

VU Research Portal

Genome-wide Association Study of Liking for Several Types of Physical Activity in the UK Biobank and Two Replication Cohorts

Klimentidis, Y.C.; Newell, M.; Van Der Zee, M.D.; Bland, V.L.; May-Wilson, S.; Arani, G.; Menni, C.; Mangino, M.; Arora, A.; Raichlen, D.A.; Alexander, G.E.; Wilson, J.F.; Boomsma, D.I.; Hottenga, J.-J.; De Geus, Eco .J.C.; Pirastu, N.

published in

Medicine and Science in Sports and Exercise
2022

DOI (link to publisher)

[10.1249/MSS.0000000000002907](https://doi.org/10.1249/MSS.0000000000002907)
[10.1249/MSS.0000000000002907](https://doi.org/10.1249/MSS.0000000000002907)

document version

Publisher's PDF, also known as Version of record

document license

Article 25fa Dutch Copyright Act

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Klimentidis, Y. C., Newell, M., Van Der Zee, M. D., Bland, V. L., May-Wilson, S., Arani, G., Menni, C., Mangino, M., Arora, A., Raichlen, D. A., Alexander, G. E., Wilson, J. F., Boomsma, D. I., Hottenga, J.-J., De Geus, E. . J. C., & Pirastu, N. (2022). Genome-wide Association Study of Liking for Several Types of Physical Activity in the UK Biobank and Two Replication Cohorts. *Medicine and Science in Sports and Exercise*, *54*(8), 1252-1260. <https://doi.org/10.1249/MSS.0000000000002907>, <https://doi.org/10.1249/MSS.0000000000002907>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

Genome-wide Association Study of Liking for Several Types of Physical Activity in the UK Biobank and Two Replication Cohorts

YANN C. KLIMENTIDIS¹, MICHELLE NEWELL¹, MATTHIJS D. VAN DER ZEE^{2,3}, VICTORIA L. BLAND⁴, SEBASTIAN MAY-WILSON⁵, GAYATRI ARANI¹, CRISTINA MENNI⁶, MASSIMO MANGINO^{6,7}, AMIT ARORA¹, DAVID A. RAICHLLEN⁸, GENE E. ALEXANDER^{9,10,11,12}, JAMES F. WILSON^{5,13}, DORRET I. BOOMSMA^{2,3}, JOUKE-JAN HOTTENGA^{2,3}, ECO J. C. DE GEUS^{2,3}, and NICOLA PIRASTU¹⁴

¹Department of Epidemiology and Biostatistics, Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, AZ; ²Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, THE NETHERLANDS; ³Amsterdam Public Health Institute, Amsterdam UMC, THE NETHERLANDS; ⁴Division of Geriatric Medicine, School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO; ⁵Centre for Global Health Research, Usher Institute, University of Edinburgh, Edinburgh, UNITED KINGDOM; ⁶Department of Twin Research and Genetic Epidemiology, King's College London, London, UNITED KINGDOM; ⁷NIHR Biomedical Research Centre at Guy's and St Thomas' Foundation Trust, London, UNITED KINGDOM; ⁸Human and Evolutionary Biology Section, Department of Biological Sciences, University of Southern California, Los Angeles, CA; ⁹Department of Psychology and Psychiatry, University of Arizona, Tucson, AZ; ¹⁰Neuroscience and Physiological Sciences Graduate Inter-Disciplinary Programs, University of Arizona, Tucson, AZ; ¹¹Evelyn F. McKnight Brain Institute, University of Arizona, Tucson, AZ; ¹²Arizona Alzheimer's Consortium, Phoenix, AZ; ¹³MRC Human Genetics Unit, Institute of Genetic and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, UNITED KINGDOM; and ¹⁴Human Technopole, Viale Rita Levi-Montalcini, Milan, ITALY

ABSTRACT

KLIMENTIDIS, Y. C., M. NEWELL, M. D. VAN DER ZEE, V. L. BLAND, S. MAY-WILSON, G. ARANI, C. MENNI, M. MANGINO, A. ARORA, D. A. RAICHLLEN, G. E. ALEXANDER, J. F. WILSON, D. I. BOOMSMA, J.-J. HOTTENGA, E. J. C. DE GEUS, and N. PIRASTU. Genome-wide Association Study of Liking for Several Types of Physical Activity in the UK Biobank and Two Replication Cohorts. *Med. Sci. Sports Exerc.*, Vol. 54, No. 8, pp. 1252–1260, 2022. **Introduction:** A lack of physical activity (PA) is one of the most pressing health issues today. Our individual propensity for PA is influenced by genetic factors. Stated liking of different PA types may help capture additional and informative dimensions of PA behavior genetics. **Methods:** In over 157,000 individuals from the UK Biobank, we performed genome-wide association studies of five items assessing the liking of different PA types, plus an additional derived trait of overall PA-liking. We attempted to replicate significant associations in the Netherlands Twin Register (NTR) and TwinsUK. Additionally, polygenic scores (PGS) were trained in the UK Biobank for each PA-liking item and for self-reported PA behavior, and tested for association with PA in the NTR. **Results:** We identified a total of 19 unique significant loci across all five PA-liking items and the overall PA-liking trait, and these showed strong directional consistency in the replication cohorts. Four of these loci were previously identified for PA behavior, including *CADM2*, which was associated with three PA-liking items. The PA-liking items were genetically correlated with self-reported ($r_g = 0.38$ – 0.80) and accelerometer ($r_g = 0.26$ – 0.49) PA measures, and with a wide range of health-related traits. Each PA-liking PGS significantly predicted the same PA-liking item in NTR. The PGS of liking for *going to the gym* predicted PA behavior in the NTR ($r^2 = 0.40\%$) nearly as well as a PGS based on self-reported PA behavior ($r^2 = 0.42\%$). Combining the two PGS into a single model increased the r^2 to 0.59%, suggesting

Address for Correspondence: Yann C. Klimentidis, Ph.D., Department of Epidemiology and Biostatistics, Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, AZ; E-mail: yann@email.arizona.edu.

Submitted for publication October 2021.

Accepted for publication February 2022.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.acsm-msse.org).

0195-9131/22/5408-1252/0

MEDICINE & SCIENCE IN SPORTS & EXERCISE®

Copyright © 2022 by the American College of Sports Medicine

DOI: 10.1249/MSS.0000000000002907

that PA-liking captures distinct and relevant dimensions of PA behavior. **Conclusions:** We have identified the first loci associated with PA-liking and extended our understanding of the genetic basis of PA behavior. **Key Words:** GENOME-WIDE ASSOCIATION STUDY, PHYSICAL ACTIVITY, LIKING, PREFERENCES, GENETIC, EXERCISE

Levels of physical activity (PA) have decreased dramatically in most parts of the world over the past several hundred years, likely contributing to a major and growing chronic disease burden (1–3). Physical inactivity has been compared with smoking in terms of its effect on disease burden, which ranges widely from cardiometabolic disease to mental health (1). As genetic factors partly explain individual differences in PA behavior (4–7), identifying specific genetic risk factors can advance our understanding of 1) important interindividual variation, 2) relevant biological pathways, and 3) the presence, direction, and strength of causal relationships between PA behaviors and health outcomes.

Several loci have already been identified as being associated with self-reported and accelerometry-measured levels of PA (8–10). These measures may each be limited in multiple ways. For example, self-reported measures may be highly influenced by social- and health-related pressures and may not be stable over time, whereas accelerometer measures may only be sensitive to certain types of PA, with wear time often limited to a single week of a person's lifetime and likely influencing behavior (11). Measures of an individual's liking of PA may more accurately capture overall life-long propensity to engage in PA and, at a minimum, serve as complementary, broader, and refined measures of PA behavior.

One's propensity to engage in habitual PA is driven by a complex set of genetic and nongenetic factors (12). Theoretically, the internal motivation and self-determination theory (SDT) of health behaviors could help explain the motivation for PA (13). SDT identifies both intrinsic and extrinsic motivators for PA and sports (14). Intrinsic motivation occurs when doing the activity provides inherent satisfaction. Extrinsic motivation is contributed by the outcome separable from the activity *per se*. Empirical studies that used SDT constructs consistently support that intrinsic motivation contributed to by competence satisfaction is a strong positive predictor of sustained exercise adherence (15). A Finnish twin study on genetic and environmental influences on motivation for leisure-time PA using a version of the Recreational Exercise Motivation Measure—a tool designed based on the self-determination theory for PA (16)—found higher heritability in the intrinsic motives than extrinsic motives. Enjoyment of PA had the highest heritability among different motivation dimensions (17). These findings suggest that measures of intrinsic motivation are strong predictors of actual PA behavior, are influenced by genetic factors, and may thus inform specific dimensions of the genetics of habitual PA behavior. In a principal component analysis of this questionnaire in the Netherlands Twin Register (NTR), in which items were mainly related to food and drink liking, it was found that the liking of PA/sport stood out as a separable factor (18).

To discover genetic loci associated with PA-related liking, we performed a genome-wide association study (GWAS) of

five individual liking items and one composite trait of overall PA-liking in over 157,000 UK Biobank participants, with replication of top loci in the NTR and TwinsUK studies. We then examined how PA-liking genetically relates to self-reported and accelerometry-measured PA (8) and to a wide range of other traits and health outcomes. Finally, we examined how polygenic scores (PGS) of PA-liking derived from the UK Biobank predicted both liking and self-reported PA in the NTR study.

METHODS

UK Biobank. The UK Biobank is a prospective cohort study of 500,000 adults (ages 37–73 yr at the baseline examination in 2006–2010) from the UK (19). All participants provided written informed consent, and ethical approval for this study was granted. Ethical approval for the UK Biobank study was obtained by the National Information Governance Board for Health and Social Care and the National Health Service North West Multicenter Research Ethics Committee.

PA-liking. In 2019, a link to a questionnaire was sent by e-mail to UK Biobank participants to assess the liking of specific foods as well as physical activities (20). This questionnaire was developed and administered mainly as a way to improve diet-related phenotyping in the UK Biobank, as it may not suffer from the same biases present in other ways of assessing dietary intake and food choice. Questionnaires about liking for foods and beverages have previously been shown to exhibit high validity and reliability (21,22), which likely applies for the liking of PA (23,24). This questionnaire consisted of 150 items, five of which were related to PA (*going to the gym, working up a sweat, exercising with others, exercising alone, and bicycling*), and assesses liking through a 9-point hedonic scale ranging from 1 for extremely dislike up to 9 for extremely like, in increments of one. Of 333,344 participants that were sent an e-mail invitation, 181,224 (54.4%) fully completed the questionnaire as of January 16, 2020 (20). The same items were asked in the two replication studies (see below).

Genetic markers. Genotypes in the UK Biobank were measured with the Affymetrix UK Biobank Axiom Array (Santa Clara, CA, USA) in 90% of participants. The remainder (10%) were genotyped with the Affymetrix UK BiLEVE Axiom Array. Further details about imputation, principal components analysis, and QC procedures can be found elsewhere (25).

Replication of sentinel single nucleotide polymorphism in NTR and TwinsUK. The NTR is a longitudinal register of twins and their relatives (26). Between December 2014 and May 2017, participants responded to the same liking questionnaire as the one administered in the UK Biobank,

including the same five PA-related questions (18). More details regarding genotyping and imputation are provided in Supplemental Methods (see Appendix, Supplemental Digital Content 1, <http://links.lww.com/MSS/C542>). Top genome-wide significant single nucleotide polymorphism (SNP) or their proxies from the UK Biobank GWAS were interrogated for each respective item and the trait of overall PA-liking in NTR GWAS results. Because we tested 25 SNP/loci, a replication was deemed successful if the P value was <0.002 .

TwinsUK is a large twin registry for the study of health that began in 1992 (27). The same liking questionnaire used in UK Biobank was previously used in the TwinsUK cohort (28). TwinsUK genotyping has been previously described in detail (29). Briefly, TwinsUK samples were genotyped with a combination of two Illumina arrays (HumanHap300, HumanHap610Q). After the Genotype QC stage, the samples from the two arrays were combined, and the imputations were performed using the Michigan Imputation Server (30) using the 1000 Genomes Phase3 v5 reference panel.

Statistical analyses. To assess associations of PA-liking with sex, age, body mass index (BMI), income, University/College degree (yes/no), and Townsend Deprivation Index, we performed linear regression after ensuring normality and homoscedasticity of residuals. In GWAS, we included only individuals of European descent. We considered participants as being of European descent if they were either among the genetically British as defined by UK Biobank or self-identified as “Irish,” “White,” or “any other White background.” We performed GWAS with fastGWA (31), which implements a mixed-effect linear regression that controls for population stratification and relatedness. We included age at time of questionnaire, sex, genotyping chip, batch, and the first 10 genetic principal components as covariates in the model used in GWAS. We used linkage disequilibrium (LD) score regression to assess test score inflation, SNP-based heritability, and to assess genetic correlations among PA-liking items and previously reported PA traits (32). To obtain a measure of overall PA-liking, we also derived a GWAS of the first principal component derived through the genetically independent phenotype (GIP) method (33,34) and starting from the genetic correlation matrix to derive the loadings of each trait on each GIP. Independent significant loci were identified as those with $P < 5 \times 10^{-8}$ with $r^2 < 0.1$ and >250 kb distance. The online LDHUB platform (35) was used to examine genetic correlations of the five individual liking items and the overall PA-liking (GIP1) trait with a wide range of traits and diseases (~800 phenotypes). We used stratified LD score regression to identify tissue-type-specific enrichment of heritability (36) from the overall PA-liking (GIP1) GWAS. PGS were calculated in NTR with the LDPRED package (37), based on UKB GWAS results for PA-liking and for strenuous sports or other exercise (SSOE) and other PA behavior measures (8), and were tested for correlation with PA-liking phenotypes and self-reported PA behavior in NTR. Before calculating the PGS, LD-adjusted beta coefficients were calculated from the summary statistics to correct for the effects of LD and to maximize predictive accuracy of

the PGS (37). These beta coefficients were calculated using an LD pruning window of 250 KB, with different cutoffs of the proportion of causal SNP (e.g., P005 for a model prior of 5% proportion of genetic variants that are causal); see Supplemental Methods, Supplemental Digital Content 1, <http://links.lww.com/MSS/C542>, for more details). Although we provide results from models across the range of priors, we report r^2 from the model prior with the highest r^2 in the text.

RESULTS

PA-liking phenotypes. Respondents to the questionnaire were on average younger, more likely to be female, have a lower BMI, and a lower self-report PA but higher acceleration average (see Supplemental Table 1 in Appendix, Supplemental Digital Content 1, <http://links.lww.com/MSS/C542>) as compared with nonrespondents. Descriptive statistics of UKB, NTR, and TwinsUK samples and mean PA-liking levels are shown in Supplemental Table 2 (see Appendix, Supplemental Digital Content 1, <http://links.lww.com/MSS/C542>). Mean score of liking was highest for *exercising with others* in the two UK samples and highest for *bicycling* in NTR. It was lowest for *going to the gym* in all three samples. In the UK Biobank, PA-liking was negatively correlated with age and was higher in males (except for *exercising with others*). We observed generally positive correlations of PA-liking items with education and income and negative correlations with Townsend Deprivation Index (e.g., *exercising with others*), such that less deprivation was associated with more PA-liking (Supplemental Table 3, Appendix, Supplemental Digital Content 1, <http://links.lww.com/MSS/C542>). Phenotypic correlations of PA-liking with PA behavior traits were strongest for self-reported vigorous PA and strenuous sports or other exercise (r between 0.27 and 0.44). Genetic correlations of PA-liking with accelerometry traits were generally strongest for *working up a sweat* liking ($r \approx 0.3$) (Supplemental Fig. 1, Appendix, Supplemental Digital Content 1, <http://links.lww.com/MSS/C542>).

GWAS. Data from up to 158,189 UK Biobank participants (mean age = 66.8 yr; 57% female) were analyzed for GWAS. SNP-based heritabilities varied from 0.054 (0.004) for *going to the gym* up to 0.075 (0.004) for *bicycling*. SNP heritability for overall PA-liking (GIP1) was 0.089 (0.004) (Supplemental Table 4, Appendix, Supplemental Digital Content 1, <http://links.lww.com/MSS/C542>). We did not observe evidence of genomic inflation beyond that explained by polygenic signal according to LD score regression estimates (Supplemental Table 3, Appendix, Supplemental Digital Content 1, <http://links.lww.com/MSS/C542>). Between 2 and 6 genome-wide significant loci were identified for each individual liking item and 8 loci for overall PA-liking (GIP1), for a total of 26 SNP-trait associations in 19 loci (Fig. 1 and Table 1). We did not observe a large degree of overlap of top loci across the five different liking items, although their level of association, regardless of P value, was generally consistent across all liking items as well as PA behavior (see Supplemental Fig. 2, Appendix, Supplemental Digital Content 1, <http://links.lww.com/MSS/C542>).

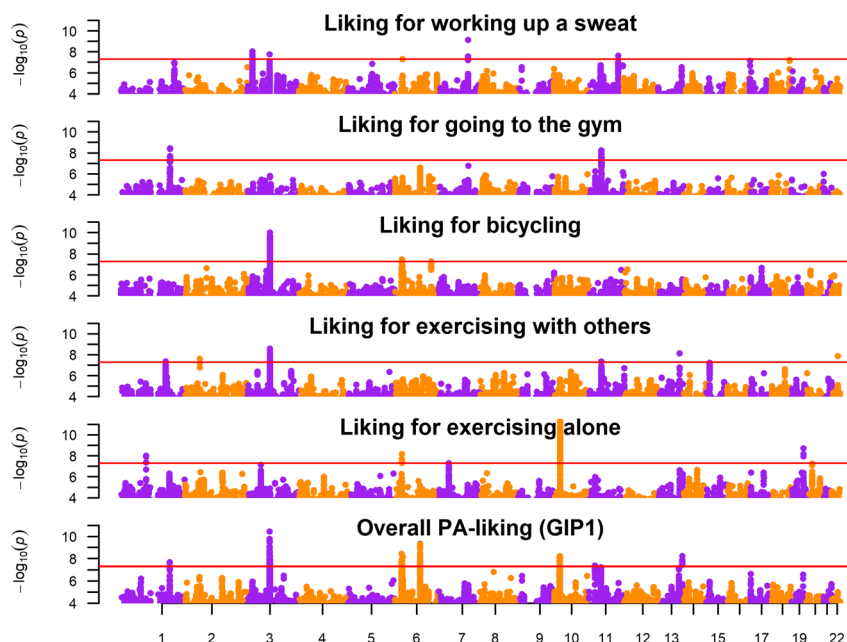


FIGURE 1—Manhattan plots from GWAS performed in the UK Biobank ($n > 140,000$) of five PA-liking items and an overall measure of PA-liking based on the genetically independent phenotype method (overall PA-liking [GIP1]).

com/MSS/C542). Only two loci (*CADM2* and the locus on chromosome 11) were significantly associated with more than one item. Sentinel SNPs were also associated with other traits such as bone mineral density, body size and body composition, educational attainment, respiratory traits, psychiatric traits, and other PA-related traits such as usual walking pace and time watching television (Supplemental spreadsheet, Supplemental Digital Content 2, <http://links.lww.com/MSS/C543>). Tissue-type enrichment analysis via stratified LD score regression identified the nucleus accumbens, hippocampus, caudate, frontal cortex, and amygdala (Supplemental Table 5, Appendix, Supplemental Digital Content 1, <http://links.lww.com/MSS/C542>).

In the NTR replication, we found directional consistency for 19 out of 26 SNPs, including 6 out of 8 PA-liking (GIP1) SNP (Table 1). Only one SNP (*CADM2*; for *exercising with others*) was nominally significant ($P < 0.05$) before multiple testing correction. In the TwinsUK replication, we found directional consistency for 15 out of 26 SNPs, including 6 out of 8 overall PA-liking (GIP1) SNPs (Table 1).

Genetic correlations. Genetic correlations among PA-liking items were moderate to strong (Fig. 2). Among the PA-liking items, strong correlations were observed for *working up a sweat* with *going to the gym* ($r_g = 0.79$), *exercising with others* ($r_g = 0.76$), and *exercising alone* ($r_g = 0.72$), and the weakest correlation was between *exercising alone* and *exercising with others* ($r_g = 0.46$). Across liking and behavior (self-reported PA and accelerometer-derived PA) traits, correlations were strongest for PA-liking with self-reported vigorous PA and strenuous sports and other exercise (r_g between 0.59 and 0.78). Genetic correlations of PA-liking with accelerometer traits were generally strongest for *exercising alone* ($r_g \approx 0.47$) and weakest with *going to the gym* ($r_g \approx 0.28$).

Genetic correlation assessments with a wide range of traits and diseases reveal correlations with UK Biobank variables related to PA including accelerometry, as well as with obesity-related traits, tiredness, and lifestyle traits such as alcohol consumption, TV watching, and taking dietary supplements, among others (Fig. 3).

PGS analyses in NTR. PGS of each PA-liking item and the overall PA-liking trait were calculated for each NTR participant, using as weights the effect sizes resulting from the GWAS of the UK Biobank data. These PGS were generally, but not always, most strongly associated with the corresponding PA-liking phenotype (Supplemental Table 6, Appendix, Supplemental Digital Content 1, <http://links.lww.com/MSS/C542>). For example, the *exercising with others* PGS was most strongly associated with *exercising with others* in NTR ($r^2 = 0.80\%$, $P = 2.6 \times 10^{-13}$). The overall PA-liking (GIP1) PGS was most strongly correlated with the liking of *exercising alone* in NTR ($r^2 = 0.82\%$, $P = 7.6 \times 10^{-15}$; Supplemental Table 6, Appendix, Supplemental Digital Content 1, <http://links.lww.com/MSS/C542>). These PGS were also significantly associated with self-reported PA behavioral phenotypes in NTR: self-reported *total exercise*, *team-based exercise*, and *solitary exercise*. Among the PA-liking PGS, the best predictors of these self-reported PA phenotypes in NTR were liking for *going to the gym* and *exercising with others* with the self-reported PA measures of total exercise and solitary exercise (but not with team-based exercise; Supplemental Fig. 2, Appendix, Supplemental Digital Content 1, <http://links.lww.com/MSS/C542>). When compared with the PGS based on the self-reported and accelerometer PA GWAS in UK Biobank, the liking PGS had a similar predictive performance. For example, a PGS of self-reported SSOE predicted self-reported total exercise in NTR only slightly better than a

TABLE 1. List of genetic variants associated with PA-liking and overall PA-liking (GIP1) in UKB, with replication results from NTR and TwinsUK.

SNP	Chr	Position	Genes with Nearest Transcription Start Site	UK Biobank				NTR Replication				TwinsUK Replication						
				Effect Allele	Other Allele	EAF	Beta	SE	P Value	n	Beta	SE	P Value	n	Beta	SE	P Value	n
Liking for working up a sweat																		
rs4645151	3	18,730,433	SATB1, TBC1D5	T	C	0.69	-0.006	0.001	8.90E-09	158,189	-0.008 ^a	0.004	0.057	7,874	0.001	0.011	0.898	1,471
rs7616022	3	85,086,321	CADM2	C	T	0.89	0.009	0.002	1.70E-09	158,189	-0.002 ^b	0.004	0.541	7,874	0.004 ^b	0.010	0.690	1,471
rs12403434	7	114,528,614	MDFC	A	G	0.06	-0.013	0.002	7.50E-10	158,189	-0.001	0.008	0.888	7,874	0.033	0.022	0.135	1,471
rs7934107	11	115,144,046	CADM1	T	C	0.1	0.009	0.002	2.30E-08	158,189	0.001	0.006	0.918	7,874	0.023	0.015	0.138	1,471
Liking for going to the gym																		
rs17370393	1	190,526,252	BRINP3	T	C	0.15	0.010	0.002	3.80E-09	140,015	0.002	0.007	0.749	7,371	-0.007	0.015	0.623	1,366
rs61882686	11	46,390,680	MDK, CHRM4	A	C	0.09	0.013	0.002	5.90E-09	140,015	-0.002	0.009	0.869	7,371	0.014	0.020	0.461	1,366
Liking for bicycling																		
rs5850689	3	85,586,903	CADM2	T	TG	0.63	-0.008	0.001	1.00E-10	148,803	-0.004 ^c	0.003	0.188	7,936	-5.0E-04 ^d	0.011	0.965	1,424
rs16897515	6	27,278,020	POMT2IL2, PRSS16	C	A	0.19	-0.008	0.001	3.20E-08	148,803	0.001	0.004	0.886	7,936	0.014	0.303	1,424	
Liking for exercising with others																		
rs25571	1	173,773,928	DARS1, CENPL	CTT	C	0.53	0.006	0.001	4.40E-08	153,358	0.004 ^e	0.004	0.383	7,738	-0.001	0.009	0.883	1,456
rs4672114	2	56,614,299	CCDC88A, EFEIMP1	C	G	0.3	-0.007	0.001	2.40E-08	153,358	-0.007	0.005	0.129	7,738	-0.003	0.010	0.740	1,456
rs13096057	3	85,342,567	CADM2	A	C	0.61	0.007	0.001	2.60E-09	153,358	0.009 ^f	0.004	0.048	7,738	0.010 ^g	0.010	0.323	1,456
rs113176495	11	46,789,346	F2, ZNF408	GT	G	0.09	0.010	0.002	4.50E-08	153,358	0.006 ^h	0.008	0.470	7,738	3.3E-04	0.017	0.985	1,456
rs9584870	13	99,245,866	STK24, FARP1-AS1	C	T	0.37	0.007	0.001	7.30E-09	153,358	0.005	0.005	0.302	7,738	0.022	0.010	0.026	1,456
rs169407	22	41,935,063	POLR3H, CSDC2	A	G	0.36	0.006	0.001	1.40E-08	153,358	-0.002	0.005	0.590	7,738	-0.018	0.010	0.063	1,456
Liking for exercising alone																		
rs10875127	1	98,527,414	DPYD	G	A	0.31	0.006	0.001	9.50E-09	156,371	-0.002	0.005	0.761	7,703	-0.001	0.011	0.928	1,454
rs149896934	6	26,230,172	HIST1H1D, HIST1H3E	AAAAAT	A	0.18	-0.008	0.001	6.80E-09	156,371	-0.010 ⁱ	0.006	0.133	7,703	-0.016	0.014	0.252	1,454
rs11275375	10	21,839,820	SKIDA1, MLLT10	insertion ^j	T	0.32	-0.007	0.001	4.50E-12	156,371	-0.004 ^k	0.005	0.413	7,703	0.011	0.010	0.287	1,454
rs429358	19	45,411,941	APOE, APOC1	C	T	0.15	0.009	0.001	1.90E-09	156,371	0.001	0.006	0.868	7,703	0.009	0.014	0.485	1,454
Overall PA-liking (GIP1)																		
rs17370393	1	190,526,252	BRINP3	T	C	0.15	0.028	0.005	2.12E-08	151,347	-0.007	0.022	0.754	7,724	-0.069	0.051	0.176	1,434
rs7616022	3	85,086,321	CADM2	C	T	0.89	0.040	0.006	3.50E-11	151,347	0.008 ^l	0.016	0.599	7,724	0.032	0.037	0.392	1,434
rs9379831	6	26,175,852	HIST1H2BE, HIST1H4D	A	C	0.31	0.023	0.004	3.55E-09	151,347	0.013	0.017	0.457	7,724	0.094	0.040	0.020	1,434
rs2263321	6	31,340,908	HLA-B, MICA	A	G	0.33	-0.022	0.004	7.00E-09	151,347	-0.006	0.017	0.706	7,724	-0.015	0.040	0.708	1,434
rs9374462	6	97,725,028	MM52L, KLHL32	T	C	0.19	0.029	0.005	4.45E-10	151,347	0.011	0.022	0.621	7,724	0.032	0.047	0.489	1,434
rs11012749	10	21,892,488	SKIDA1, MLLT10	C	A	0.53	-0.021	0.004	6.04E-09	151,347	0.012	0.016	0.460	7,724	-0.030	0.038	0.418	1,434
rs3993050	11	22,548,634	GAS2, FANCF	C	A	0.68	0.022	0.004	4.13E-08	151,347	0.029 ^m	0.030	0.336	7,724	0.050 ⁿ	0.042	0.237	1,434
rs4468472	13	111,528,251	ANKRD10, CAR52	C	T	0.41	-0.022	0.004	6.14E-09	151,347	-0.026	0.016	0.106	7,724	0.001	0.038	0.976	1,434

The beta coefficient units refer to the units used in the scale, from 1 to 9. n = sample size used in analysis; SNP position based on GRCh37.

^aA allele of rs4390955 used as proxy.
^bC allele of rs76508707 used as proxy.
^cA allele of rs2167046 used as proxy.
^dG allele of rs12494446 used as proxy.
^eG allele of rs6660825 used as proxy.
^fA allele of rs7642661 used as proxy.
^gC allele of rs2326128 used as proxy.
^hC allele of rs3136434 used as proxy.
ⁱA allele of rs1417998 used as proxy.
^jTAATAAAAAATTATGAAAAAC.
^kG allele of rs1802669 used as proxy.
^lC allele of rs76508707 used as proxy.
^mA allele of rs3136469 used as proxy.
ⁿG allele of rs7120717 used as proxy.
EAF, effect allele frequency.

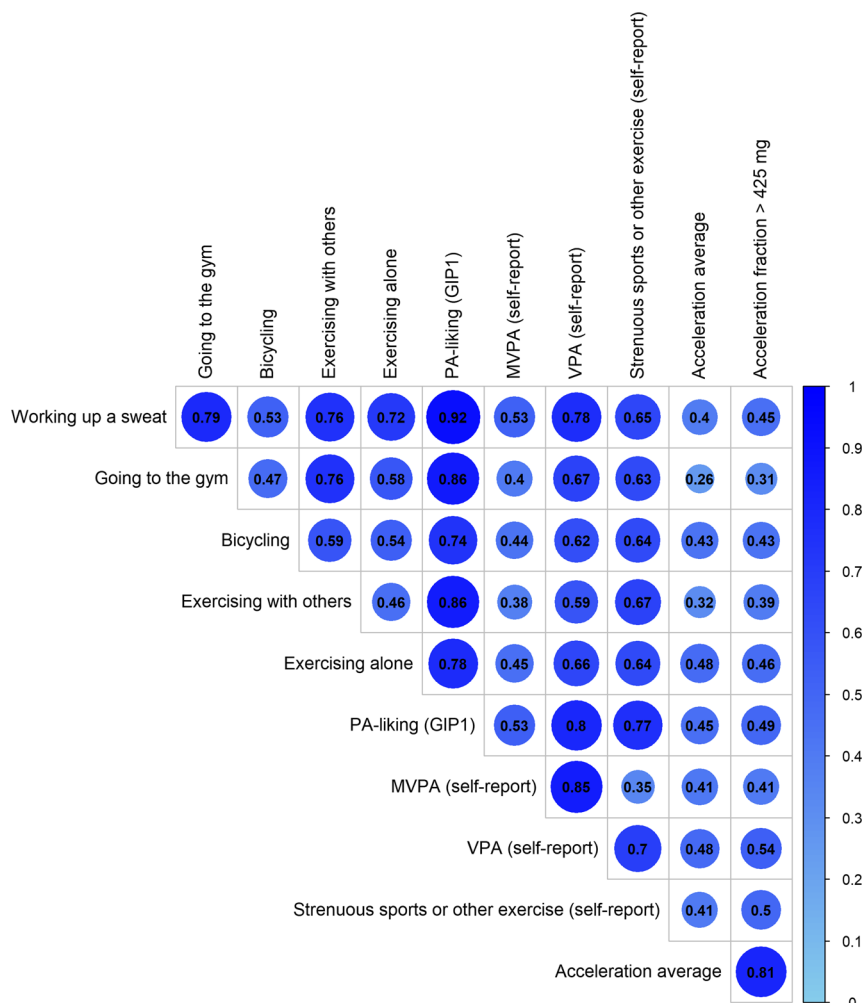


FIGURE 2—Genetic correlations (r_g) across PA from self-report, PA from accelerometry, and PA-liking items including the overall PA-liking trait, using LD score regression with summary statistics from GWAS performed in the UK Biobank.

PGS of liking *going to the gym* (Supplemental Fig. 3, Appendix, Supplemental Digital Content 1, <http://links.lww.com/MSS/C542>). When the best predicting liking PGS (*going to the gym*) was combined with the best predicting self-reported PA PGS (SSOE), the prediction of self-reported PA in NTR improved from $r^2 = 0.42\%$ to $r^2 = 0.59\%$, corresponding to a 40% improvement in prediction, suggesting that these two measures are capturing distinct and complementary components of PA (Supplemental Fig. 4, Appendix, Supplemental Digital Content 1, <http://links.lww.com/MSS/C542>).

DISCUSSION

With the advent of large-scale biobanks with genetic data, we are now able to identify genomic loci associated with complex behavioral and health-related traits such as PA. Here, we broaden and deepen our nascent understanding of the genetics of PA behavior by identifying genetic variants associated with PA-liking. We found some but minimal overlap of loci with those identified for self-reported and accelerometer measures of PA, found genetic correlations with other health-related

traits, including behaviors and health outcomes, and found that PGS of PA-liking adds substantially to the prediction of PA in an independent sample, beyond a prediction based on a self-reported PA PGS.

SNP-based heritabilities of PA-liking (ranging from 5.4% to 7.5%; PA-liking (GIP1) = 8.9%) were generally higher than those of self-reported behaviors at the baseline examination (ranging from 4.6% to 5.6%), but lower than accelerometry measures (14.3% and 11.0%) (8), likely due to lower measurement error of accelerometry. Of all 19 loci identified for PA-liking, only *APOE*, *CADM2*, *HIST1H1D*, and *SKIDA1* have been previously found to be associated with self-reported or accelerometry-measured PA (8,9). However, it is likely that some of the other 20 loci are associated with PA behavior at a less stringent significance threshold. This relatively small degree of overlap likely suggests that these measures of liking are reflecting additional and more specific dimensions of PA such as motivations and perceptions about PA and personality traits, all of which can play a role in the type and amount of PA one engages in. There is also a relatively small degree of overlap of top hits among the PA-liking items, which points to the

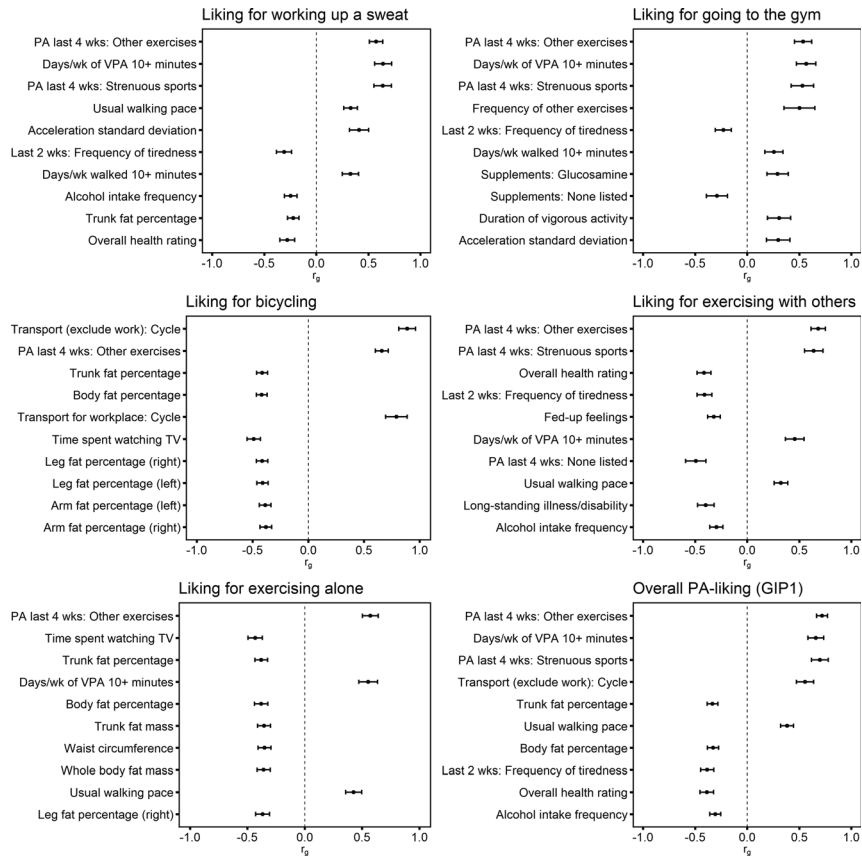


FIGURE 3—Genetic correlations of PA-liking with other traits and diseases assessed through the LDHUB resource. Top 10 genetic correlations according to lowest *P* values among 800 tested phenotypes/diseases are shown. The order of traits shown in each panel from top to bottom is in ascending order of *P* value. Error bars represent 95% CI.

importance and strength of considering multiple facets of PA-liking. It is possible that because accelerometers measure overall PA, and not just purposeful exercise, PA-liking is more strongly genetically correlated with self-reported PA than with accelerometer measures of PA. On the other hand, the relatively low genetic correlation of *going to the gym* with accelerometer-measured PA may reflect limitations of accelerometer measurements in the context of certain types of PA such as resistance exercise. Genetic correlation analyses across a wider set of traits and diseases mainly revealed correlations of PA-liking with self-reported PA and body fat measures. However, several other notable findings such as negative genetic correlations with frequency of tiredness, fed-up feelings, and alcohol intake frequency, and positive associations with usual walking pace, supplement intakes, and variance in accelerometer measurements in UKB, were also observed. Some of these correlations may represent causal effects in one or both directions.

Among the individual loci identified, *CADM2* has previously been identified in GWAS of BMI (38), risk taking, and other behavioral traits (39–41), including PA (8). However, the pattern of association with *CADM2* variants is particularly interesting because alleles associated with higher BMI are associated with higher levels of PA and PA-liking, in the opposite direction of the phenotypic association. As the most consistently identified genetic locus across self-reported PA

and PA-liking measures, we will benefit from future work to understand the molecular mechanisms underlying this association. We found a SNP (rs7934107-T) in another cell adhesion molecule gene, *CADMI*, which was associated with *working up a sweat*. A SNP in the same gene was found to be associated with anorexia nervosa (SNP-risk allele: rs6589488-A) (42). These alleles at these two SNP are positively and moderately correlated ($r^2 = 0.39$).

Other identified loci share associations with several other traits such as social and emotional characteristics (e.g., *MMS22L–KLHL32*), lung function (e.g., *POM121L2–PRSS16*), food/drink intake (e.g., *DARS1–CENPL*), and cognitive traits (e.g., *MDK–CHRM4*). However, our finding of an association of the *APOE* variant with one of the PA-liking items (*exercising alone*) is possibly the result of selection/survival bias due to older individuals with the $\epsilon 4$ risk allele being relatively enriched for healthy behaviors that have offset their genetic risk and enabled their survival and participation in the study (8). It is possible that our estimates for other identified loci are subject to this bias, although it is likely to be minimal.

Although it turns out that we were underpowered to detect statistically significant associations in our replication cohorts, we did observe a nominally significant association of the *CADM2* variant with the liking of *exercising with others* in the NTR, further reinforcing this locus directly or indirectly

in PA behavior, in addition to other traits, as mentioned above. In PGS analyses, we found that these PA-liking measures were genetically consistent across the UK Biobank and NTR samples, and that they could contribute substantially to the prediction and genetic understanding of self-reported PA behavior. It should be noted that the proportion of phenotypic variance explained by these PGS is still extremely small. This is at least partly attributable to the degree of measurement error one would expect from any self-report measure.

The strengths of this study include the relatively large sample size, the multiple measures of PA-liking, the ability to examine correlations with both self-reported and accelerometry-derived PA behaviors, and the availability of two additional cohorts for replication and testing of PGS. The inclusion of middle- to older-age adults of European-descent individuals is a limitation of our study as these results may not generalize to other groups. Furthermore, UK Biobank participants and the subset of them that responded to this questionnaire are not representative of the UK population (43). However, based on previous studies in the UK Biobank, we anticipate that this participation bias will not strongly influence the loci identified at the genome-wide significant level, but it may influence genetic correlations. In this study, we have considered the individual PA-liking items on their own, without consideration of other liking items. Furthermore, although we have genetically derived an overall PA-liking factor using information from the five individual PA-liking items, further work in this area may consider other approaches to capture overall PA-liking.

CONCLUSIONS

In conclusion, we have identified genetic variants associated with PA-liking, further refining our understanding of the genetics of PA behavior. Our results show that PA-liking can capture additional elements of PA not captured by either self-report or accelerometry. Future work is needed to understand the mechanisms linking the identified loci to PA behavior, and how these may vary over the life-course. Examining the genetic correlates of stated liking for different facets of PA and how they associate with self-reported and device-based measures of PA can provide insight into both

the genetic and nongenetic determinants of PA behavior, potentially help plan more effective interventions to improve PA habits, and bridge gaps that may exist between liking and engaging in PA.

This research was conducted using the UK Biobank Resource under Application Number 15678 and 19655. The authors thank the funders, organizers, and participants of the UK Biobank, the Netherlands Twin Register, and TwinsUK. Y. C. K. acknowledges support from the NIH/NHLBI (R01-HL136528). J. F. W. acknowledges support from the MRC Human Genetics Unit quinquennial program grant “QTL in Health and Disease” (MC_UU_00007/10). TwinsUK receives funding from the Wellcome Trust (212904/Z/18/Z), Medical Research Council (AIMHY; MR/M016560/1), and European Union (H2020 contract #733100). TwinsUK and M. M. are supported by the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy’s and St Thomas’ NHS Foundation Trust in partnership with King’s College London. C. M. is funded by the Chronic Disease Research Foundation and by the Medical Research Council (MRC)/British Heart Foundation Ancestry and Biological Informative Markers for Stratification of Hypertension (AIMHY; MR/M016560/1). G. E. A. acknowledges support from the National Institute on Aging, NIH (AG072980, AG067200, AG019610, AG064587), State of Arizona and Arizona Department of Health Services, and McKnight Brain Research Foundation. Phenotyping in NTR was funded by BBRMI-CP2011-38: Enrichment of NTR with information on dietary intake and aspects of eating behavior (PI: Boomsma and Feskens); genotyping in NTR was funded by multiple grants from the Netherlands Organization for Scientific Research (NWO) and The Netherlands Organization for Health Research and Development (ZonMW), the Biobanking and Biomolecular Resources Research Infrastructure (BBMRI-NL, 184.021.007 and 184.033.111), the European Community’s Framework Programs (FP5-LIFE QUALITY-CT-2002-2006, FP7-HEALTH-F4-2007-2013, grant 01254: GenomEUtwin, grant 01413: ENGAGE); the European Research Council (ERC Starting 284167, ERC Consolidator 771057, ERC Advanced 230374), Rutgers University Cell and DNA Repository (NIMH U24 MH068457-06), the National Institutes of Health (NIH, R01D0042157-01A1, MH081802, DA018673, R01 DK092127-04, Grand Opportunity grant 1RC2 MH089951), and the Avera Institute for Human Genetics, Sioux Falls, South Dakota (USA). The results of this study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation. The results of the present study do not constitute endorsement by the American College of Sports Medicine. The authors have no relevant conflicts of interest to disclose.

Y. C. K. and N. P. conceived and designed the study. N. P., Y. C. K., A. A., S. M. W., M. N., V. B., M. M., C. M., M. D. V. Z., J. J. H., E. J. C. dG., and G. A. performed data analyses. Y. C. K. and N. P. drafted the manuscript. Y. C. K., N. P., A. A., S. M. W., M. N., V. B., M. M., C. M., D. A. R., G. E. A., M. D. V. Z., J. J. H., J. F. W., E. J. C. dG., and D. I. B. critically revised the manuscript for important intellectual content.

Access to UK Biobank data is available to registered researchers upon application. GWAS summary statistics data for the UK Biobank PA-liking items and the overall PA-liking trait will be made available through the GWAS catalog portal at the time of publication.

REFERENCES

- Lee I-M, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *Lancet*. 2012;380(9838):219–29.
- Blair SN. Physical inactivity: the biggest public health problem of the 21st century. *Br J Sports Med*. 2009;43(1):1 LP–2.
- van der Ploeg HP, Chey T, Korda RJ, Banks E, Bauman A. Sitting time and all-cause mortality risk in 222 497 Australian adults. *Arch Intern Med*. 2012;172(6):494–500.
- Stubbe JH, Boomsma DI, Vink JM, et al. Genetic influences on exercise participation in 37,051 twin pairs from seven countries. *PLoS One*. 2006;1(1):e22.
- den Hoed M, Brage S, Zhao JH, et al. Heritability of objectively assessed daily physical activity and sedentary behavior. *Am J Clin Nutr*. 2013;98(5):1317–25.
- Franks PW, Ravussin E, Hanson RL, et al. Habitual physical activity in children: the role of genes and the environment. *Am J Clin Nutr*. 2005;82(4):901–8.
- Aaltonen S, Ortega-Alonso A, Kujala UM, Kaprio J. A longitudinal study on genetic and environmental influences on leisure time physical activity in the Finnish twin cohort. *Twin Res Hum Genet*. 2010; 13(5):475–81.
- Klimentidis YC, Raichlen DA, Bea J, et al. Genome-wide association study of habitual physical activity in over 377,000 UK Biobank

- participants identifies multiple variants including CADM2 and APOE. *Int J Obes (Lond)*. 2018;42(6):1161–76.
9. Doherty A, Smith-Byrne K, Ferreira T, et al. GWAS identifies 14 loci for device-measured physical activity and sleep duration. *Nat Commun*. 2018;9(1):5257.
 10. Qi G, Dutta D, Leroux A, et al. Genome-wide association studies of 27 accelerometry-derived physical activity measurements identified novel loci and genetic mechanisms. *Genet Epidemiol*. 2022;46(2):122–38.
 11. Prince SA, Adamo KB, Hamel ME, Hardt J, Connor Gorber S, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. *Int J Behav Nutr Phys Act*. 2008;5:56.
 12. Bauman AE, Reis RS, Sallis JF, Wells JC, Loos RJF, Martin BW. Correlates of physical activity: why are some people physically active and others not? *Lancet*. 2012;380(9838):258–71.
 13. Deci EL, Ryan RM. *Intrinsic Motivation and Self-Determination in Human Behavior*. New York: Plenum Press; 1985.
 14. Ryan RM, Williams GC, Patrick H, Deci EL. Self-determination theory and physical activity: the dynamics of motivation in development and wellness. *Hell J Psychol*. 2009;6(2):107–24.
 15. Teixeira PJ, Carraça EV, Markland D, Silva MN, Ryan RM. Exercise, physical activity, and self-determination theory: a systematic review. *Int J Behav Nutr Phys Act*. 2012;9:78.
 16. Rogers H, Morris T, Moore M. A qualitative study of the achievement goals of recreational exercise participants. *Qual Rep*. 2008;13(4):706–34.
 17. Aaltonen S, Kaprio J, Vuoksimaa E, Huppertz C, Kujala UM, Silventoinen K. Genetic architecture of motives for leisure-time physical activity: a twin study. *Scand J Med Sci Sports*. 2017;27(11):1431–41.
 18. Vink JM, van Hooijdonk KJM, Willemsen G, Feskens EJM, Boomsma DI. Causes of variation in food preference in the Netherlands. *Twin Res Hum Genet*. 2020;23(4):195–203.
 19. Sudlow C, Gallacher J, Allen N, 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*. 2015;12(3):e1001779.
 20. Biobank U. Food preferences web questionnaire [Internet]. 2020. Available from: <http://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/foodpref.pdf>
 21. Duffy VB, Hayes JE, Sullivan BS, Faghri P. Surveying food and beverage liking: a tool for epidemiological studies to connect chemosensation with health outcomes. *Ann N Y Acad Sci*. 2009;1170:558–68.
 22. Deglaire A, Méjean C, Castetbon K, et al. Development of a questionnaire to assay recalled liking for salt, sweet and fat. *Food Qual Prefer*. 2012;23(2):110–24.
 23. Brunelle-Greene D. *The Physical Activity Screening Tool and Liking of Physical Activity in Young Adults*. CT, USA: University of Connecticut Storrs; 2017.
 24. Hubert PA, Mahoney M, Huedo-Medina TB, Leahey TM, Duffy VB. Can assessing physical activity liking identify opportunities to promote physical activity engagement and healthy dietary behaviors? *Nutrients*. 2021;13(10):3366.
 25. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203–9.
 26. Ligthart L, van Beijsterveldt CEM, Kevenaar ST, et al. The Netherlands Twin Register: longitudinal research based on twin and twin-family designs. *Twin Res Hum Genet*. 2019;22(6):623–36.
 27. Verdi S, Abbasian G, Bowyer RCE, et al. TwinsUK: the UK adult twin registry update. *Twin Res Hum Genet*. 2019;22(6):523–9.
 28. Pallister T, Sharafi M, Lachance G, et al. Food preference patterns in a UK twin cohort. *Twin Res Hum Genet*. 2015;18(6):793–805.
 29. Roederer M, Quaye L, Mangino M, et al. The genetic architecture of the human immune system: a bioresource for autoimmunity and disease pathogenesis. *Cell*. 2015;161(2):387–403.
 30. Das S, Forer L, Schönherr S, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;48(10):1284–7.
 31. Jiang L, Zheng Z, Qi T, et al. A resource-efficient tool for mixed model association analysis of large-scale data. *Nat Genet*. 2019;51(12):1749–55.
 32. Bulik-Sullivan BK, Loh PR, Finucane HK, et al, Consortium SWG of the PG. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47(3):291–5.
 33. Tsepilov YA, Freidin MB, Shadrina AS, et al. Analysis of genetically independent phenotypes identifies shared genetic factors associated with chronic musculoskeletal pain conditions. *Commun Biol*. 2020;3(1):329.
 34. Timmers PRHJ, Wilson JF, Joshi PK, Deelen J. Multivariate genomic scan implicates novel loci and haem metabolism in human ageing. *Nat Commun*. 2020;11(1):3570.
 35. Zheng J, Erzurumluoglu AM, Elsworth BL, et al. LD hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics*. 2017;33(2):272.
 36. Finucane HK, Bulik-Sullivan B, Gusev A, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet*. 2015;47(11):1228–35.
 37. Vilhjálmsson BJ, Yang J, Finucane HK, et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am J Hum Genet*. 2015;97(4):576–92.
 38. Speliotes EK, Willer CJ, Berndt SI, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet*. 2010;42(11):937–48.
 39. Boutwell B, Hinds D, Tielbeek J, Ong KK, Day FR, Perry JRB. Replication and characterization of CADM2 and MSRA genes on human behavior. *Heliyon*. 2017;3(7):e00349.
 40. Day FR, Helgason H, Chasman DI, et al. Physical and neurobehavioral determinants of reproductive onset and success. *Nat Genet*. 2016;48(6):617–23.
 41. Strawbridge RJ, Ward J, Cullen B, et al. Genome-wide analysis of self-reported risk-taking behaviour and cross-disorder genetic correlations in the UK Biobank cohort. *Transl Psychiatry*. 2018;8(1):39.
 42. Watson HJ, Yilmaz Z, Thornton LM, et al. Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. *Nat Genet*. 2019;51(8):1207–14.
 43. Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am J Epidemiol*. 2017;186(9):1026–34.