

## Genome-wide association study of liking of physical activity in the UK Biobank

Yann C. Klimentidis <sup>1\*</sup>, Michelle Newell <sup>1</sup>, Matthijs D. van der Zee <sup>2,3</sup>, Victoria L. Bland <sup>4</sup>, Sebastian May-  
Wilson <sup>5</sup>, Cristina Menni <sup>6</sup>, Massimo Mangino <sup>6,7</sup>, Amit Arora <sup>1</sup>, David A. Raichlen <sup>8</sup>, Gene E. Alexander  
<sup>9,10,11,12</sup>, James F. Wilson <sup>5,13</sup>, Dorret I. Boomsma <sup>2,3</sup>, Jouke-Jan Hottenga <sup>2,3</sup>, Eco J.C. de Geus <sup>2,3</sup>, Nicola  
Pirastu <sup>5\*</sup>

<sup>1</sup> Department of Epidemiology and Biostatistics, Mel and Enid Zuckerman College of Public Health,  
University of Arizona, Tucson, AZ, USA

<sup>2</sup> Department of Biological Psychology, Vrije Universiteit Amsterdam, The Netherlands.

<sup>3</sup> Netherlands Twin Register, The Netherlands.

<sup>4</sup> Division of Geriatric Medicine, School of Medicine, University of Colorado Anschutz Medical Campus,  
Aurora, CO 80045 USA

<sup>5</sup> Centre for Global Health Research, Usher Institute, University of Edinburgh, Edinburgh, United Kingdom.

<sup>6</sup> Department of Twin Research and Genetic Epidemiology, King's College London, London SE1 7EH, UK

<sup>7</sup> NIHR Biomedical Research Centre at Guy's and St Thomas' Foundation Trust, London SE1 9RT, UK

<sup>8</sup> Human and Evolutionary Biology Section, Department of Biological Sciences, University of Southern  
California, CA, USA

<sup>9</sup> Department of Psychology and Psychiatry, University of Arizona, Tucson, AZ, USA.

<sup>10</sup> Neuroscience and Physiological Sciences Graduate Inter-Disciplinary Programs, University of Arizona,  
Tucson, AZ, USA

<sup>11</sup> Evelyn F. McKnight Brain Institute, University of Arizona, Tucson, AZ, USA

<sup>12</sup> Arizona Alzheimer's Consortium, AZ, USA

24 <sup>13</sup> MRC Human Genetics Unit, Institute of Genetic and Molecular Medicine, University of Edinburgh,  
25 Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, Scotland

26

27 \* Department of Epidemiology and Biostatistics, Mel and Enid Zuckerman College of Public Health,  
28 University of Arizona, Tucson, AZ, USA; [yann@email.arizona.edu](mailto:yann@email.arizona.edu)

29 Centre for Global Health Research, Usher Institute for Population Health Sciences and Informatics,  
30 University of Edinburgh, Edinburgh, United Kingdom; [nicola.pirastu@ed.ac.uk](mailto:nicola.pirastu@ed.ac.uk)

31

32 **Keywords:** genome-wide association study, physical activity, liking, preferences, genetic, exercise

## 33 ABSTRACT

34 A lack of physical activity (PA) is one of the most pressing health issues facing society today. Our individual  
35 propensity for PA is partly influenced by genetic factors. Stated liking of various PA behaviors may capture  
36 additional dimensions of PA behavior that are not captured by other measures, and contribute to our  
37 understanding of the genetics of PA behavior. Here, in over 157,000 individuals from the UK Biobank, we  
38 sought to complement and extend previous findings on the genetics of PA behavior by performing genome-  
39 wide association studies of self-reported liking of several PA-related behaviors plus an additional derived  
40 trait of overall PA-liking. We identified a total of 19 unique genome-wide significant loci across all traits,  
41 only four of which overlap with loci previously identified for PA behavior. The PA-liking traits were  
42 genetically correlated with self-reported ( $r_g$ : 0.38 – 0.80) and accelerometry-derived ( $r_g$ : 0.26 – 0.49) PA  
43 measures, and with a wide range of health-related traits and dietary behaviors. Replication in the Netherlands  
44 Twin Register (NTR;  $n > 7,300$ ) and the TwinsUK ( $n > 1,300$ ) study revealed directionally consistent  
45 associations. Polygenic risk scores (PRS) were then trained in UKB for each PA-liking trait and for self-  
46 reported PA behavior. The PA-liking PRS significantly predicted the same liking trait in NTR. The PRS for  
47 liking of *going to the gym* predicted PA behavior in NTR ( $r^2 = 0.40\%$ ) nearly as well as the one constructed  
48 based on self-reported PA behavior ( $r^2 = 0.42\%$ ). Combining the two PRS into a single model increased the  
49  $r^2$  to 0.59%, suggesting that although these PRS correlate with each other, they are also capturing distinct  
50 dimensions of PA behavior. In conclusion, we have identified the first loci associated with PA-liking, and  
51 extended and refined our understanding of the genetic basis of PA behavior.

## 52 INTRODUCTION

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

60 Several loci have already been identified as being associated with self-reported and accelerometry-  
61 measured levels of PA<sup>8,9</sup>. These measures may each be limited in multiple ways. For example, self-reported  
62 measures may be highly influenced by social- and health-related pressures and may not be stable over time,  
63 while accelerometer measures may only be sensitive to certain types of PA, with wear-time often limited to a  
64 single week of a person's lifetime and likely influencing behavior<sup>10</sup>. Measures of an individual's liking of  
65 PA may more accurately capture overall life-long propensity to engage in PA, and at a minimum, serve as  
66 complementary, broader, and refined measures of PA behavior.

67 To discover genetic loci associated with PA-related liking, we performed GWAS of five individual  
68 liking traits and one composite trait in over 157,000 UK Biobank participants, with replication of top loci in  
69 the Netherlands Twin Register (NTR) and TwinsUK studies. We then examined how these traits genetically  
70 relate to self-reported and accelerometry-measured PA<sup>8</sup>, and to a wide range of other traits and health  
71 outcomes. Finally, we examined how polygenic scores (PGS) of PA-liking derived from the UK Biobank  
72 predicted both liking and self-reported PA in the NTR study.

73

74 **METHODS**

75 *UK Biobank*

76 The UK Biobank is a prospective cohort study of 500,000 adults (ages 37-73 at the baseline  
77 examination in 2006-2010) from the UK <sup>11</sup>. All participants provided written informed consent, and ethical  
78 approval for this study was granted. Ethical approval for the UK Biobank study was obtained by the National  
79 Information Governance Board for Health and Social Care, and the National Health Service North West  
80 Multicenter Research Ethics Committee.

81

82 *Physical activity liking*

83 In 2019, a link to a questionnaire was sent by email to UK Biobank participants to assess liking of  
84 specific foods as well as physical activities <sup>12</sup>. This online questionnaire consisted of 150 items, five of which  
85 were related to PA (*going to the gym, working up a sweat, exercising with others, exercising alone, and*  
86 *bicycling*), and assesses liking through a 9-point hedonic scale ranging from 1 for extremely dislike up to 9  
87 for extremely like, in increments of one. The same items were asked in the two replication studies (see  
88 below).

89

90 *Genetic markers*

91 Genotypes in the UK Biobank were measured with the Affymetrix UK Biobank Axiom Array (Santa  
92 Clara, CA, USA) in 90% of participants. The remainder (10%) were genotyped with the Affymetrix UK  
93 BiLEVE Axiom Array. Further details about imputation, principal components analysis, and QC procedures  
94 can be found elsewhere <sup>13</sup>.

95

96 *Replication of sentinel SNPs in Netherlands Twin Register (NTR) and TwinsUK*

97 The NTR is a longitudinal register of twins and their relatives <sup>14</sup>. Between December 2014 and May  
98 2017, participants responded to the same liking questionnaire as the one administered in the UK Biobank,  
99 including the same 5 PA-related questions <sup>15</sup>. Details regarding genotyping are provided in Supplementary  
100 Methods. Top genome-wide significant SNPs or their proxies from the UK Biobank GWASs were  
101 interrogated for each respective trait in NTR GWAS results. Since we tested 25 SNPs/loci, a replication was  
102 deemed successful if the p-value was  $< 0.002$ . Polygenic scores (PGS) were calculated in NTR with the  
103 LDpred package <sup>16</sup>, based on UKB GWAS results for PA-liking and for SSOE (strenuous sports or other  
104 exercise) and other PA behavior measures <sup>8</sup>, and were tested for correlation with PA-liking phenotypes and  
105 self-reported PA behavior in NTR (see Supplementary Methods for more details).

106 TwinsUK is a large twin registry for the study of health that began in 1992 <sup>17</sup>. The same liking  
107 questionnaire used in UK Biobank was previously used in the TwinsUK cohort <sup>18</sup>. TwinsUK genotype has  
108 been previously described in detail <sup>19</sup>. Briefly, TwinsUK samples were genotyped with a combination of two  
109 Illumina arrays (HumanHap300, HumanHap610Q). After the Genotype QC stage, the samples from the two  
110 arrays were combined and the imputations were performed using the Michigan Imputation Server <sup>20</sup> using  
111 the 1000 Genomes Phase3 v5 reference panel.

112

113 *Statistical analyses*

114 To assess associations of PA liking with sex, age, body mass index (BMI), income, University/College  
115 degree (yes/no), and Townsend Deprivation Index, we performed linear regression after ensuring normality  
116 and homoscedasticity of residuals. In GWAS, we included only individuals of European descent. We

117 considered participants as being of European descent if they were either among the genetically British as  
118 defined by UK Biobank or self-identified as “Irish”, “White”, or “Any other White background”. We  
119 performed GWAS with fastGWA<sup>21</sup>, which implements a mixed-effect linear regression that controls for  
120 population stratification and relatedness. We included age at time of questionnaire, sex, genotyping chip,  
121 batch, and the first 10 genetic principal components, as covariates. We used LD-score regression to assess  
122 test score inflation, SNP-based heritability, and to assess genetic correlations among PA-liking traits and  
123 previously reported PA traits<sup>22</sup>. To obtain a measure of overall PA liking, we also derived a GWAS of the  
124 first principal component (PC) derived through the genetically-independent phenotype (GIP) method<sup>23,24</sup>  
125 and starting from the genetic correlation matrix to derive the loadings of each trait on each GIP. Independent  
126 significant loci were identified as those with  $p < 5 \times 10^{-8}$  with  $r^2 < 0.1$ , and  $> 250$  kb distance. The online  
127 LDHUB platform<sup>25</sup> was used to examine genetic correlations of the five individual liking traits and the  
128 overall PA-liking (GIP1) trait with a wide range of traits and diseases (~ 800 phenotypes). We used stratified  
129 LD score regression to identify cell type-specific enrichment of heritability<sup>26</sup> from the overall PA-liking  
130 (GIP1) GWAS.

131

## 132 RESULTS

### 133 *PA-liking phenotypes*

134 Descriptive statistics of UKB, NTR, and TwinsUK samples and mean PA-liking levels are shown in  
135 **Supplementary Table 1**. Mean score of liking was highest for *exercising with others* in the two UK  
136 samples, and highest for *bicycling* in NTR. It was lowest for *going to the gym* in all three samples. In the UK  
137 Biobank, PA liking was negatively correlated with age, and was higher in males (except for *exercising with*  
138 *others*). We observed generally positive correlations of PA-liking traits with education and income, and

139 negative correlations with Townsend Deprivation Index (e.g. *exercising with others*), such that less  
140 deprivation was associated with more PA-liking (**Supplementary Table 2**).

141

#### 142 *Genome-wide association study*

143 Data from up to 158,189 UK Biobank participants (mean age=66.8; 57% female) were analyzed for  
144 GWAS. SNP-based trait heritabilities varied from 0.054 (0.004) for *going to the gym*, up to 0.075 (0.004) for  
145 *bicycling*. SNP heritability for overall PA-liking (GIP1) was 0.089 (0.004) (**Supplementary Table 3**). We  
146 did not observe evidence of genomic inflation beyond that explained by polygenic signal according to LD-  
147 score regression estimates (**Supplementary Table 2**). Between 2 and 6 genome-wide significant loci were  
148 identified for each individual liking trait, and 8 loci for overall PA-liking (GIP1), for a total of 25 SNP-trait  
149 associations in 19 loci (**Figure 1 and Table 1**). We did not observe a large degree of overlap of top loci  
150 across the five different liking traits. Only two loci (*CADM2* and the locus on chromosome 11) were  
151 significantly associated with more than one trait. Sentinel SNPs were also associated with other traits such as  
152 with bone mineral density, body size and body composition, educational attainment, respiratory traits,  
153 psychiatric traits, as well as other PA-related traits such as usual walking pace and time watching television  
154 (**Supplementary spreadsheet**). Cell-type enrichment analysis via stratified LD score regression identified  
155 the nucleus accumbens, hippocampus, caudate, frontal cortex, and amygdala cell types (**Supplementary**  
156 **Table 4**).

157 In the NTR replication, we found directional consistency for 19 out of 25 SNPs, including 6 out of 8  
158 PA-liking (GIP1) SNPs (**Table 1**). Only one SNP (*CADM2*; for *exercising with others*) was nominally  
159 significant ( $p < 0.05$ ) before multiple testing correction. In the TwinsUK replication, we found directional  
160 consistency for 14 out of 25 SNPs, including 6 out of 8 overall PA-liking (GIP1) SNPs (**Table 1**).



161

162 *Genetic correlations*

163 Genetic correlations among PA liking traits were moderate-to-strong (**Figure 2**). Among the PA  
164 liking traits, strong correlations were observed for *working up a sweat* with *going to the gym* ( $r_g=0.79$ ),  
165 *exercising with others* ( $r_g=0.76$ ), and *exercising alone* ( $r_g=0.72$ ), and the weakest correlation was a positive  
166 one between *exercising alone* and *exercising with others* ( $r_g=0.46$ ). Across liking and behavior (self-reported  
167 PA and accelerometer-derived PA) traits, correlations were strongest for PA-liking with self-reported  
168 vigorous physical activity and strenuous sports and other exercise ( $r_g$  between 0.59 and 0.78). Genetic  
169 correlations of PA-liking with accelerometry traits were generally strongest for *exercising alone* ( $r_g\approx 0.47$ ),  
170 and weakest with *going to the gym* ( $r_g\approx 0.28$ ). Genetic correlation assessments with a wide range of traits and  
171 diseases reveal correlations with UK Biobank variables related to physical activity including accelerometry,  
172 as well as with obesity-related traits, tiredness, and lifestyle traits such as alcohol consumption, TV  
173 watching, and taking dietary supplements, among others (**Figure 3**).

174

175 *PRS analyses in NTR*

176 PRSs of each PA-liking trait were calculated for each NTR participant, using as weights the effect  
177 sizes resulting from the GWAS of the UK Biobank data. These PRSs were generally, but not always, most  
178 strongly associated with the corresponding PA-liking phenotype (**Supplementary Table 4**). For example,  
179 the *exercising with others* PRS was most strongly associated with *exercising with others* in NTR ( $r^2=0.79\%$ ,  
180  $p=2.7 \times 10^{-13}$ ). The overall PA-liking (GIP1) PRS was most strongly correlated with liking of *exercising*  
181 *alone* ( $r^2=0.83\%$ ,  $p=7.0 \times 10^{-15}$ ; **Supplementary Table 4**). These PRSs were also significantly associated  
182 with self-reported PA behavioral phenotypes in NTR: self-reported *total exercise*, *team-based exercise*, and

183 *solitary exercise*. Among the PA-liking PRSs, the best predictors of these self-reported PA phenotypes were  
184 *going to the gym*, and *exercising with others* in association with the self-reported PA measures of *total*  
185 *exercise* and *solitary exercise* (but not with *team-based exercise*; **Supplementary Figure 1**). When  
186 compared with the PRS based on the self-reported and accelerometer PA GWAS in UK Biobank, the liking  
187 PRSs had a similar predictive performance. For example, a PRS of self-reported SSOE (strenuous sports or  
188 other exercise) predicted self-report *total exercise* in NTR only slightly better than a PRS of liking *going to*  
189 *the gym* (**Supplementary Figure 1**). When the best predicting liking PRS (*going to the gym*) was combined  
190 with the best predicting self-reported PA PRS (SSOE), the prediction of self-reported PA in NTR improved  
191 from  $r^2=0.42\%$  to  $r^2=0.59\%$ , corresponding to a 40% improvement in prediction, suggesting that these two  
192 measures are capturing distinct and complementary components of PA (**Supplementary Figure 2**).

193

## 194 **DISCUSSION**

195 With the advent of large-scale biobanks with genetic data, we are now able to identify genomic loci  
196 associated with complex behavioral and health-related traits such as PA. Here, we broaden and deepen our  
197 nascent understanding of the genetics of PA behavior by identifying genetic variants associated with PA-  
198 liking. We found some but minimal overlap of loci with those identified for self-reported and accelerometer  
199 measures of PA, found genetic correlations with other health-related traits, including behaviors and health  
200 outcomes, and find that PRS of PA-liking adds substantially to prediction of PA in an independent sample,  
201 beyond a prediction based on a self-reported PA PRS.

202 SNP-based heritabilities of PA-liking traits (ranging from 5.4 to 7.5 %; PA-liking (GIP1) = 8.9 %)   
203 were generally higher than those of self-reported behaviors at the baseline exam (ranging from 4.6 to 5.6 %),  
204 but lower than accelerometry measures (14.3 and 11.0 %) <sup>8</sup>, likely due to lower measurement error of

205 accelerometry. Of all 24 loci identified for liking traits, only *APOE*, *CADM2*, *HIST1H1D*, and *SKIDA1* were  
206 significantly associated with self-reported or accelerometry-measured PA<sup>8,9</sup>. However, it is likely that some  
207 of the other 20 loci are associated with PA behavior at a less stringent significance threshold. This relatively  
208 small degree of overlap likely suggests that these measures of liking are reflecting different and specific  
209 facets of PA, and attitudes and perceptions about PA, as opposed to PA behavior assessed by self-report or  
210 accelerometry. It is possible that since accelerometers measure overall PA, and not just purposeful exercise,  
211 the PA-liking traits are more strongly genetically correlated with self-reported PA than with accelerometer  
212 measures of PA. On the other hand, the relatively low genetic correlation of *going to the gym* with  
213 accelerometer-measured PA may reflect limitations of accelerometer measurements in the context of certain  
214 types of PA such as resistance exercise. Genetic correlation analyses across a wider set of traits and diseases  
215 mainly revealed correlations of PA-liking with self-reported PA and body fat measures. However, several  
216 other notable findings such as negative genetic correlations with frequency of tiredness, fed-up feelings, and  
217 alcohol intake frequency, and positive associations with usual walking pace, supplement intakes, and  
218 variance in accelerometer measurements in UKB, were also observed. Some of these correlations may  
219 represent causal effects in one or both directions.

220         Among the individual loci identified, *CADM2* has previously been identified in GWAS of BMI<sup>27</sup>,  
221 risk taking and other behavioral traits<sup>28-30</sup>, including PA<sup>8</sup>. However, the pattern of association with *CADM2*  
222 variants is particularly interesting since alleles associated with higher BMI are associated with *higher* levels  
223 of PA and PA-liking, in the opposite direction of the phenotypic association. As the most consistently  
224 identified genetic locus across self-reported PA and PA-liking measures, we will benefit from future work to  
225 understand the molecular mechanisms underlying this association. We found a SNP (rs7934107-T) in  
226 another cell adhesion molecule gene, *CADMI*, which was associated with *working up a sweat*. A SNP in the

227 same gene was found to be associated with anorexia nervosa (SNP-risk allele: rs6589488-A)<sup>31</sup>. These alleles  
228 at these two SNPs are positively and moderately correlated ( $r^2=0.39$ ).

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

236 Although it turns out that we were under-powered to detect statistically significant associations in our  
237 replication cohorts, we did observe a nominally significant association of the *CADM2* variant with liking of  
238 *exercising with others* in the NTR, further reinforcing this locus directly or indirectly in PA behavior, in  
239 addition to other traits, as mentioned above. In PRS analyses, we found that these PA-liking measures were  
240 genetically consistent across the UK Biobank and NTR samples, and that they could contribute substantially  
241 to the prediction and genetic understanding of (at least) self-reported PA behavior. It should be noted that the  
242 proportion of phenotypic variance explained by these PRS is still extremely small. This is at least partly  
243 attributable to the degree of measurement error one would expect from any self-report measure.

244 The strengths of this study include the relatively large sample size, the multiple measures of PA-  
245 liking, the ability to examine correlations with both self-reported and accelerometry-derived PA behaviors,  
246 and the availability of two additional cohorts for replication and testing of PRS. The inclusion of middle-to-  
247 older-age adults of European-descent individuals is a limitation of our study as these results may not  
248 generalize to other groups.

249 In conclusion, we have identified genetic variants associated with PA-liking, further refining our  
250 understanding of the genetics of PA behavior. Our results show that PA-liking can capture additional  
251 elements of PA not captured by either self-report or accelerometry. Future work is needed to understand the  
252 mechanisms linking the identified loci to PA behavior, and how these may vary over the life-course.  
253 Examining the genetic correlates of stated liking for different facets of PA and how they associate with self-  
254 reported and device-based measures of PA can provide insight into both the genetic and non-genetic  
255 determinants of PA behavior, potentially help plan more effective interventions to improve PA habits, and  
256 bridge gaps that may exist between liking and engaging in PA.

257

## 258 **ACKNOWLEDGMENTS**

259 This research was conducted using the UK Biobank Resource under Application Number 15678 and  
260 19655. The authors would like to thank the funders, organizers and participants of the UK Biobank, the  
261 Netherlands Twin Register, and TwinsUK. J.F.W. acknowledges support from the MRC Human Genetics  
262 Unit quinquennial programme grant “QTL in Health and Disease” (MC\_UU\_00007/10). TwinsUK receives  
263 funding from the Wellcome Trust (212904/Z/18/Z), Medical Research Council (AIMHY; MR/M016560/1)  
264 and European Union (H2020 contract #733100). TwinsUK and M.M. are supported by the National Institute  
265 for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research  
266 Centre based at Guy’s and St Thomas’ NHS Foundation Trust in partnership with King’s College London.  
267 C.M. is funded by the Chronic Disease Research Foundation and by the Medical Research Council  
268 (MRC)/British Heart Foundation Ancestry and Biological Informative Markers for Stratification of  
269 Hypertension (AIMHY; MR/M016560/1). G.E.A. acknowledges support from the National Institute on  
270 Aging, NIH (AG072980, AG067200, AG064587), State of Arizona and Arizona Department of Health  
271 Services, and McKnight Brain Research Foundation. Phenotyping in NTR was funded by BBRMI-CP2011-

272 38: Enrichment of NTR with information on dietary intake and aspects of eating behavior (PI: Boomsma &  
273 Feskens); genotyping in NTR was funded by multiple grants from the Netherlands Organization for  
274 Scientific Research (NWO) and The Netherlands Organisation for Health Research and Development  
275 (ZonMW); the Biobanking and Biomolecular Resources Research Infrastructure (BBMRI –NL, 184.021.007  
276 and 184.033.111), the European Community's Framework Programs (FP5- LIFE QUALITY-CT-2002-2006,  
277 FP7- HEALTH-F4-2007-2013, grant 01254: GenomEUtwin, grant 01413: ENGAGE); the European  
278 Research Council (ERC Starting 284167, ERC Consolidator 771057, ERC Advanced 230374), Rutgers  
279 University Cell and DNA Repository (NIMH U24 MH068457-06), the National Institutes of Health (NIH,  
280 R01D0042157-01A1, MH081802, DA018673, R01 DK092127-04, Grand Opportunity grant 1RC2  
281 MH089951); and the Avera Institute for Human Genetics, Sioux Falls, South Dakota (USA).

282

## 283 **AUTHOR CONTRIBUTIONS**

284       Conceived and designed study: YCK & NP. Performed data analyses: NP, YCK, AA, SMW, MN,  
285 VB, MM, CM, MDVZ, JJH & EJCdG. Drafting of manuscript: YCK & NP. Critically revised the manuscript  
286 for important intellectual content: YCK, NP, AA, SMW, MN, VB, MM, CM, DAR, GEA, MDVZ, JJH,  
287 JFW, EJCdG & DIB.

288

## 289 **DATA AVAILABILITY**

290       Access to UK Biobank data is available to registered researchers upon application. GWAS summary  
291 statistics data for the UK Biobank PA-liking traits will be made available through the GWAS catalog portal  
292 at the time of publication.

## 293 REFERENCES

- 294
- 295 1. Lee, I.-M. *et al.* Effect of physical inactivity on major non-communicable diseases worldwide: an  
296 analysis of burden of disease and life expectancy. *Lancet (London, England)* **380**, 219–229 (2012).
  - 297 2. Blair, S. N. Physical inactivity: the biggest public health problem of the 21st century. *Br. J. Sports*  
298 *Med.* **43**, 1 LP – 2 (2009).
  - 299 3. van der Ploeg, H. P., Chey, T., Korda, R. J., Banks, E. & Bauman, A. Sitting Time and All-Cause  
300 Mortality Risk in 222 497 Australian Adults. *Archives of Internal Medicine* vol. 172 494–500 (2012).
  - 301 4. Stubbe, J. H. *et al.* Genetic influences on exercise participation in 37,051 twin pairs from seven  
302 countries. *PLoS ONE* vol. 1 (2006).
  - 303 5. den Hoed, M. *et al.* Heritability of objectively assessed daily physical activity and sedentary behavior.  
304 *The American journal of clinical nutrition* vol. 98 1317–1325 (2013).
  - 305 6. Franks, P. W. *et al.* Habitual physical activity in children: the role of genes and the environment. *Am J*  
306 *Clin Nutr* (2005).
  - 307 7. Aaltonen, S., Ortega-Alonso, A., Kujala, U. M. & Kaprio, J. A longitudinal study on genetic and  
308 environmental influences on leisure time physical activity in the Finnish Twin Cohort. *Twin Res.*  
309 *Hum. Genet.* (2010) doi:10.1375/twin.13.5.475\r10.1375/twin.13.5.475 [pii].
  - 310 8. Klimentidis, Y. C. *et al.* Genome-wide association study of habitual physical activity in over 377,000  
311 UK Biobank participants identifies multiple variants including CADM2 and APOE. *Int. J. Obes.* **42**,  
312 (2018).
  - 313 9. Doherty, A. *et al.* GWAS identifies 14 loci for device-measured physical activity and sleep duration.  
314 *Nat. Commun.* **9**, 5257 (2018).
  - 315 10. Prince, S. A. *et al.* A comparison of direct versus self-report measures for assessing physical activity  
316 in adults: a systematic review. *Int. J. Behav. Nutr. Phys. Act.* **5**, 56 (2008).
  - 317 11. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range  
318 of Complex Diseases of Middle and Old Age. *PLoS Med.* **12**, (2015).
  - 319 12. Biobank, U. Food preferences web questionnaire.  
320 <http://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/foodpref.pdf> (2020).
  - 321 13. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**,  
322 203–209 (2018).
  - 323 14. Ligthart, L. *et al.* The Netherlands Twin Register: Longitudinal Research Based on Twin and Twin-  
324 Family Designs. *Twin Res. Hum. Genet. Off. J. Int. Soc. Twin Stud.* **22**, 623–636 (2019).
  - 325 15. Vink, J. M., van Hooijdonk, K. J. M., Willemsen, G., Feskens, E. J. M. & Boomsma, D. I. Causes of  
326 Variation in Food Preference in the Netherlands. *Twin Res. Hum. Genet. Off. J. Int. Soc. Twin Stud.*  
327 **23**, 195–203 (2020).
  - 328 16. Vilhjálmsson, B. J. *et al.* Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk  
329 Scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
  - 330 17. Verdi, S. *et al.* TwinsUK: The UK Adult Twin Registry Update. *Twin Res. Hum. Genet. Off. J. Int.*  
331 *Soc. Twin Stud.* **22**, 523–529 (2019).
  - 332 18. Pallister, T. *et al.* Food Preference Patterns in a UK Twin Cohort. *Twin Res. Hum. Genet. Off. J. Int.*  
333 *Soc. Twin Stud.* **18**, 793–805 (2015).
  - 334 19. Roederer, M. *et al.* The genetic architecture of the human immune system: a bioresource for  
335 autoimmunity and disease pathogenesis. *Cell* **161**, 387–403 (2015).
  - 336 20. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287  
337 (2016).
  - 338 21. Jiang, L. *et al.* A resource-efficient tool for mixed model association analysis of large-scale data. *Nat.*

- 339 *Genet.* **51**, 1749–1755 (2019).
- 340 22. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in  
341 genome-wide association studies. *Nat Genet* **47**, 291–295 (2015).
- 342 23. Tsepilov, Y. A. *et al.* Analysis of genetically independent phenotypes identifies shared genetic factors  
343 associated with chronic musculoskeletal pain conditions. *Commun. Biol.* **3**, 329 (2020).
- 344 24. Timmers, P. R. H. J., Wilson, J. F., Joshi, P. K. & Deelen, J. Multivariate genomic scan implicates  
345 novel loci and haem metabolism in human ageing. *Nat. Commun.* **11**, 3570 (2020).
- 346 25. Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that  
347 maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation  
348 analysis. *Bioinformatics* **33**, 272 (2017).
- 349 26. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association  
350 summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
- 351 27. Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with  
352 body mass index. *Nat Genet* **42**, 937–948 (2010).
- 353 28. Boutwell, B. *et al.* Replication and characterization of CADM2 and MSRA genes on human behavior.  
354 *Heliyon* **3**, e00349 (2017).
- 355 29. Day, F. R. *et al.* Physical and neurobehavioral determinants of reproductive onset and success. *Nat*  
356 *Genet* **48**, 617–623 (2016).
- 357 30. Strawbridge, R. J. *et al.* Genome-wide analysis of self-reported risk-taking behaviour and cross-  
358 disorder genetic correlations in the UK Biobank cohort. *Transl. Psychiatry* **8**, 39 (2018).
- 359 31. Watson, H. J. *et al.* Genome-wide association study identifies eight risk loci and implicates metabo-  
360 psychiatric origins for anorexia nervosa. *Nat. Genet.* **51**, 1207–1214 (2019).
- 361



**362 Table 1:** List of genetic variants associated with physical activity (PA) liking and overall PA-liking (GIP1)  
**363** in UKB, with replication results from NTR and Twins UK.

SNP	UK Biobank										NTR replication				Twins UK replication			
	Chr	Position	Nearest gene(s)	Effect allele	Other allele	EAF	Beta	SE	p-value	n	Beta	SE	P-value	n	Beta	SE	P-value	n
<i>Going to gym</i>																		
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
<i>Working up a sweat</i>																		
rs4645151	3	18,730,433	SATB1, TBC1D5	T	C	0.69	-0.006	0.001	8.90E-09	158,189	-0.008 <sup>a</sup>	0.004	0.057	7,874	0.001	0.011	0.898	1,471
rs112403434	7	114,528,614	MDFIC	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
<i>Exercising with others</i>																		
rs25571	1	173,773,928	DARSI, CENPL	CTT	C	0.53	0.006	0.001	4.40E-08	153,358	0.004 <sup>b</sup>	0.004	0.383	7,738	-0.001	0.009	0.883	1,456
rs4672114	2	56,614,299	CCDC85A, EFEMP1	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 <sup>c</sup>	0.004	0.048	7,738	0.010 <sup>j</sup>	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 <sup>d</sup>	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
<i>Exercising alone</i>																		
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	AAAAT	A	0.18	-0.008	0.001	6.80E-09	156,371	-0.010 <sup>e</sup>	0.006	0.133	7,703	-0.016	0.014	0.252	1,454
rs11275375	10	21,839,820	SKIDA1, MLLT10	insertion <sup>*</sup>	T	0.32	-0.007	0.001	4.50E-12	156,371	-0.004 <sup>f</sup>	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
<i>Bicycling</i>																		
rs5850689	3	85,586,903	CADM2	T	TG	0.63	-0.008	0.001	1.00E-10	148,803	-0.004 <sup>g</sup>	0.003	0.188	7,936	-5.0E-04 <sup>k</sup>	0.011	0.965	1,424
rs16897515	6	27,278,020	POM12L2, PRSS16	C	A	0.19	-0.008	0.001	3.20E-08	148,803	0.001	0.004	0.886	7,936	0.014	0.014	0.303	1,424
<i>PA-liking (GIP1)</i>																		
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 <sup>h</sup>	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	MMS22L, 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 <sup>i</sup>	0.030	0.336	7,724	0.050 <sup>l</sup>	0.042	0.237	1,434
rs4468472	13	111,528,251	ANKRD10, CARS2	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

**364**

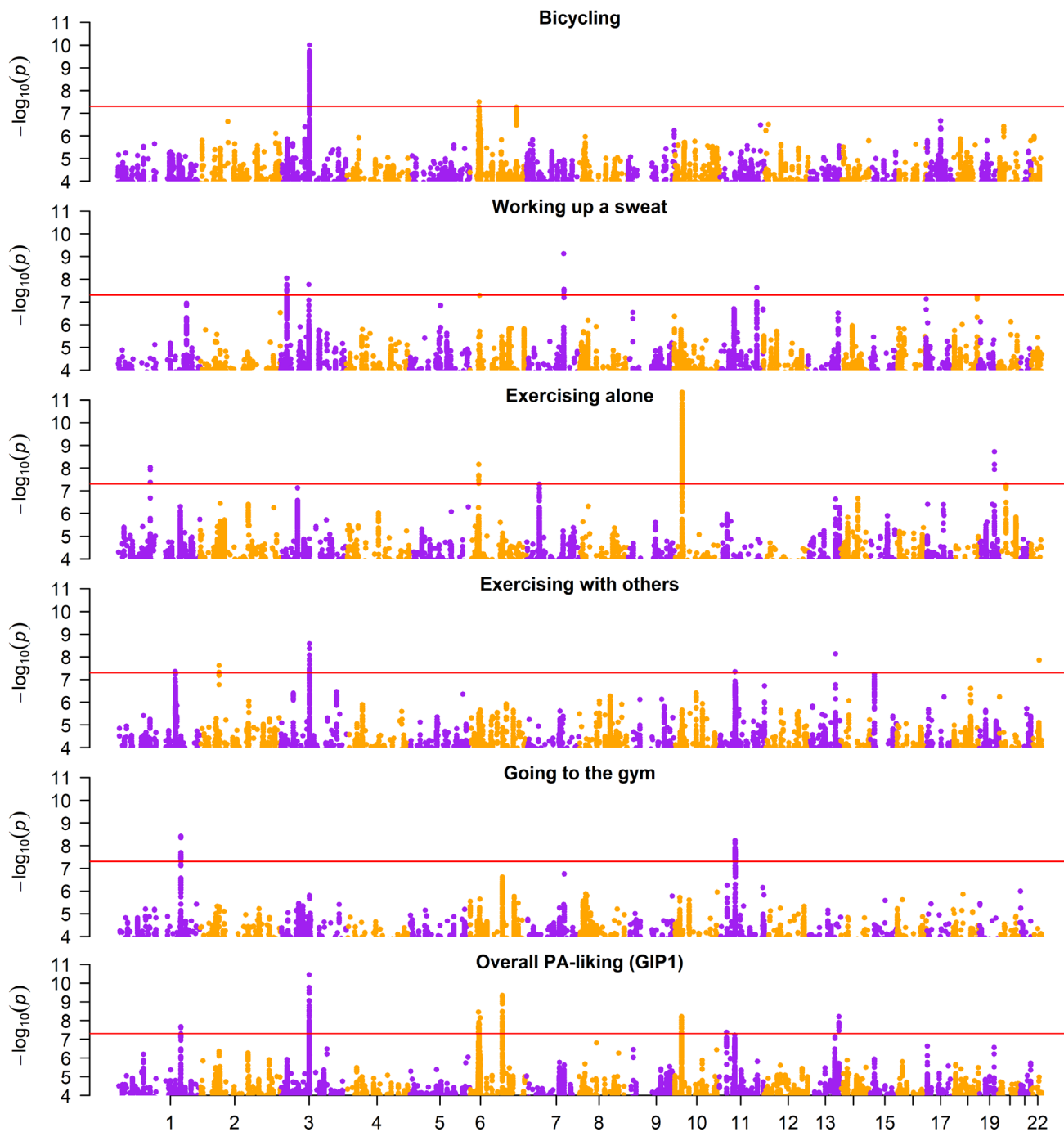
The beta coefficient units refer to the units used in the scale, from 1 to 9. EAF: effect allele frequency; SE: standard error; n= sample size used in analysis; SNP position based on GRCh37

\* TAATTAATAATTATGTGAAAAC; <sup>a</sup> A allele of rs4390955 used as proxy, <sup>b</sup> G allele of rs6660825 used as proxy, <sup>c</sup> A allele of rs7642661 used as proxy, <sup>d</sup> C allele of rs3136434 used as proxy, <sup>e</sup> A allele of rs1417998 used as proxy, <sup>f</sup> G allele of rs1802669 used as proxy, <sup>g</sup> A allele of rs2167046 used as proxy, <sup>h</sup> C allele of rs76508707 used as proxy, <sup>i</sup> A allele of rs3136469 used as proxy, <sup>j</sup> G allele of rs2326128 used as proxy, <sup>k</sup> G allele of rs12494446 used as proxy, <sup>l</sup> G allele of rs7120717 used as proxy.

**365**

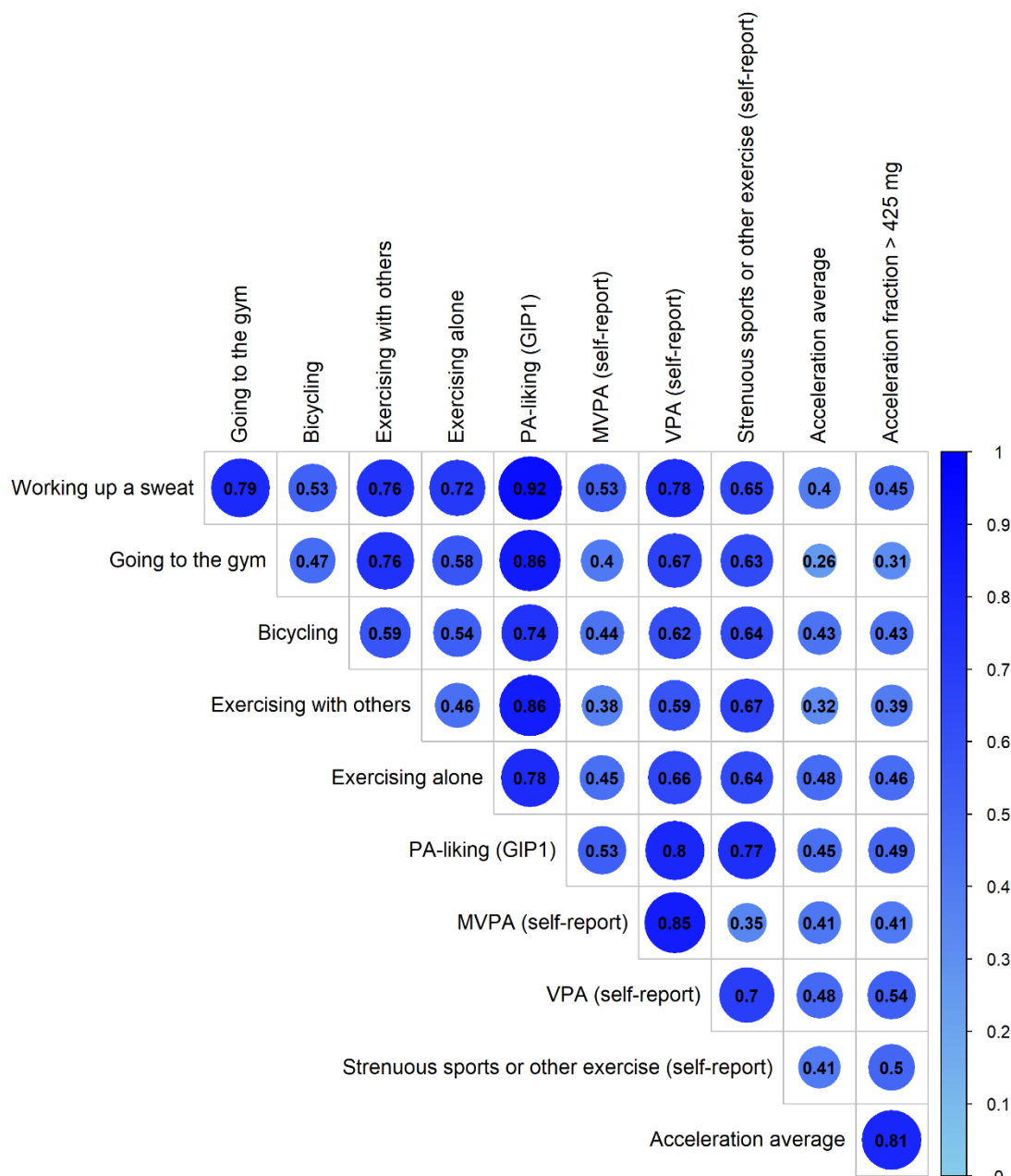
366  
367

**Figure 1:** Manhattan plots of five physical activity liking traits and overall PA-liking (GIP1).



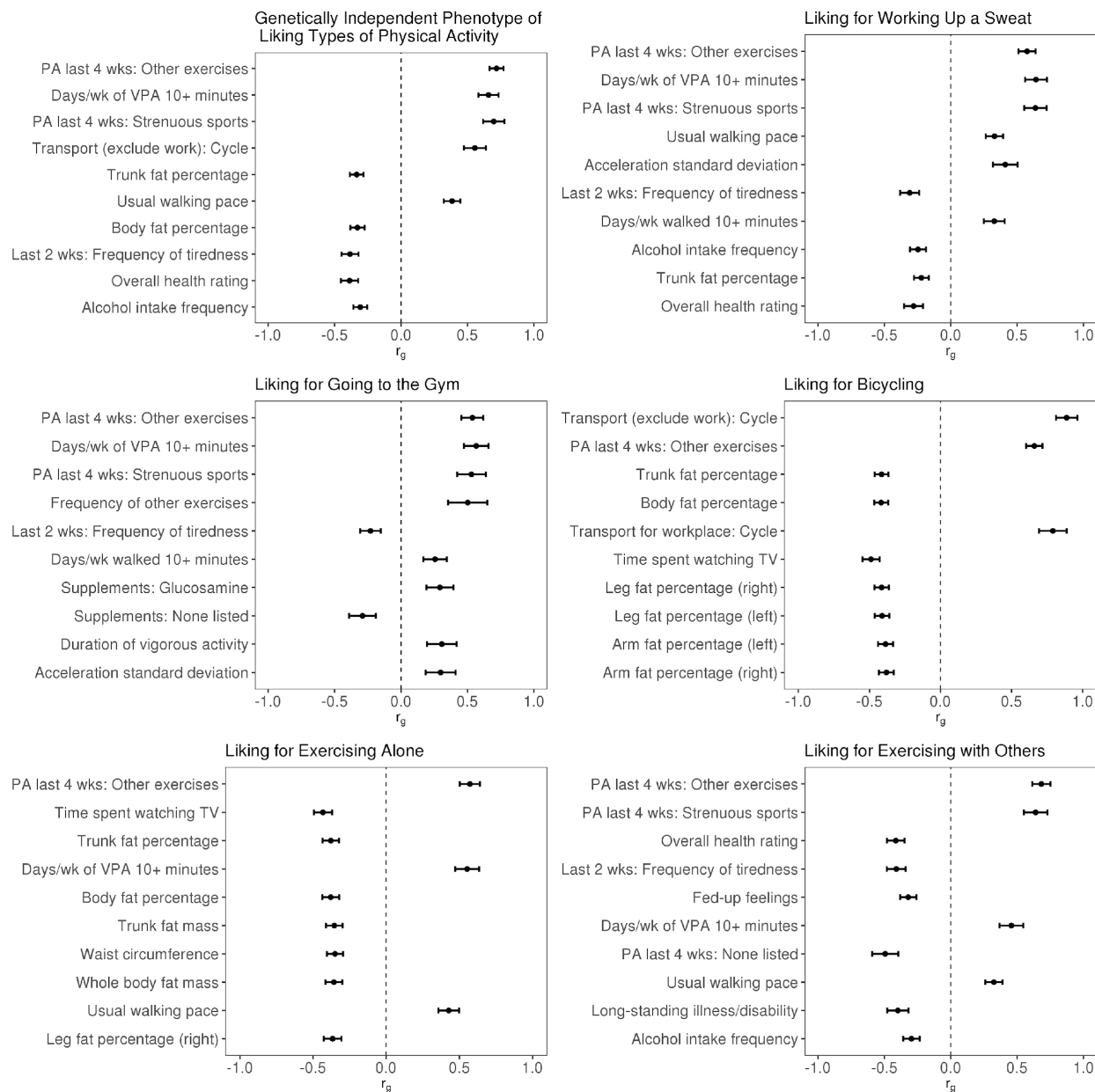
369

370 **Figure 2:** Genetic correlations across PA self-report, PA from accelerometry, and PA-liking traits.



372  
373  
374  
375

**Figure 3:** Genetic correlations of PA-liking traits with other traits and diseases. Top ten correlations are shown according to p-value. The order of traits shown in each panel is in ascending order by p-value. Error bars represent 95 % CI.



377