Genome-Wide Association Study of Exercise Behavior in Dutch and American Adults

MARLEEN H.M. DE MOOR[1], YONG-JUN LIU[2], DORRET I. BOOMSMA[1], JIAN LI[2], JAMES J. HAMILTON[2], JOUKE-JAN HOTTENGA[1], SHAWN LEVY[7], XIAO-GANG LIU[2], YU-FANG PEI[2], DANIELLE POSTHUMA[1], ROBERT R. RECKER[6], PATRICK F. SULLIVAN[3], LIANG WANG[2], GONNEKE WILLEMSEN[1], HAN YAN[2], ECO J.C. DE GEUS[1], and HONG-WEN DENG[2,4,5,8,9]

[1]Department of Biological Psychology, VU University Amsterdam, Amsterdam, THE NETHERLANDS; [2]School of Medicine, University of Missouri, Kansas City, MO; [3]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC; [4]The Key Laboratory of Biomedical Information Engineering of Ministry of Education and Institute of Molecular Genetics, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, PR CHINA; [5]Laboratory of Molecular and Statistical Genetics, College of Life Sciences, Hunan Normal University, Changsha, Hunan, PR CHINA; [6]Osteoporosis Research Center, Creighton University, Omaha, NE; [7]Vanderbilt Microarray Shared Resource, Vanderbilt University, Nashville, TN; [8]Center of System Biomedical Sciences, Shanghai University of Science and Technology, Shanghai, PR CHINA; and [9]College of Life Sciences and Engineering, Beijing Jiao Tong University, Beijing, CHINA

ABSTRACT

DE MOOR, M. H., Y. LIU, D. I. BOOMSMA, J. LI, J. J. HAMILTON, J. HOTTENGA, S. LEVY, X. LIU, Y. PEI, D. POSTHUMA, R. R. RECKER, P. F. SULLIVAN, L. WANG, G. WILLEMSEN, H. YAN, E. J. DE GEUS, and H. DENG. Genome-Wide Association Study of Exercise Behavior in Dutch and American Adults. *Med. Sci. Sports Exerc.*, Vol. 41, No. 10, pp. 1887–1895, 2009. **Introduction**: The objective of this study was to identify genetic variants that are associated with adult leisure time exercise behavior using genome-wide association (GWA) in two independent samples. **Methods**: Exercise behavior was measured in 1644 unrelated Dutch and 978 unrelated American adults of European ancestry with detailed questions about type, frequency, and duration of exercise. Individuals were classified into regular exercisers or nonexercisers using a threshold of 4 MET·h (metabolic equivalents·hours per week). GWA analyses of ~1.6 million observed and imputed Single Nucleotide Polymorphism (SNP) were conducted in both samples independently using logistic regression in SNPTEST, including sex, age, and body mass index as covariates. A meta-analysis of the results was performed using the weighted inverse variance method in METAL. **Results**: Thirty-seven novel SNPs in the *PAPSS2* gene and in two intergenic regions on chromosomes 2q33.1 and 18p11.32 were associated with exercise participation (pooled P values $<1.0 \times 10^{-5}$). Previously reported associations (*ACE*, *CASR*, *CYP19A1*, *DRD2*, *LEPR*, and *MC4R* genes) or linkage findings (2p22.3, 4q28, 4q31.21 7p13, 9q31, 11p15, 13q22, 15q13, 18q12.2, 18q21.1, 19p13.3, and 20q12) were not replicated, although suggestive evidence was found for association to rs12405556 in the *LEPR* gene (pooled P value 9.7×10^{-4}; American sample, P value 9.8×10^{-5}) and for association to rs8036270 in the *GABRG3* gene (pooled P value 4.6×10^{-5}) in the linkage region 15q12–13. **Conclusions**: The heritability of leisure time exercise behavior is likely to be accounted for by many genetic variants with small effect size. These can be detected by GWA as was shown here for the *PAPSS2* gene, but larger samples with genome-wide genotypes and high-quality exercise data are needed for further progress.

Key Words: PHYSICAL ACTIVITY, SPORTS PARTICIPATION, GENETICS, GENOTYPE IMPUTATION

A sedentary lifestyle is an important risk factor for a variety of physical health problems, such as obesity, cardiovascular disease, type II diabetes, and osteoporosis (2,4,24,37). Although lack of exercise participation is generally considered as an environmental risk factor, twin studies have found that genetic factors play a substantial role in adult exercise participation (5,34), suggesting that not all individuals have the same intrinsic drive to participate and persist in exercise. On the basis of a study conducted in more than 85,000 adult twins from seven different countries, Stubbe et al. (34) reported that between 48% and 71% of the variance in adult exercise behavior is explained by genetic factors. The remaining variance is accounted for by environmental factors that are not shared within families.

The heritability of exercise behavior in adults has been well established, but not much is known about the genetic variants that are associated with this trait. So far, three genome-wide linkage studies and seven candidate gene association studies have been conducted for exercise behavior or related physical activity phenotypes. The first linkage study for exercise behavior was conducted in 767 white adults from 207 families (32). Four physical

Address for correspondence: Marleen H.M. De Moor, M.Sc., VU University Amsterdam, Department of Biological Psychology, van der Boechorststraat 1, 1081 BT, Amsterdam, The Netherlands; E-mail: mhm.de.moor@psy.vu.nl.
Submitted for publication November 2008.
Accepted for publication February 2009.

0195-9131/09/4110-1887/0
MEDICINE & SCIENCE IN SPORTS & EXERCISE®

DOI: 10.1249/MSS.0b013e3181a2f646

activity phenotypes were measured, of which three reflected daily physical activity and one past-year physical activity. For past-year physical activity, suggestive linkage ($P < 0.01$) was found on chromosomes 11p15 and 15q13.3. For daily physical activity, promising linkage ($P < 0.0023$) was found on chromosome 2p22–p16 and suggestive linkages were found for different loci on chromosomes 4q28.2, 7p11.2, 9q31.1, 13q22–q31, and 20q13.1. The second study consisted of 1030 children from 319 Hispanic American families (9). Significant linkage was found on chromosome 18q12–q21 (logarithm of odds = 4.09) for daily physical activity. The third study was conducted in 1432 adult Dutch sibling pairs from 622 families (12). Suggestive linkage was found for regular exercise participation on chromosome 19p13.3 (logarithm of odds = 2.18).

Seven studies tested for association of genetic variation in six candidate genes with exercise or physical activity phenotypes. In a study of adolescent girls (21), the calcium-sensing receptor (*CASR*) gene was associated ($P = 0.01$) with weekly hours spent on physical activities. In a study of Pima Indians (34), the leptin receptor (*LEPR*) gene was associated ($P = 0.007$) with 24-h energy expenditure and physical activity levels. In a study of 222 prepubertal boys (27), *LEPR* was also associated with physical activity energy expenditure ($P = 0.016$). In another study (31), the dopamine 2 receptor (*DRD2*) gene was associated ($P = 0.016$) with past-year physical activity in women reporting European ancestry but not in subjects reporting African ancestry. In a sample of postmenopausal women (29), the aromatase (*CYP19*) gene was associated ($P = 0.04$) with physical activity. In a study of mild hypertensives (40), the angiotensin-converting enzyme (*ACE*) gene was associated ($P = 0.001$) with leisure time physical activity. Finally, a study conducted in adults (20) showed that the melanocortin-4 receptor (*MC4R*) gene was associated ($P = 0.005$) with daily physical activity levels, independent of sex, age, and body mass index (BMI). The *MC4R* gene is located on chromosome 18, in the same region for which a significant linkage with child physical activity has been found (9).

This study is the first to report the results of a genome-wide association (GWA) study for leisure time exercise behavior, conducted in two independent samples comprising 1644 Dutch and 978 American subjects genotyped on 1,607,535 observed and imputed SNP markers that passed quality controls. The aims of the study were 1) to identify new genetic variants that are associated with leisure time exercise behavior and 2) to replicate the associations to previously reported candidate genes and linkage regions.

MATERIALS AND METHODS

Subjects

The Netherlands. Dutch data on leisure time exercise behavior were obtained from an ongoing longitudinal study (1991–2004) on health, lifestyle, and personality in twins and their family members registered at the Netherlands Twin Register (NTR) (6). A total of 1860 unrelated individuals registered at the NTR were selected to be genotyped as part of the Genetic Association Information Network (GAIN) initiative (22), of which 1703 served as controls and 160 as cases in a GWA study for major depressive disorder (GAIN-MDD) (7,36). The study was approved by the Central Ethics Committee on Research Involving Human Subjects. All subjects provided written informed consent. After exclusion of all depression cases and rigorous quality control of the genotype data (36) (see discussion later), data from 1649 individuals were left for analysis, of whom 5 did not have valid data on exercise. For the remaining 1644 individuals, we used their most recent exercise data. For 1112 individuals (67.6%), data came from a survey sent out in 2004; for 307 individuals (18.7%), data came from a survey from 2002; and for the remaining individuals, data came from earlier surveys (1991–2000). There were 620 men (37.7%) and 1024 women (62.3%). Mean age of the participants was 43.5 yr (SD = 14.6, range = 14.5–79.8 yr) at the time of the survey collection. Mean BMI (defined as weight divided by squared height) was 24.3 $kg \cdot m^{-2}$ (SD = 3.6, range = 15.8–42.0 $kg \cdot m^{-2}$).

United States of America. American data on leisure time exercise behavior were collected as part of a larger study into the genetics of common human complex diseases/traits (e.g., osteoporosis, obesity, and height) in normal healthy subjects (13). The study was approved by the necessary institutional review boards of involved institutions. Signed informed consent was obtained from all study subjects before they entered the study. A random sample consisting of 978 unrelated white subjects was identified from our ongoing study currently consisting more than 6000 individuals. All of the chosen subjects were of US whites of Northern European origin living in Omaha, NE, and its surrounding areas. The inclusion and exclusion criteria were well defined (13). Briefly, subjects with chronic diseases and conditions involving vital organs (heart, lung, liver, kidney, and brain) and severe endocrinological, metabolic, and nutritional diseases were excluded from this study. Mean age of the participants was 49.96 yr (SD = 18.3, range = 19.1–87.2 yr) at the time of the survey collection. Mean BMI was 27.3 $kg \cdot m^{-2}$ (SD = 5.2, range = 14.2–49.4 $kg \cdot m^{-2}$). There were 494 men (50.5%) and 484 women (49.5%).

Phenotypes

Leisure time exercise behavior was measured in a comparable way in the Dutch and American samples, except for a minor difference in the first question asked. In the Dutch sample, the first question was "Do you participate in exercise regularly?," and in the American question, this was "Do you take exercise for 60 minutes per week?." These questions could be answered with "Yes" or "No." If the participants

responded affirmative, they were asked to list all voluntary leisure time exercise activities and to indicate the type, frequency, and duration of each activity. Activities such as biking to work, gardening, and work-related physical activity were not counted as leisure time exercise. All selected activities were assigned a MET value according to the widely accepted Ainsworth's Compendium of physical activity (1), and the total MET·h were computed as MET hours per week. A MET score of 1 corresponds to the rate of energy expenditure when at rest (1 $kcal \cdot kg^{-1} \cdot h^{-1}$). Scores of individuals who reported no leisure time exercise activities were coded as zero. To keep consistent with existing epidemiological studies (15,34), we classified individuals as regular versus nonexercisers on the basis of a minimal threshold of at least 4 MET·h weekly. The 6-month test-retest reliability of this dichotomous measure is 0.91 (35).

In the Dutch sample, the 1772 individuals were classified as regular exercisers (878 individuals, 49.5%) or as nonexercisers (894 individuals, 50.5%). From the 878 regular exercisers, 817 individuals had valid data on the type, frequency, and duration of exercise [4–12 MET·h (N = 385, 47.1%), 13–21 MET·h (N = 218, 26.7%), 22–30 MET·h (N = 98, 12.0%), 31–39 MET·h (N = 44, 5.4%), and ≥40 MET·h (N = 72, 8.8%)]. Exercisers were, on average, leaner than nonexercisers [mean difference = 0.87, $t(1608) = 4.92$, $P < 0.001$]. There were no significant sex differences in exercise participation [$\chi^2(1) = 2.82$, $P = 0.09$].

In the American sample, the 978 individuals were classified as regular exercisers (612 individuals, 62.6%) or as nonexercisers (366 individuals, 37.4%). All the 612 regular exercisers had valid data on the type, frequency, and duration of exercise [4–12 MET·h (N = 57, 9.3%), 13–21 MET·h (N = 329, 53.6%), 22–30 MET·h (N = 194, 31.7%), 31–39 MET·h (N = 19, 3.1%), and ≥40 MET·h (N = 14, 2.3%)]. Exercisers were, on average, leaner than nonexercisers [mean difference = 1.20, $t(647) = 3.26$, $P = 0.001$]. There were no significant sex differences in exercise participation [$\chi^2(1) = 0.05$, $P = 0.83$].

Genotypes

The Netherlands. DNA was extracted from frozen whole blood samples using the Puregene DNA Isolation kit (Gentra Systems, Minneapolis, MN). All procedures were performed according to the manufacturer's protocols. Genotyping was conducted by Perlegen Sciences (Mountain View, CA) using a high-density oligonucleotide array-based platform (35). A total of 599,156 genotyped SNPs from 98.5% of all individuals participating in the GAIN-MDD study were returned. After quality control, 435,291 SNP remained, of which 427,024 were autosomal SNPs. SNPs were excluded because of gross mapping errors (1487 SNP), duplicate errors (1143 SNPs), mendelian inconsistencies (536 SNPs), minor allele frequency (MAF) <0.01 (41,495 SNPs), and missing genotypes >0.05 (156,673 SNPs) or a combination of these reasons. Population stratification effects were examined with a subset of ~127,000 SNP in linkage equilibrium. First, a nearest-neighbor approach on the basis of genome-wide IBS estimates in PLINK was used to identify sample outliers. Second, a principal components analysis was carried out in the "smartpca" module in EigenSoft to compute two principal components that contrasted individuals with some Asian or African ancestry from individuals with clear European ancestry. On the basis of these two approaches, 58 individuals were removed. More details on genotyping procedures, genotype calling, and quality control checks of SNP and samples in the GAIN-MDD sample can be found in a study by Sullivan et al. (36).

United States of America. Genomic DNA was extracted from whole human blood using a commercial isolation kit (Gentra Systems) following the protocols detailed in the kit. Genotyping with the Affymetrix Mapping 250,000 Nsp and 250,000 Sty arrays was performed using the standard protocol recommended by the manufacturer. Of the initial full set of 500,568 SNPs, we discarded 32,961 SNPs with sample call rate <95%, another 36,965 SNPs with allele frequencies deviating from Hardy–Weinberg equilibrium (HWE; $P < 0.001$), and 51,323 SNP with MAF <1%. Thus, 381,100 SNP remained. Population stratification effects were tested with the Structured Association method (26), performing nine independent analyses assuming 2, 3, or 4 population strata and a set of 200, 2000, or 6000 randomly selected unlinked markers. Most of the subjects (>98%) were tightly clustered together, with the exception of only six subjects, suggesting that there is essentially no population stratification in this sample. Exclusion of these six subjects resulted in almost identical association results (data not shown).

Genotype Imputation

To be able to compare results at the SNP level, we imputed all ~2.5×10^6 common SNP that are included in the Haplotyping Mapping Project sample of Utah residents with ancestry from Northern and Western Europe (HapMap CEU sample). Imputation was carried out in IMPUTE (23) using the HapMap phase 2 data available on the IMPUTE Web site: http://www.stats.ox.ac.uk/~marchini/software/gwas/impute.html#. IMPUTE computes the probabilities of each of the three possible genotypes for each unobserved SNP for each individual in the sample using the information of the surrounding observed genotypes for that individual and the linkage disequilibrium (LD) information available from the HapMap data. In the Dutch sample, 2,135,543 SNPs were imputed, and in the American sample, 2,184,310 SNPs were imputed.

IMPUTE also gives information about how well each SNP is imputed, quantified as the maximum posterior call averaged over all individuals for each SNP. SNPs with an average maximum posterior call below 0.80 in either sample were excluded from analysis, leading to the exclusion of 17,935 SNPs (0.7%) in the Dutch sample and

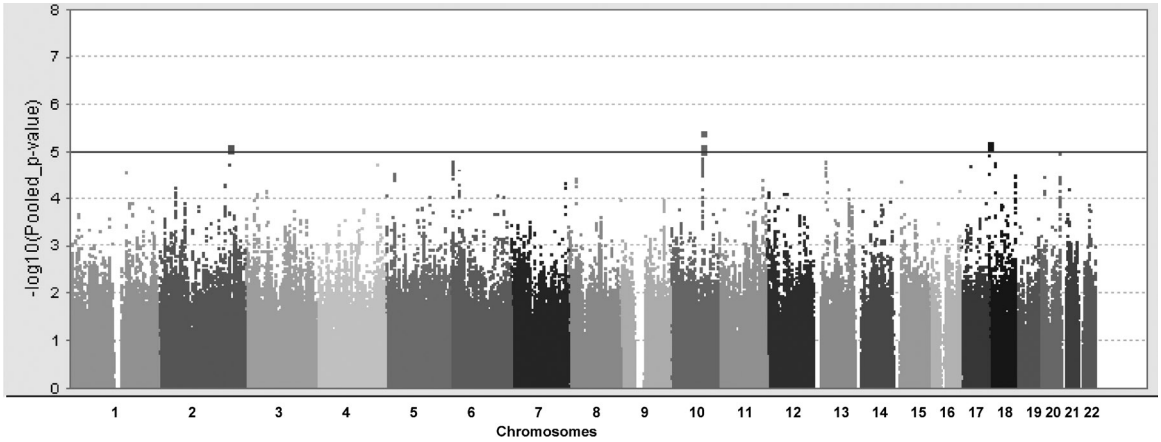

FIGURE 1—Manhattan plot of exercise participation for the pooled *P* values. *Blue line* indicates the threshold for genome-wide significant association.

165,167 SNP (6.4%) in the American sample. The maximum posterior call averaged over all individuals over all remaining imputed SNPs was 0.99 (SD = 0.023) in the Dutch sample and 0.97 (SD = 0.045) in the American sample. We further excluded within each sample all SNP with less or equal than five individuals in at least one of the genotype groups on the basis of a genotype calling threshold of 0.90 for the imputed SNP. Given the size of our data sets, this roughly corresponded to excluding all SNPs with MAF smaller than 0.05 in the Dutch sample (19.4%) and 0.07 in the American sample (31.9%). This resulted in 2,051,750 SNPs in the Dutch sample and 1,636,636 SNPs in the American sample. The overlap between the two samples was 1,607,535 SNPs, and these formed the basis of all further analyses. Of these, 281,199 SNPs were observed and 1,326,336 SNPs were imputed in the Dutch sample. In the American sample, these numbers were 303,963 and 1,303,572, respectively.

Statistical Analysis

GWA analyses were conducted using logistic regression in the SNPTEST (23). Analyses were performed in both samples independently while taking the uncertainty of the imputed genotypes into account and including sex and age as covariates. Association of each SNP with the exercise participation phenotype was tested using the 1-degree of freedom additive test. A meta-analysis of the results was performed using the weighted inverse variance method in METAL (http://www.sph.umich.edu/csg/abecasis/Metal/index.html), which computes a pooled effect estimate [ln(odds ratio)], its SE, and its *P* value by weighting the effect estimates of the individual samples by the inverse of its variance and by taking into account the direction of effect. In this method, we also automatically corrected for any population stratification effects by applying genomic control.

We used Haploview to create a Manhattan plot and WGAViewer to compute λs and to annotate selected SNP (3,14). The Manhattan plot depicts the genomic location on the *x*-axis and the -10 log (*P* value) on the *y*-axis and provides an overview of the distribution of *P* values across the genome. The λ is a measure of the enhanced number of very low *P* values on the basis of what is expected under chance. These inflated significant results may be due to the true association signals but could also be due to the population stratification effects. A λ of 1 indicates no inflation, and a $\lambda > 1$ indicates that genomic control is needed. We used a threshold of a pooled *P* value $< 1.0 \times 10^{-5}$ to identify SNPs that might contribute to the heritability of exercise behavior. To bolster our confidence in the relevance of the SNP for exercise behavior, independent of obesity, we performed an additional association analysis with BMI added as a covariate.

TABLE 1. SNP for exercise participation that reach threshold for genome-wide significant association (combined $P < 1.0 \times 10^{-5}$).

Closest Gene	Chromosome	Location of SNP	Most Significant SNP in Region[a]	Base Pair Position (kb)	Pooled *P*[a]	Pooled OR	Pooled 95% CI	Alleles[b]	*P*: Dutch Sample	*P*: American Sample	Pooled *P*: Corrected for BMI
DNAPTP6	2q33.1	Intergenic	rs12612420[c,d]	200,984	7.61×10^{-6}	1.43	1.22–1.67	**A**G	0.000881	0.00224	7.65×10^{-5}
PAPSS2	10q23.2	Intron	rs10887741[c,d]	89,433	3.81×10^{-6}	1.32	1.17–1.49	**T**C	0.00337	0.000134	6.26×10^{-6}
C18orf2	18p11.32	Intergenic	rs8097348[c,d]	1585	6.68×10^{-6}	1.36	1.19–1.56	A**G**	0.000367	0.00535	6.99×10^{-5}

[a] Other significant SNP in same regions (pooled *P* value $< 1.0 \times 10^{-5}$): Region 2q33.1: rs17592517[c,d]. Region 10q23.2: rs1980647[c,d], rs4934355[c,d], rs1358864[c,d], rs7903568[c,d], rs2077695[c,d], rs10218939[c,d], rs11202501[c,d], rs12412482[c,d], rs7908056[d]. Region 18p11.32: rs4502301[c,d], rs2345036[c,d], rs2111926[c,d], rs11080871[c,d], rs2111925[c,d], rs2160961[c,d], rs2052420[c,d], rs2052419[c,d], rs12605872[c,d], rs9646479[c,d], rs9646456[c,d], rs8095975[c,d], rs8083526[c,d], rs4389203[c,d], rs8099756[c,d], rs4797926[c,d], rs4797928[c,d], rs4797929[c,d], rs4797930[c,d], rs1869532[c,d], rs1051393[c,d], rs1869533[c,d], rs16941818[c,d], rs07229607[c,d].

[b] Effect allele in bold; the increase in one allele of the effect allele corresponds to an increased risk to be an exerciser as indicated by the OR.

[c] Imputed in the Dutch sample.

[d] Imputed in the American sample.

Guided by previous reports (8), we explicitly inspected the association of all SNP in or near (<10 kb) all candidate genes for physical activity phenotypes from previous candidate gene association studies, including the associated SNP reported in those studies if available. We used a replication strategy at the gene level, i.e., *any* SNP (not just the original SNP reported on) on a pooled *P* value reaching smaller than 1.0×10^{-5} was considered to constitute a replication "in the broad sense." We also inspected association of all SNP in the 95% confidence interval (CI) of all linkage peaks previously reported for exercise behavior and physical activity phenotypes. Again, finding at least one SNP with a pooled *P* value $< 1.0 \times 10^{-5}$ in one of these regions was considered to be a replication of the linkage signal. For all significant SNP, we performed checks for HWE (tested at $P < 1.0 \times 10^{-5}$), inspection of genotype calling cluster plots, and (if imputed) average maximum posterior call.

RESULTS

Novel gene finding. The Manhattan plot for exercise participation is given in Figure 1. This figure shows that there are some significant SNP that cluster in several regions on different chromosomes. The *l* was 0.99, showing that there are no more significant *P* values than would be expected under chance, which indicates that the genomic control was successfully applied to the combined sample in the meta-analysis procedure. Table 1 displays the results of all SNP that reach the threshold for GWA (pooled *P* value $< 1.0 \times 10^{-5}$). The odds ratio (OR) for these SNP ranged between 1.29 and 1.43. The lowest pooled

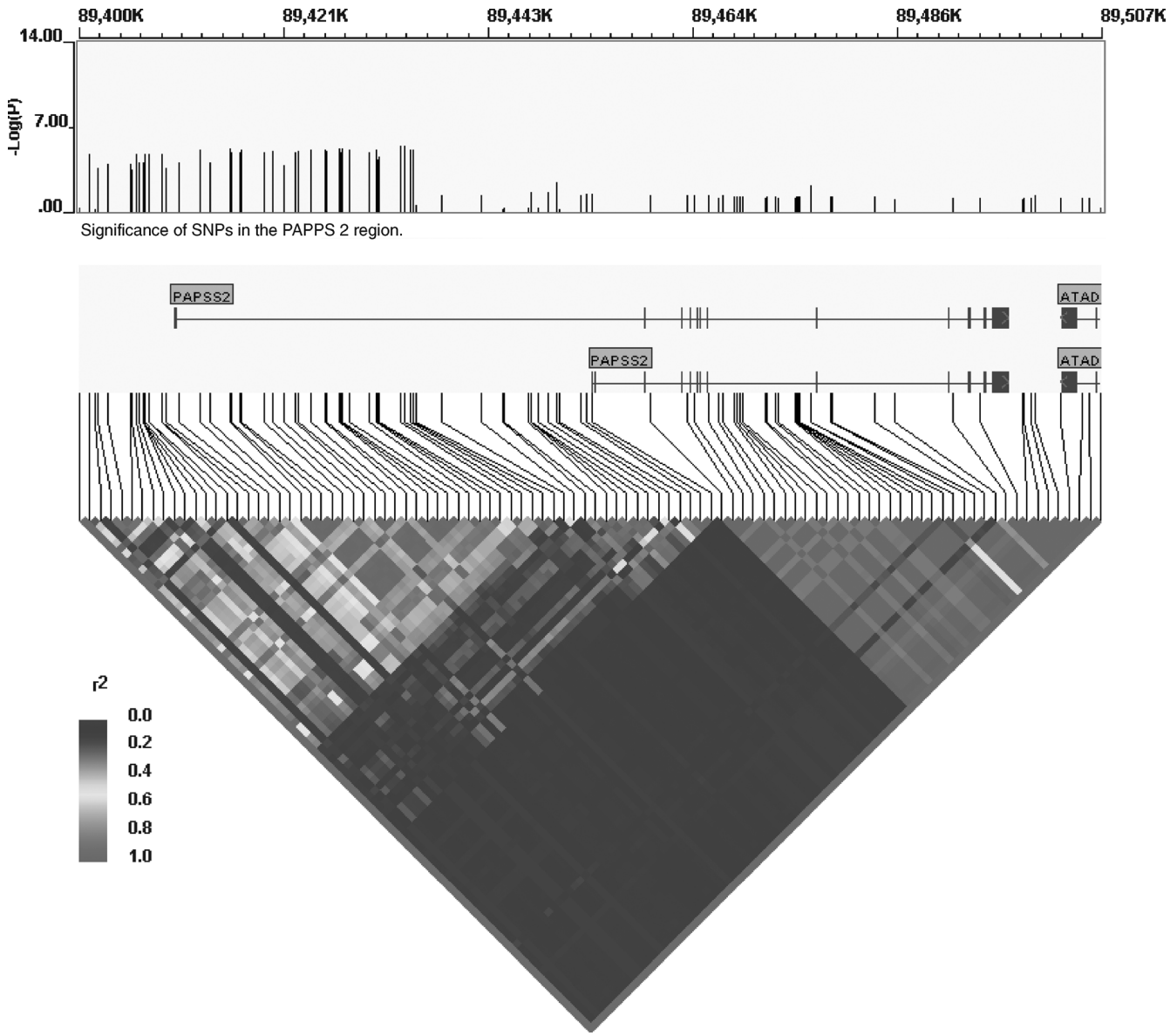

FIGURE 2—LD plot of the *PAPSS2* gene region on chromosome 10.

TABLE 2. Testing association with six candidate genes for exercise participation.

Candidate Gene	Study	Chrom	SNP with Lowest Pooled P[a]	Location of SNP	Pooled P	Pooled OR	Pooled 95% CI	Alleles[b]	No. SNP in Gene	Coverage	P: Dutch Sample	P: American Sample	Pooled P: Corrected for BMI
ACE	Winnicki et al. (24)	17q23.3	rs12451328[c,d]	Intron	0.0420	1.13	1.00–1.18	A**C**	17	0.58	0.1015	0.2234	0.1275
CASR	Lorentzon et al. (21)	3q21.1	rs9811123[c,d]	Intron	0.0284	1.16	1.02–1.32	**A**G	157	0.72	0.0285	0.4459	0.0665
CYP19A1	Salmen et al. (29)	15q21.2	rs6493487[c,d]	Intron	0.0208	1.16	1.02–1.33	**A**G	115	0.76	0.0777	0.1305	0.0679
DRD2	Simonen et al. (31)	11q23	rs1800497[c,d]	Non-Synonymous coding	0.0439	1.16	1.00–1.34	**A**G	69	0.94	0.1712	0.1214	0.0973
LEPR	Stefan et al. (33), Richert et al. (27)	1p31.3	rs12405556[d]	Intron	0.00097	1.24	1.09–1.40	**T**G	277	0.92	0.2265	9.79×10^{-5}	0.00020
MC4R	Loos et al. (20)	18q21.32	rs17066829[d]	Intergenic	0.2879	1.07	0.95–1.21	**A**T	7	0.69	0.2210	0.8692	0.3250

Sex and age were included as covariates in these analyses.
[a] In gene or 10 kb around gene.
[b] Effect allele in bold; the increase in one allele of the effect allele corresponds to an increased risk to be an exerciser as indicated by the OR.
[c] Imputed in the Dutch sample.
[d] Imputed in the American sample.
Chrom, chromosome; Coverage, number of observed SNP in gene + number of tagged common SNP ($R^2 > 0.80$) divided by the total number of common SNP in HapMap.

P value was 3.8×10^{-6} for SNP rs10887741 (OR = 1.32). This SNP is part of an LD block that includes an intron of the 3′-phosphoadenosine 5′-phosphosulfate synthase 2 (*PAPSS2*) gene and the region just upstream of this gene (Fig. 2). There are 10 of 44 SNP in this LD block that have a P value smaller than 1.0×10^{-5}. Forty SNP have a P value smaller than 1.0×10^{-3}. Two other regions with genome-wide significant SNP are located on chromosomes 2q33.1 and 18p11.32 and are intergenic. The gene closest to the two significant SNP on 2q33.1 is the DNA polymerase-transactivated protein 6 (*DNAPTP6*) gene. The other region on 18p11.32, with 25 significant SNP that are all in tight LD, is closest to the chromosome 18 open reading frame 2 (*C18orf2*) gene. Both genes are known to be protein coding, but their specific functions are unknown.

Candidate gene replication. Table 2 presents the association of SNP in six candidate genes with exercise participation. No SNP in these genes reach our *a priori* criterion for replication. The lowest pooled P value is found in the *LEPR* gene for SNP rs12405556 ($P = 9.7 \times 10^{-4}$, OR = 1.24). This SNP is associated with exercise participation in the American sample but not in the Dutch sample. SNP rs12405556 is located only 4604-bp away from the SNP that was reported to be significant in the two previous candidate gene studies (rs1137101) (27,33). SNP rs12405556 is in moderate LD with SNP rs1137101 ($r^2 = 0.44$, on the basis of the HapMap reference sample). The association of SNP rs1137101 with exercise in our sample was not significant (pooled P value = 0.15, P value American sample = 0.032, P value Dutch sample = 0.84).

Linkage region replication. Table 3 presents an overview of the most significant SNP in the linkage regions that have been reported for exercise and physical activity phenotypes in previous studies (9,12,32). None of the SNP in the linkage regions reached our *a priori* criterion for replication. The lowest pooled P value was found for SNP rs8036270, which is in an intron of the γ-aminobutyric acid (GABA) receptor subunit γ-3 precursor (*GABRG3*) gene ($P = 4.6 \times 10^{-5}$). This gene is located around 4 Mb away from the marker peak reported in the previous linkage study (32).

DISCUSSION

This study is the first to report the results from a GWA study for leisure time exercise behavior, using data from 1644 Dutch and 978 American adults of European ancestry genotyped on 1,607,535 common SNP (both observed and successfully imputed). The GWA analyses revealed 37 novel SNP for exercise participation that cluster in three

TABLE 3. Testing association in 10 linkage regions for exercise participation.

Linkage Region	Study	Marker at Peak	Flanking Markers[a]	Lowest Pooled P	SNP	Distance SNP to Marker (kb)	No. SNP in Region	Closest Gene of SNP	P: Dutch Sample	P: American Sample	Pooled P: Corrected for BMI
19p13.3	De Moor et al. (12)	*D19S247*	D19S591–D19S865	0.00649	rs11085873[b]	2.617	1269	*ADAMTS10*	0.0431	0.0648	0.0159
18q12.2	Cai et al. (9)	*D18S1102*	D18S1102–D18S474	0.000146	rs12959140[b,c]	3.155	9462	—	0.0117	0.0027	0.00070
18q21.2	Cai et al. (9)	*D18S474*	D18S474–D18S61	0.000233	rs12458537[b,c]	1.780	13472	*MALT1*	0.0019	0.0444	0.00130
2p22.3	Simonen et al. (32)	*D2S2347*	—	0.000530	rs11888625[b,c]	2.625	4588	*AC068274.1*	0.00257	0.0765	0.0057
4q31.21	Simonen et al. (32)	*UCP1*	—	0.00264	rs10440457[c]	359	3205	*MAML3*	0.000411	0.94789	0.0039
7p13-p12	Simonen et al. (32)	*IGFBP1*	—	0.000771	rs12537130[c]	1.306	4216	*TNS3*	0.003925	0.07939	0.0041
9q31	Simonen et al. (32)	*D9S938*	—	0.001358	rs947122[b]	2.273	5104	*KLF4*	0.03876	0.00920	0.0027
13q22	Simonen et al. (32)	*D13S317*	—	0.004656	rs9545319[b,c]	2.070	8768	*SPRY2*	0.01561	0.13443	0.0475
15q13	Simonen et al. (32)	*D15S165*	—	0.0000461	rs8036270[c]	4.001	6376	*GABRG3*	0.00299	0.00342	0.0003
20q12	Simonen et al. (32)	*PLCG1*	—	0.0003982	rs1548244[b]	186	6696	*TOP1*	0.00382	0.03869	

[a] In the study by Simonen et al. (32), no CI or flanking regions are given. For regions reported in this study, we used a 6-Mb interval around the peak. Note that the peak at the marker *C11P15_3* reported by Simonen et al. (32) is omitted from this table because this marker could not be identified. Sex and age were included as covariates in these analyses.
[b] Imputed in the Dutch sample.
[c] Imputed in the American sample.
—, not available.

different genomic regions (pooled P value $< 1.0 \times 10^{-5}$): in the *PAPSS2* gene on chromosome 10q23.2 and in two intergenic regions on chromosomes 2q33.1 and 18p11.32. Previously reported associations in candidate genes (*ACE*, *CASR*, *CYP19A1*, *DRD2*, *LEPR*, and *MC4R* genes) or linkage findings (2p22.3, 4q28, 4q31.21 7p13, 9q31, 11p15, 13q22, 15q13, 18q12.2, 18q21.1, 19p13.3, and 20q12) were not replicated, although suggestive evidence was found for association to rs12405556 in the *LEPR* gene (pooled P value 9.7×10^{-4}; American sample P value 9.8×10^{-5}) and for association to rs8036270 in the *GABRG3* gene (pooled P value 4.6×10^{-5}) in the linkage region 15q12–13.

The *PAPSS2* gene encodes a protein that is involved in the sulfation of compounds such as lipids, carbohydrates, proteins, and exogenous drugs. The gene is widely expressed in skeletal and smooth muscles and the brain and in several other organs. The *PAPSS2* gene has been related to skeletal development and arthrosis (16,38). The 10q23 region that harbors the *PAPSS2* gene has been linked to maximal exercise capacity in a genome-wide linkage study of 453 sib pairs (28). This suggests that exercise ability may be an important determinant of leisure time exercise behavior as has been suggested from a theoretical point of view (11). The two other regions with significant SNP (on 2q33.1 and 18p11.32) do not have a known function and have not been related to exercise before (8). Replication studies using large samples sizes and similar phenotyping should further clarify the possible role of these regions in exercise behavior.

The *LEPR* gene is expressed in the hypothalamus (25) and codes for the leptin receptor. Leptin is a hormone that has an important role in the regulation of energy balance. The *LEPR* gene has frequently been linked to obesity and type 2 diabetes mellitus and is thought to play a physiological role in the hypothalamic neuronal systems that promote positive energy balance and weight gain (18,25). The genetic effects of SNP in the *LEPR* gene were independent of BMI. This corroborates a study in a sample of 268 nondiabetic Pima Indians, in which the *LEPR* gene was found to be associated with 24-h energy expenditure and physical activity, independent of adiposity (33). This suggests that the *LEPR* gene may be involved in the drive to exercise through the hypothalamic regulation of energy balance.

The GABA is the major inhibitory neurotransmitter in the human brain where it acts at the $GABA_A$ receptors. The *GABRG3* gene is located in chromosome 15 in a cluster of two other genes that together code for subunits of the $GABA_A$ receptor. A previous linkage study found that this region was suggestively linked to past years' physical activity (32). A study on physical exercise-associated gene expression found higher expression levels of the *GABRG3* gene after a bout of exhaustive exercise but not after aerobic exercise at 60% $\dot{V}O_{2max}$ (17). These results tentatively suggest that the *GABRG3* gene may have an impact on exercise behavior through individual differences in exercise-induced fatigue.

A major limitation of the current study is the low power to detect polygenic effects of small to very small effect size, which may well prove to account for the heritability of leisure time exercise behavior. Although we combined full GWA data sets from two relatively large samples with highly similar phenotyping, and used a fairly liberal P value threshold in view of the huge number of tests, we still only identified three regions that account for only a minor portion of the heritability previously estimated for this same exercise phenotype (48%–71%) (34). This concurs with several recent GWA studies showing that the effect sizes of most loci that are identified for complex human diseases so far are small, and very large samples including 10 thousands of individuals instead of thousands of individuals are needed to detect these effects (10,19,39). A second limitation of our study is that parts of the relevant genomic variation may not have been captured. Variation in the final SNP set used may not have tagged all of the common variation in the genome, and the GWA chip used in the Dutch sample did not allow reliable extraction of another set of genetic variants, copy number variation, which may be important for exercise. Moreover, our analyses were restricted to the autosomal genome, which excludes genes on the sex chromosomes and the mitochondrial DNA. The mitochondrial DNA is especially of interest because of its well-known involvement in energy metabolism and adenosine triphosphate generation (30).

To conclude, this study found an association between genetic variants in the *PAPSS2* gene and two intergenic regions and leisure time exercise behavior. In the next phase, a large collaborative GWA consortium is needed to identify the full palette of genetic variants influencing leisure time exercise behavior.

De Moor and Liu are joint first authors.
De Geus and Deng are joint last authors.

For the Dutch sample, we acknowledge financial support from the Netherlands Organization for Scientific Research (NWO): Database Twin Register (575-25-006), Twin-family database for behavior genetics and genomic studies (480-04-004), genetic basis of anxiety and depression (904-61-090); resolving cause and effect in the association between exercise and well-being (904-61-193); twin-family database for behavior genomics studies (480-04-004); twin research focusing on behavior (400-05-717); Center for Medical Systems Biology (NWO Genomics); Spinozapremie (SPI 56-464-14192); Centre for Neurogenomics and Cognitive Research (CNCR-VU); genome-wide analyses of European twin and population cohorts (EU/QLRT-2001-01254); genome scan for neuroticism (National Institute of Mental Health (NIMH) grant R01 MH059160); Geestkracht program of ZonMW (10-000-1002); matching funds from universities and mental health care institutes involved in NESDA (GGZ Buitenamstel-Geestgronden, Rivierduinen, University Medical Center Groningen, GGZ Lentis, GGZ Friesland, GGZ Drenthe). Genotyping was funded by the Genetic Association Information Network (GAIN) of the Foundation for the US National Institutes of Health, and analysis was supported by grants from GAIN and the NIMH (MH081802). Genotype data were obtained from dbGaP (http://www.ncbi.nlm.nih.gov/dbgap, accession number phs000020.v1.p1). Statistical analyses

were carried out on the Genetic Cluster Computer (http://www.geneticcluster.org), which is financially supported by the NWO (480-05-003).

For the American sample, we acknowledge financial support from the National Institutes of Health (NIH) of USA (R01 AR050496, R21 AG027110, R01 AG026564, P50 AR055081, and R21 AA015973). The study was also benefited from grants from National Science Foundation of China, Huo Ying Dong Education Foundation, HuNan Province, Xi'an Jiaotong University, and the Ministry of Education of China.

The results of this study do not constitute endorsement by ACSM.

REFERENCES

1. Ainsworth BE, Haskell WL, Whitt MC, et al. Compendium of physical activities: an update of activity codes and MET intensities. *Med Sci Sports Exerc*. 2000;32(Suppl 9):S498–516.
2. Albright A, Franz M, Hornsby G, et al. Exercise and type 2 diabetes. *Med Sci Sports Exerc*. 2000;32(7):1345–60.
3. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005; 21(2):263–5.
4. Berlin JA, Colditz GA. A meta-analysis of physical activity in the prevention of coronary heart disease. *Am J Epidemiol*. 1990; 132(4):612–28.
5. Beunen G, Thomis M. Genetic determinants of sports participation and daily physical activity. *Int J Obes Relat Metab Disord*. 1999;23:S55–63.
6. Boomsma DI, De Geus EJC, Vink JM, et al. Netherlands Twin Register: from twins to twin families. *Twin Res Hum Genet*. 2006;9(6):849–57.
7. Boomsma DI, Willemsen G, Sullivan PF, et al. Genome-wide association of major depression: description of samples for the GAIN Major Depressive Disorder Study: NTR and NESDA biobank projects. *Eur J Hum Genet*. 2008;16(3):335–42.
8. Bray MS, Hagberg JM, Perusse L, et al. The human gene map for performance and health-related fitness phenotypes: the 2006–2007 update. *Med Sci Sports Exerc*. 2009;41(1):34–72.
9. Cai GW, Cole SA, Butte N, et al. A quantitative trait locus on chromosome 18q for physical activity and dietary intake in Hispanic children. *Obes Res*. 2006;14(9):1596–604.
10. Craddock N, O'Donovan MC, Owen MJ. Genome-wide association studies in psychiatry: lessons from early studies of non-psychiatric and psychiatric phenotypes. *Mol Psychiatry*. 2008; 13(7):649–53.
11. De Geus EJC, De Moor MHM. A genetic perspective on the association between exercise and mental health. *Mental Health Phys Act*. 2008;1(2):53–61.
12. De Moor MHM, Posthuma D, Hottenga JJ, Willemsen AHM, Boomsma DI, De Geus EJC. Genome-wide linkage scan for exercise participation in Dutch sibling pairs. *Eur J Hum Genet*. 2007;15:1252–9.
13. Deng HW, Deng HY, Liu YJ, et al. A genomewide linkage scan for quantitative-trait loci for obesity phenotypes. *Am J Hum Genet*. 2002;70(5):1138–51.
14. Ge DL, Zhang KL, Need AC, et al. WGAViewer: software for genomic annotation of whole genome association studies. *Genome Res*. 2008;18(4):640–3.
15. Haase A, Steptoe A, Sallis JF, Wardle J. Leisure-time physical activity in university students from 23 countries: associations with health beliefs, risk awareness, and national economic development. *Prev Med*. 2004;39(1):182–90.
16. Ikeda T, Mabuchi A, Fukuda A, et al. Identification of sequence polymorphisms in two sulfation-related genes, *PAPSS2* and *SLC26A2*, and an association analysis with knee osteoarthritis. *J Hum Genet*. 2001;46(9):538–43.
17. Kawai T, Morita K, Masuda K, et al. Physical exercise-associated gene expression signatures in peripheral blood. *Clin J Sport Med*. 2007;17(5):375–83.
18. Liu YJ, Rocha-Sanchez SMS, Liu PY, et al. Tests of linkage and/or association of the *LEPR* gene polymorphisms with obesity phenotypes in Caucasian nuclear families. *Physiological Genomics*. 2004;17(2):101–6.
19. Loos RJF, Lindgren CM, Li SX, et al. Common variants near *MC4R* are associated with fat mass, weight and risk of obesity. *Nat Genet*. 2008;40(6):768–75.
20. Loos RJF, Rankinen T, Tremblay A, Perusse L, Chagnon Y, Bouchard C. Melanocortin-4 receptor gene and physical activity in the Quebec Family Study. *Int J Obes Relat Metab Disord*. 2005;29(4):420–8.
21. Lorentzon M, Lorentzon R, Lerner UH, Nordstrom P. Calcium sensing receptor gene polymorphism, circulating calcium concentrations and bone mineral density in healthy adolescent girls. *Eur J Endocrinol*. 2001;144(3):257–61.
22. Manolio TA, Rodriguez LL, Brooks L, et al. New models of collaboration in genome-wide association studies: the Genetic Association Information Network. *Nat Genet*. 2007;39(9):1045–51.
23. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet*. 2007;39(7):906–13.
24. Martinez-Gonzalez MA, Martinez JA, Hu FB, Gibney MJ, Kearney J. Physical inactivity, sedentary lifestyle and obesity in the European Union. *Int J Obes Relat Metab Disord*. 1999;23(11): 1192–201.
25. Park KS, Shin HD, Park BL, et al. Polymorphisms in the leptin receptor (*LEPR*)—putative association with obesity and T2DM. *J Hum Genet*. 2006;51(2):85–91.
26. Pritchard JK, Rosenberg NA. Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet*. 1999;65(1):220–8.
27. Richert L, Chevalley T, Manen D, Bonjour JP, Rizzoli R, Ferrari S. Bone mass in prepubertal boys is associated with a Gln223Arg amino acid substitution in the leptin receptor. *J Clin Endocrinol Metab*. 2007;92(11):4380–6.
28. Rico-Sanz J, Rankinen T, Rice T, et al. Quantitative trait loci for maximal exercise capacity phenotypes and their responses to training in the HERITAGE Family Study. *Physiol Genomics*. 2004;16(2):256–60.
29. Salmen T, Heikkinen AM, Mahonen A, et al. Relation of aromatase gene polymorphism and hormone replacement therapy to serum estradiol levels, bone mineral density, and fracture risk in early postmenopausal women. *Ann Med*. 2003;35(4): 282–8.
30. Saxena R, de Bakker PIW, Singer K, et al. Comprehensive association testing of common mitochondrial DNA variation in metabolic disease. *Am J Hum Genet*. 2006;79(1):54–61.
31. Simonen RL, Rankinen T, Perusse L, et al. A dopamine D_2 receptor gene polymorphism and physical activity in two family studies. *Physiol Behav*. 2003;78(4–5):751–7.
32. 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.
33. 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;26:1629–32.

34. Stubbe JH, Boomsma DI, Vink JM, et al. Genetic influences on exercise participation: a comparative study in adult twin samples from seven countries. *PLoS ONE*. 2006;1(1):e22. doi:10.1371/journal.pone.0000022. Available from: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0000022.
35. Stubbe JH, De Moor MHM, Boomsma D, De Geus EJ. The association between exercise participation and well-being: a co-twin study. *Prev Med*. 2007;44(2):148–52.
36. Sullivan PF, De Geus EJC, Willemsen G, et al. Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. *Mol Psychiatry*. 2009;14(4):359–75.
37. Todd JA, Robinson RJ. Osteoporosis and exercise. *Postgrad Med J*. 2003;79(932):320–3.
38. Ul Haque MF, King LM, Krakow D, et al. Mutations in orthologous genes in human spondyloepimetaphyseal dysplasia and the brachymorphic mouse. *Nat Genet*. 1998;20(2):157–62.
39. Weedon MN, Lango H, Lindgren CM, et al. Genome-wide association analysis identifies 20 loci that influence adult height. *Nat Genet*. 2008;40(5):575–83.
40. Winnicki M, Accurso V, Hoffmann M, et al. Physical activity and angiotensin-converting enzyme gene polymorphism in mild hypertensives. *Am J Med Genet*. 2004;125A(1):38–44.