OPEN

Genomewide Association Study of Leisure-Time Exercise Behavior in Japanese Adults

MEGUMI HARA[1], TSUYOSHI HACHIYA[2], YOICHI SUTOH[2], KEITARO MATSUO[3,4], YUICHIRO NISHIDA[1], CHISATO SHIMANOE[1], KEITARO TANAKA[1], ATSUSHI SHIMIZU[2], KEIZO OHNAKA[5], TAKAHISA KAWAGUCHI[6], ISAO OZE[3], FUMIHIKO MATSUDA[6], HIDEMI ITO[3,4,7], SAYO KAWAI[8,9], ASAHI HISHIDA[8], RIEKO OKADA[8], TAE SASAKABE[8,9], AKIE HIRATA[3], RIE IBUSUKI[10], YORA NINDITA[10,11], NORIHIRO FURUSYO[12], HIROAKI IKEZAKI[12], NAGATO KURIYAMA[13], ETSUKO OZAKI[13], HARUO MIKAMI[14], YOHKO NAKAMURA[14], SADAO SUZUKI[15], AKIHIRO HOSONO[15], SAKURAKO KATSUURA-KAMANO[16], KOKICHI ARISAWA[16], KIYONORI KURIKI[17], KAORI ENDOH[17], NAOYUKI TAKASHIMA[18], AYA KADOTA[18,19], MASAHIRO NAKATOCHI[20], YUKIHIDE MOMOZAWA[21], MICHIAKI KUBO[22], MARIKO NAITO[8,23], and KENJI WAKAI[8]

[1]Department of Preventive Medicine, Faculty of Medicine, Saga University, Saga, JAPAN; [2]Division of Biomedical Information Analysis, Iwate Tohoku Medical Megabank Organization, Disaster Reconstruction Center, Iwate Medical University, Iwate, JAPAN; [3]Division of Molecular and Clinical Epidemiology, Aichi Cancer Center Research Institute, Nagoya, JAPAN; [4]Department of Epidemiology, Nagoya University Graduate School of Medicine, Nagoya, JAPAN; [5]Department of Geriatric Medicine, Graduate School of Medical Sciences, Kyushu University, Fukuoka, JAPAN; [6]Center for Genomic Medicine, Graduate School of Medicine, Kyoto University, Kyoto, JAPAN; [7]Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, JAPAN; [8]Department of Preventive Medicine, Nagoya University Graduate School of Medicine, Nagoya, JAPAN; [9]Department of Public Health, Aichi Medical University, School of Medicine, Aichi, JAPAN; [10]Department of International Island and Community Medicine Kagoshima University, Graduate School of Medical and Dental Sciences, Kagoshima, JAPAN; [11]Department of Pharmacology and Therapeutic, Faculty of Medicine, Diponegoro University, Semarang, INDONESIA; [12]Department of General Internal Medicine, Kyushu University Hospital, Fukuoka, JAPAN; [13]Department of Epidemiology for Community Health and Medicine, Kyoto Prefectural University of Medicine Graduate School of Medical Science, Kyoto, JAPAN; [14]Cancer Prevention Center, Chiba Cancer Center Research Institute, Chiba, JAPAN; [15]Department of Public Health, Nagoya City University Graduate School of Medical Sciences, Nagoya, JAPAN; [16]Department of Preventive Medicine, Institute of Biomedical Sciences, Tokushima University Graduate School, Tsukuba, JAPAN; [17]Laboratory of Public Health, School of Food and Nutritional Sciences, University of Shizuoka, Shizuoka, JAPAN; [18]Department of Public Health, Shiga University of Medical Science, Shiga, JAPAN; [19]Center for Epidemiologic Research in Asia, Shiga University of Medical Science, Shiga, JAPAN; [20]Statistical Analysis Section, Center for Advanced Medicine and Clinical Research, Nagoya University Hospital, Nagoya, JAPAN; [21]Laboratory for Genotyping Development, Riken Center for Integrative Medical Sciences, Yokohama, JAPAN; [22]RIKEN Center for Integrative Medical Sciences, Yokohama, JAPAN; and [23]Department of Oral Epidemiology, Graduate School of Biomedical and Health Sciences, Hiroshima University, Hiroshima, JAPAN

ABSTRACT

HARA, M., T. HACHIYA, Y. SUTOH, K. MATSUO, Y. NISHIDA, C. SHIMANOE, K. TANAKA, A. SHIMIZU, K. OHNAKA, T. KAWAGUCHI, I. OZE, F. MATSUDA, H. ITO, S. KAWAI, A. HISHIDA, R. OKADA, T. SASAKABE, A. HIRATA, R. IBUSUKI, Y. NINDITA, N. FURUSYO, H. IKEZAKI, N. KURIYAMA, E. OZAKI, H. MIKAMI, Y. NAKAMURA, S. SUZUKI, A. HOSONO, S. KATSUURA-KAMANO, K. ARISAWA, K. KURIKI, K. ENDOH, N. TAKASHIMA, A. KADOTA, M. NAKATOCHI, Y. MOMOZAWA, M. KUBO, M. NAITO, and K. WAKAI. Genomewide Association Study of Leisure-Time Exercise Behavior in Japanese Adults. *Med. Sci. Sports Exerc.*, Vol. 50, No. 12, pp. 2433–2441, 2018. **Purpose**: Although several genetic factors may play a role in leisure-time

Address for correspondence: Megumi Hara, M.D., Ph.D., Department of Preventive Medicine, Faculty of Medicine, Saga University 5-1-1 Nabeshima, Saga 849-8501, Japan; E-mail: harameg@cc.saga-u.ac.jp.

Submitted for publication September 2017.
Accepted for publication June 2018.
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).

0195-9131/18/5012-2433/0
MEDICINE & SCIENCE IN SPORTS & EXERCISE®

DOI: 10.1249/MSS.0000000000001712

exercise behavior, there is currently no evidence of a significant genomewide association, and candidate gene replication studies have produced inconsistent results. **Methods**: We conducted a two-stage genomewide association study and candidate single-nucleotide polymorphisms (SNP) association study on leisure-time exercise behavior using 13,980 discovery samples from the Japan Multi-Institutional Collaborative Cohort (J-MICC) study, and 2036 replication samples from the Hospital-based Epidemiologic Research Program at Aichi Cancer Center-2 study. Leisure-time physical activity was measured using a self-administered questionnaire that inquired about the type, frequency and duration of exercise. Participants with ≥4 MET·h·wk^{-1} of leisure-time physical activity were defined as exhibiting leisure-time exercise behavior. Association testing using mixed linear regression models was performed on the discovery and replication samples, after which the results were combined in a meta-analysis. In addition, we tested six candidate genetic variants derived from previous genomewide association study. **Results**: We found that one novel SNP (rs10252228) located in the intergenic region between *NPSR1* and *DPY19L1* was significantly associated with leisure-time exercise behavior in discovery samples. This association was also significant in replication samples (combined *P* value by meta-analysis = 2.2×10^{-9}). Several SNP linked with rs10252228 were significantly associated with gene expression of *DPY19L1* and *DP19L2P1* in skeletal muscle, heart, whole blood, and the nervous system. Among the candidate SNP, rs12612420 in *DNAPTP6* demonstrated nominal significance in discovery samples but not in replication samples. **Conclusions**: We identified a novel genetic variant associated with regular leisure-time exercise behavior. Further functional studies are required to validate the role of these variants in exercise behavior. **Key Words:** PHYSICAL ACTIVITY, SPORTS PARTICIPATION, GENOTYPE

Many epidemiological studies have shown that physical activity (PA) reduces the risk of all-cause mortality, coronary heart disease, hypertension, stroke, metabolic syndrome, type 2 diabetes, breast cancer, colon cancer, and depression (1). Inactivity was estimated to have caused 9.4% of 57 million premature deaths worldwide in 2008 (2). Although most countries have national recommendations for PA (3), the general population remains physically inactive or only rarely participates in leisure-time exercise (4).

Although environmental factors are well known to influence exercise participation, recent studies have shown that genetic factors also play a role (5–7). The heritability of exercise behavior in adults has been well documented (8,9). According to twin studies, between 48% and 71% of the variance in exercise behavior among adults can be explained by genetic factors (8). Early linkage and candidate gene studies have reported associations between several genomic regions and exercise behavior (10–14). Of these candidates or linkage regions, rs12405556 in the leptin receptor (*LEPR*) gene and rs8036270 on 15q13 are significantly associated with leisure-time exercise behavior in Dutch and American populations (15), while other reported associations have not been replicated.

To date, only two genomewide association studies (GWAS) have been carried out on PA (15,16). The first study, which examined Dutch and American populations, reported that 3 single-nucleotide polymorphisms (SNP), located in the DNA polymerase-transactivated protein 6 (*DNAPTP6*) gene, 3′-phosphoadenosine 5′-phosphosulfate synthase 2 (*PAPSS2*) gene, and chromosome 18 open reading frame 2 (*C18orf2*) gene, were associated with exercise participation (combined $P < 1 \times 10^{-5}$) (15). Another GWAS of the Korean population reported that even the most significant association between a SNP (rs7023003) and exercise participation did not reach genomewide significance (16). It is also important to note that not all significant associations between SNP and exercise participation have been replicated for all candidate SNP (15,16). Because the effect size of most SNP is small, very large sample sizes are needed to detect significant effects, and the number of detected variants increases with increasing sample size (17). Previous studies have shown that relatively small sample sizes, such as those less than 10,000, may constitute a factor in the ability of GWAS studies (15,16) to detect significant genomewide associations or replicate previous findings. In addition, differences in the ancestry of participants among studies might also influence the replication of results. Therefore, it remains unclear which genes or genetic regions are responsible for the physically active phenotype.

Here, we conducted a GWAS study to detect new genetic variants that are associated with leisure-time exercise behavior, and to replicate associations for previously reported candidate SNP in a Japanese population.

MATERIALS AND METHODS

Study Population

Japan Multi-institutional Collaborative Cohort study. The discovery phase used data from participants from the Japan Multi-institutional Collaborative Cohort (J-MICC) Study (18). Briefly, the J-MICC study includes volunteers age 35 to 69 yr from 13 sites in Japan: Aichi, Chiba, Fukuoka, Iga, Kagoshima, KOPS, Kyoto, Okazaki, Sakuragaoka, Saga, Shizuoka-Daiko, Takashima, and Tokushima subcohorts. The study had recruited 92,647 participants by the end of March 2014. All participants gave written informed consent; answered a questionnaire that inquired about lifestyle-related factors, past medical history, medication status, and anthropometric characteristics; and donated a blood sample. The study protocol was approved by the ethics committees of Nagoya University School of Medicine and other participating institutions. A total of 14,539 participants were randomly selected to be genotyped from 47,163 participants in 12 sites (except for the Iga subcohort, where the survey was conducted from 2013 to 2014) who were recruited between 2004 and 2013. Participants with inconsistent baseline information on sex between the questionnaire and genotyping results were excluded (n = 26). Quality control (QC) (described below) was conducted for the remaining 14,513 participants, of whom 422 participants whose genotype data did not meet QC filters were excluded. Of the 14,091 individuals who passed genotype QC filtering, three were excluded because leisure-time

PA data was not available. In addition, 108 subjects were excluded because body mass index (BMI) data were not available. The remaining 13,980 subjects were used in the GWAS analysis.

Hospital-based Epidemiologic Research Program at Aichi Cancer Center-2 study. The replication phase of this GWAS study was conducted using data from participants of the Hospital-based Epidemiologic Research Program at Aichi Cancer Center (HERPACC)-2 study (19) conducted between January 2001 and November 2005 at the Aichi Cancer Center. Briefly, all first-visit outpatients (n = 29,736) during the study period were asked to complete a self-administered questionnaire and provide blood samples. Of these, 28,776 (96.7%) participated in the study by giving written informed consent and 13,824 subjects (48.0% of participants) further provided blood samples. Among the 13,824 subjects, 7053 were confirmed to have had no detectable cancer and no history of neoplasia within a 1-yr period from participation. The study protocol was approved by the ethics committees of the Aichi Cancer Center.

A total of 2074 participants were randomly selected from the 7053 noncancer subjects, genotyped using Illumina Human Core Exome 12 chips, and selected for this replication phase. After exclusion of subjects lacking baseline information on BMI and leisure-time PA (n = 35), and those with low imputation quality (n = 3), 2036 subjects were used for the replication analysis.

Leisure-time Exercise Behavior

Leisure-time PA was determined using a self-administered questionnaire, similar to a short format of the International Physical Activity Questionnaire, for both the discovery and replication phases (20). Physical activity was assessed by METs of leisure-time exercise (21). Participants were asked about the frequency and average duration of exercise behavior according to three broad categories of intensity (vigorous, moderate and light). Vigorous activities, defined as those that cause a person to breathe more heavily than normal (to the extent that they cannot talk), were allocated 8 METs; moderate activities, defined as those that cause a person to breathe somewhat more heavily than normal (to the extent that they can still talk), were allocated 4.0 METs; and light activities, defined as those that cause a person to breathe normally (e.g., walking), were allocated 3.3 METs. The frequency categories (assigned average days per week) for leisure-time PA were: almost none (0), one to three times per month (0.1), one to two times per week (0.2), three to four times per week (0.5), and five to six times per week (0.8). The average duration categories (assigned average hours per activity) were: <30 min (0.3), 30 min to <1 h (0.8), 1 to <2 h (1.5), 2 to <3 h (2.5), 3 to <4 h (3.5), and ≥4 h (4.5). A MET-hour per week of leisure-time PA for each category of intensity was calculated using the weekly frequency, duration, and intensity of leisure time PA, according to the formula: Leisure time PA ($MET \cdot h \cdot wk^{-1}$) = intensity of PA (METs)*duration (h)*frequency ($d \cdot wk^{-1}$). The resulting values for the three intensities were summed and reported as an outcome. Although the use of a continuous outcome results in a higher statistical power than a dichotomous outcome, the distribution of leisure-time physical activity levels was extremely right-skewed (P = 0.005, Anderson-Darling test; see Figure, Supplemental Digital Content 1, the distribution of leisure-time physical activity levels, http://links.lww.com/MSS/B331), making it difficult to handle this parameter as a continuous outcome. Thus, we defined participants who reported ≥4 or more $MET \cdot h \cdot wk^{-1}$ of leisure-time PA as exhibiting regular leisure-time exercise behavior and used a dichotomous outcome (i.e., with or without regular leisure-time exercise behavior) in our analyses.

Genotyping and QC Filtering

J-MICC study. Buffy coat fractions and DNA were prepared from blood samples and stored at −80°C at the central J-MICC study office. DNA was extracted from all buffy coat fractions using a BioRobot M48 Workstation (Qiagen Group, Tokyo, Japan) at the central study office. For samples from two sites (Fukuoka and Kyushu-KOPS), DNA was extracted from samples of whole blood using an automatic nucleic acid isolation system (NA-3000, Kurabo, Co., Ltd, Osaka, Japan) at the original study site. The 14,539 study participants from the 12 sites of the J-MICC study were genotyped at the RIKEN Center for Integrative Medicine using a HumanOmniExpressExome-8 v1.2 BeadChip array (Illumina Inc., San Diego, CA). Twenty-six samples with inconsistent information regarding sex between the questionnaire and genotyping results were excluded. The identity-by-descent method implemented in PLINK software (22) found 388 close relationship pairs (pi-hat >0.1875); one sample of each pair was excluded. Principal component analysis (23,24) with the 1000 Genomes reference panel (phase 3) (25) detected 34 subjects with non-Japanese ancestry (26); their samples were excluded. The remaining 14,091 samples met the sample-wise genotype call rate criterion (≥0.99). Single-nucleotide polymorphism with a genotype call rate <0.98 and/or a Hardy-Weinberg equilibrium exact test P value $<1 \times 10^{-6}$ were removed. This QC filtering left 14,091 individuals and 873,254 autosomal variants.

HERPACC-2 study. DNA was prepared from buffy coat fractions using a Qiagen Blood Mini Kit (Qiagen Group, Tokyo, Japan) at the Aichi Cancer Center. Genotyping of the 2074 HERPACC-2 study subjects was conducted at the Center for Genomic Medicine, Kyoto University Graduate School of Medicine using a Human Core Exome-12 v1.1 BeadChip array (Illumina Inc.). The same QC procedure used in the discovery phase was performed. Data for rs10252228, rs187522732, rs374914476, rs565411857, 10:120175179, rs11350613 were extracted from imputed data for validation.

Heritability Estimation

Narrow-sense heritability of exhibiting leisure-time exercise behavior was estimated from genotyping data using a

mixed linear model with adjustments for age, sex, and site. Briefly, the model assumes that a genetic random effect for each individual is drawn from a multivariate normal distribution with a mean 0 and variance–covariance matrix calculated from genotype data (i.e., genetic relationship matrix [GRM]). Similarly, a nongenetic random effect is assumed to be normally distributed with mean 0 and identity variance-covariance matrix. Adjustment variables were modeled as fixed-effect variables. The relative contributions of genetic and nongenetic effects were estimated using the restricted maximum likelihood method. For calculation of the GRM, additional QC filtering was applied according to a previous study (Hardy-Weinberg exact test *P*-value ≥0.05, and minor allele frequency [MAF] ≥ 0.01) (27,28), and the remaining 482,567 directly genotyped SNP on autosomal chromosomes were used. Calculation of the GRM and heritability estimation was performed using GCTA software (29) version 1.24.2. Observed-scale heritability was transformed into liability-scale heritability according to a previously proposed formula (27).

Genotype Imputation

Genotype imputation was performed using SHAPIT (30) and Minimac3 (31) software based on the 1000 Genomes reference panel (phase 3) (25) in both the discovery and replication phases. After genotype imputation, variants with an imputation quality $R^2 < 0.8$ for discovery samples and 0.3 for replication samples, and a MAF < 0.01 were excluded, leaving 7,094,228 variants in discovery samples and 11,070,774 variants in replication samples.

Association Tests between Genetic Variants and Leisure-time Exercise Behavior

Statistical analysis for novel findings. To calculate the GRM, genotyped SNP were excluded using the QC criteria proposed in a previous study (genotype call rate ≥ 0.95, Hardy-Weinberg exact test *P*-value ≥0.05, and MAF ≥ 0.01) (27). The remaining 482,567 SNP on autosomal chromosomes were used to calculate the GRM. Calculations for the GRM, heritability estimation, and GWAS tests were performed using GCTA software (29) version 1.24.2. Observed-scale heritability was transformed into liability-scale heritability using a previously proposed formula (27). The association between 7,094,228 imputed genetic variants and leisure-time exercise behavior (defined as ≥4 MET·h·wk^{-1} of leisure time PA) was analyzed using a mixed linear model association method (32) adjusted for age, sex, and site. The mixed linear model used adjustment covariates as fixed-effect variables and GRM as a variance-covariance matrix for random-effects. The genomewide significance level was set at $P < 5 \times 10^{-8}$, and the suggestive significance level was set at $P < 1 \times 10^{-6}$ for all analyses. Because high BMI has been suggested as a factor preventing people from exercising, we also conducted genomewide association tests with adjustment for age, sex, site, and BMI as a sensitivity analysis. To seek genetic variants that are associated with leisure time PA specifically in younger or older adults, we stratified our population based on the median age (56 yr) into younger (age ≤ 56) and older (age > 56) strata, and performed genomewide association tests for each stratum. The summary statistics derived from the present GWAS are available upon request to the authors.

To further validate our GWAS findings in the discovery phase, we performed multiple linear regression analyses adjusted for age, sex, and site for the imputed data from HERPACC-2 samples in the replication phase. Subsequently, meta-analysis was performed using discovery and replication samples. Heterogeneity across the two data sets was evaluated using the I^2 index.

Statistical analysis for replication of candidate SNP. For replication analysis of previously reported SNP associated with exercise behavior, we used discovery samples (N = 13,980) for association tests. We examined the three SNP in *DNAPTP6, PAPSS2,* and *C18orf2* found in Dutch and American populations (combined $P < 1 \times 10^{-5}$) (11), and a SNP in the RNA, 7SK small nuclear (*RN7SK*) gene-solute carrier family member 1 (*SLC44A1*) gene found in the Korean population ($P < 1 \times 10^{-6}$) (16). We also examined SNP in *LEPR* and *GABRG3* because these were candidates from earlier studies (12,14), and their associations were replicated in the Dutch and American populations (15). To further validate our replication of the association between previously identified SNP and exercise behavior in the discovery phase, we also performed the analysis in replication samples and conducted a meta-analysis.

Expression Quantitative Trait Locus Analysis

To explore the functional consequences of SNP identified in our GWAS, we used data from Genotype-Tissue Expression (GTEx), a publicly available database for expression Quantitative Trait Locus (eQTL) (33).

Enrichment Analysis

Based on GWAS summary data (chromosomal position and *P* value) for the directly genotyped SNP, gene- and pathway-based analyses were conducted using the MAGMA software (version 1.06) (34). The details have been described elsewhere (35). Briefly, variants were mapped onto protein-coding genes based on gene annotations downloaded from the NCBI Gene database. Gene-based *P* values were then calculated by aggregating variant-based *P* values after accounting for the linkage-disequilibrium structure based on the East Asian population (25,36). Pathway-based *P* values were calculated by aggregating the gene-based *P* values.

RESULTS

Baseline characteristics of the subjects in the discovery and replication phases are shown in Table 1. The mean age of individuals in the discovery and replication populations was 54.8 ± 9.4 yr and 51.8 ± 11.1 yr, and the percentage of female participants was 54.8% and 50.8%, respectively. Mean

TABLE 1. Baseline characteristics of the study subjects.

Site	*N*	Age (yr) (mean ± SD)	Female (%)	Leisure-Time PA (MET·h·wk^{-1}) (mean ± SD)	Leisure-Time Exercise Behavior (%)	BMI (kg·m^{-2}) (mean ± SD)
Discovery samples						
Aichi	1152	55.1 ± 9.4	50.4	1.2 ± 1.9	8.8	22.4 ± 3.2
Chiba	1099	53.8 ± 9.8	65.9	1.8 ± 2.4	14.9	22.5 ± 3.0
Fukuoka	1560	60.5 ± 5.4	56.2	1.8 ± 2.3	14.6	23.2 ± 3.0
Kagoshima	1224	55.2 ± 8.3	57.7	1.6 ± 2.7	14.1	24.5 ± 3.3
KOPS	1144	55.0 ± 9.7	64	1.0 ± 1.9	7.4	23.8 ± 3.5
Kyoto	1111	49.9 ± 9.8	52	1.5 ± 2.4	11.9	22.4 ± 3.3
Okazaki	1061	55.7 ± 9.1	45.2	2.0 ± 2.6	19.6	23.3 ± 3.1
Sakuragaoka	573	50.1 ± 9.4	38.2	1.7 ± 2.4	14.8	23.1 ± 3.4
Saga	1864	56.8 ± 8.2	56.6	1.5 ± 2.1	11.6	23.1 ± 3.0
Shizuoka-Daiko	1979	52.9 ± 9.7	56	1.6 ± 2.1	11.4	22.4 ± 3.1
Takashima	536	56.6 ± 9.7	70.7	1.8 ± 2.8	19.6	23.1 ± 3.3
Tokushima	677	50.4 ± 8.9	33.1	1.4 ± 2.1	9.9	23.9 ± 3.6
Total	13,980	54.8 ± 9.4	54.8	1.6 ± 2.3	12.8	23.1 ± 3.3
Replication samples						
HERPACC-2	2036	51.8 ± 11.1	50.8	1.6 ± 2.7	12.3	22.7 ± 3.2

leisure-time PA was 1.6 ± 2.3 MET·h·wk^{-1} for the discovery population and 1.6 ± 2.7 MET·h·wk^{-1} for the replication population. The proportion of subjects that exhibited regular leisure-time exercise behavior was 12.8% for the discovery population and 12.3% for the replication population.

Heritability Estimation

We estimated the narrow-sense heritability of target phenotype variables using data from the discovery samples (N = 13,980). The observed-scale heritability was 1.3% (standard error [SE], 2.3%), whereas the transformed liability-scale heritability was 3.3% (SE = 5.9%); the difference was not statistically significant (P = 0.28). Therefore, the overall SNP heritability of regular leisure-time exercise behavior was considered small.

Novel associations and their replications. We performed genomewide scans for leisure-time exercise behavior-associated genetic variants using the discovery samples (N = 13,980). The quantile–quantile plot for P values (see Figure, Supplemental Digital Content 2, quantile–quantile plot for P values, http://links.lww.com/MSS/B332) shows an inflation factor for the genomewide scan of 0.996 (95% confidence interval, 0.994–0.998). The Manhattan plot of genetic variants and leisure-time exercise behavior for discovery samples is shown in Figure 1. Table 2 summarizes the two-stage GWAS and meta-analysis findings. Results from the discovery and replication phases were analyzed using a mixed linear model association method adjusted for age, sex, and site, and multiple linear regression analysis adjusted for age, sex, and site, respectively. We found that rs10252228, located in the intergenic region between neuropeptide S receptor 1 (*NPSR1*) and dpy-19 like C-mannosyltransferase 1 (*DPY19L1*), had genomewide significance (effect size = 0.0269 [SE = 0.0048], $P = 2.5 \times 10^{-8}$) in discovery samples. This variant also showed significant associations in replication samples (P = 0.004). The combined P value for this variant in the meta-analysis ($P = 2.2 \times 10^{-9}$) was lower than that observed in the discovery phase, suggesting that this association is robust. In contrast, rs187522732, located in the intergenic region between

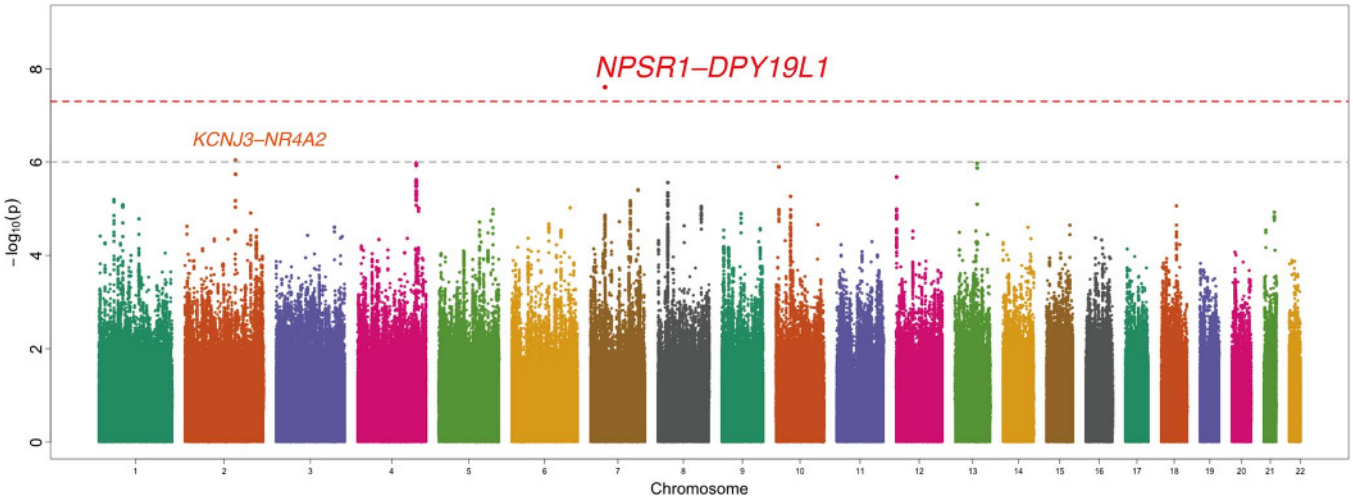

FIGURE 1—Genomewide association signals for discovery samples (N = 13,980). The x-axis represents chromosomal position and the y-axis represents $-\log_{10}$ P value calculated using mixed linear model association analysis. The gray dotted horizontal line indicates the suggestive significance level ($P = 1 \times 10^{-6}$).

TABLE 2. Leisure-time exercise behavior (≥4 MET·h·wk^{-1} of leisure-time physical activity)-associated SNP in discovery samples (J-MICC study, N = 13,980), replication samples (HERPACC-2 study, N = 2036), and combined samples (N = 16,016) adjusted for age, sex, and site.

SNP	Chr[a]	Position[b]	Gene(s)	EA[c]	NEA[d]	Population	Rsq[e]	AF[f]	Beta[g]	SE (Beta)[h]	P	I^2
rs10252228[i]	**7**	**34,940,039**	***NPSR1–DPY19L1***	**G**	**A**	**J-MICC**	**0.995**	**0.213**	**0.0269**	**0.0048**	**2.5×10^{-8}**	—
						HERPACC-2	**0.598**	**0.175**	**0.0278**	**0.0133**	**0.04**	—
						Meta-analysis	—	**0.209**	**0.0270**	**0.0045**	**2.2×10^{-9}**	0.0
rs187522732	2	156,864,816	*KCNJ3–NR4A2*	A	T	J-MICC	0.872	0.021	0.0713	0.0145	9.2×10^{-7}	—
						HERPACC-2	0.783	0.014	−0.0293	0.0439	0.51	—
						Meta-analysis	—	0.209	0.0614	0.0138	8.2×10^{-6}	78.8

Results in bold emphasis showed a consistent association in discovery and replication samples.
[a]Chromosome.
[b]Chromosomal position (GRCh37/hg19).
[c]Effect allele.
[d]Noneffect allele.
[e]Inputation quality in terms of R^2 calculated by the Minimac3 software version 1.0.11.
[f]Allele frequency of effect allele.
[g]Effect size.
[h]Standard error of effect size.
[i]Copy number variation.
[j]These SNP were directly genotyped.

KCNJ3 and *NR4A2*, showed a trend toward a significant association (effect size = 0.0713 [SE = 0.0145], $P = 9.2 \times 10^{-7}$) in the discovery phase, but not in the replication phase (P = 0.51).The combined P value for this variant in the meta-analysis ($P = 8.2 \times 10^{-6}$) was higher than that in the discovery phase, indicating that this association is likely a false positive. Adjustment for BMI did not change these findings (see Table, Supplemental Digital Content 3, Sensitivity analyses of primary analyses, http://links.lww.com/MSS/B333).

Association signals around rs10252228 (in the intergenic region between *NPSR1* and *DPY19L1*) for the discovery samples (N = 13,980) are shown in Figure 2. The eQTL analysis found no significant functional consequences for rs10252228 in the GTEx database; however, six SNP linked with rs10252228 were significantly associated with gene expression of *DPY19L1* and *DPY19L2P1* in skeletal muscle, heart, whole blood, and the nervous system.

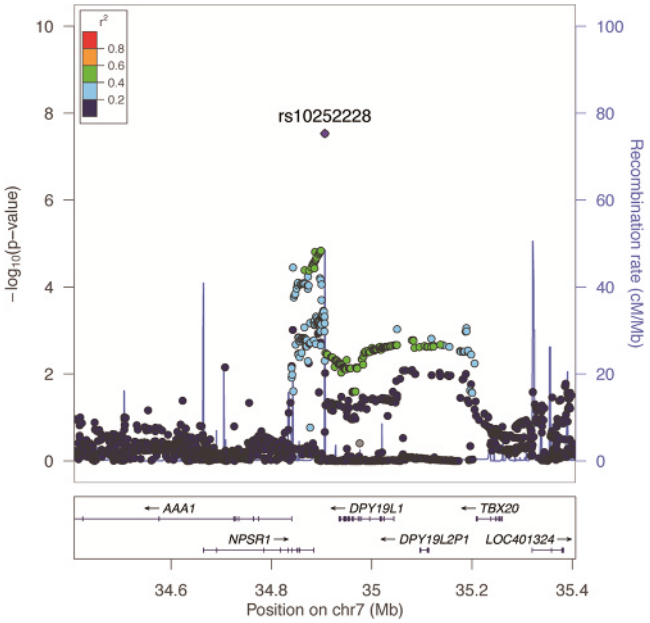

FIGURE 2—Association signals around rs10252228 for discovery samples (N = 13,980). The x-axis represents the chromosomal position around the rs10252228 variant and the y-axis represents $-\log_{10} P$ value. The top signal in this locus (rs10252228) is shown in purple. Dot color for a variant represents the estimated degree of linkage disequilibrium (R^2) between each variant and rs10252228.

Genomewide association tests on subjects stratified by age strata did not identify any loci with even suggestive significance in younger subjects; however, the *KCNJ3-NR4A2* locus showed genomewide significance in older subjects ($P = 1.9 \times 10^{-8}$). An additional three SNP showed suggestive significance in older subjects ($P = 3.5 \times 10^{-7}$ to 6.3×10^{-7}). However, none of these associations could be confirmed in the replication samples and combined samples (see Table, Supplemental Digital Content 4, Sensitivity analyses of primary analyses, http://links.lww.com/MSS/B334).

Replication of associations for candidate SNP. Table 3 shows the association between six previously reported SNP and leisure-time exercise behavior in the discovery samples, replication samples, and meta-analysis findings. rs12612420 in *DNAPTP6* showed a trend toward an association in the meta-analysis (effect size, −0.0107; SE, 0.0046; P = 0.0199), but the association was not significant after Bonferroni correction ($P > 0.05/6$). The eQTL analysis showed that the polymorphism in rs12612420 was significantly associated with gene expression in the neighboring *DNAPTP6* gene, also known as spermatogenesis associated serine rich 2-like (*SPATS2L*). *SPATS2L* expression was lower in the left ventricle with the effect allele (A) than with the noneffect allele (G), but was higher in sun-exposed/not sun-exposed skin with the effect allele.

Enrichment analysis. Our gene-based analyses did not identify any significant genes associated with regular leisure-time exercise behavior. We also searched for molecular pathways that were collectively associated with regular leisure-time exercise behavior. Although no significant pathways were detected after Bonferroni correction based on the PANTHER pathway database (37), KEGG pathway database (38) or Ingenuity Pathway Database (http://www.ingenuity.com/index.html) (see Table, Supplemental Digital Content 5, Enrichment analysis

TABLE 3. Association signals for previously-reported SNP in discovery samples (J-MICC study, $N = 13,980$), replication samples (HERPACC-2 study, $N = 2,036$), and combined samples ($N = 16,016$) adjusted for age, sex, and site.

SNP	Chr[a]	Position[b]	Gene(s)	EA[c]	NEA[d]	Population	Rsq[e]	AF[f]	Beta[g]	SE(Beta)[h]	P	I^2
rs12612420[i]	2	201,158,122	*DNAPTP6*	G	A	J-MICC	1.000	0.772	−0.0123	0.0050	0.0092	—
						HERPACC-2	1.000	0.774	−0.0018	0.0120	0.8800	—
						Meta-analysis	—	0.772	−0.0107	0.0046	0.0199	0.0
rs10887741[i]	10	89,443,310	*PAPSS2*	T	C	J-MICC	1.000	0.652	−0.0050	0.0040	0.2200	—
						HERPACC-2	1.000	0.651	0.0000	0.0110	1.0000	—
						Meta-analysis	—	0.652	−0.0044	0.0038	0.2401	0.0
rs12405556[i]	1	66,063,117	*LEPR*	G	T	J-MICC	1.000	0.204	0.0040	0.0050	0.4200	—
						HERPACC-2	0.951	0.207	0.0074	0.0120	0.5500	—
						Meta-analysis	—	0.204	0.0045	0.0046	0.3292	0.0
rs7023003	9	107,880,670	*RN7SK–SLC44A1*	A	G	J-MICC	0.963	0.699	0.0033	0.0040	0.4500	—
						HERPACC-2	0.969	0.687	−0.0149	0.0110	0.1900	—
						Meta-analysis	—	0.697	0.0012	0.0038	0.7547	58.6
rs8097348	18	1,595,021	*C18orf2*	A	G	J-MICC	0.999	0.953	0.0038	0.0090	0.6800	—
						HERPACC-2	1.000	0.951	−0.0396	0.0240	0.1000	—
						Meta-analysis	—	0.953	−0.0016	0.0084	0.8540	65.1
rs8036270[i]	15	27,463,874	*GABRG3*	A	G	J-MICC	0.999	0.555	0.0016	0.0040	0.6900	—
						HERPACC-2	0.648	0.629	−0.0094	0.0110	0.3900	—
						Meta-analysis	—	0.564	0.0003	0.0038	0.9331	0.0

[a]Chromosome.
[b]Chromosomal position (GRCh37/hg19).
[c]Effect allele.
[d]Noneffect allele.
[e]Inputation quality in terms of R^2 calculated by the Minimac3 software version 1.0.11.
[f]Allele frequency of effect allele.
[g]Effect size.
[h]Standard error of effect size.
[i]These SNP were directly genotyped.

for PANTHER pathway, http://links.lww.com/MSS/B335; Table, Supplemental Digital Content 6, Enrichment analysis for KEGG pathway, http://links.lww.com/MSS/B336; and Table, Supplemental Digital Content 7, Enrichment analysis for ipath, http://links.lww.com/MSS/B337), many pathways and enrichments with $P < 0.05$ tended to be related to metabolism and degradation.

DISCUSSION

This was a large GWAS for leisure-time exercise behavior which examined 13,980 Japanese adults with 7,094,228 SNP. We found a novel association between an SNP in the *NPSR1-DPY19L1* intergenic region and confirmed it in replication samples (rs10252228, combined $P = 2.2 \times 10^{-9}$). Of the only two GWAS studies for leisure-time exercise behavior conducted to date, neither reported significant genomewide associations ($P < 5 \times 10^{-8}$). Therefore, to our knowledge, the present study is one of the first to show significant genomewide associations between genetic polymorphisms and leisure-time exercise behavior. In addition, we found that a candidate SNP (rs12612420) in *DNAPTP6* (also known as *SPATS2L*) showed a trend toward an association with leisure-time exercise behavior, although the association was not significant after Bonferroni correction. This latter association might warrant further investigation.

We found that several SNP linked with rs10252228 were significantly associated with gene expression of *DPY19L1* and *DPY19L2P1* in skeletal muscle, heart, whole blood and the nervous system, although no significant functions for rs10252228 were identified by eQTL. *DPY19L1* and *DPY19L2P1* are homologs of dumpy (DPY)-19 in *Caenorhabditis elegans* and mammals, and mediate C-mannosylation at thrombospondin type 1 repeats (TSRs) in target proteins to control the localization of proteins (39). *DPY19L1*-knockdown causes a defect in the radial migration of glutamatergic neurons in mice (40). *DPY19L2P1* is considered a pseudogene; however, it is expressed and is a suggested binding protein of amyloid beta (41,42). Further functional studies are required to unravel the molecular mechanisms behind the association between rs10252228 and leisure-time exercise behavior.

The Manhattan plot showed that, despite being a common variant, the chromosome 7 locus had only one variant (rs10252228) that showed a significant association (Fig. 1). This finding is unusual for common variants but is not surprising on this occasion because rs10252228 showed low correlations with neighboring SNP ($R2 < 0.6$; Fig. 2). Despite detailed analyses using available databases including GTEx, there was no evidence indicating that rs10252228 affected the expression of any genes. Although the mechanism is unclear, our findings indicate a robust association between rs1025228 and physical activity. This new finding may be worthy of scrutiny in future studies.

In our replication study of candidate genes, we did not detect any significant associations. Several reasons may explain the discrepancy between our results and those of previous studies. Differences in the ancestry of study participants might have influenced the effect size. Carlson et al. (43) reported that while most European ancestry (EA) GWAS findings can be expected to show an effect in the same direction as for non-EA populations, the effects tend to be lower or almost zero. Small effect sizes make detection of significant genomewide associations difficult. Differences in age distribution may also be important because both genetic and social factors can influence leisure-time exercise behavior at certain ages. For example, retired people are more likely to participate in

exercise (44,45); therefore, social factors are suspected to have a greater influence on exercise behaviors than genetic factors. We also observed an increased effect between the effect allele of the SNP in the *NPSR1-DPY19L1* intergenic region on leisure-time PA between the ages of 40 and 59, and a weaker association in participants age 60 to 69, most of who might be retired (see Table, Supplemental Digital Content 8, Sensitivity analyses according to age, http://links.lww.com/MSS/B338). The mean age of participants in this study was higher (mean, 55 yr; range, 35–69 yr) than that in the Dutch and American study (mean, 50 yr; range, 19–87 yr). Therefore, we hypothesize that the strong effect of social factors in the present study may have affected the ability to replicate associations for candidate genes. Discrepancies between our results and those from the Korean population (16) may be due to the use of different evaluation methods for PA because inconsistencies in estimates of genetic contribution may reflect differences in definitions of exercise behavior (8,46). Although the Korean study evaluated total amount of daily PA, the present study and the Dutch and American study (15) assessed leisure-time exercise behavior, which was defined as more than 4 MET·h·wk^{-1} of leisure-time PA. We suspect that the contributing genetic factors might differ between total daily PA and leisure-time exercise participation.

The estimated heritability of leisure-time exercise behavior due to common SNP was unexpectedly small in this study. Because of the large standard error of this estimation, however, the magnitude of heritability remains inconclusive. For example, the SNP-based heritability of conscientiousness, which is one of the Big Five personality traits (47), was estimated to be 0.01 (SE = 0.08) in an earlier study (N = 5011) (48). However, in a later study with a large sample size (N = 59,176), it was estimated to be 0.096 (SE = 0.009) (49). Additionally, the latter study successfully identified SNP significantly associated with conscientiousness. Therefore, we propose that it may be possible to identify loci that are associated with leisure-time exercise behavior and that such studies are worth performing.

In contrast to our low heritability (3.3%), previous European twin studies reported that between 48% and 71% of the variance in exercise behaviors among adults can be explained by genetic factors (8). Various reasons may explain the large difference between the findings in the present study and previous twin studies. First, differences in the measurement of leisure-time physical activity (e.g., categorical in this study vs continuous in previous studies) may have influenced the results. Second, differences in study populations (e.g., Japanese vs European) may be relevant. Third, possible interactions between genes and environmental factors (e.g., birth cohorts) may be important factors. A previous study reported that the heritability of different samples was more reflective of the heterogeneity in phenotypic measurements and gene–environment interactions than genetic heterogeneity (50). Finally, rare variants that were not accounted for in this study may contribute to the heritability of leisure-time exercise behavior. Although the contribution of common variants may not be large, our study identified a common SNP which influences leisure-time exercise behavior, indicating that common SNP may constitute a genetic factor for such behaviors.

This study has some limitations. First, a self-reported questionnaire for evaluating leisure-time exercise behavior is not an objective measurement method. However, the criteria of 4 MET·h·wk^{-1} for leisure-time PA is often used in epidemiological studies (8) and is used in the Japanese PA guideline (51). Second, because analysis was restricted to individuals of Japanese ancestry, our results may not be generalizable to other ethnic populations.

In conclusion, this study found a novel association between an SNP in the *NPSR1-DPY19L1* intergenic region (rs10252228, combined $P = 2.2 \times 10^{-9}$) and leisure-time exercise behavior, although further functional studies are needed to validate these findings. Our results suggest that genetic variants may affect leisure-time exercise behavior in adults.

This study was supported by Grants-in-Aid for Scientific Research for Priority Areas of Cancer (17015018) and Innovative Areas (221S0001) and by JSPS KAKENHI grants JP15H02524, JP16H06277, JP16K09058 from the Japanese Ministry of Education, Culture, Sports, Science and Technology. This study was supported in part by funding for the BioBank Japan Project from the Japan Agency for Medical Research and Development since April 2015, and the Ministry of Education, Culture, Sports, Science and Technology from April 2003 to March 2015.

The authors declare that the results of the study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation.

The authors declare that there are no conflicts of interest. The results of the present study do not constitute endorsement by the American College of Sports Medicine.

REFERENCES

1. Warburton DE, Charlesworth S, Ivey A, Nettlefold L, Bredin SS. A systematic review of the evidence for Canada's Physical Activity Guidelines for Adults. *Int J Behav Nutr Phys Act*. 2010;7:39.
2. Lee IM, 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.
3. Kahlmeier S, Wijnhoven TM, Alpiger P, Schweizer C, Breda J, Martin BW. National physical activity recommendations: systematic overview and analysis of the situation in European countries. *BMC Public Health*. 2015;15:133.
4. Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U. Global physical activity levels: surveillance progress, pitfalls, and prospects. *Lancet*. 2012;380(9838):247–57.
5. de Vilhena e Santos DM, Katzmarzyk PT, Seabra AF, Maia JA. Genetics of physical activity and physical inactivity in humans. *Behav Genet*. 2012;42(4):559–78.
6. Kelly SA, Pomp D. Genetic determinants of voluntary exercise. *Trends Genet*. 2013;29(6):348–57.
7. Lightfoot JT, DE Geus EJC, Booth FW, et al. Biological/genetic regulation of physical activity level: consensus from GenBioPAC. *Med Sci Sports Exerc*. 2018;50(4):863–73.

8. 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:e22.
9. Stubbe JH, de Moor MH, Boomsma DI, de Geus EJ. The association between exercise participation and well-being: a co-twin study. *Prev Med*. 2007;44(2):148–52.
10. Cai G, Cole SA, Butte N, et al. A quantitative trait locus on chromosome 18q for physical activity and dietary intake in Hispanic children. *Obesity*. 2006;14(9):1596–604.
11. De Moor MH, Posthuma D, Hottenga JJ, Willemsen G, Boomsma DI, De Geus EJ. Genome-wide linkage scan for exercise participation in Dutch sibling pairs. *Eur J Hum Genet*. 2007;15(12):1252–9.
12. Simonen RL, Rankinen T, Perusse L, et al. Genome-wide linkage scan for physical activity levels in the Quebec Family study. *Med Sci Sports Exerc*. 2003;35(8):1355–9.
13. Simonen RL, Rankinen T, Perusse L, et al. A dopamine D2 receptor gene polymorphism and physical activity in two family studies. *Physiol Behav*. 2003;78(4–5):751–7.
14. Stefan N, Vozarova B, Del Parigi A, et al. The Gln223Arg polymorphism of the leptin receptor in Pima Indians: influence on energy expenditure, physical activity and lipid metabolism. *Int J Obes Relat Metab Disord*. 2002;26(12):1629–32.
15. De Moor MH, Liu YJ, Boomsma DI, et al. Genome-wide association study of exercise behavior in Dutch and American adults. *Med Sci Sports Exerc*. 2009;41(10):1887–95.
16. Kim J, Kim J, Min H, et al. Joint identification of genetic variants for physical activity in Korean population. *Int J Mol Sci*. 2014; 15(7):12407–21.
17. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461(7265):747–53.
18. Hamajima N. The Japan Multi-Institutional Collaborative Cohort Study (J-MICC Study) to detect gene–environment interactions for cancer. *Asian Pac J Cancer Prev*. 2007;8(2):317–23.
19. Hamajima N, Matsuo K, Saito T, et al. Gene–environment Interactions and Polymorphism Studies of Cancer Risk in the Hospital-based Epidemiologic Research Program at Aichi Cancer Center II (HERPACC-II). *Asian Pac J Cancer Prev*. 2001;2(2):99–107.
20. Craig CL, Marshall AL, Sjöström M, et al. International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc*. 2003;35(8):1381–95.
21. Hara M, Higaki Y, Taguchi N, et al. Effect of the PPARG2 Pro12Ala polymorphism and clinical risk factors for diabetes mellitus on HbA1c in the Japanese general population. *J Epidemiol*. 2012;22(6):523–31.
22. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559–75.
23. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38(8):904–9.
24. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet*. 2006;2(12):e190.
25. Abecasis GR, Auton A, Brooks LD, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491(7422):56–65.
26. Yamaguchi-Kabata Y, Nakazono K, Takahashi A, et al. Japanese population structure, based on SNP genotypes from 7003 individuals compared to other ethnic groups: effects on population-based association studies. *Am J Hum Genet*. 2008;83(4):445–56.
27. Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet*. 2011;88(3):294–305.
28. Hachiya T, Kamatani Y, Takahashi A, et al. Genetic predisposition to ischemic stroke: a polygenic risk score. *Stroke*. 2017;48(2):253–8.
29. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011;88(1):76–82.
30. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nat Methods*. 2011;9(2):179–81.
31. Fuchsberger C, Abecasis GR, Hinds DA, et al. *Bioinformatics*. 2015; 31(5):782–4.
32. Yang J, Benyamin B, McEvoy BP, et al. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet*. 2010;42(7):565–9.
33. GTEx C. Genetic effects on gene expression across human tissues. *Nature*. 2017;550:204–13.
34. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol*. 2015;11(4):e1004219.
35. Hachiya T, Komaki S, Hasegawa Y, et al. Genome-wide meta-analysis in Japanese populations identifies novel variants at the TMC6-TMC8 and SIX3-SIX2 loci associated with HbA1c. *Sci Rep*. 2017;7(1):16147.
36. Goals galore. *Nature*. 2015;526(7571):6.
37. Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc*. 2013;8(8):1551–66.
38. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res*. 2017;45(D1):D353–D61.
39. Shcherbakova A, Tiemann B, Buettner FF, Bakker H. Distinct C-mannosylation of netrin receptor thrombospondin type 1 repeats by mammalian DPY19L1 and DPY19L3. *Proc Natl Acad Sci U S A*. 2017;114(10):2574–9.
40. Watanabe K, Takebayashi H, Bepari AK, Esumi S, Yanagawa Y, Tamamaki N. Dpy19l1, a multi-transmembrane protein, regulates the radial migration of glutamatergic neurons in the developing cerebral cortex. *Development*. 2011;138(22):4979–90.
41. Virok DP, Simon D, Bozsó Z, et al. Protein array based interactome analysis of amyloid-β indicates an inhibition of protein translation. *J Proteome Res*. 2011;10(4):1538–47.
42. Olah J, Vincze O, Virok D, et al. Interactions of pathological hallmark proteins: tubulin polymerization promoting protein/p25, beta-amyloid, and alpha-synuclein. *J Biol Chem*. 2011;286(39):34088–100.
43. Carlson CS, Matise TC, North KE, et al. Generalization and dilution of association results from European GWAS in populations of non-European ancestry: the PAGE study. *PLoS Biol*. 2013;11(9): e1001661.
44. Popham F, Mitchell R. Leisure time exercise and personal circumstances in the working age population: longitudinal analysis of the British household panel survey. *J Epidemiol Community Health*. 2006;60(3):270–4.
45. Bauman A, Ma G, Cuevas F, et al. Cross-national comparisons of socioeconomic differences in the prevalence of leisure-time and occupational physical activity, and active commuting in six Asia-Pacific countries. *J Epidemiol Community Health*. 2011;65(1): 35–43.
46. Kim J, Oh S, Min H, Kim Y, Park T. Practical issues in genome-wide association studies for physical activity. *Ann N Y Acad Sci*. 2011;1229:38–44.
47. Costa PT, McCrae RR. Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI): psychologial assessment resources: Odessa, FL; 1992.
48. Power RA, Pluess M. Heritability estimates of the Big Five personality traits based on common genetic variants. *Transl Psychiatry*. 2015;5:e604.
49. Lo MT, Hinds DA, Tung JY, et al. Genome-wide analyses for personality traits identify six genomic loci and show correlations with psychiatric disorders. *Nature Genet*. 2017;49(1):152–6.
50. Tropf FC, Lee SH, Verweij RM, et al. Hidden heritability due to heterogeneity across seven populations. *Nat Hum Behav*. 2017; 1(10):757–65.
51. Exercise and physical activity reference for health promotion 2013. Japan, Ministry of Health and Wealfare of Japan (in Japansese) available at http://www.nibiohn.go.jp/files/sintai2013.pdf. Accessed 18 May 2018.