heritable covariates can
745 bias effect estimates in genome-wide association studies. Am J Hum Genet. United States; 2015
746 Feb;96(2):329–39.

747 75. Vink JM, Boomsma DI, Medland SE, de Moor MHM, Stubbe JH, Cornes BK, et al. Variance
748 components models for physical activity with age as modifier: a comparative twin study in seven
749 countries. Twin Res Hum Genet. England; 2011 Feb;14(1):25–34.

750 76. Shungin D, Winkler T, Croteau-Chonka D, Ferreira T, Mägi R, Lakka T, et al. New genetic loci
751 link adipose and insulin biology to body fat distribution. Nature. 2015;518(7538):187–96.

752 77. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility

753 locus near ODZ4. Nat Genet. United States; 2011 Sep;43(10):977–83.

754 78. Biological insights from 108 schizophrenia-associated genetic loci. Nature. England; 2014
755 Jul;511(7510):421–7.

756 79. Sniekers S, Stringer S, Watanabe K, Jansen PR, Coleman JRI, Krapohl E, et al. Genome-wide
757 association meta-analysis of 78,308 individuals identifies new loci and genes influencing human
758 intelligence. Nat Genet. United States; 2017 Jul;49(7):1107–12.

759 80. Jones SE, Tyrrell J, Wood AR, Beaumont RN, Ruth KS, Tuke MA, et al. Genome-Wide
760 Association Analyses in 128,266 Individuals Identifies New Morningness and Sleep Duration
761 Loci. PLoS Genet. United States; 2016 Aug;12(8):e1006125.

762

763

Figure Legends

Figure 1: Manhattan plot of GWAS for self-reported MVPA and VPA, strenuous sports or other exercises (abbreviated as SS or Other Exer.), and for accelerometer-based average accelerations and fraction of accelerations > 425 mg. Negative log10-transformed p-value for each SNP is plotted by chromosome and position (x-axis). The red horizontal line represents the threshold for genome-wide significant associations ($p<5 \times 10^{-9}$).

Figure 2: Genetic correlation of self-reported PA variables with other traits and diseases across the three statistical models employed. Traits/diseases shown are those that are in the top 10 of genetically correlated traits/diseases (according to p-value) for at least one of the 3 models. Traits/diseases are ordered from top to bottom in order of increasing p-value for Model 1. Horizontal position of bars corresponds to the genetic correlation (r_g) between PA and the respective trait/disease. Error bars represent 95% confidence intervals for r_g estimates. Bright green bars represent traits that showed a correlation with p-value $<2.5 \times 10^{-4}$, and light green bars represent traits with genetic correlation $p<0.05$. We excluded highly redundant traits (e.g. obesity, overweight) after leaving higher ranked one in.

Figure 3: Genetic correlation of accelerometry-based PA variables with other traits and diseases across the three statistical models employed. Traits/diseases shown are those that are in the top 10 of genetically correlated traits/diseases (according to p-value) for at least one of the 3 models. Traits/diseases are ordered from top to bottom in order of increasing p-value for Model 1. Horizontal position of bars corresponds to the genetic correlation (r_g) between PA and the respective trait/disease. Error bars represent 95% confidence intervals for r_g estimates. Bright green bars represent traits that showed a correlation with p-value $<2.5 \times 10^{-4}$, and light green bars represent traits with genetic correlation $p<0.05$. We excluded highly redundant traits (e.g. obesity, overweight) after leaving higher ranked one in.

788

789 **Figure 4:** Results of gene-based enrichment analysis for 30 general tissue types for PA-associated loci.

790 Dashed line represents the Bonferroni-corrected significance threshold.

Supplementary Table 1: Association of covariates with five measures of physical activity. Other covariates included in the model (PCs, center, and genotyping chip) are not shown here. The beta coefficient units for MVPA are standard deviations of MVPA (MET-minutes/week), after inverse normalization. Beta coefficients for VPA and Stren. Sports or other Exerc. are log-OR. The beta coefficient unit for acceleration average is milli-gravities. The beta coefficient for "Fraction Accel. > 425 mg" is fraction of time.

	MVPA		VPA		Stren. Sports or other Exerc.		Accel. Average		Fraction Accel. > 425 mg	
	Beta	p-value	Beta	p-value	Beta	p-value	Beta	p-value	Beta	p-value
Age	0.012	<2E-16	-0.025	<2E-16	-0.04	<2E-16	-0.25	<2E-16	-0.035	<2E-16
Sex	0.055	<2E-16	0.47	<2E-16	0.21	<2E-16	-0.28	1.68E-8	0.30	<2E-16
BMI	-0.028	<2E-16	-0.068	<2E-16	-0.066	<2E-16	-0.47	<2E-16	-0.047	<2E-16
Season	0.047	<2E-16	0.078	<2E-16	0.066	<2E-16	0.67	<2E-16	0.033	<2E-16
Work - Walk/Stand	7.69E-03	0.0002	-0.048	<2E-16	-0.058	<2E-16	0.49	<2E-16	0.031	4.80E-15
Work - Physical activity	0.400	<2E-16	0.499	<2E-16	-0.17	<2E-16	0.98	<2E-16	0.031	2.74E-07
TDI	-5.35E-03	<2E-16	-0.046	<2E-16	-0.086	<2E-16	-0.12	<2E-16	-0.019	<2E-16

BMI: body mass index, TDI: Townsend deprivation index

Supplementary Table 2: Phenotypic correlation (Pearson's product-moment correlation/Spearman's rho are shown for continuous variable correlations, phi coefficient is shown for correlation between two dichotomous variables, and point-biserial correlations are shown for correlations between a continuous and a dichotomous variable) among the five different PA measures. All correlations were highly significant.

	MVPA	VPA	Stren. Sports or Other Ex.	Accel. Average	Fract. Accel.>425mg
MVPA		0.65	0.28	0.19/0.22	0.15/0.20
VPA			0.73	0.29	0.32
Stren. Sports or Other Ex.				0.23	0.27
Accel. Average					0.68/0.71

Supplementary Table 3: 'Chip Heritability' and inflation estimates from LD score regression analysis. Ratio is (intercept-1)/(mean(χ^2)-1), which measures the proportion of the inflation in the mean χ^2 that the LD Score regression intercept ascribes to causes other than polygenic heritability.

	h^2	Lambda GC	Mean χ^2	Ratio (s.d.)	Intercept
Self-reported measures					
MVPA	0.046 (0.0023)	1.31	1.35	0.040 (0.009)	1.014
VPA	0.054 (0.0029)	1.25	1.30	0.099 (0.030)	1.030
Stren. Sports or Other Exerc.	0.056 (0.0025)	1.31	1.41	0.058 (0.022)	1.024
Accelerometry					
Average Acceleration	0.143 (0.0083)	1.20	1.26	0.030 (0.032)	1.008
Fraction Accelerations > 425 mg	0.110 (0.0067)	1.15	1.19	0.010 (0.037)	1.002

h^2: heritability; GC: genomic control.

Supplementary Table 4: Summary of polymorphisms identified in the UK Biobank using a model additionally adjusted for Townsend Deprivation index, walk or standing at work, and physical activity at work (Model 2). The beta coefficient units for MVPA are standard deviations of MVPA (MET-minutes/week), after inverse normalization. The beta coefficient unit for acceleration average is milli-gravities. The beta coefficient unit for "Fraction Accel. > 425 mg" is fraction of time. The beta coefficient units for BMI correspond to residualized BMI after inverse normalization.

									BMI association	
SNP ID	**Chr.**	**Gene/Nearest Gene**	**Position**	**EA**	**EAF**	**Beta/OR**	**p-value**	**n**	**Beta**	**p-value**
MVPA										
rs429358	19	*APOE*	45,411,941	T	0.85	-0.022	2.4E-14	376,789	2.5E-02	2.4E-21
rs169504	6	*PBX2*	32,153,406	C	0.81	0.018	4.9E-11	376,789	-1.1E-02	1.8E-05
rs4129572	7	*EXOC4*	133,636,888	T	0.40	-0.14	2.9E-10	376,789	3.4E-03	8.0E-02
rs3094622	6	*RPP21*	30,327,952	A	0.86	0.019	3.1E-10	376,789	-9.9E-03	3.8E-04
rs181220614	3	*ARHGEF26-AS1*	153,806,914	C	0.99	-0.07	8.8E-10	376,789	-1.2E-02	2.6E-01
rs149943	6	*ZNF165*	28,002,388	G	0.85	0.018	9.9E-10	376,789	-1.2E-02	1.4E-05
rs2988004	9	*PAX5*	37,044,388	T	0.55	-0.013	2.8E-09	376,789	-1.1E-02	5.2E-08
Vigorous PA: ≥ 3 vs. 0 days/week										
rs1248860	3	*CADM2*	85,015,779	G	0.48	0.96	1.1E-13	260,738	-9.5E-03	6.8E-07
rs2764261	6	*FOXO3*	108,927,842	A	0.37	1.04	9.3E-11	260,738	-1.8E-02	8.7E-20
rs3781411	10	*CTBP2*	126,715,436	C	0.88	1.06	2.3E-10	260,738	1.7E-03	5.6E-01
rs12707131	7	*EXOC4*	133,620,982	A	0.61	1.04	4.6E-10	260,738	-3.9E-03	4.7E-02
rs328919	7	*DPY19L1*	35,033,983	A	0.69	0.96	9.3E-10	260,738	-6.2E-03	2.7E-03
Strenuous sports or other exercises: ≥ 2-3 vs. 0 days/week										
rs62253088	3	*CADM2*	85,400,801	T	0.33	1.05	1.5E-20	350,492	1.1E-02	1.2E-07
rs166840	17	*AKAP10*	19,799,698	G	0.59	1.03	3.0E-10	350,492	5.4E-03	5.8E-03
rs10946808	6	*HIST1H1D*	26,233,387	A	0.73	0.96	5.5E-10	350,492	1.2E-02	3.6E-08
rs75930676	14	*SIPA1L1*	71,826,547	T	0.95	0.93	8.5E-10	350,492	2.1E-03	6.4E-01
rs4865656	5	*LOC642366*	50,659,788	G	0.62	1.03	3.0E-09	350,492	-9.0E-03	4.6E-06
Accelerometry – Average acceleration										
rs55657917	17	*CRHR1*	43,844,560	T	0.78	-0.31	1.1E-12	90,986	-3.7E-03	1.1E-01
rs185829646	10	*ANKRD22*	90,583,705	A	0.998	-2.7	4.8E-09	90,986	-2.6E-02	3.0E-01
Accelerometry – Fraction accelerations > 425 milli-gravities										
rs743580	15	*PML*	74,328,116	A	0.51	0.024	2.5E-09	90,569	-1.2E-02	1.4E-09
rs6433478	2	*CIR1*	175,241,482	T	0.46	-0.024	4.9E-09	90,569	9.9E-03	2.8E-07

EA refers to effect allele that Beta/OR corresponds to. EAF: effect allele frequency

Supplementary Table 5: Summary of polymorphisms identified in the UK Biobank using a model additionally adjusted for Townsend Deprivation index, walk or standing at work, physical activity at work, and BMI. The beta coefficient units for MVPA are standard deviations of MVPA (MET-minutes/week), after inverse normalization. The beta coefficient unit for acceleration average is milli-gravities. The beta coefficient for "Fraction Accel. > 425 mg" is fraction of time. The beta coefficient units for BMI correspond to residualized BMI after inverse normalization. The collider bias quantification columns refer to estimates for the respective SNP, with (corrected) and without correction (uncorrected) for collider bias resulting from adjustment for BMI. The uncorrected and corrected estimates are from a linear model (for all PA measures) without incorporating the other SNP effects as is implemented in the GWAS analysis by the BOLT-LMM software.

											Collider Bias Quantification			
									BMI association		Uncorrected		Corrected	
SNP ID	Chr.	Gene/Nearest Gene	Position	EA	EAF	Beta/OR	p	n	Beta	p	Beta	p	Beta	p
MVPA														
rs2988004	9	*PAX5*	37,044,388	T	0.56	-0.014	2.4E-11	375,862	-1.1E-02	5.2E-08	-0.013	1.1E-09	-0.012	2.6E-08
rs429358	19	APOE	45,411,941	T	0.85	-0.019	7.3E-11	375,862	2.5E-02	2.4E-21	-0.019	1.2E-10	-0.022	7.1E-14
rs181220614	3	*ARHGEF26-AS1*	153,806,914	C	0.99	-0.072	2.3E-10	375,862	-1.2E-02	2.6E-01	-0.076	7.9E-09	-0.073	3.8E-08
rs7804463	7	*EXOC4*	133,447,651	T	0.53	0.013	4.1E-10	375,862	-2.9E-04	8.8E-01	0.013	7.6E-10	0.013	1.2E-09
rs4576826	11	*C11orf80*	66,560,982	T	0.74	0.015	5.5E-10	375,862	6.3E-03	3.7E-03	0.015	2.3E-10	0.014	4.0E-09
rs13081745	3	*CADM2*	85,063,983	G	0.33	-0.014	9.6E-10	375,862	-1.5E-02	1.0E-12	-0.014	8.9E-10	-0.012	4.3E-08
rs11913445	22	*CCDC188*	20,142,513	C	0.83	0.017	2.1E-09	375,862	5.4E-03	3.6E-02	0.017	5.5E-09	0.016	6.1E-08
rs169504	6	*PBX2*	32,153,406	C	0.81	0.017	1.4E-09	375,862	-1.1E-02	1.8E-05	0.016	2.2E-09	0.018	4.6E-11
Vigorous PA: ≥ 3 vs. 0 days/week														
rs653481	3	*CADM2*	85,005,685	A	0.33	0.95	1.9E-15	259,991	-1.4E-02	1.5E-12	-0.011	9.3E-16	-0.010	8.3E-14
rs3781411	10	*CTBP2*	126,715,436	C	0.88	1.06	1.0E-10	259,991	1.7E-03	5.6E-01	0.013	5.0E-11	0.013	5.6E-11
rs328902	7	*DPY19L1*	35,020,843	C	0.69	0.96	1.3E-10	259,991	-6.2E-03	2.7E-03	-0.009	8.7E-11	-0.009	4.2E-10
rs3101340	1	*NEGR1*	72,744,144	G	0.49	1.04	1.1E-09	259,991	1.7E-02	8.0E-20	0.008	1.7E-09	0.007	1.6E-07
rs12707131	7	*EXOC4*	133,620,982	A	0.61	1.04	2.0E-09	259,991	-3.9E-03	4.7E-02	0.008	1.9E-09	0.008	3.3E-10
rs6689056	1	*MYOM3*	24,364,531	G	0.67	1.04	4.9E-09	259,991	2.0E-03	3.3E-01	0.008	9.8E-09	0.008	8.2E-09
Strenuous sports or other exercises: ≥ 2-3 vs. 0 days/week														
rs1691471	3	*CADM2*	85,011,013	C	0.62	0.95	6.5E-23	349,082	-1.3E-02	8.0E-12	-0.013	1.7E-22	-0.011	7.0E-20
rs55872725	16	*FTO*	53,809,123	C	0.6	0.96	3.0E-13	349,082	-7.4E-02	3.4E-312	-0.008	7.0E-13	-0.003	3.0E-03
rs166840	17	AKAP10	19,799,698	G	0.59	1.03	7.3E-11	349,082	5.4E-03	5.8E-03	0.007	2.0E-10	0.007	8.5E-10
rs75930676	14	*SIPA1L1*	71,826,547	T	0.95	0.93	8.5E-10	349,082	2.1E-03	6.4E-01	-0.016	1.3E-10	-0.016	1.2E-09
rs159544	5	*CTC-436P18.1*	60,489,247	A	0.61	0.97	2.7E-09	349,082	-5.5E-03	5.3E-03	-0.007	2.4E-09	-0.007	7.1E-09
Accelerometry – Average acceleration														
rs55915917	17	*CRHR1*	43,892,784	T	0.78	-0.28	1.5E-11	90,845	-4.0E-03	8.4E-02	-0.28	5.0E-11	-0.31	6.7E-13
rs62398709	5	*LINC01470*	152,055,235	A	0.65	0.24	7.1E-11	90,845	8.9E-03	1.1E-05	0.23	3.0E-10	0.21	1.6E-08
rs185829646	10	*ANKRD22*	90,583,705	A	1	-2.69	1.5E-09	90,845	-2.6E-02	3.0E-01	-2.94	2.0E-09	2.97	1.4E-09
rs17551090	5	*MEF2C*	88,037,209	T	0.89	0.34	3.0E-09	90,845	1.3E-02	2.0E-05	0.34	3.2E-09	0.32	2.6E-08
Accelerometry – Fraction accelerations > 425 milli-gravities														
No significant associations														

EA refers to effect allele that Beta/OR corresponds to. EAF: effect allele frequency

Supplementary Table 6: Meta-analysis of UK Biobank MVPA top hits with ARIC PA. Beta coefficient units are standard deviations of MVPA MET-minutes/week, after inverse normalization.

					ARIC (n=8,556)		Meta-analysis (n=385,790)	
SNP	**Chr.**	**Gene/Nearest Gene**	**Position**	**EA**	**Beta**	**p**	**Beta**	**p**
rs429358	19	*APOE*	45,411,941	T	-0.035	0.22	-0.002	2.80E-13
rs7804463	7	*EXOC4*	133,447,651	T	0.003	0.87	0.015	1.30E-11
rs2854277	6	*HLA-DQB1*	32,628,084	C	0.009	0.72	0.031	5.00E-10
rs7791992	7	*C7orf72/SPATA48*	50,237,784	C	-0.005	0.73	-0.014	2.20E-10
rs3094622	6	*RPP21*	30,327,952	A	0.035*	0.17	0.020	1.84E-10
rs149943	6	*ZNF165*	28,002,388	G	0.031	0.20	0.019	4.25E-10
rs2035562	3	*CADM2*	85,056,521	A	-0.014	0.38	-0.014	3.68E-10
rs2988004	9	*PAX5*	37,044,388	T	-0.033	0.03	-0.013	1.46E-09
rs1043595	7	*CALU*	128,410,012	G	0.015	0.37	0.014	7.27E-09

*rs3094034-A allele used as a proxy for rs3094622-A

Supplementary Table 7: Association with PA in ARIC of SNPs identified in Model 1 of UK Biobank for PA phenotypes other than MVPA. Beta coefficients for ARIC correspond to standardized residualized units of the sport/exercise index. Meta-analysis using only p-values is only reported for SNPs that have consistent directions of effect in UK Biobank and ARIC.

						UK Biobank		ARIC (n=8,556)		Meta-analysis
SNP	**Chr.**	**Gene/Nearest Gene**	**Position**	**EA**	**EAF**	**Beta/OR**	**p**	**Beta**	**p**	**p**
Vigorous PA: ≥ 3 vs. 0 days/week										
rs1248860	3	*CADM2*	85,015,779	G	0.48	0.96	1.10E-13	0.008	0.614	NA
rs2764261	6	*FOXO3*	108,927,842	A	0.37	1.04	2.00E-11	0.010	0.500	0.16
rs3781411	10	*CTBP2*	126,715,436	C	0.88	1.06	3.00E-10	0.029	0.193	0.09
rs12707131	7	*EXOC4*	133,620,982	A	0.61	1.04	9.00E-11	-0.002	0.917	NA
rs328919	7	*DPY19L1*	35,033,983	A	0.69	0.96	5.50E-10	-0.011	0.503	0.16
Strenuous sports or other exercises: ≥ 2-3 vs. 0 days/week										
rs62253088	3	*CADM2*	85,400,801	T	0.33	1.05	1.00E-19	-0.007	0.660	NA
rs166840	17	*AKAP10*	19,799,698	G	0.59	1.03	3.10E-11	-0.011	0.500	NA
rs10946808	6	*HIST1H1D*	26,233,387	A	0.73	0.97	9.90E-10	-0.006	0.727	0.21
rs159544	5	*CTC-436P18.1*	60,489,247	A	0.61	0.97	1.30E-09	0.010	0.536	NA
rs75930676	14	*SIPA1L1*	71,826,547	T	0.95	0.93	2.00E-09	-0.077	0.059	0.04
rs111901094	19	*GATAD2A*	19,513,570	G	0.82	1.04	3.00E-09	0.001	0.972	0.31
Accelerometry – Average acceleration										
rs55657917	17	*CRHR1*	43,844,560	T	0.78	-0.3	5.00E-12	0.020	0.296	NA
rs59499656	18	*RIT2/SYT4*	40,768,309	A	0.66	-0.23	2.40E-09	0.015	0.343	NA
Accelerometry – Fraction accelerations > 425 milli-gravities										
rs743580	15	*PML*	74,328,116	A	0.51	0.025	1.30E-09	0.019	0.208	0.09

NA indicates not applicable since direction of effect is not consistent.

Supplementary Table 8: Loci consistently associated with PA across self-report and accelerometry measures (each $p<5 \times 10^{-5}$ and in consistent direction) using Model 1. The beta coefficient units for MVPA are standard deviations of MVPA (MET-minutes/week), after inverse normalization. The beta coefficient unit for acceleration average is milli-gravities. The beta coefficient for "Fraction Accel. > 425 mg" is fraction of time.

					MVPA		VPA		SSOE		AA		AF>425	
SNP	Chr.	Nearest Gene	Position	Effect Allele	Beta	p	OR	p	OR	p	Beta	p	Beta	p
MVPA and AA														
rs447801	5	*MEF2C*	88002653	T	-0.010	7.5E-06	0.98	5.0E-01	1.02	1.2E-01	-0.164	8.5E-06	-0.014	7.2E-04
rs8013957*	14	*RCOR1*	103140254	C	0.010	1.4E-05	1.03	2.0E-04	1.02	2.4E-04	0.155	4.8E-05	0.015	5.6E-04
rs7174985	15	*STOML1*	74274948	T	0.011	2.5E-06	1.03	7.0E-06	1.02	1.0E-05	0.175	6.4E-06	0.021	1.0E-06
rs55915917*	17	*CRHR1*	43892784	T	-0.012	8.7E-06	0.98	1.2E-02	0.98	1.7E-04	-0.303	5.5E-12	-0.025	7.9E-07
VPA, SSOE, and AF>425														
rs1248860*	3	*CADM2*	85015779	G	-0.011	2.0E-07	0.96	1.1E-13	0.96	7.8E-17	-0.113	1.8E-03	-0.018	8.9E-06
rs5742915	15	*PML*	74336633	T	-0.011	6.2E-07	0.96	5.2E-07	0.97	3.4E-05	-0.191	1.7E-07	-0.024	2.9E-09
rs17599450	19	*CCNE1*	30328753	C	0.008	1.0E-03	1.03	1.5E-05	1.02	1.4E-06	0.185	1.9E-06	0.019	1.5E-05

SSOE: Strenuous sports or other exercises; AA: Average acceleration, AF>425: acceleration fraction greater than 425mg.
* indicates SNPs or loci that remain significant in Model 3.

Supplementary Table 9: Loci consistently associated with PA across self-report and accelerometry measures (each $p<5 \times 10^{-5}$ and in consistent direction) using Models 2 and 3. The beta coefficient units for MVPA are standard deviations of MVPA (MET-minutes/week), after inverse normalization. The beta coefficient unit for acceleration average is milli-gravities. The beta coefficient for "Fraction Accel. > 425 mg" is fraction of time.

					MVPA		VPA		SSOE		AA		AF>425	
SNP	Chr.	Nearest Gene	Position	Effect Allele	Beta	p	OR	p	OR	p	Beta	p	Beta	p
Model 2														
MVPA and average acceleration														
rs55915917	17	*CRHR1*	43892784	T	-0.011	3.8E-05	0.97	8.9E-03	0.97	1.8E-05	-0.310	1.2E-12	-0.025	4.2E-07
rs5742915	15	*PML*	74336633	T	-0.010	2.5E-06	0.96	3.5E-06	0.97	6.1E-05	-0.187	2.2E-07	-0.024	4.4E-09
rs580241	11	*TMEM151A*	66066349	G	-0.013	3.9E-07	0.96	9.3E-05	0.98	9.3E-03	-0.189	1.3E-05	-0.014	5.4E-03
rs8013957	14	*RCOR1*	103140254	C	0.010	1.6E-05	1.02	2.9E-04	1.02	1.4E-04	0.162	1.8E-05	0.015	4.7E-04
VPA, SSOE, and fraction acceleration > 425 mg														
rs1248860	3	*CADM2*	85015779	G	-0.010	1.8E-06	0.96	1.1E-13	0.96	1.8E-17	-0.112	1.9E-03	-0.018	7.1E-06
rs7174985	15	*STOML1*	74274948	T	0.011	1.5E-06	1.02	2.3E-05	1.02	1.7E-05	0.172	7.0E-06	0.021	1.6E-06
rs17599450	19	*CCNE1*	30328753	C	0.008	7.4E-04	1.03	2.8E-05	1.03	1.8E-06	0.179	3.4E-06	0.019	2.4E-05
Model 3														
MVPA and average acceleration														
rs7559547	2	*TMEM18*	615627	C	-0.014	5.2E-07	0.96	9.5E-06	0.97	6.0E-07	-0.232	3.6E-07	-0.021	6.3E-05
rs580241	11	*TMEM151A*	66066349	G	-0.013	3.9E-07	0.97	9.8E-05	0.98	1.0E-02	-0.176	2.4E-05	-0.012	1.0E-02
rs8013957	14	*RCOR1*	103140254	C	0.010	1.3E-05	1.02	2.4E-04	1.02	9.7E-05	0.150	3.9E-05	0.014	9.0E-04
rs8044769	16	*FTO*	53839135	T	-0.009	6.6E-06	0.97	1.3E-06	0.96	1.1E-08	-0.187	4.5E-08	-0.015	1.8E-04
rs55915917	17	*CRHR1*	43892784	T	-0.011	1.1E-05	0.97	4.0E-03	0.97	6.1E-06	-0.284	1.5E-11	-0.022	3.2E-06
VPA, SSOE, and fraction acceleration > 425 mg														
rs7567570	2	*TMEM18*	615140	T	-0.014	1.0E-06	0.98	1.4E-05	0.97	1.0E-06	-0.236	2.5E-07	-0.021	4.2E-05
rs4856584	3	*CADM2*	85580328	A	0.012	3.1E-08	1.05	1.4E-14	1.06	6.0E-21	0.110	2.4E-03	0.021	5.3E-07

Supplementary Table 10: Top ten significant gene sets from MAGMA-based gene-set analysis using Model 1 for each of 5 PA phenotypes.

Gene Set	N genes	Beta	Beta STD	SE	P	P_{bon}
MVPA						
GO_bp:go_positive_regulation_of_histone_methylation	32	0.694	0.029	0.176	4.10E-05	0.4461
GO_cc:go_mhc_class_ii_protein_complex	14	1.33	0.0369	0.339	4.36E-05	0.4753
GO_mf:go_mhc_class_ii_receptor_activity	10	1.43	0.0335	0.368	4.99E-05	0.5430
GO_bp:go_neuron_migration	105	0.362	0.0274	0.1	1.42E-04	1
GO_bp:go_adaptive_immune_response_based_on_somatic_recombination_of_immune_receptors_built_from_immunoglobulin_superfamily_domains	113	0.326	0.0255	0.09	1.47E-04	1
GO_mf:go_pdz_domain_binding	88	0.377	0.0261	0.106	1.81E-04	1
GO_bp:go_cell_morphogenesis_involved_in_neuron_differentiation	354	0.185	0.0254	0.052	1.99E-04	1
GO_mf:go_nucleic_acid_binding_transcription_factor_activity	1134	0.107	0.0259	0.03	2.09E-04	1
GO_bp:go_neuron_projection_guidance	198	0.246	0.0255	0.07	2.10E-04	1
GO_cc:go_voltage_gated_calcium_channel_complex	39	0.569	0.0263	0.161	2.10E-04	1
VPA						
GO_bp:go_dendrite_development	72	0.473	0.0296	0.115	2.13E-05	0.2314
GO_bp:go_neurogenesis	1355	0.107	0.0281	0.027	2.92E-05	0.3182
GO_bp:go_neuron_projection_development	522	0.157	0.0262	0.042	7.83E-05	0.8523
GO_bp:go_dendrite_morphogenesis	39	0.601	0.0277	0.159	7.99E-05	0.8701
GO_bp:go_regulation_of_nervous_system_development	722	0.127	0.0248	0.036	1.76E-04	1
GO_bp:go_cell_morphogenesis_involved_in_neuron_differentiation	354	0.178	0.0245	0.05	1.91E-04	1
GO_bp:go_negative_regulation_of_neuron_differentiation	183	0.232	0.0231	0.066	2.07E-04	1
GO_bp:go_neuron_projection_morphogenesis	386	0.169	0.0243	0.048	2.11E-04	1
GO_bp:go_negative_regulation_of_nitric_oxide_metabolic_process	10	1.11	0.026	0.323	2.92E-04	1
GO_bp:go_negative_regulation_of_nitric_oxide_biosynthetic_process	10	1.11	0.026	0.323	2.92E-04	1
SSOE						
GO_cc:go_synapse	718	0.188	0.0364	0.039	5.90E-07	0.0064
GO_cc:go_synapse_part	580	0.181	0.0318	0.043	1.13E-05	0.1227
Curated_gene_sets:myllykangas_amplification_hot_spot_3	4	1.84	0.0272	0.504	1.30E-04	1
GO_bp:go_t_cell_mediated_immunity	26	0.675	0.0255	0.186	1.45E-04	1
GO_cc:go_postsynapse	355	0.195	0.027	0.054	1.60E-04	1
Curated_gene_sets:cheok_response_to_mercaptopurine_and_ld_mtx_up	8	1.26	0.0263	0.349	1.63E-04	1
GO_bp:go_presynaptic_process_involved_in_synaptic_transmission	109	0.335	0.0258	0.096	2.51E-04	1
GO_cc:go_protein_phosphatase_type_2a_complex	19	0.821	0.0264	0.236	2.55E-04	1
Curated_gene_sets:dacosta_uv_response_via_ercc3_dn	834	0.128	0.0268	0.037	2.77E-04	1
GO_bp:go_glutamate_receptor_signaling_pathway	40	0.527	0.0246	0.153	2.84E-04	1
AA						
GO_bp:go_regulation_of_transporter_activity	191	0.303	0.0308	0.071	9.85E-06	0.1073
GO_cc:go_synapse	718	0.15	0.0292	0.036	1.31E-05	0.1426
Curated_gene_sets:basso_hairy_cell_leukemia_dn	77	0.421	0.0273	0.11	6.46E-05	0.7031
GO_bp:go_sterol_catabolic_process	11	1.15	0.0283	0.303	6.76E-05	0.7356
GO_bp:go_cholesterol_catabolic_process	11	1.15	0.0283	0.303	6.76E-05	0.7356
GO_cc:go_synapse_part	580	0.15	0.0262	0.04	8.13E-05	0.8847
Curated_gene_sets:schaeffer_sox9_targets_in_prostate_development_dn	44	0.545	0.0267	0.145	8.26E-05	0.8986
GO_mf:go_transition_metal_ion_binding	1313	0.097	0.025	0.026	1.15E-04	1
Curated_gene_sets:sesto_response_to_uv_c7	66	0.413	0.0247	0.113	1.39E-04	1
GO_bp:go_phenol_containing_compound_biosynthetic_process	31	0.596	0.0245	0.166	1.63E-04	1
AF						
GO_mf:go_transcriptional_activator_activity_rna_polymerase_ii_core_promoter_proximal_region_sequence_specific_binding	218	0.258	0.028	0.065	3.13E-05	0.3406
Curated_gene_sets:nikolsky_breast_cancer_20q11_amplicon	31	1.31	0.0541	0.331	3.68E-05	0.4009
GO_bp:go_cell_fate_determination	43	0.532	0.0258	0.137	5.47E-05	0.5961
GO_mf:go_transcription_factor_activity_rna_polymerase_ii_core_promoter_proximal_region_sequence_specific_binding	320	0.197	0.0258	0.052	7.45E-05	0.8110
Curated_gene_sets:cui_tcf21_targets_2_dn	789	0.125	0.0253	0.034	1.14E-04	1
GO_bp:go_homophilic_cell_adhesion_via_plasma_membrane_adhesion_molecules	145	0.342	0.0303	0.093	1.24E-04	1
GO_bp:go_zinc_ii_ion_transport	26	0.645	0.0243	0.176	1.25E-04	1
Curated_gene_sets:reactome_adherens_junctions_interactions	27	0.665	0.0255	0.182	1.30E-04	1
GO_bp:go_neuron_projection_morphogenesis	386	0.165	0.0237	0.046	1.75E-04	1
GO_bp:go_cell_junction_organization	179	0.234	0.023	0.065	1.79E-04	1

Supplementary Table 11: Gene expression and previous GWAS associations for each of the genes (nearest genes) confirmed across Models 1 & 3.

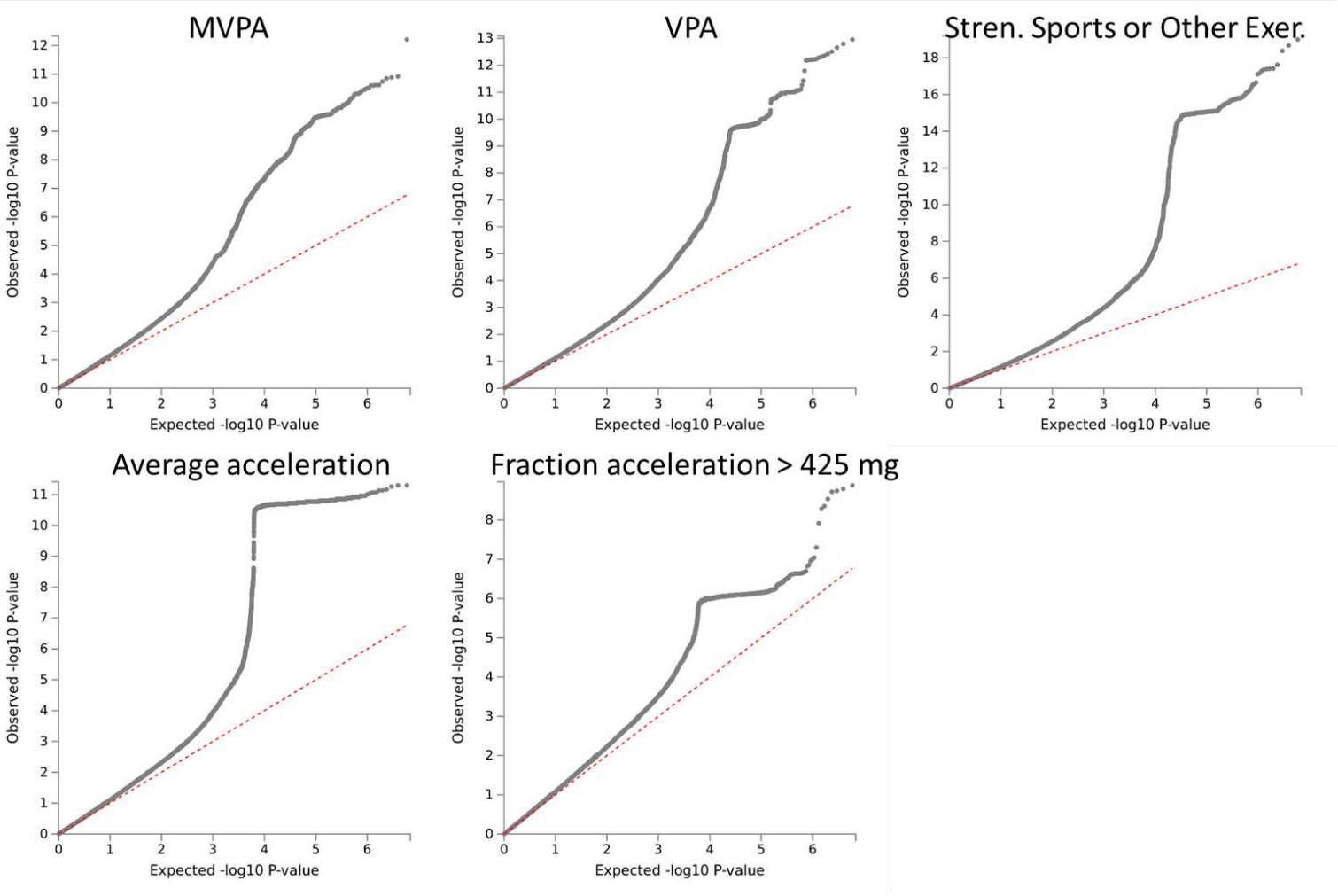
Gene	Tissue with highest median expression (from GTeX v6)	Previous GWAS associations (from NHGRI-EBI GWAS catalog)
APOE	adrenal gland	Alzheimer's disease; cognitive decline; longevity; coronary artery disease; lipid levels
EXOC4	transformed fibroblasts	educational attainment; intelligence; cognitive decline; blood pressure; schizophrenia
PAX5	EBV-transformed lymphocytes	HIV-1 susceptility; response to tocilizumab; intelligence; obesity related traits
CADM2	brain	obesity; alcohol consumption; risk taking; rubella; longevity; information processing speed; temperament; educational attainment; executive function; Alzheimer's disease; cognitive function; blood toxins; age at menarche; photic sneeze reflex; FEV1
CTBP2	cervix	prostate cancer; height; body mass index
DPY19L1	testis	FEV1
AKAP10	spleen	schizophrenia; blood cell traits
CTC-436P18.1	EBV-transformed lymphocytes	schizophrenia
SIPA1L1	brain	response to alcohol consumption; obesity; heart rate variability traits; FEV1
CRHR1	brain	Parkinson's disease; blood cell traits; bone mineral density; neuroticism; Alzheimer's disease; neurodegeneration
RCOR1	esophagus-mucosa	allergic disease; blood cell traits

FEV1: forced expiratory volume in first second of forced breath

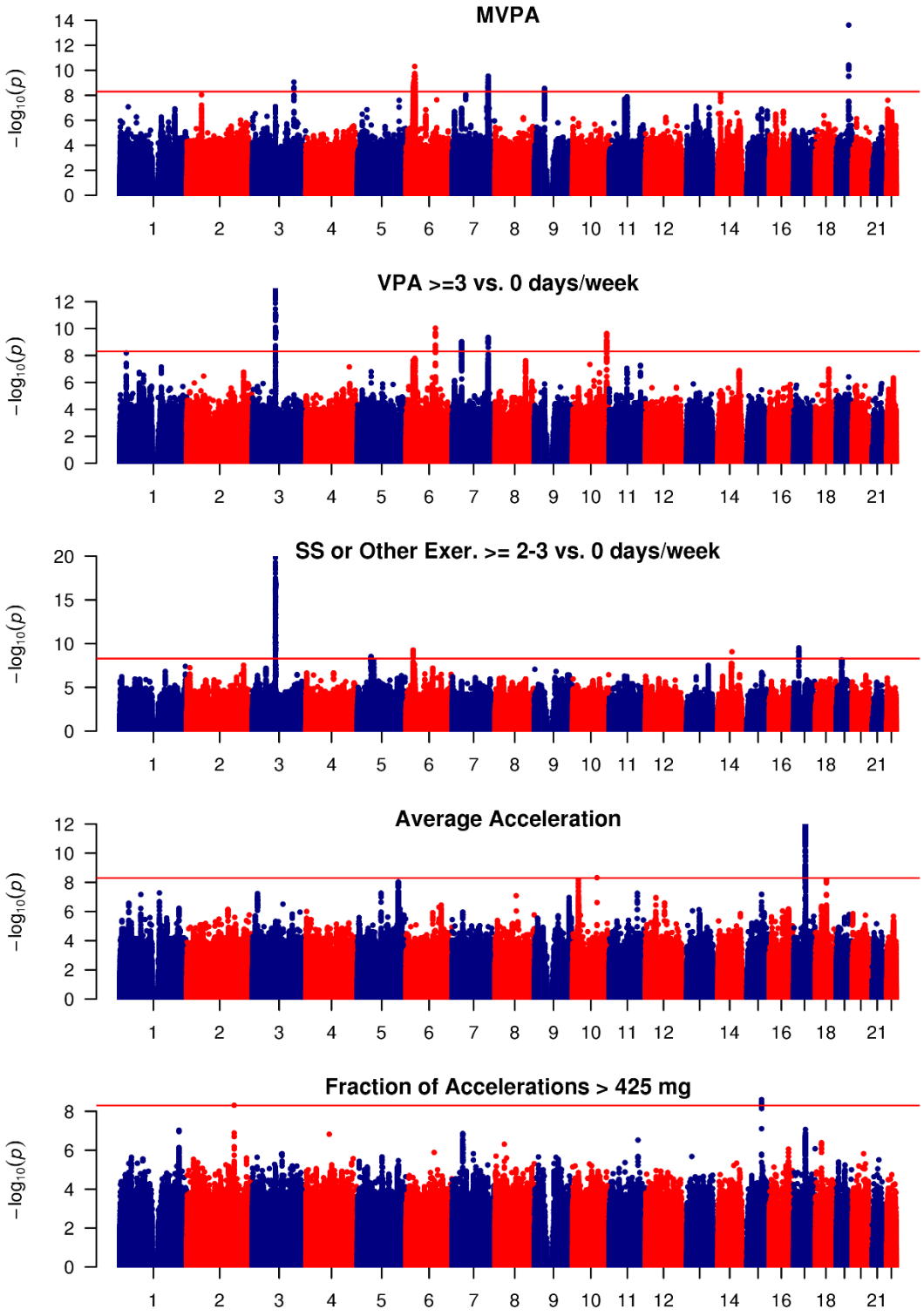
Supplementary Table 12: Expression quantitative trait locus analysis using GTEx v7 data and GTEx Portal online server, for each of the top 10 PA-associated SNPs identified across the five PA phenotypes. Only top five eQTLs are listed here in cases where there are more than five.

SNP	Nearest gene	Expressed Gene	Tissue	P-value
rs429358	APOE	No significant eQTLs found		
rs7804463	*EXOC4*	No significant eQTLs found		
rs62253088	*CADM2*	*CADM2*	Lung	7.40E-07
		CADM2	Adipose-Subcutaneous	4.20E-06
		CADM2	Adipose-Visceral	6.50E-05
rs2988004	*PAX5*	*RP11-220I1.1*	Muscle - Skeletal	3.90E-10
		RP11-220I1.1	Adrenal gland	1.80E-08
		RP11-220I1.1	Testis	1.30E-06
rs3781411	*CTBP2*	No significant eQTLs found		
rs328902	*DPY19L1*	*DPY19L1*	Brain - Frontal cortex (BA9)	4.80E-08
		DPY19L1	Brain - Cortex	1.60E-07
		DPY19L1	Muscle - Skeletal	7.60E-07
		DPY19L1P1	Brain - Cerebellum	1.60E-06
		DPY19L1P1	Thyroid	8.80E-06
rs166840	*AKAP10*	*AKAP10*	Whole Blood	3.30E-17
		RP11-209D14.2	Testis	6.00E-14
		KRT16P3	Esophagus - Mucosa	4.50E-13
		LGALS9B	Esophagus - Mucosa	3.50E-10
		RP11-78O7.2	Esophagus - Muscularis	2.60E-09
rs159544	*CTC-436P18.1*	*ERCC8*	Cells - Transformed fibroblasts	4.00E-06
		NDUFAF2	Brain - Nucleus Accumbens	4.10E+06
		ERCC8	Muscle - Skeletal	6.50E-06
		ELOVL7	Stomach	7.40E-06
		CTC-436P18.1	Testis	9.50E-06
rs75930676	*SIPA1L1*	No significant eQTLs found		
rs55657917	*CRHR1*	*CRHR1-IT1*	Skin - Sun Exposed	2.30E-137
		CRHR1-IT1	Adipose - Subcutaneous	3.80E-130
		CRHR1-IT1	Artery - Tibial	6.20E-122
		CRHR1-IT1	Nerve - Tibial	1.00E-121
		LRRC37A4P	Thyroid	1.10E-120

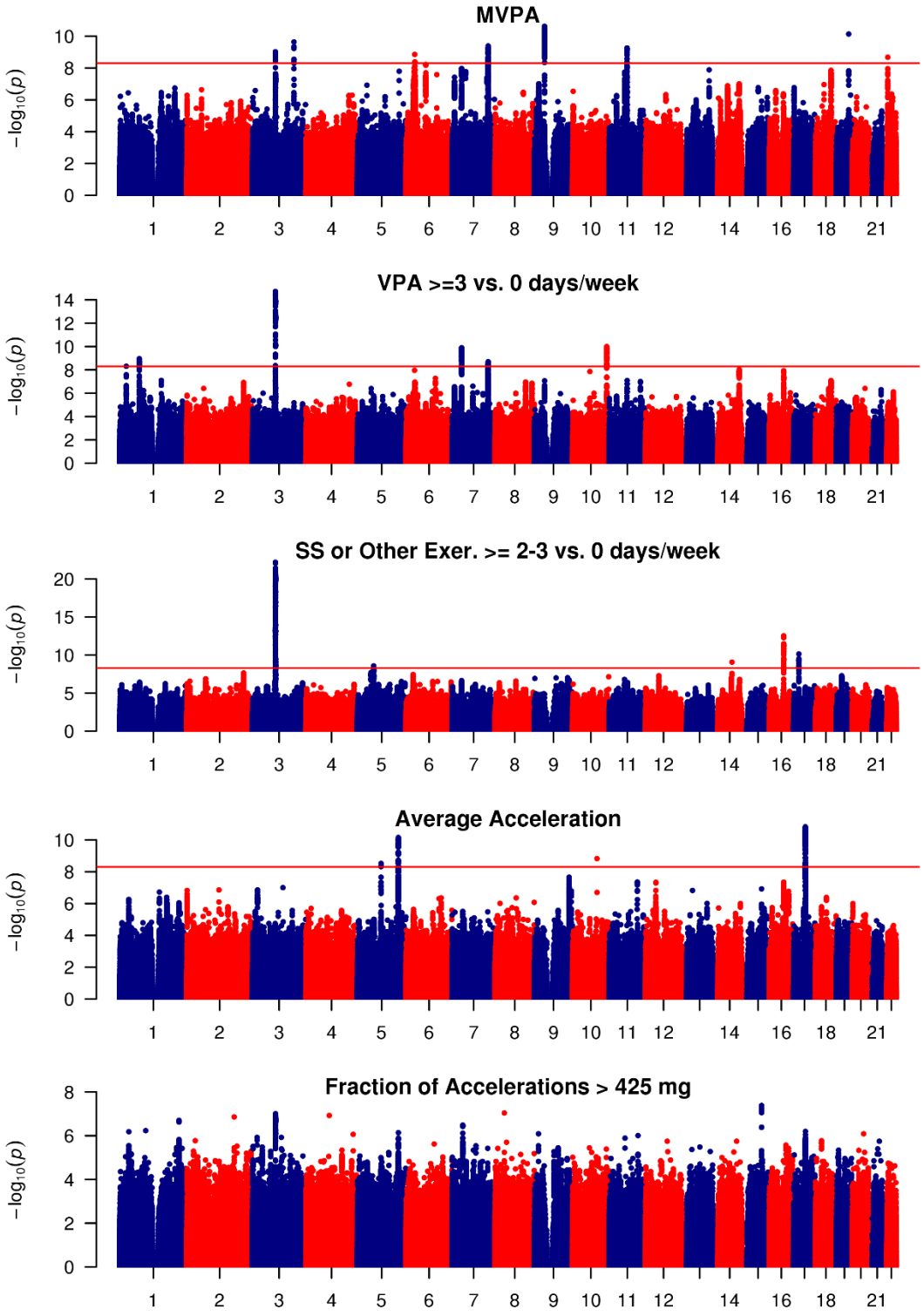
Supplementary Figure 1: Q-Q plots of Model 1 GWAS of all 5 PA phenotypes in the UK Biobank. Plots were drawn after uploading summary statistics into the FUMA GWAS online tool. Overlapping data points are not drawn (filtering was performed only for SNPs with P-value ≥ 1e-5)



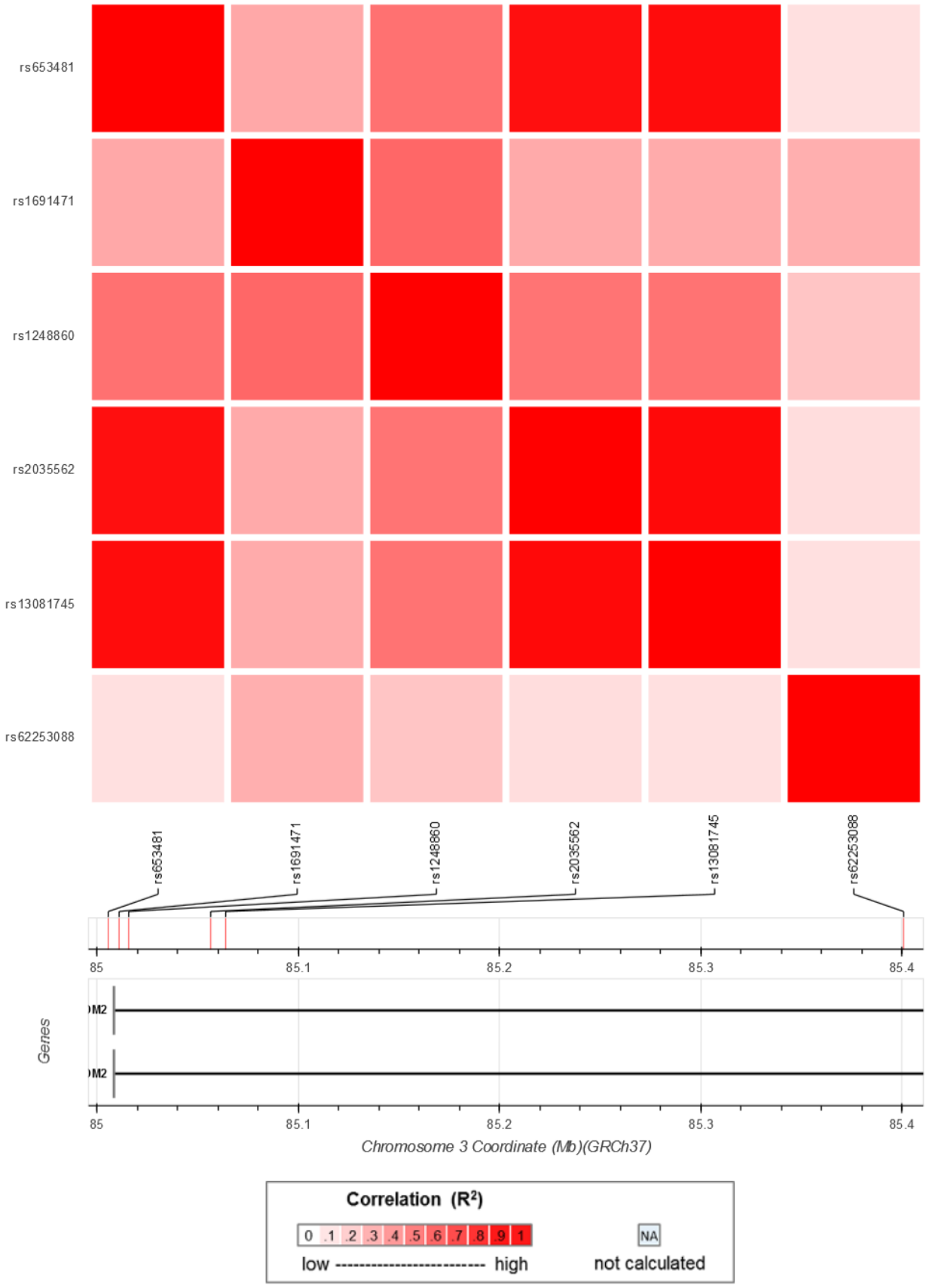
Supplementary Figure 2: Manhattan plot for PA phenotypes using a model additionally adjusted for Townsend Deprivation index, walk or standing at work, and physical activity at work.



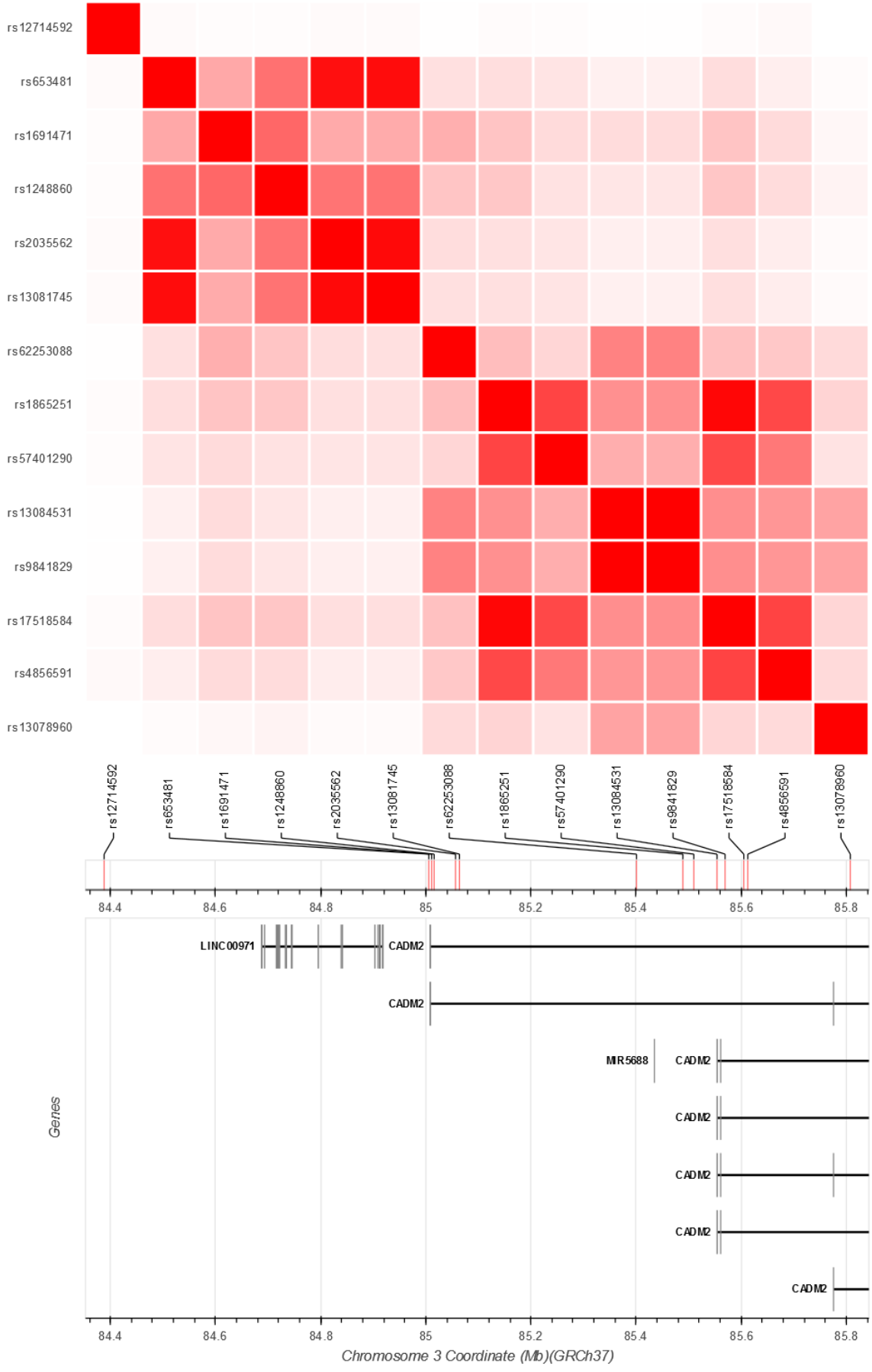
Supplementary Figure 3: Manhattan plot for PA phenotypes using a model additionally adjusted for Townsend Deprivation index, walk or standing at work, physical activity at work, and BMI.



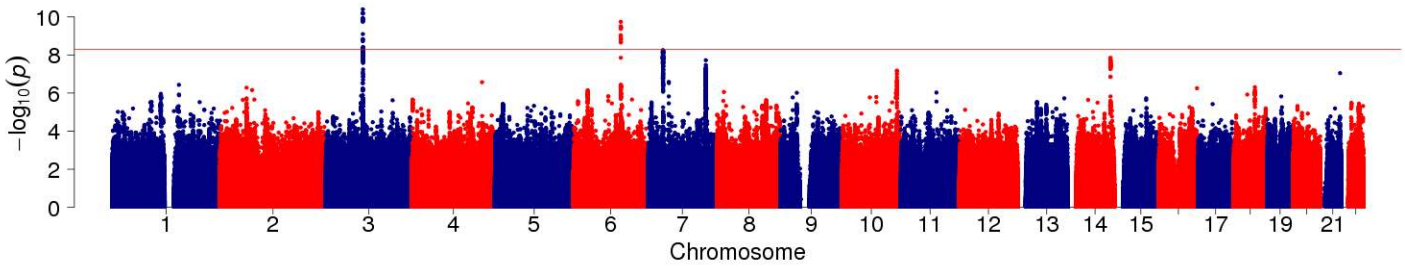
Supplementary Figure 4: Linkage Disequilibrium of PA-associated SNPs at the *CADM2* locus (from data on CEU and GBR from 1,000 Genomes data, phase 3. rs2035562, rs1248860, and rs62253088 were associated with MVPA, VPA and SSOE, respectively in Model 1 and/or Model 2. rs1308175, rs653481, and rs1691471 were associated with MVPA, VPA, and SSOE, respectively, in Model 3.



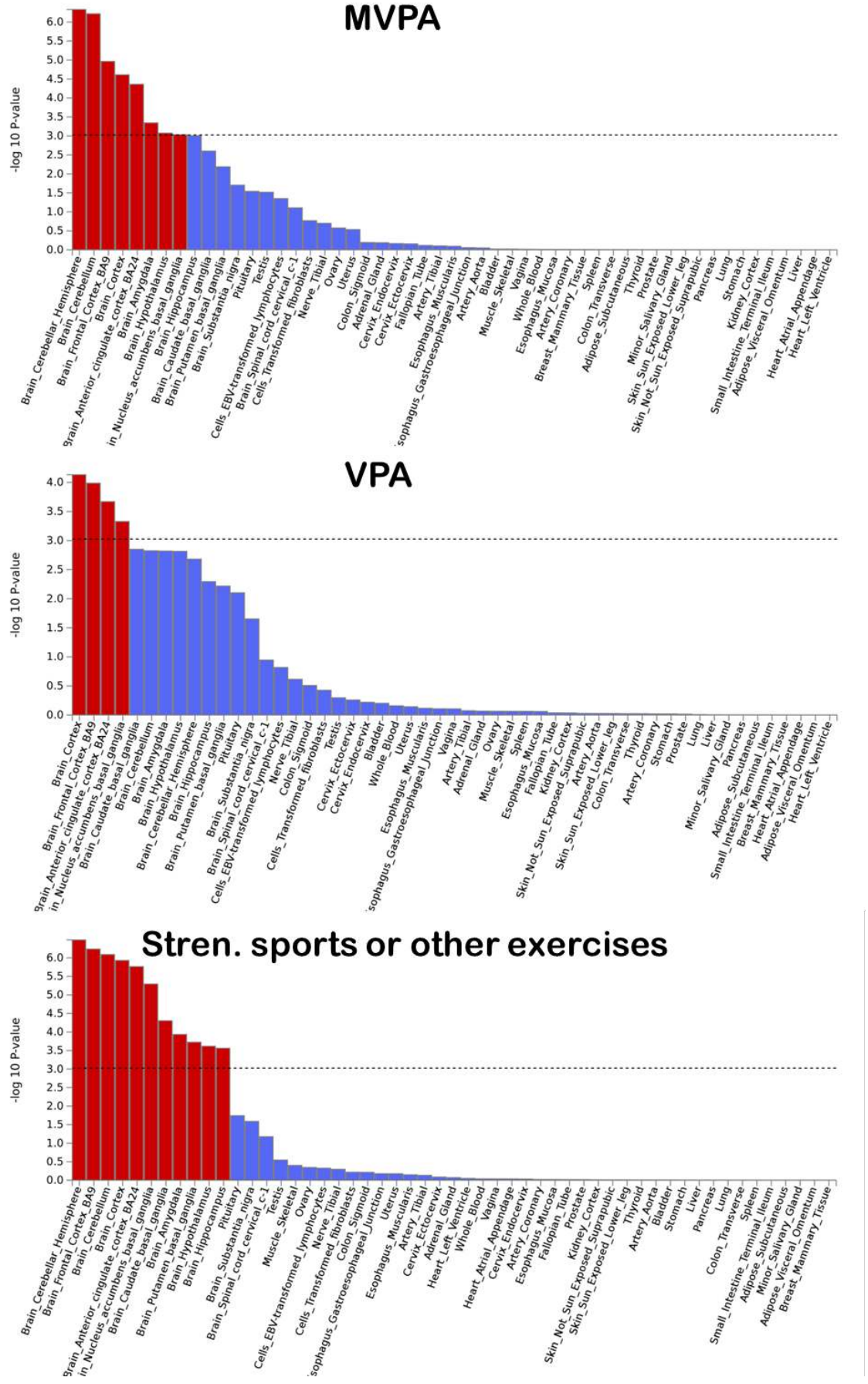
Supplementary Figure 5 (next page): Linkage disequilibrium among PA-associated SNPs and previously identified *CADM2* SNPs for BMI, and behavioral traits (from 1,000 Genomes data, phase 3 data on CEU and GBR). rs2035562, rs1248860, rs62253088, rs1308175, rs653481, and rs1691471 were associated with PA measures in the present study. rs13078960 was previously associated with BMI (Locke et al., 2015). rs12714592 and rs57401290 were previously associated with age at first sexual intercourse (Day et al. 2016), rs1865251 was associated with risk-taking behavior (Boutwell et al., 2017), rs9841829 with alcohol consumption (Clarke et al., 2017). rs17518584 was associated with cognitive processing speed (Ibrahim-Verbaas et al, 2016). Rs4856591 was associated with risk taking propensity (Day et al., 2016).


rs12714592
rs653481
rs1691471
rs1248860
rs2035562
rs13081745
rs62253088
rs1865251
rs57401290
rs13084531
rs9841829
rs17518584
rs4856591
rs13078960
84.4
84.6
84.8
85
85.2
85.4
85.6
85.8
LINC00971
CADM2
MIR5688
Genes
Chromosome 3 Coordinate (Mb)(GRCh37)
Correlation (R²)
0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1
low ---------- high
NA
not calculated

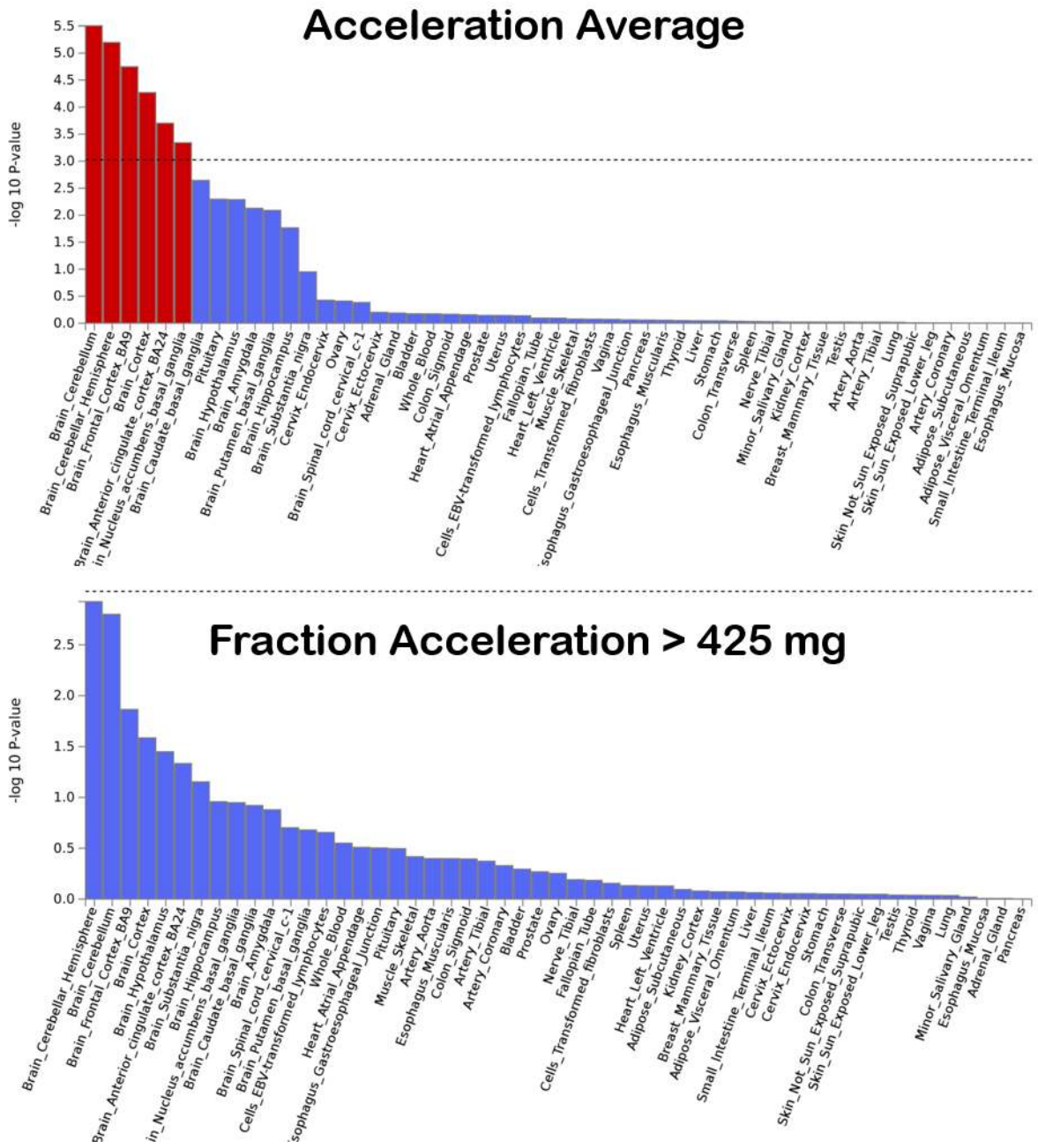
Supplementary Figure 6: Manhattan plot for VPA phenotype, using Model 1, not excluding individuals with intermediate levels of VPA.



Supplementary Figure 7: Results of gene-based tissue enrichment analysis for 53 tissue types, using Model 1 for self-reported PA phenotypes.

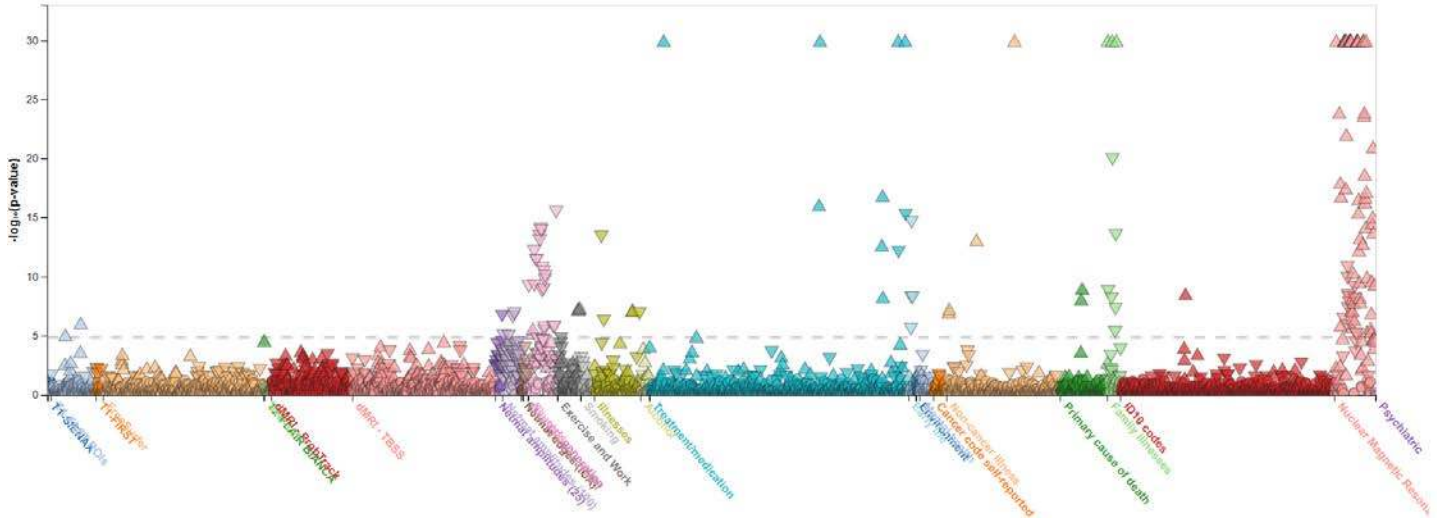


Supplementary Figure 8: Results of gene-based tissue enrichment analysis for 53 tissue types, using Model 1 for accelerometry-based PA phenotypes.

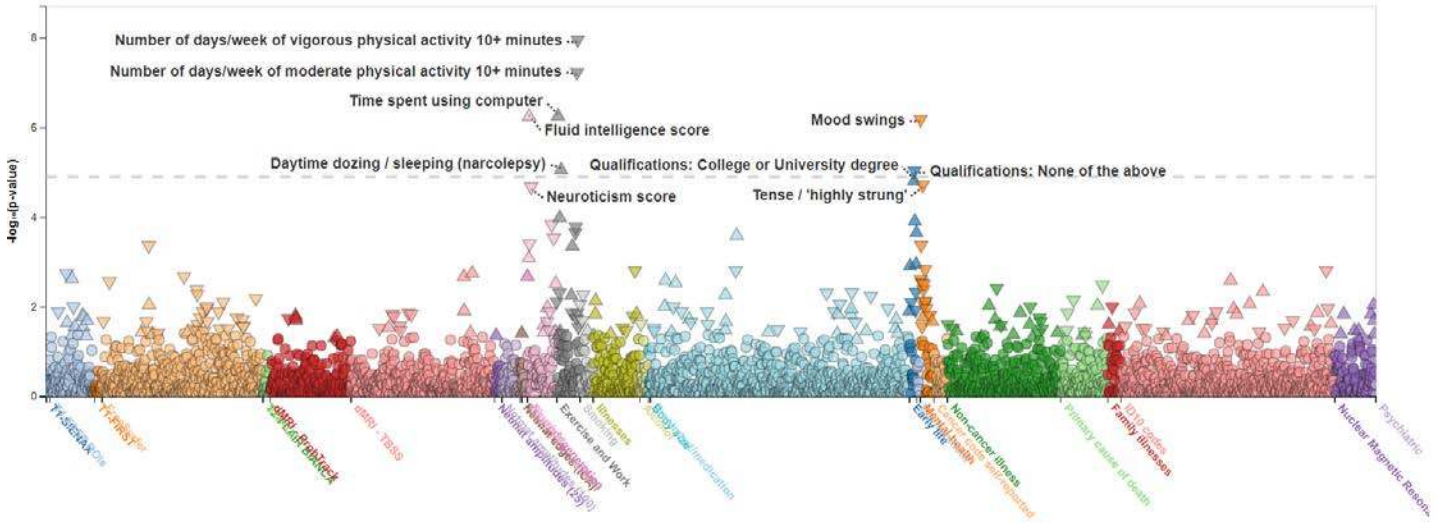


Supplementary Figure 9: Association of top PA-associated SNPs with other phenotypes in the UK Biobank (Images from the Oxford Brain Imaging Genetics (BIG) Server - version 2.0; http://big.stats.ox.ac.uk/)

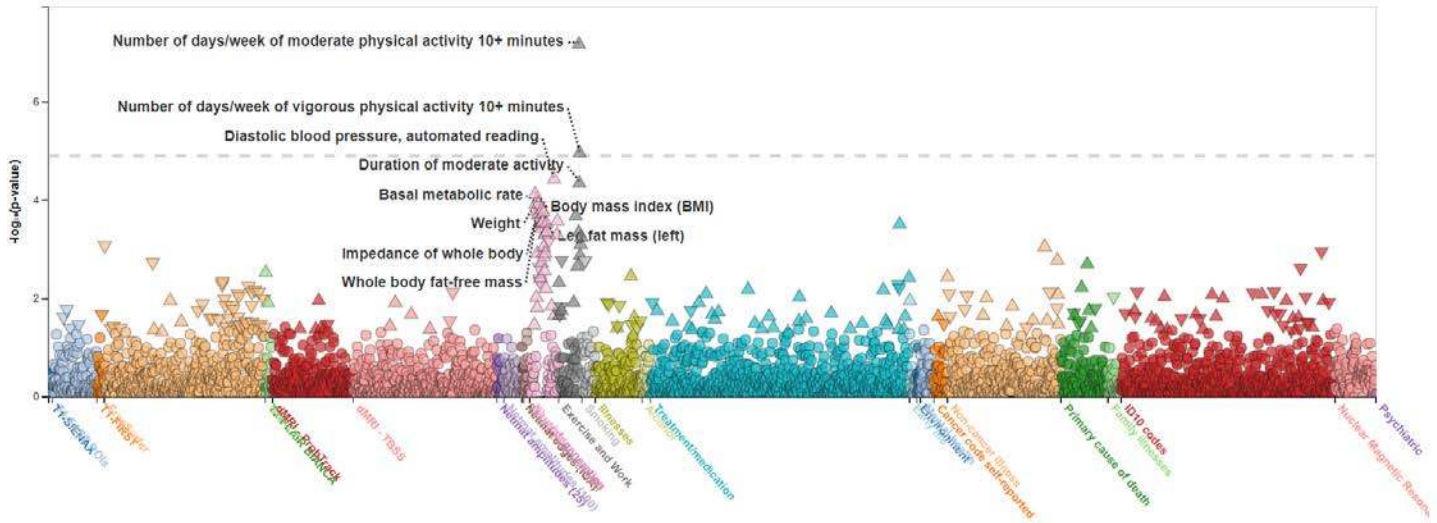
rs429358-C (*APOE*)



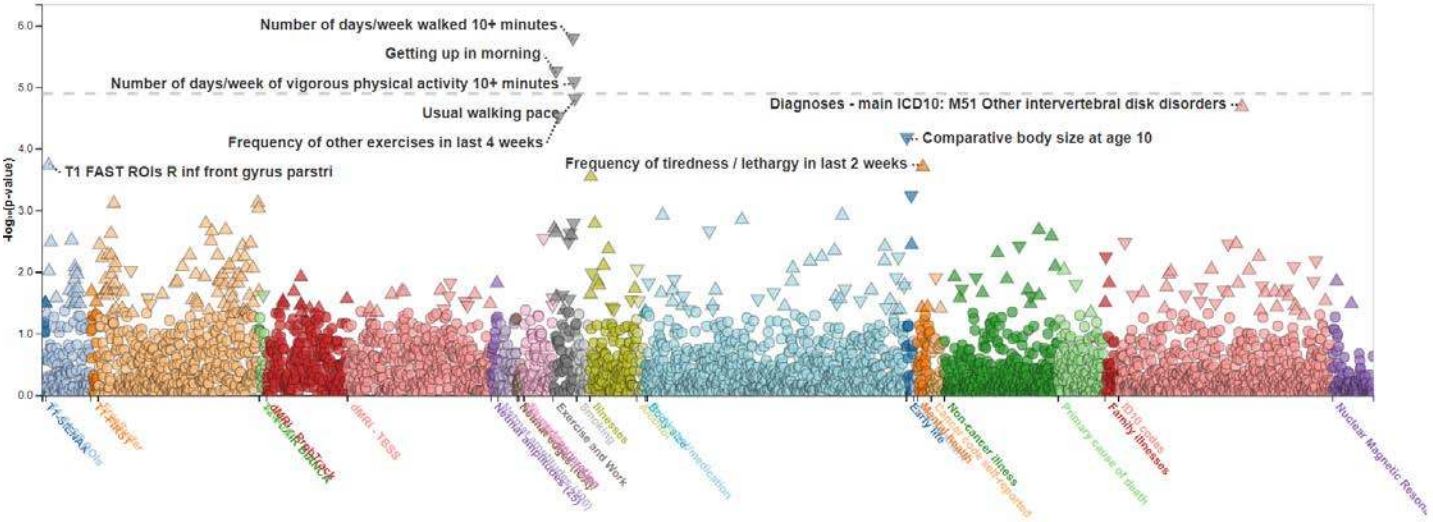
rs7804463-C (*EXOC4*)



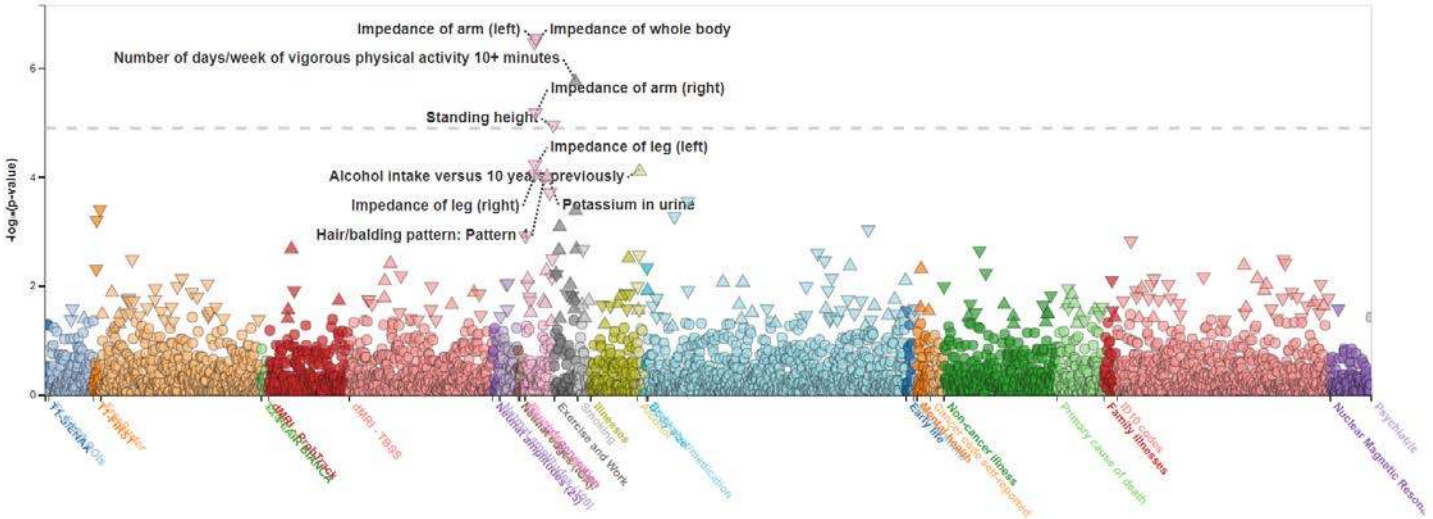
rs2988004-G (*PAX5*)



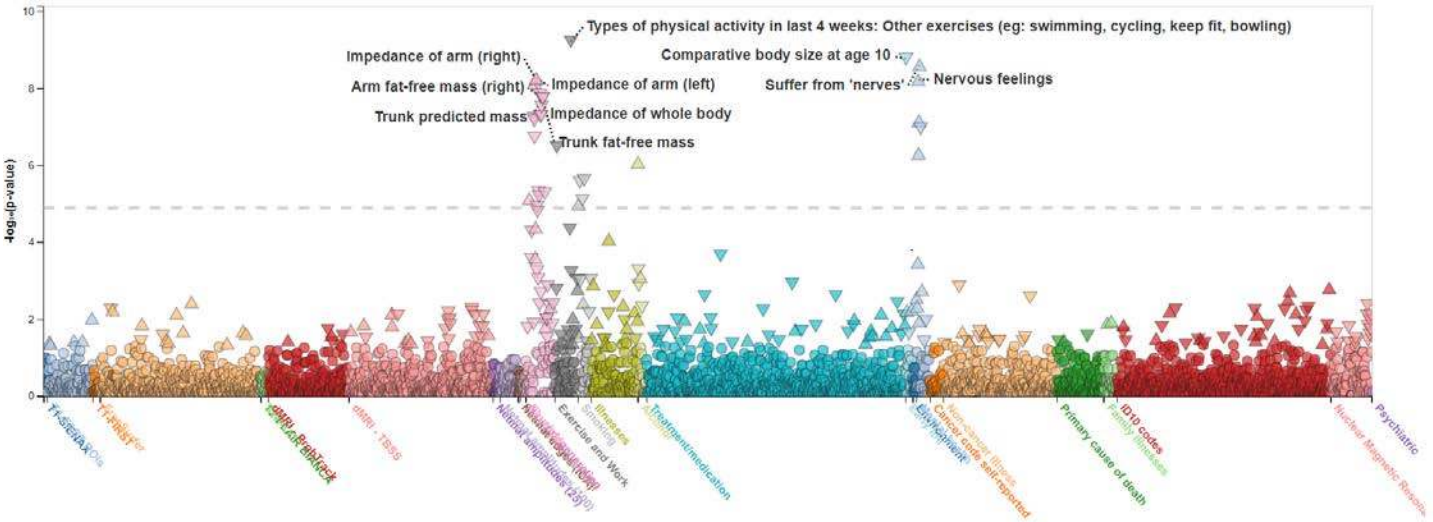
rs3781411-T (*CTBP2*)



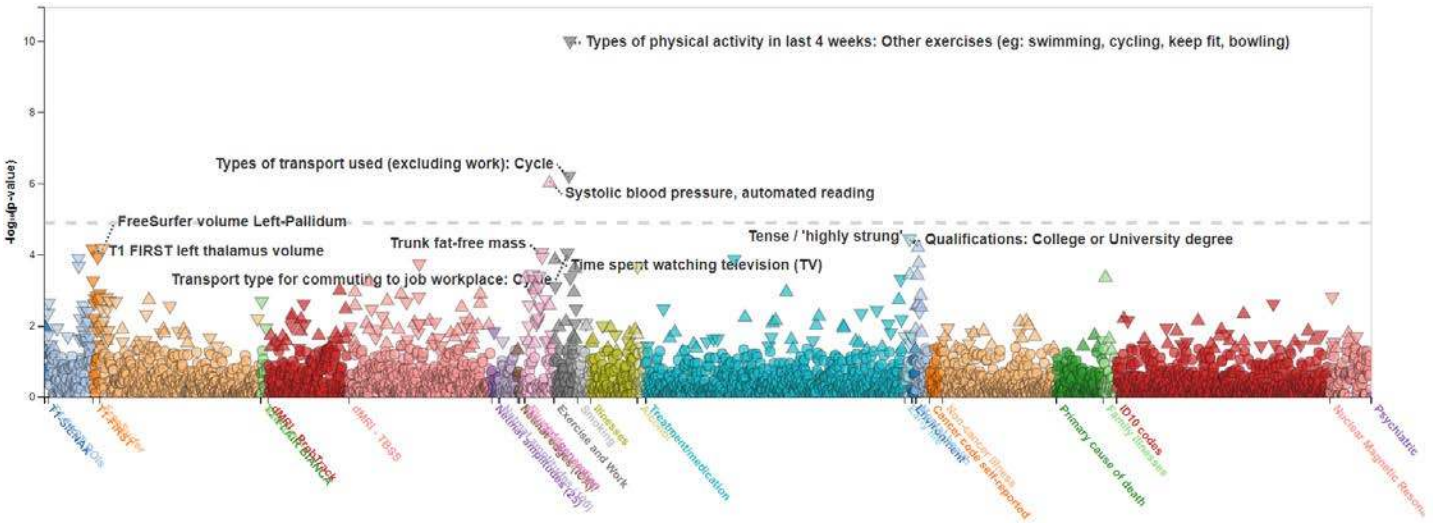
rs328902-T (*DPY19L1*)



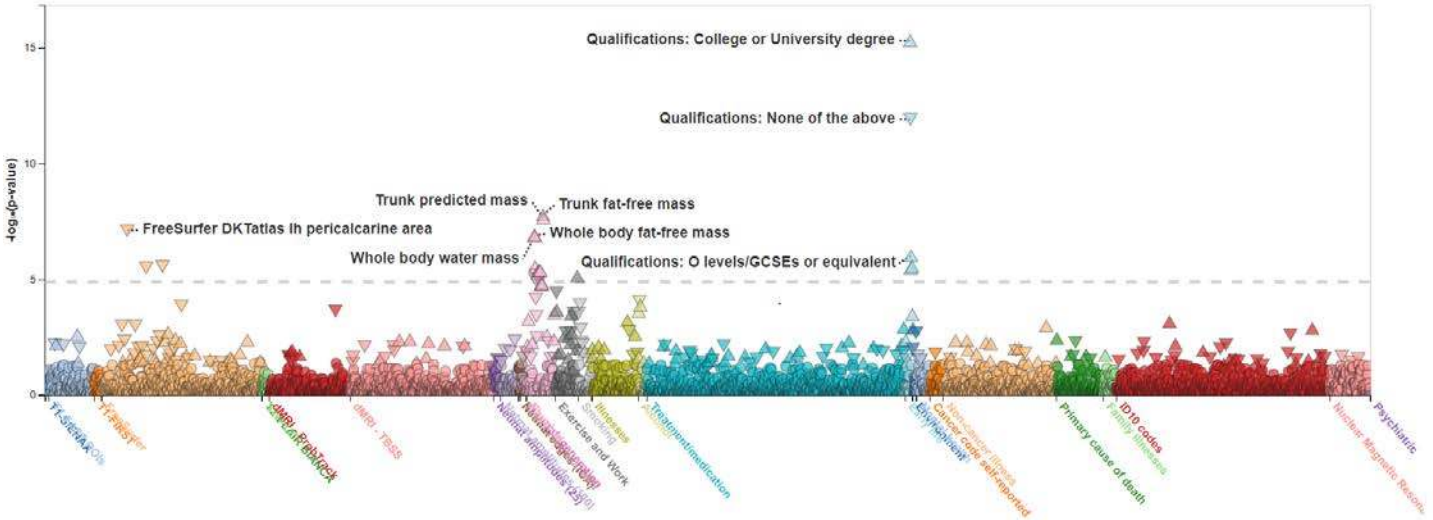
rs62253088-C (*CADM2*)



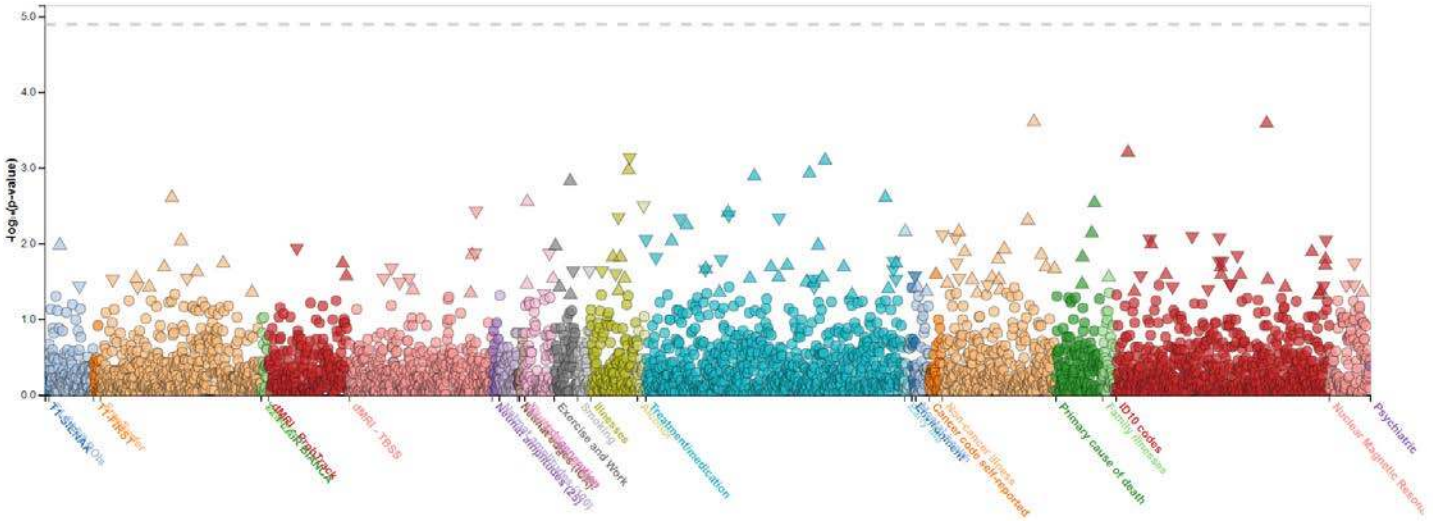
rs166840-A (*AKAP10)*



rs159544-G (*CTC-436P18.1)*



rs75930676-C (*SIPA1L1)*



rs55657917-G (*CRHR1*)