



Published in final edited form as:

Med Sci Sports Exerc. 2018 August ; 50(8): 1620–1628. doi:10.1249/MSS.0000000000001607.

Genetic Determinants for Leisure-Time Physical Activity

Xiaochen Lin^{1,2}, Katie Kei-hang Chan^{1,2,3}, Yen-Tsung Huang^{1,4,5}, Xi Luo⁴, Liming Liang⁶, James Wilson⁷, Adolfo Correa⁸, Daniel Levy⁹, and Simin Liu^{1,2,10}

¹Department of Epidemiology, Brown University, Providence, RI

²Center for Global Cardio-metabolic Health, Brown University, Providence, RI

³Hong Kong Institute of Diabetes and Obesity, Department of Medicine and Therapeutics, the Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, Hong Kong Special Administrative Region, China

⁴Department of Biostatistics, Brown University, Providence, RI

⁵Institute of Statistical Science, Academia Sinica, Taipei, Taiwan

⁶Department of Epidemiology and Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA

⁷Department of Physiology and Biophysics and Department of Medicine, University of Mississippi Medical Center, Jackson, MS

⁸The Jackson Heart Study and the Departments of Medicine and Pediatrics, University of Mississippi Medical Center, Jackson, MS

⁹The Framingham Heart Study, Framingham, MA and the Population Sciences Branch, Nation Heart, Lung, and Blood Institute of the National Institute, Bethesda, MD

¹⁰Division of Endocrinology, Department of Medicine, Alpert School of Medicine and Rhode Island Hospital, Providence, RI

Abstract

Purpose—Leisure-time physical activity (LTPA) is a well-established modifiable lifestyle determinant for multiple cardio-metabolic outcomes. However, current understanding of the genetic architecture that may determine LTPA remains very limited. Therefore, we aimed to examine the role of genetic factors in affecting LTPA, which has yet to be investigated comprehensively and in-depth.

Methods—We conducted a genome-wide analysis using 1000 Genomes Project imputed data from the Women’s Health Initiative (n=11,865), the Jackson Heart Study (n=3,015) and the Framingham Heart Study (n=7,339). A series of secondary analyses, including candidate gene

Corresponding Author: Simin Liu, M.D., Sc.D.; Professor of Epidemiology and Medicine; Department of Epidemiology and Medicine, Brown University; 121 South Main St, Providence, RI 02903; Simin_liu@brown.edu.

Disclosures:

The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institute of Health; or the U.S. Department of Health and Human Services. No conflicts of interest exist.

analysis, sequence kernel association tests, pathway analysis, functional annotation and expression quantitative trait loci analysis, were performed to follow up on the primary findings.

Results—Ethnicity-specific genetic signals were investigated respectively for African Americans (AA) and European Americans (EA). Two variants, rs116550874 (meta-analysis: $P = 1.63 \times 10^{-7}$) and rs3792874 (meta-analysis: $P = 8.33 \times 10^{-7}$), were associated with LTPA in AA; rs28524846 (meta-analysis: $P = 1.30 \times 10^{-6}$) was identified for EA. We also replicated four previously reported loci (*GABRG3*, *CYP19A1*, *PAPSS2* and *CASR*; P for lead SNPs < 0.005). Further fine-mapping and functional annotation suggested that several identified loci (novel and replicated) are involved in 1) the homeostatic drive coupled with the reward system and 2) the development and regulation of the capacity to perform LTPA.

Conclusions—To our knowledge, our analysis is the first to comprehensively investigate the genome-wide signals for LTPA in multiple ethnicities. These findings support the notion that genetic predisposition plays a critical role in determining LTPA, of which the biological and clinical implications warrants further investigation.

Keywords

Genetic determinants; leisure-time physical activity; cardio-metabolic health; lifestyle intervention

INTRODUCTION

Leisure-time physical activity (LTPA) is a well-established modifiable lifestyle factor related to multiple cardio-metabolic outcomes, including obesity, type 2 diabetes, and cardiovascular diseases(1). Various psychological, biological, social, and environmental correlates of LTPA in humans have been identified(2), and many studies have examined the extent to which these correlates influence LTPA. Nevertheless, the role of genetic factors in affecting LTPA has yet to be investigated comprehensively and in-depth.

Twin studies have estimated the heritability for LTPA in the range from 35% to 83%(3). Recent work of candidate gene analyses(4–9), linkage studies(10–12), and genome-wide association study (GWAS)(13), primarily among European-derived populations, have identified eight potentially relevant genes: the leptin receptor gene (*LEPR*), the melanocortin-4 receptor gene (*MC4R*), the dopamine 2 receptor gene (*DRD2*), the gamma-aminobutyric acid type A receptor-gamma 3 (*GABRG3*), cytochrome P450 family 19 subfamily A member 1 (*CYP19A1*), the calcium-sensing receptor gene (*CASR*), the 3'-Phosphoadenosine 5'-Phosphosulfate Synthase 2 gene (*PAPSS2*), and the angiotensin-converting enzyme gene (*ACE*). The only GWAS that has been conducted to date used a binary variable for leisure-time exercise participation (≥ 4 metabolic equivalents (MET)-hour per week vs. < 4 MET-hour per week) and was based on a HapMap imputation training set with only $\sim 2.5 \times 10^{-6}$ SNPs(13). Because the phenotype of LTPA is likely to be determined by a large number of genetic variants with modest independent effects, statistical power of this GWAS was apparently insufficient in detecting these potential signals of importance (Dutch sample: $n=1644$; US sample: $n=978$). With the emerging techniques of genotyping, imputation and next-generation sequencing, along with the increasing resources

allocated to genetic studies, the ability to investigate the genetic architecture underlying LTPA has substantially improved.

Our current understanding of the genetic architecture underpinning LTPA is still limited, especially compared to other phenotypes, such as obesity and diabetes. Further, relative to European Americans, other ethnic groups are disproportionately affected by physical inactivity and yet under-represented in almost all genetic studies. Defining the underlying genetic architecture responsible for physical inactivity in multiple ethnicities will not only generate novel insights into the etiology of many complex diseases related to insufficient PA and potential new molecular targets for lifestyle interventions, but will also have significant implications for the development of personalized and precision lifestyle intervention programs. Thus, we aimed to identify the genetic components of human LTPA by augmenting standard GWAS with the 1000 Genomes Project imputed data from multiple well-established cohorts, advanced statistical analysis methodologies, comprehensive functional annotation and strategies of systems biology.

METHODS

Study Participants

The WHI is the major national study of women's health that began in 1993 when 161,808 postmenopausal women of diverse ethnicities, aged 50 to 79 years were enrolled and followed. The WHI is still ongoing with funding committed by NIH until 2018. The WHI SNP Health Association Resource (WHI-SHARe) was conducted in ~13,000 AA and Hispanic American (HA) participants who had consented. Among them, a random subsample of 8,515 AA and 3,642 HA women were selected for the genetic study due to budget constraints. DNA samples were collected at the time of enrollment and extracted by the Specimen Processing Laboratory at the Fred Hutchinson Cancer Research Center. In addition, the participants of the WHI Genomics and Randomized Trials Network (WHI-GARNET) were women enrolled in the WHI Hormone Therapy (HT) Trial who meet eligibility requirements for this study and eligibility for submission to dbGaP, and who provided DNA samples. In total, the WHI-GARNET included ~4,900 participants, predominantly EA. All participants of the WHI-SHARe and the WHI-GARNET provided written informed consent as approved by local human subjects committees. After quality control and excluding those participants without data on LTPA, the current study includes 8,092 AA participants of the WHI-SHARe (WHI-AA) and 3,773 EA participants of the WHI-GARNET (WHI-EA) with LTPA data as the discovery samples for AA and EA respectively.

The JHS is a community-based, observational study that consists of 5,301 African American women and men aged between 35 and 84 years, who were followed from the baseline examination in 2000–2004 and at the 12-month intervals after the baseline. Major components of each exam include medical history, physical examination, blood/urine analytes and interviews. In total, 3,001 AA with available GWAS and LTPA data are included for the proposed research. The JHS was used for the replication and validation of the findings in the WHI-AA.

The FHS is a large prospective, community-based, three-generation family-based cohort study. The Original Cohort consists of 5,209 participants. The Offspring Cohort consists of 5,124 participants and includes biological descendants of the Original Cohort as well as spouses and adopted offspring. The Third Generation Cohort consists of 4,095 participants and is the biological descendants of the Offspring Cohort and adopted offspring. After excluding the participants with no genetic data or LTPA data, 6,911 EA participants are included in the current study. The FHS was used for the replication and validation of the discoveries in the WHI-EA, accounting for the family structure by using generalized estimation equation models with each pedigree treated as a cluster.

The current study includes four study populations: the African Americans participated in the WHI (WHI-AA, n=8,092), the Jackson Heart Study (JHS, n=3,001), the European Americans participated in the WHI (WHI-EA, n=3,773), and the Framingham Heart Study (FHS, n=6,911). Each study was approved by the Institutional Review Boards (IRBs). Detailed information regarding demographics, genotyping and imputation for each study population are provided in Supplementary Table 1 (see Table, SDC1, Summary of genotyping, imputation and quality control procedures in each cohort). WHI-AA and WHI-EA were used in the discovery stage (Stage 1), and in the validation stage (Stage 2) the ethnicity-specific SNPs discovered in Stage 1 were examined in JHS and FHS respectively. The confirmed SNPs will be taken forward to the ethnicity-specific meta-analysis by using the data from the discovery and the validation cohorts combined.

Phenotype Measurements

LTPA levels were ascertained by questionnaires and interviews in each cohort. For WHI, detailed questionnaires were administered to collect data on different types of activities (strenuous, moderate, and mild) done in leisure time, including the frequency and duration. Examples of strenuous activities included activities that result in increased heart rate and sweating, such as aerobics, aerobic dancing, jogging, tennis, and swimming laps. Examples of moderate activity included biking outdoors, using an exercise machine (such as stationary bicycle or a treadmill), calisthenics, easy swimming, and popular or folk dancing. Examples of mild exercise were slow dancing, bowling, and golf. Standard METs were taken from the most current version of the national compendium(14) and assigned to different activity categories. Then the total leisure-time physical activity related energy expenditure (LTPA-EE) was obtained by multiplying standard METs for each activity by its frequency and duration, summing over activities, and then multiplying the sum by the body weight(15, 16). For the JHS participants, the 40-item Jackson Heart Physical Activity Cohort (JPAC) survey was administered by trained interviewers to assess physical activity over the past 12 months(17). The JPAC provides index scores summarized for different domains of physical activity, and the Sport and Exercise Index, which captures LTPA, was computed by taking the sum of the product of intensity and time engaged in each activity and then mapped onto a scale with lower scores indicating lower activity. For FHS, physical activity during the past year was assessed using questionnaires. Participants were asked about their time spent on different categories of physical activity on an average day. Time spent at each activity in hours per week was recorded. Then a composite score, the Physical Activity Index (PAI), was calculated by summing up the products of hours at each activity level multiplied by a

weight based on required oxygen consumptions required for that activity(18). Questionnaires used for PA assessment have been validated in the original cohorts(17, 19). Although each LTPA measure used for the study cohorts has been validated previously, there is no perfect correspondence across cohorts. To accommodate this issue, we will use a P -value based approach, detailed in **Statistical Analysis**, for results synthesis across cohorts.

Genotyping and Imputation

Study samples were genotyped on either Illumina or Affymetrix platforms. Each study imputed the genetic data to the 1000 Genomes Project. The details regarding genotyping and imputation procedures for each study are provided in Supplementary Table 1 (see Table, SCD1, Summary of genotyping, imputation and quality control procedures in each cohort). Several quality control filters have been applied to the genotyped and imputed data from each study. In brief, single nucleotide polymorphisms (SNPs) with a low sample call rate, a low SNP call rate, a minor allele frequency < 0.01 , or violating Hardy-Weinberg Equilibrium were excluded, and the details regarding specific filters in each study are also provided in Supplementary Table 1 (see Table, SDC1, Summary of genotyping, imputation and quality control procedures in each cohort). Markov Chain Haplotyping software, MACH(20), was adopted for imputation. For each participating cohort, only SNPs with sufficient imputation quality scores (r^2 of MACH > 0.3) were included. In total, 13,892,960 and 8,258,952 SNPs were included in the discovery GWAS for AA and EA respectively.

Statistical Analysis

The primary analysis has two analytical stages. Stage 1 is the discovery stage where standard GWAS were conducted separately for WHI-AA and WHI-EA. Stage 2 is the validation stage where the ethnicity-specific SNPs discovered in Stage 1 were examined in JHS and FHS respectively. The confirmed SNPs were taken forward to the ethnicity-specific meta-analyses by using the data from the discovery and the validation cohorts combined. Subsequently, we performed an extensive series of follow-up analyses to explore the potential pathways and further fine map the genomic areas of interest.

Stage 1—Standard GWAS was conducted for the WHI-AA ($n=8,092$) and WHI-EA ($n=3,773$), respectively. Values of LTPA-EE were natural logarithm-transformed before analysis, and outliers (> 3 S.D.) were removed if existed. Assuming an additive genetic mode, linear regression models, adjusting for age, region and the first four principal components for population stratification, were used to determine the association of each SNP with LTPA-EE. For WHI-EA, the analysis was adjusted for the matching factors (age and hysterectomy status) used for selecting participants into the original cohorts, in addition to the first four principal components. Genomic control correction was applied to each ancestral group separately. Ethnicity-specific candidate SNPs reached the suggestive significance level of $P < 1 \times 10^{-5}$ were identified and taken forward to the subsequent ethnicity-specific meta-analysis in Stage 2.

Stage 2—Ethnicity-specific SNPs discovered in Stage 1 were then examined in JHS and FHS as the validation samples by fitting linear regression models, adjusting for age, sex and the first four principal components. For FHS, the family structure was additionally

accounted for by using generalized estimation equation models with each pedigree treated as a cluster. The ethnicity-specific candidate SNPs passed the threshold of $P < 0.05$ in the validation sample were then taken forward to the ethnicity-specific meta-analysis. To account for the fact that each cohort used slightly different LTPA measures, the ethnicity-specific meta-analyses, combining WHI-AA with JHS and WHI-EA with FHS, were performed using the P -value based method, taking into account the sample size and the direction of effect in METAL(21). We also conducted a sequence kernel association test (SKAT) based analysis to test for association between the genetic variants in each novel region discovered in the primary analysis as a set and LTPA levels, adjusting for covariates. SKAT, as a score-based variance-component test, has the advantage to quickly calculate p values analytically by fitting the null model containing only the covariates(22).

Follow-up Analyses—In addition to the SNPs discovered in the current study, SNPs located in eight genes (*LEPR*, *CASR*, *PAPSS2*, *DRD2*, *GABRG3*, *CYP19A1*, *ACE* and *MC4R*) reported in previous studies(4–9, 13) were also examined by using the data available in the current study. The replicated loci, along with all novel SNPs identified were further scrutinized based on previous evidence and biological plausibility.

As another follow-up analysis, we fine mapped the potentially relevant regions (within 300 kb upstream and downstream to the lead SNP confirmed in the current study). Linkage-disequilibrium (LD) information from the 1000 Genomes Project reference populations (AFR and EUR) were incorporated in the fine-mapping analysis of the top signals identified. LocusTrack (<http://gump.qimr.edu.au/general/gabrieC/LocusTrack/>) was used to generate the regional association and LD plot and annotate the genomic regions of interest(23). Publicly available expression quantitative trait loci (eQTLs) data from the Genotype-Tissue Expression (GTEx) Project (<http://gtexportal.org/home/>) was accessed to examine tissue-specific eQTLs(24).

To explore the biological mechanisms involving those genes located near the confirmed SNPs, a hypothesis-based pathway analysis was performed using MAGENTA(25). Using curated biological pathways from Gene Ontology databases, Protein Analysis Through Evolutionary Relationships (PANTHER), Ingenuity, Kyoto Encyclopedia of Genes and Genomes (KEGG), Reactome and Biocarta, we tested for the enrichment of genetic signals in pathways that contain at least one gene from the loci confirmed in the current study. Specifically, a corrected gene association P value for each gene was calculated in the genome based on the most significant P value of SNPs within 110 kb upstream and 40 kb downstream to each gene's most extreme transcript start and end sites, adjusting for gene size, number of SNPs per kb, number of recombination hotspots per kb, LD units per kb and genetic distance. For each pathway, the potential enrichment of highly ranked gene scores was evaluated by comparing the proportions of genes within each set whose corrected P values is more significant than the 75th percentile of gene P values of 10,000 randomly sampled gene sets of identical size from the genome.

RESULTS

Stage 1 Discovery Stage

We firstly performed the ethnicity-specific GWAS of LTPA-EE for both AA and EA participants in the WHI. The demographic characteristics of the discovery samples used in Stage 1, WHI-AA and WHI-EA, are summarized according to quartiles of LTPA-EE in Supplementary Table 2 [see Table, SDC2, Characteristics of WHI-SHARe African American participants (N=8,092) according to quartiles of leisure time physical activity related energy expenditure] **and** Supplementary Table 3 [see Table, SDC3, Characteristics of WHI-GARNET European American participants (N=3,773) according to quartiles of leisure time physical activity related energy expenditure]. In total, 8,092 African Americans and 3,773 European Americans are included in the current study. The results from the GWAS are shown in the Manhattan plots (see Figure, SDC4, Stage 1 Manhattan plots for the African Americans (Panel A; N=8,092) and the European America). Stage 1 analysis revealed one significant ($P < 5 \times 10^{-8}$) and 130 suggestive ($P < 1 \times 10^{-5}$) SNPs related to LTPA-EE for the African Americans and 369 suggestive ($P < 1 \times 10^{-5}$) SNPs for the European Americans participated in the WHI [see Table, SDC5, Suggestive SNPs ($P < 1 \times 10^{-5}$) related to LTPA-EE identified in the Stage 1 GWAS for WHI-AA; see Table, SDC6, Suggestive SNPs ($P < 1 \times 10^{-5}$) related to LTPA-EE identified in the Stage 1 GWAS for WHI-EA].

Stage 2 Validation Stage

In Stage 2, we firstly examined the identified ethnicity-specific SNPs (131 in WHI-AA and 369 in WHI-EA) in the JHS (n=3,001) and the FHS (n=6,911) participants. The detailed demographic characteristics of JHS and FHS are summarized in Supplementary Table 6 [see Table, SDC7, Characteristics of JHS participants (N=3,001) according to simple sport score] **and** Supplementary Table 7 [see Table, SDC8, Characteristics of FHS participants (N=6,911) according to quartiles of physical activity index]. Among the 131 SNPs discovered for AA in Stage 1, 5 SNPs reached the threshold of $P < 0.05$ in the JHS; among the 369 SNPs discovered for EA, 100 SNPs concentrated on Chromosome 14 reached the threshold of $P < 0.05$ in the FHS. Then we conducted the ethnicity-specific meta-analyses for those SNPs, combining JHS with WHI-AA and FHS with WHI-EA. In total, there are 11,093 (8,092+3,001) participants included in the ethnicity-specific meta-analysis for African Americans and 10,684 (3,773+6,911) participants for European Americans. For African Americans, all five SNPs (rs116550874, rs3792874, rs3792877, rs3792878, rs79173796) showed suggestive ($P < 1 \times 10^{-5}$) evidence with a combined meta-analysis P value (P values = 1.63×10^{-7} , 8.33×10^{-7} , 8.48×10^{-7} , 8.61×10^{-7} , and 8.49×10^{-7}) (Table 1). The effect of each SNP was in the same direction in the discovery sample and the validation sample. For European Americans, we found genome-wide suggestive evidence ($P < 1 \times 10^{-5}$) for 90 SNPs in high LD on Chromosome 14 (Table 1). The lead SNP was rs28524846 with a combined P value of 1.30×10^{-6} . For all the confirmed SNPs, the directions of the effects were consistent across WHI-EA and FHS.

Candidate Gene Analysis

A candidate gene analysis was performed on eight candidate genes (*LEPR*, *CASR*, *PAPSS2*, *DRD2*, *GABRG3*, *CYP19A1*, *ACE* and *MC4R*) that have been reported previously for

human physical activity. SNPs in several previously reported loci, including *GABRG3*, *CYP19A1*, *PAPSS2* and *CASR*, reached $P < 5 \times 10^{-3}$ in the meta-analysis (Table 2). In particular, SNPs in two loci, *GABRG3* and *CASR*, were found associated with LTPA for both ethnic groups.

Pathway Analysis

To further explore the underlying biological pathways, we assessed whether genetic signals are enriched in pathways that contain at least one gene from the loci confirmed in the primary analysis and previously reported, using multiple databases of curated biological pathways (Gene Ontology databases, PANTHER, Ingenuity, KEGG, Reactome and Biocarta). Pathways with evidence for enrichment ($P < 0.05$) are summarized in Supplementary Table 8 [see Table SDC9, Gene set enrichment analysis (MAGENTA) for African Americans of biological pathways with at least one gene from the confirmed loci; see Table, SDC10, Gene set enrichment analysis (MAGENTA) for European Americans of biological pathways with at least one gene from the confirmed loci]. Some of the identified loci include genes with established connections to glucose metabolisms, including glucose transport, glucose transmembrane transporter and glycolysis. In addition, we also found suggestive evidence of enrichment for other pathways, such as ATPase activity coupled to transmembrane movement of substances, apoptosis, NKT pathway and oxidoreductase activity.

Fine-mapping

To further fine map and functionally annotate our findings, we interrogated the regions around the top SNPs identified from the primary analysis. Genomic regions with concentrated association signals (within 300 kb upstream and downstream to the lead SNP) are shown in Figure 1. In the plot for each locus of interest, the upper panel shows the associated SNPs colored by their correlation to the lead variant, which was calculated on the 1000 Genome Project reference population (AFR for AA; EUR for EA). The lower panel corresponds to a region 10-time zoom into the upper panel region. The gene track shows the gene models from the GENCODE version 19. The other annotation tracks in the lower panel, which highlight the genomic region that probably contains the causal variant(s), are displayed to indicate potential functionality. Figure 1A shows that the region around the lead SNP (rs3792874) identified among AA contains several transcription factor binding sites. The computationally inferred chromatin state by the ENCODE/Broad indicates that the SNP is very close to a promoter region in human embryonic stem cells (hESC), where a histone modification site is present. Figure 1B shows the LD block around the lead SNP (rs28524846) identified in EA covers probable transcription factor binding sites and enhancer/promoter regions. By examining the available eQTL data from the GTEx Portal, we found significant associations of rs28524846 with the expression levels of *MPP5* and *APT6VID* in the nerve tissue (Figure 2A and 2B). For two confirmed SNPs from previously reported candidate genes, we found evidence for differential gene expression in the skeletal muscle tissue: rs1819162 is associated with the expression level of the *PAPSS2* gene (Figure 2C); rs62020072 appears to be significantly associated with *AP4E1* expression level (Figure 2D).

DISCUSSION

The genetic architecture underlying human LTPA remain challenging to delineate. By conducting the first ethnicity-specific meta-analysis of GWAS of LTPA, we identified three novel loci and confirmed four previously reported loci. The genes located in these loci provide intriguing insights into the potential mechanisms that regulate human LTPA.

It has been speculated that the heritability of PA behavior reflects genes influencing individual differences in the engagement of 1) the homeostatic drive coupled with the reward system for physical activity (similar to eating and drinking) and 2) the development and regulation of the capacity to perform physical activity(3) (Figure 3A). This speculation has been partly supported by previous genetic studies. Among previously reported genes, the *GABRG3* gene could influence the reward or aversive experience of physical activity through dopaminergic actions directly(3, 12, 13, 26); the *CYP19A1* gene has been associated to insulin resistance and obesity(27); *PAPSS2* has been reported to be involved in initial skeletal development, which may influence the capacity to perform physical activity later in life, and interestingly it has also been related to longevity in a recent study(28, 29).

The current study substantially deepens and expands our understanding of those mechanistic pathways potentially implicated in the regulation of LTPA. In the replication of previously reported genes, we confirmed the effects of *GABRG3*, *CYP19A1*, *PAPSS2* and *CASR* on LTPA, with multiple SNPs located in those loci reached $P < 0.005$. The analysis also yielded that another gene in close vicinity of *CYP19A1*, *AP4E1*, may also relate LTPA levels (rs62020072: P value from the meta-analysis for AA = 1.57×10^{-3} ; Table 2). This gene encodes a component of the adaptor protein complex 4. This protein complex mediates the processes of vesicle formation and also sorting of integral membrane proteins, which are both essential for the trafficking of neurotransmitters. Evidence has linked the *AP4E1* gene to diseases such as cerebral palsy, which essentially is neuromuscular mobility impairment, and also the development of the Alzheimer's disease(30–32). Although no novel variants shared by AA and EA were discovered from the current genome-wide single-SNP association analysis, it does not necessarily mean that the regulatory mechanisms of LTPA are completely different across ethnic groups. In fact, two previously reported genes (*GABRG3* and *CASR*) were replicated in both AA and EA, which support the notion that the genetic architecture of human LTPA may be partly shared by different ethnic groups.

In addition, we discovered three new loci that may potentially affect human LTPA through those speculated mechanisms. rs116550874 is located upstream (< 1 kb) to the *ENO1* gene and downstream (< 1 kb) to the *ENO1-AS1* gene (*ENO1* antisense RNA1). The *ENO1* gene encodes a glycolytic enzyme, which has been related to glucose metabolism. The pathway analysis in the current study also found that association signals were enriched in the pathway of glycolysis, which contains the *ENO1* gene. rs3792874 is located in the intron region of the *SLC22A4* gene, which has been associated with joint erosion and rheumatoid arthritis(33, 34). rs28524846 is an intron variant located on the *MPP5* gene. The encoded proteins participate in cell polarization and regulate myelinating Schwann cells, which wrap around axons of motor and sensory neurons to form the myelin sheath, essentially a protective covering serving as an electrical insulator. Myelination greatly increases nerve

conduction velocity and saves energy(35). Another gene covered by the LD block around rs28524846, *ATP6VID*, is related to synaptic vesicle cycle and ATPase activity, a pathway for which the formal pathway analysis in the current study also showed evidence of enrichment. In addition, by examining the eQTL data from the GTEx Portal, we found differential expression levels of *MPP5* and *ATP6VID* associated with the lead SNP rs28524846 in the nerve tissue, which further support the notion that variants in these genes may contribute to the regulation of human PA by affecting neuro-muscular junction, which is critical for motor ability. In summary, our findings, corroborated by previous studies, suggest that human LTPA may be regulated by the genetic factors through 1) homeostasis and neural reward system, both of which are under the constant control of the nervous system; 2) the structural and metabolic constraints exerted on the capacity to actually perform LTPA. These potential mechanisms are illustrated in Figure 3A and 3B.

There are several potential limitations to this study. First, gene-environment and gene-gene interaction is not considered in the current study. The discovery samples from the WHI are community-living ambulatory post-menopausal women, and the results may have limited generalizability to men if there is gender-difference in the genetic effects on LTPA. Compared to other complex diseases or phenotypes, such as obesity and type 2 diabetes, our knowledge of PA genetics is relatively limited. In fact, our study is the first one investigating the genetic components of LTPA among different ethnicities on the genome-wide scale, and the one with the largest sample size to date. Serving as one of the first attempts to provide genome-wide evidence, our study was aimed to explore the genetic component of human physical activity, which generates potentially interesting hypotheses for future investigations. Second, as in some GWAS for other complex diseases/phenotypes, not all previously suspected loci were replicated in the current study even though our study successfully confirmed several candidate genes, supporting the potential engagement of the homeostatic drive coupled with the reward system as well as the structural and metabolic constraints on the capacity to perform physical activity. This reflects the limited power of the current meta-analysis to fully clarify the genetic components of LTPA and call for individual GWAS from additional cohorts and further fine mapping of the identified signals. Collaborative research across more cohorts and the use of sequence data will facilitate further investigations. Third, some genes located in the vicinity of the SNPs identified by our study have no obvious biological links to the regulation of PA. However, we cannot exclude the possibility that they may have effects on the genes from the region and influence LTPA levels through the biological mechanisms that have not been well established. Fourth, there is no perfect correspondence between LTPA measures across cohorts. However, this issue should be addressed as possible by the use of the *P*-value based method adopted by METAL in the meta-analysis. Additionally, all the instruments used for PA measurement in WHI, JHS, and FHS have been objectively validated and shown to have reasonably high quality. Further, because all germline mutations/variants were objectively measured before PA assessment, any measurement error in the genetics-LTPA relation should be non-differential and independent, as such will not lead to spurious findings. Another limitation is the relatively high age range of the study populations. It has been shown that the heritability of physical activity declines with age, therefore, the current findings may not be generalizable to younger populations.

In conclusion, utilizing GWAS-based meta-analyses and 1000 Genomes Project imputed data from two ancestral groups in multiple large well-established cohorts, we identified several potential loci for human LTPA at the genome-wide scale. Most of the identified and confirmed SNPs are located in loci with biologically plausible candidate genes involved in different facets of human physical activity. Our study is a first step to exploit PA genetics to inform future personalized intervention strategies, although it is important to note that the putative genes identified in the current study require substantial further investigation and validation before firm links can be established.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Acknowledgements and Funding:

The results of the study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation, and statement that results of the present study do not constitute endorsement by ACSM.

The Jackson Heart Study is supported by contracts HHSN268201300046C, HHSN268201300047C, HHSN268201300048C, HHSN268201300049C, and HHSN268201300050C from the National Heart, Lung, and Blood Institute and the National Institute on Minority Health and Health Disparity. The Framingham Heart Study is conducted and supported by the National Heart, Lung, and Blood Institute in collaboration with Boston University (Contract No. N01 HC25195 and N01 HHSN268201500001). The Women's Health Initiative program is funded by the National Heart, Lung and Blood Institute; National Institutes of Health, US Department of Health and Human Services, through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C. The authors thank the participants and data collection staff of the Jackson Heart Study, the Framingham Heart Study, and the Women Health Initiative.

References

1. US Department of Health and Human Services. Physical activity guidelines advisory committee report. Vol. 2008. Washington, DC: US Department of Health and Human Services; 2008.
2. Bauman AE, Reis RS, Sallis JF, Wells JC, Loos RJ, Martin BW. Correlates of physical activity: why are some people physically active and others not? *Lancet*. 2012; 380(9838):258–71. [PubMed: 22818938]
3. De Geus EJ, De Moor MH. Genes, exercise, and psychological factors. Genetic and molecular aspects of sport performance. 2011:294–305.
4. Loos RJ, 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 (Lond)*. 2005; 29(4):420–8. [PubMed: 15597110]
5. Lorentzon M, Lorentzon R, Lerner UH, Nordstrom P. Calcium sensing receptor gene polymorphism, circulating calcium concentrations and bone mineral density in healthy adolescent girls. *European journal of endocrinology / European Federation of Endocrine Societies*. 2001; 144(3):257–61.
6. 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. *Annals of medicine*. 2003; 35(4):282–8. [PubMed: 12846271]
7. Simonen RL, Rankinen T, Perusse L, et al. A dopamine D2 receptor gene polymorphism and physical activity in two family studies. *Physiology & behavior*. 2003; 78(4–5):751–7. [PubMed: 12782232]
8. Stefan N, Vojarova 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. *International*

journal of obesity and related metabolic disorders : journal of the International Association for the Study of Obesity. 2002; 26(12):1629–32.

9. Winnicki M, Accurso V, Hoffmann M, et al. Physical activity and angiotensin-converting enzyme gene polymorphism in mild hypertensives. *Am J Med Genet A*. 2004; 125A(1):38–44. [PubMed: 14755464]
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 (Silver Spring)*. 2006; 14(9):1596–604. [PubMed: 17030971]
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. *European journal of human genetics : EJHG*. 2007; 15(12):1252–9. [PubMed: 17700629]
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. [PubMed: 12900690]
13. 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. [PubMed: 19727025]
14. Ainsworth BE, Haskell WL, Herrmann SD, et al. 2011 Compendium of Physical Activities: a second update of codes and MET values. *Med Sci Sports Exerc*. 2011; 43(8):1575–81. [PubMed: 21681120]
15. Sallis JF, Haskell WL, Wood PD, et al. Physical activity assessment methodology in the Five-City Project. *American journal of epidemiology*. 1985; 121(1):91–106. [PubMed: 3964995]
16. Pereira MA, FitzerGerald SJ, Gregg EW, et al. A collection of Physical Activity Questionnaires for health-related research. *Med Sci Sports Exerc*. 1997; 29(6 Suppl):S1–205. [PubMed: 9243481]
17. Smitherman TA, Dubbert PM, Grothe KB, et al. Validation of the Jackson Heart Study Physical Activity Survey in African Americans. *Journal of physical activity & health*. 2009; 6(Suppl 1):S124–32. [PubMed: 19998858]
18. Kannel WB, Sorlie P. Some health benefits of physical activity. The Framingham Study. *Archives of internal medicine*. 1979; 139(8):857–61. [PubMed: 464698]
19. Albanes D, Conway JM, Taylor PR, Moe PW, Judd J. Validation and comparison of eight physical activity questionnaires. *Epidemiology*. 1990; 1(1):65–71. [PubMed: 2081242]
20. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic epidemiology*. 2010; 34(8):816–34. [PubMed: 21058334]
21. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010; 26(17):2190–1. [PubMed: 20616382]
22. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet*. 2011; 89(1):82–93. [PubMed: 21737059]
23. Cuellar-Partida G, Renteria ME, MacGregor S. LocusTrack: Integrated visualization of GWAS results and genomic annotation. *Source code for biology and medicine*. 2015; 10:1. [PubMed: 25750659]
24. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, Hasz R, Walters G, Garcia F, Young N, Foster B. The Genotype-Tissue Expression (GTEx) project. *Nature genetics*. 2013; 45(6):580–5. [PubMed: 23715323]
25. Segre AV, Groop L, Mootha VK, Daly MJ, Altshuler D. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS genetics*. 2010; 6(8)
26. Lin X, Eaton CB, Manson JE, Liu S. The Genetics of Physical Activity. *Curr Cardiol Rep*. 2017; 19(12):119. [PubMed: 29046975]
27. Coban N, Onat A, Guclu-Geyik F, Can G, Erginel-Unaltuna N. Sex- and Obesity-specific Association of Aromatase (CYP19A1) Gene Variant with Apolipoprotein B and Hypertension. *Archives of medical research*. 2015; 46(7):564–71. [PubMed: 26415088]
28. Iida A, Simsek-Kiper PO, Mizumoto S, et al. Clinical and radiographic features of the autosomal recessive form of brachyolmia caused by PAPSS2 mutations. *Human mutation*. 2013; 34(10):1381–6. [PubMed: 23824674]

29. Yerges-Armstrong LM, Chai S, O'Connell JR, et al. Gene Expression Differences Between Offspring of Long-Lived Individuals and Controls in Candidate Longevity Regions: Evidence for PAPS2 as a Longevity Gene. *The journals of gerontology Series A, Biological sciences and medical sciences*. 2016; 71(10):1295–9.
30. Abou Jamra R, Philippe O, Raas-Rothschild A, et al. Adaptor protein complex 4 deficiency causes severe autosomal-recessive intellectual disability, progressive spastic paraplegia, shy character, and short stature. *Am J Hum Genet*. 2011; 88(6):788–95. [PubMed: 21620353]
31. Moreno-De-Luca A, Helmers SL, Mao H, et al. Adaptor protein complex-4 (AP-4) deficiency causes a novel autosomal recessive cerebral palsy syndrome with microcephaly and intellectual disability. *J Med Genet*. 2011; 48(2):141–4. [PubMed: 20972249]
32. Burgos PV, Mardones GA, Rojas AL, et al. Sorting of the Alzheimer's disease amyloid precursor protein mediated by the AP-4 complex. *Dev Cell*. 2010; 18(3):425–36. [PubMed: 20230749]
33. Tokuhira S, Yamada R, Chang X, et al. An intronic SNP in a RUNX1 binding site of SLC22A4, encoding an organic cation transporter, is associated with rheumatoid arthritis. *Nature genetics*. 2003; 35(4):341–8. [PubMed: 14608356]
34. Han TU, Lee HS, Kang C, Bae SC. Association of joint erosion with SLC22A4 gene polymorphisms inconsistently associated with rheumatoid arthritis susceptibility. *Autoimmunity*. 2015; 48(5):313–7. [PubMed: 25707686]
35. Waxman SG, Bennett MV. Relative conduction velocities of small myelinated and non-myelinated fibres in the central nervous system. *Nature: New biology*. 1972; 238(85):217–9. [PubMed: 4506206]

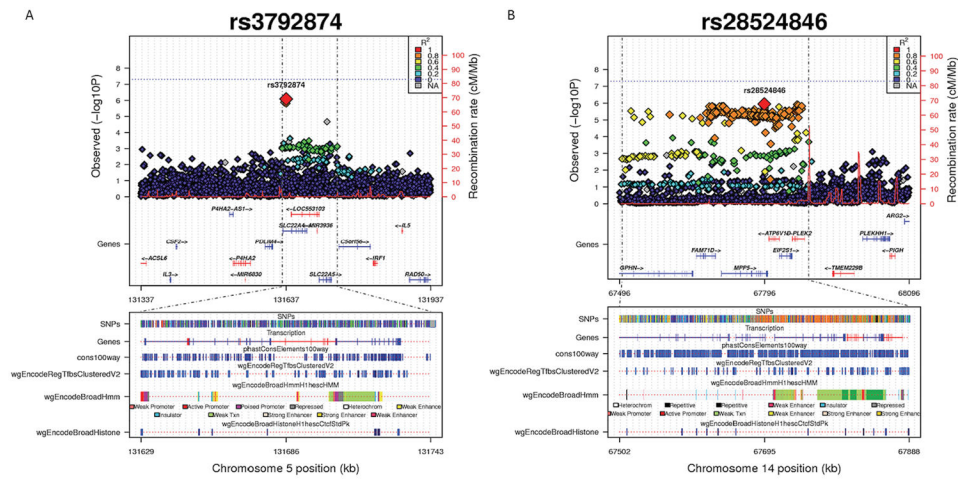


Figure 1. Regional association, linkage disequilibrium (LD) and functional annotation of the ethnicity-specific loci identified

The upper panel of each plot shows the associated SNPs colored by their correlation to the lead variant, which was calculated on the 1000 Genome Project population (AFR for AA; EUR for EA). The lower panel corresponds to a region 10X zoom into the upper panel region. The gene track shows the gene models from the GENCODE version 19. The other annotation tracks in the lower panel illustrate regulatory elements, including transcription factor binding site clusters, inferred chromatin state and histone modification sites in human embryonic stem cells. The SKAT-based analysis P -values for the genetic variants within ± 300 kb of rs3792874 are 0.004 for WHI-AA and 0.023 for JHS; the SKAT-based analysis P -values for the genetic variants within ± 300 kb of rs28524846 are 0.004 for WHI-EA and 0.384 for FHS.

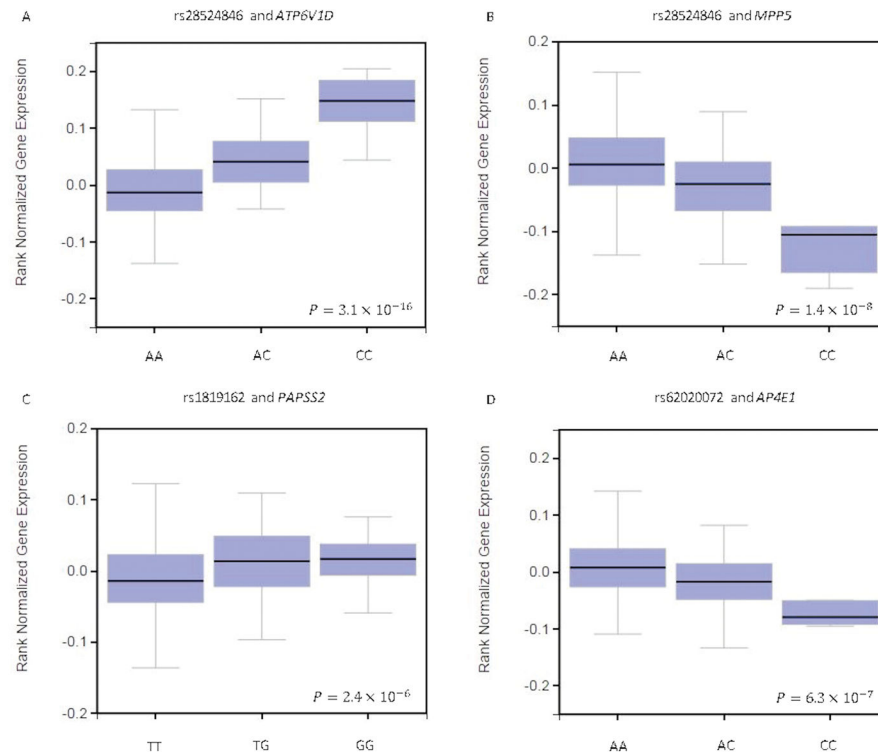


Figure 2. Significant associations between the genotypes of identified SNPs and gene expression levels

The associations and the P values are shown according to the data from the GTEx Portal (<http://gtexportal.org/home/>). A: association between rs28524846 and the expression levels of *ATP6V1D*; B: association between rs28524846 and the expression levels of *MPP5*; C: association between rs1819162 and the expression levels of *PAPS2*; D: association between rs62020072 and the expression levels of *AP4E1*.

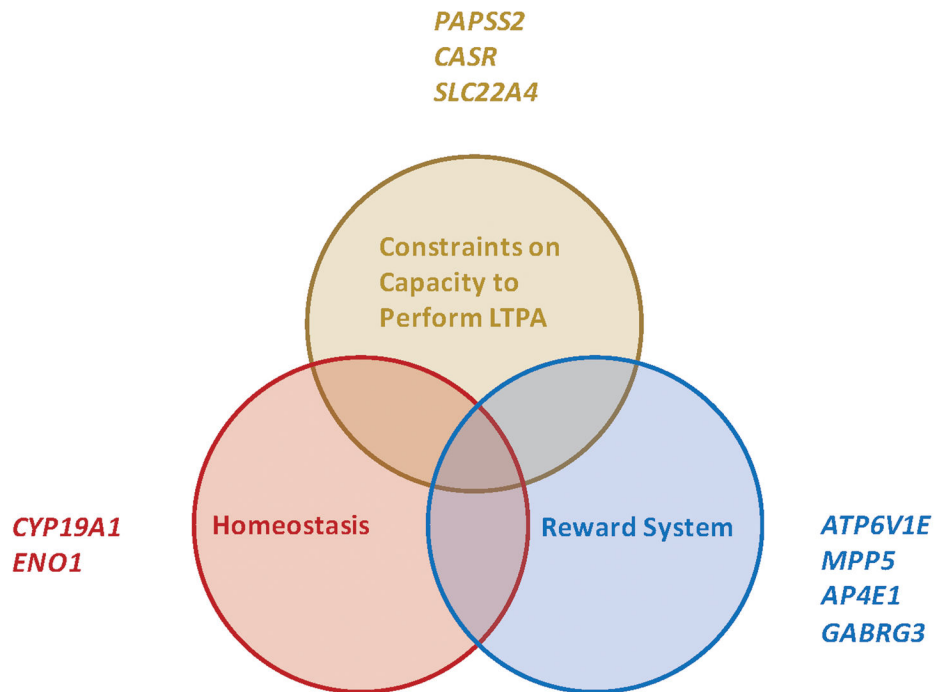


Figure 3. Potential mechanisms perturbed by genetic variation affecting LTPA levels in humans Genes replicated and identified in the current study (*PAPSS2*, *CASR*, *SLC22A4*, *ATP6V1E*, *MPP5*, *AP4E1*, *GABRG3*, *CYP19A1*, and *ENO1*) may affect the homeostasis, the neural reward system, and the structural and metabolic constraints exerted on the capacity to perform LTPA.

Table 1

Stage 1 and Stage 2 results for the lead ethnicity-specific SNPs at loci that were associated with LTPA.

Lead SNP	CHR	Position	MAF	Genes ¹	Effect allele	Ref. allele	Stage 1: Discovery			Stage 2: Validation			Ethnicity-Specific Meta-analysis		
							β	S.E.	P value	β	S.E.	P value	z-score	P value	Direction
African Americans (Discovery: WHI-AA; Replication: JHS) ²															
rs116550874	1	8940666	0.01	RERE ENO1 ENO1-AS1 CA6 SLC2A7 SLC2A5 GPR157	T	A	1.72	0.35	9.92×10 ⁻⁷	0.41	0.20	4.20×10 ⁻²	-5.24	1.63×10 ⁻⁷	--
rs3792874 ³	5	131636975	0.13	ACSL6 LOC101927693 IL3 CSF2 P4HA2 LOC101927705 PDLIM4 SLC22A4 LOC553103 SLC22A5 C5orf56 LOC101927732 IRF1 IL5 RAD50	A	G	0.31	0.07	6.32×10 ⁻⁶	0.06	0.03	3.95×10 ⁻²	4.93	8.33×10 ⁻⁷	++
European Americans (Discovery: WHI-EA; Replication: FHS) ⁴															
rs28524846 ⁵	14	67796184	0.06	GPHN FAM71D MPP5 ATP6VID EIF2S1 PLEK2 TMEM229B PIGH LOC100419668 ARG2	A	C	0.81	9.18	9.10×10 ⁻⁶	1.09	0.40	6.16×10 ⁻³	4.84	1.30×10 ⁻⁶	++

¹ Genes within ± 300 kb of the lead SNPs.

² *r*² = 8092 and 3001 for WHI-AA and JHS.

³ At this locus, three other SNPs (rs3792877, rs3792878 and rs79173796) in the LD block around the lead SNP also have a P value < 1×10⁻⁵.

⁴ *r*² = 3773 and 6911 for WHI-EA and FHS.

At this locus, 89 other SNPs in the LD block around the lead SNP also have a P value $< 1 \times 10^{-5}$.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Replication of previously identified loci^{1/}.

Candidate Gene	SNP	CHR	Position	MAF	Effect allele	Ref. allele	Stage 1			Stage 2			Ethnicity-Specific Meta-analysis					
							β	S.E.	P value	β	S.E.	P value	z-score	P value	Direction			
African Americans (WHI-AA+JHS) ²																		
<i>GABRG3</i>	rs72707657	15	27725497	0.07	T	C	-0.23	0.11	3.55×10^{-2}	-0.13	0.04	1.27×10^{-3}	3.47	5.16×10^{-4}	++			
<i>GABRG3</i>	rs12438610	15	27189768	0.08	A	G	-0.22	0.09	1.10×10^{-2}	-0.07	0.03	2.49×10^{-2}	3.34	8.40×10^{-4}	++			
<i>GABRG3</i>	rs12902711	15	27222824	0.14	A	G	0.17	0.07	1.10×10^{-2}	0.06	0.03	4.47×10^{-2}	3.23	1.22×10^{-3}	++			
<i>CYP19A1</i>	rs62020072	15	51475652	0.05	T	G	0.29	0.10	3.86×10^{-3}	0.06	0.04	1.82×10^{-1}	3.16	1.57×10^{-3}	++			
<i>PAPSS2</i>	rs1819162	10	89527210	0.19	A	C	-0.17	0.06	2.96×10^{-3}	-0.03	0.02	2.21×10^{-1}	3.17	1.50×10^{-3}	++			
<i>CASR</i>	rs7650960	3	121876304	0.27	C	G	0.10	0.05	5.47×10^{-2}	0.06	0.02	5.19×10^{-3}	3.09	1.97×10^{-3}	++			
<i>CASR</i>	rs112909877	3	121883987	0.26	A	G	0.11	0.06	4.09×10^{-2}	0.05	0.02	3.72×10^{-2}	2.83	4.66×10^{-3}	++			
European Americans (WHI-EA+FHS) ³																		
<i>GABRG3</i>	rs12595253	15	27312290	0.08	A	G	0.25	0.12	4.04×10^{-2}	1.19	0.44	6.45×10^{-3}	-3.41	6.53×10^{-4}	--			
<i>CASR</i>	rs14655373	3	121985666	0.12	A	G	0.17	0.10	9.12×10^{-2}	0.80	0.30	8.97×10^{-3}	-3.11	1.90×10^{-3}	--			
<i>CASR</i>	rs55716378	3	12202958	0.16	T	C	0.26	0.09	2.50×10^{-3}	NA	NA	NA	-3.02	2.50×10^{-3}	-?			

^{1/} SNPs with a P value < 0.005.

^{2/} n= 8092 and 3001 for WHI-AA and JHS.

^{3/} n= 3773 and 6911 for WHI-EA and FHS.