Received: 12 March 2022 | Revised: 13 November 2022 | Accepted: 29 November 2022

DOI: 10.1111/gbb.12834

ORIGINAL ARTICLE

Genes, Brain and Behavior

Genome-wide assessment reveals a significant association between *ACSS3* and physical activity

Jinyeon Jo[1] | Youngkyu Song[2] | Dankyu Yoon[3] | Chung Gun Lee[4,5] | Sungho Won[1,6,7]

[1]Department of Public Health Sciences, Institute of Health & Environment, Seoul National University, Seoul, South Korea

[2]NGenebio Corps, Seoul, South Korea

[3]Division of Endocrine and Kidney Disease Research, Department of Chronic Disease Convergence Research, National Institute of Health, Korea Disease Control and Prevention Agency, Cheongju, South Korea

[4]Department of Physical Education, College of Education, Seoul National University, Seoul, South Korea

[5]Institute of Sport Science, Seoul National University, Seoul, South Korea

[6]Institute of Health and Environment, Seoul National University, Seoul, South Korea

[7]RexSoft Corps, Seoul, South Korea

Correspondence
Jinyeon Jo, Department of Public Health Sciences, Institute of Health & Environment, Seoul National University, Seoul 08826, Korea.
Email: jyjo43043@gmail.com

Sungho Won, Department of Public Health Sciences, Seoul National University, 1, Gwanak-ro, Gwanak-gu, Seoul 08826, South Korea.
Email: won1@snu.ac.kr

Abstract

Recent genetic studies have identified physical activity (PA)-susceptible loci in European ancestry subjects; however, due to considerable genetic differences, these findings are not likely extendable to East Asian populations. Therefore, the present study aimed to identify significantly associated PA-susceptible loci using genome-wide association studies (GWASs) with East Asian (EAS) subjects and to generalize the findings to European (EUR) ancestries. The mRNA levels of genes located near the genome-wide significantly associated single-nucleotide polymorphisms (SNP) were compared under PA and control conditions. Rs74937256, located in ACSS3 (chromosome 12), which primarily functions in skeletal muscle tissues, was identified as a genome-wide significant variant ($P = 6.06 \times 10^{-9}$) in EAS. Additionally, the rs2525840, also in *ACSS3* satisfied the Bonferroni corrected significance ($P = 3.77 \times 10^{-5}$) in EUR. We found that rs74937256 is an expressed trait locus of *ACSS3* ($P = 10^{-4}$), and *ACSS3* mRNA expression significantly differs after PA, based on PrediXcan ($P = 7 \times 10^{-8}$) and the gene expression omnibus database ($P = 0.043$).

KEYWORDS

ACSS3, genome-wide association study, ketone bodies, physical activities, single-nucleotide polymorphism

1 | INTRODUCTION

A moderate level of physical activity (PA) is a requisite for a healthy lifestyle. However, PA has also been defined as a risk predictor of fatal diseases associated with metabolic syndrome.[1–7] In fact, brain-related diseases, particularly cerebrovascular diseases, are inversely associated with PA.[8–11] It is speculated that genetic pleiotropy may account for the frequently observed associations between high baseline PA and subsequent reduced mortality.[12]

Multiple studies have evaluated the importance of genetic components in PA and have shown it to be affected by various genetic factors. For instance, heritability estimation of leisure time exercise was 0.013 [standard error (SE) 0.023] in Japanese populations,[13] and that of PA was 0.046 (SE 0.002) using the UK Biobank data.[14] Furthermore, multiple genome-wide association studies (GWASs) identified variants in *APOE* as associated with greater moderate-to-vigorous PA, *CADM2* for habitual PA,[14] and *DRD2* for sporting activities.[15] Although recent studies have found a genetic predisposition for PA in

individuals of with European (EUR) ancestry, few reports have screened disease susceptibility loci of PA in East Asians (EAS).

Furthermore, multiple studies reported that PA have a greater effect on brain-related diseases, including neurodegenerative diseases or mental illnesses,[9,16] compared with other health-related outcomes. However, although cerebrovascular diseases are fatal forms of brain diseases, there is a current dearth of data regarding whether the genetic factors associated with PA impact these diseases.

The current study seeks to identify loci associated with PA using the Korean Genome Epidemiology Study (KoGES), and to elucidate the biological mechanisms underlying the roles of these loci. Additionally, the genetic effects of PA on cerebrovascular diseases are assessed within this population. Importantly, given that trends in genetic expression associated with PA can differ according to the PA definition, hereafter PA is defined as any bodily movement produced by skeletal muscles that results in energy expenditure and opposes an inactive lifestyle.

2 | MATERIALS AND METHODS

2.1 | Study population

Discovery analyses were conducted using the KoGES cohort (https://nih.go.kr/contents.es?mid=a50401010100) comprising participants residing in Ansan (an urban area) and Ansung (a rural area) in Gyeonggi Province, South Korea.[17] It was designed to investigate the long-term follow-up of genetic, environmental, and behavioral risk factors of common complex diseases and causes of death in Koreans.[18] The baseline survey was completed during 2001–2002. A total of 10,030 participants (4758 men and 5272 women) aged 40–69 years were recruited and followed up for 14 years. Each participant was assessed every 2 years. The number of measurements for each subject ranged from one to eight. All participants provided written informed consent, and the study was approved by the ethics committee of the Korean Center for Disease Control and Institutional Review Boards of the Korea University Ansan Hospital and Ajou University School of Medicine (IRB No. E2011/001-014).

Generalization analyses were conducted using the United Kingdom Biobank (UKB) prospective cohort (https://www.ukbiobank.ac.uk).[19] Over 500,000 participants between the ages of 40 and 69 years were recruited across the UK to collect their genetic and phenotypic data ranging from social demographic to environmental factors. The baseline data survey began in 2006 and follow-up survey is ongoing. All participants provided electronically signed consent upon recruitment.

2.2 | PA definition

We define PA as the daily sum of bodily movements produced by skeletal muscles resulting in energy expenditure. In the KoGES, subjects provided answers on a questionnaire regarding the number of hours in a day spent in high-intensity level activities ranging from labor to exercise. Among eight times measurements, the PA measurement in second stage was absent. The mean age at the baseline was 52.22 (SD = 8.91) and PA levels gradually decreased throughout follow-up, approaching 0 h at the last measurement. The subject-specific average PA was calculated and transformed using the rank-based inverse normalization method (Table 1).

In the UKB, the data field 22,035 indicates whether each participant met the 2017 UK PA guidelines of 150 min of moderate activity or 75 min of vigorous (or shorter time of intense vigorous) activity, per week, and was considered as a response variable in the current study. Subjects belonging to the physically active group ($n = 172,435$) and physically inactive group ($n = 144,515$) were defined as cases and controls, respectively.

2.3 | Genotyping, quality control, and imputation

The KoGES cohort was genotyped using the Affymetrix Genome-Wide Human SNP Array 5.0 (Santa Clara, CA, USA). SNPs were removed if the missing genotype call rates were >0.05, minor allele

TABLE 1 Demographic summary of response and covariates

Traits	Male		Female		Total	
Phenotype	Mean ± SD	*N*	Mean ± SD	*N*	Mean ± SD	*N*
Stage 1: PA	2.60 ± 2.07	3757	2.91 ± 1.94	4347	2.77 ± 2.01	8104
Stage 2: PA	NA	NA	NA	NA	NA	NA
Stage 3: PA	1.03 ± 2.19	2903	0.68 ± 1.86	3321	0.84 ± 2.03	6224
Stage 4: PA	0.20 ± 0.97	1547	0.05 ± 0.39	1415	0.13 ± 0.76	2962
Stage 5: PA	0.22 ± 0.92	1535	0.09 ± 0.44	1431	0.15 ± 0.73	2966
Stage 6: PA	0.13 ± 0.71	1410	0.02 ± 0.24	1367	0.09 ± 0.54	2777
Stage 7: PA	0.21 ± 0.79	1377	0.08 ± 0.40	1342	0.15 ± 0.63	2719
Stage 8: PA	0.14 ± 0.66	1450	0.03 ± 0.32	1355	0.09 ± 0.53	2805
PA	1.51 ± 1.62	3817	1.58 ± 1.56	4435	1.54 ± 1.59	8252
INV-PA	−0.07 ± 1.03	3817	0.04 ± 0.93	4435	−0.01 ± 0.98	8252

Abbreviations: PA, average PA score measured in hours; INV-PA, Rank-based inverse-normalized transformed PA.

frequencies (MAFs) were <0.05, or the *P*-values for Hardy–Weinberg equilibrium (HWE) were < 10^{-5}. Subjects were excluded if they had missing genotype call rates >0.05 or sex inconsistency was observed. The remaining genotypes were imputed with IMPUTE2[20] using the cosmopolitan reference panel from the 1000 Genome Project (1000G) Phase 3.[21] For further quality trimming, imputed SNPs were removed if information quality was <0.5, SNP exclusion process with missing genotype call rate was >0.03, MAF was <0.05, or HWE *P*-value was <10^{-5}. Subjects were also filtered out if SNP heterozygosity was >3 × IQR, estimated genetic relationship was >0.125, or mean difference of principal components (PCs) was >5 × IQR. The detailed procedure is shown in Figure 1A. The final dataset comprised of 8252 subjects (comprising 3817 men and 4435 women) and their genotypes for 2,915,187 SNPs.

The UKB data, which was used for generalization analysis, consists of 488,377 subjects for which two genotyping arrays were performed and 95% shared marker information was reported. A subset of 49,950 subjects, from the 438,427 included in the UK Biobank Axiom Array, was genotyped using UK Biobank Lung Exome Variant Evaluation (UK BiLEVE) Axiom Array by Affymetrix. SNPs were removed if missing genotype call rate was >0.03. Subjects were then filtered out if the missing rate was >0.05, or if a sex mismatch was detected between self-reported sex and sex estimated with X-linked SNPs. The untyped SNPs were imputed with IMPUTE4[22] using UK10K, 1000G Genome phase 3, and the HRC panel, as reference data, resulting in 93,095,623 SNPs. The detailed quality control protocol is described in Figure 1B. The remaining 316,950 subjects and 4,300,578 SNPs were applied for our association analyses.

From the discovery analyses, a genome-wide significant SNP located in *ACSS3* was identified and considered a candidate for generalization analyses using the UKB data. Due to the absence of candidate SNPs from discovery analyses, SNPs located in the *ACSS3* region between 81,331,594 and 81,650,533 base pairs in chromosome 12 were imputed using IMPUTE2 with the 1000G Phase 3 reference panel. Imputed SNPs were quality controlled with INFO >0.5, MAF <0.01, and HWE *P*-value <10^{-5}. Accordingly, 580 SNPs were considered for the generalization analyses.

All data management was conducted using PLINK,[23] GCTA,[24] and ONETOOL.[25]

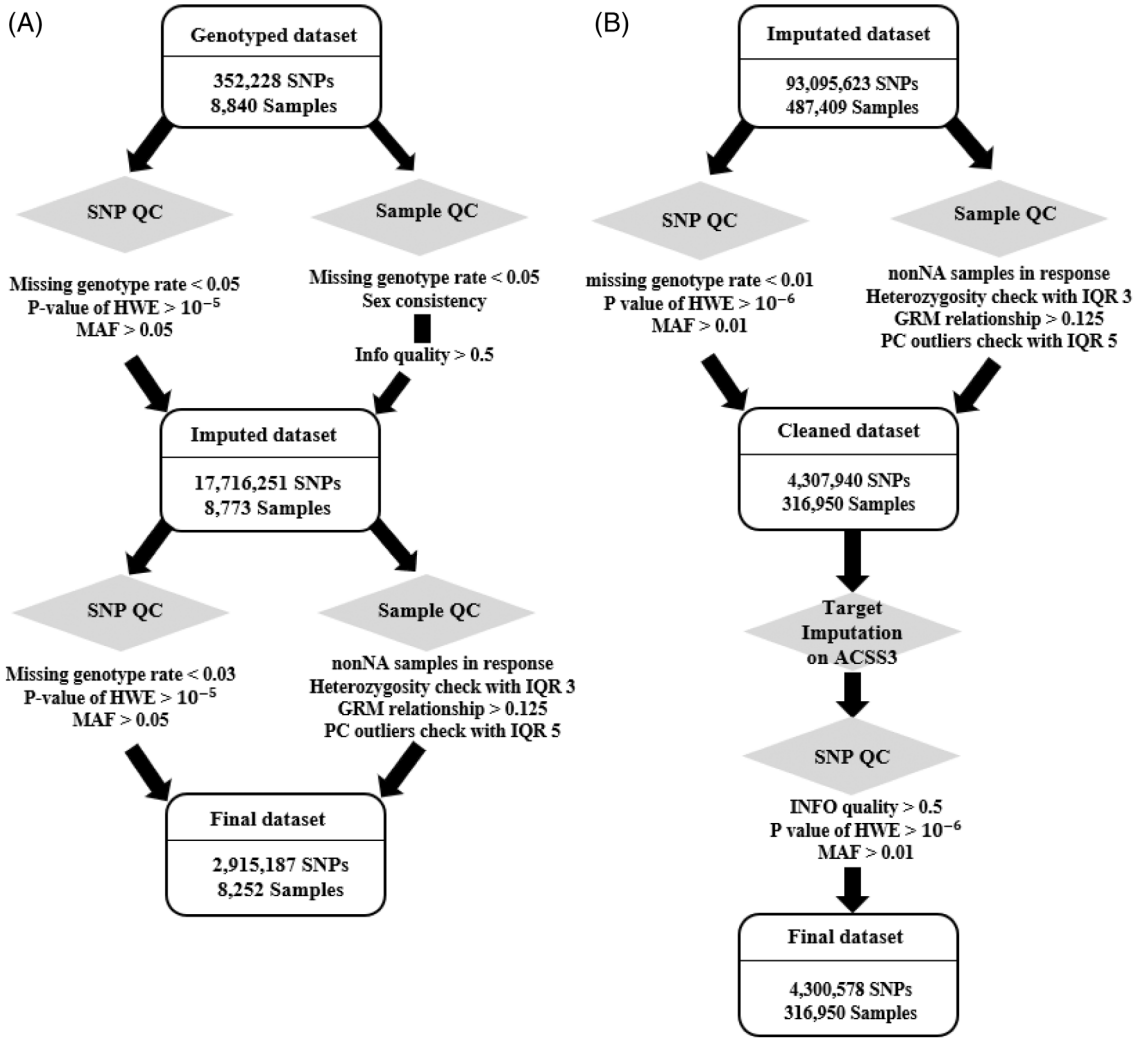

FIGURE 1 Genotype imputation and quality control of discovery and generalization datasets. (A) Discovery dataset imputation and quality control flowchart. (B) Generalization dataset quality control flowchart.

2.4 | Geographic genetic groups

The MDS plot was generated using the 1000G phase3, KoGES and UKB cohort. The data for 1007 subjects (504 EAS and 503 EUR) were extracted from 1000G. Their genotypes for 1,236,948 SNPs were combined with those of KoGES and UKB subjects, and PCs were calculated after pruning SNPs based on $r^2 = 0.2$.

2.5 | Genome-wide associations

The number of vigorous activity measurements differed among subjects, resulting in an unstable distribution. For the discovery analyses, PA was first averaged as a daily score and then transformed using the inverse-normalization method; the transformed PA was considered a response variable. Thereafter, linear regression was performed for GWAS after adjusting for the effects of age, sex, and five PCs corresponding to the top five largest eigenvalues. GWAS was performed using PLINK at the significance level $\alpha = 5 \times 10^{-8}$, significant SNPs were annotated using ANNOVAR.[26] A regional plot was generated using LocusZoom.[27] Significant SNPs were generalized using the UKB data, for which the PA was coded as a binary variable, and generalization analyses were conducted using logistic regression.

2.6 | Heritability analyses

GCTA was used to estimate SNP heritability with the discovery data by setting a 0.5 pruning level.

2.7 | Gene-based analyses

Gene-based analyses were conducted using the discovery dataset. The SNP2GENE function in FUMA[28] was used with 1000G Phase 3 EAS as a reference panel population with a total of 15,280 genes. Significant results were further tested using the UKB release2b 10 k European dataset.

2.8 | Differentially expressed gene analyses

For the genome-wide significant SNPs identified during the discovery analyses, we identified the expression quantitative trait locus (eQTL) genes using the GTEx portal,[29] and differentially expressed gene (DEG) analyses were conducted using LIMMA[30] and BALLI.[31] The mRNA expression levels were predicted for each subject in the discovery dataset using PrediXcan[32] (version 7), and DEG analyses were performed with the predicted values. We also downloaded the gene expression omnibus (GEO) microarray dataset GSE1786[33] and conducted DEG analyses on this dataset. The GSE1786 dataset comprises six healthy sedentary men aged 68.0 ± 2.7 years. Participants were nonsmokers and free of significant cardiovascular, metabolic, or musculoskeletal disorders, and led a sedentary lifestyle with no participation in regular exercise more than once weekly. Subjects were trained on a cycle ergometer thrice per week for 12 weeks at 80% of the predetermined maximal heart rate.

2.9 | Mediation analyses

We conducted mediation analyses to determine the mediation effect of PA on cerebrovascular disease. The cerebrovascular disease status

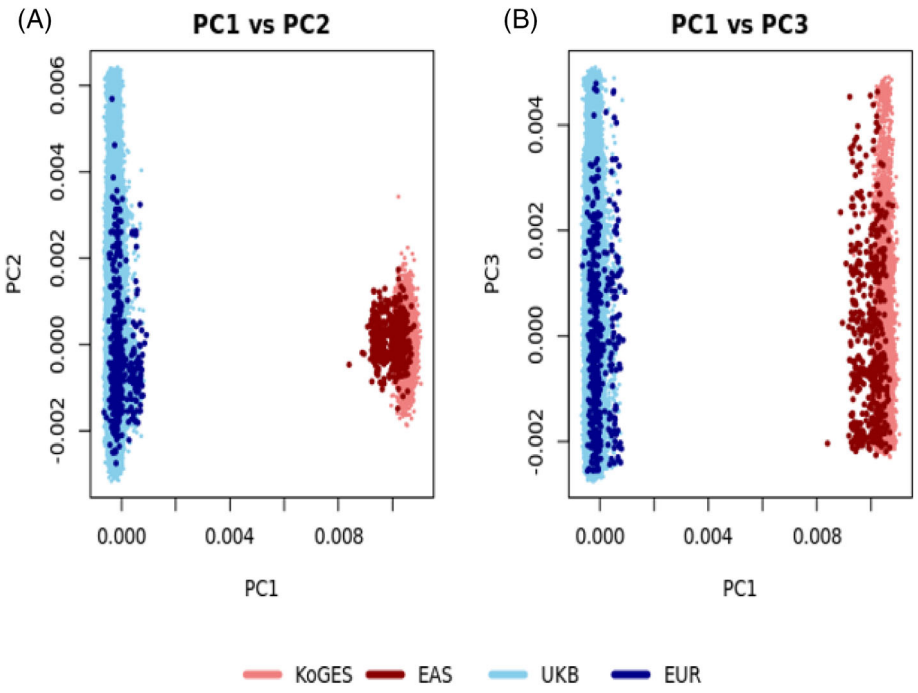

FIGURE 2 MDS plot of the 1000G–KoGES-UKB combined dataset. (A) MDS plot with principal component 1 on the x-axis and principal component 2 on the y-axis. (B) MDS plot with principal component 1 on the x-axis and principal component 3 on the y-axis. EAS, East Asian; EUR, European.

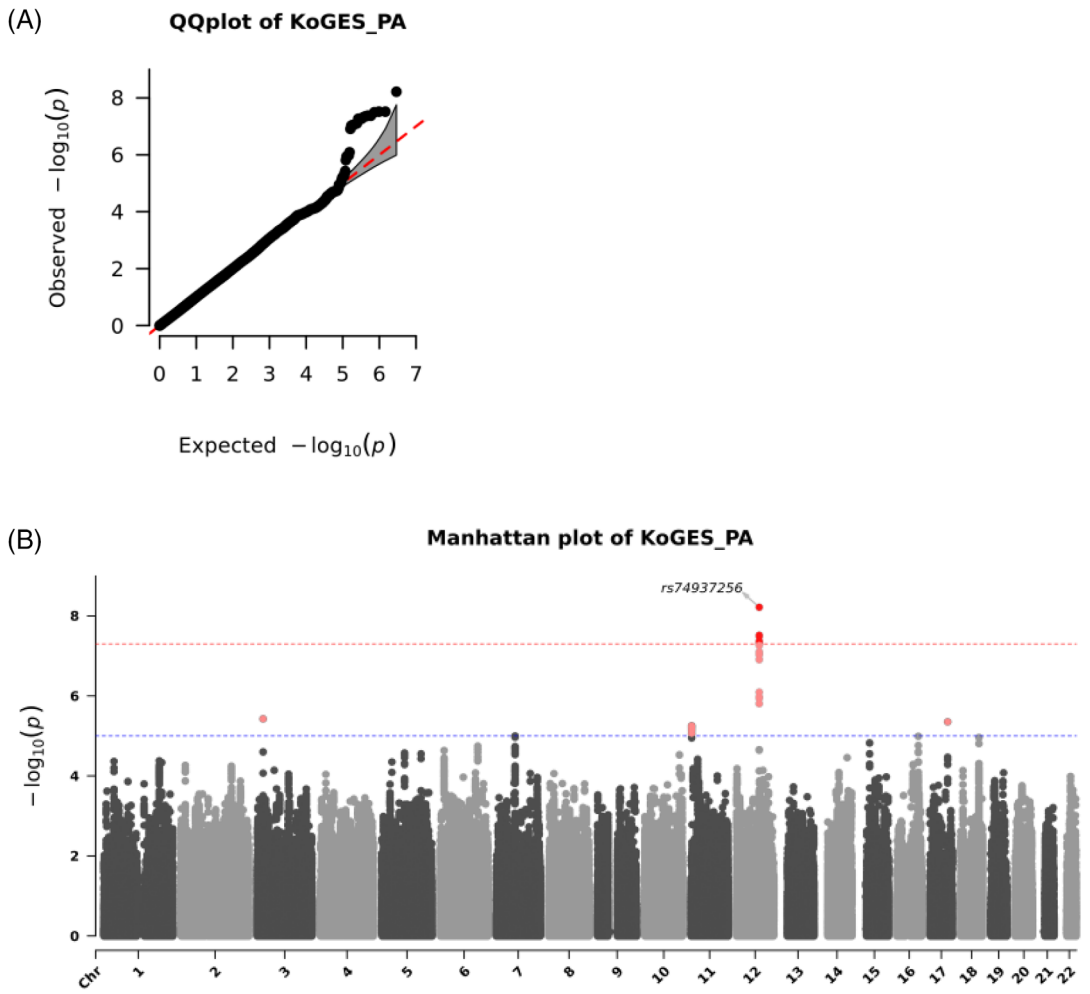

FIGURE 3 QQ and Manhattan plots of GWAS for PA. (A) QQ plot from GWAS summary statistics with a genomic inflation value of 0.98. est. vif. estimated genomic inflation factor. (B) Manhattan plot from GWAS summary statistics. The blue line represents the suggestive significance level and the red line represents the genome-wide significance level.

was available for the KoGES dataset. Subjects diagnosed with stroke, palsy, cerebral infraction or cerebral hemorrhage was defined as cases, while the others were defined as controls. Among the 8252 subjects in a discovery dataset, 7930 were controls and 322 cases. The rs74937256 and predicted ACSS3 expression levels with PrediXcan were considered as predictor (X). Cerebrovascular disease and PA were defined as outcome (Y) and mediators (M), respectively. Analyses were performed using the mediation package in R, version 4.5.0,[34] and mediator significance was evaluated with 1000 bootstrap samples.

3 | RESULTS

3.1 | Descriptive statistics

The discovery dataset comprised 8252 subjects (3817 men and 4435 women) with an average age was 52 years (SD = 8.87) at baseline, and no significant difference in average age between men and women (51.52 ± 8.69 and 52.48 ± 9.00, respectively). Subjects in the KoGES cohort spent 1.5 h on PA, which was not normally distributed. Thus, the subject-specific means of repeated measurements were considered as a response variable and was transformed using the rank-based inverse normal transformation method. The detailed descriptive statistics are presented in Table 1.

3.2 | Geographic genetic groups

MDS plots were constructed based on the three PCs, corresponding to the three largest eigenvalues as in Figure 2A,B, the two geographic genetic groups in 1000G were distinctly separated. The KoGES cohort and the UKB cohort were in accordance with EAS and EUR in 1000G, respectively, indicating that there was no evidence of population stratification.

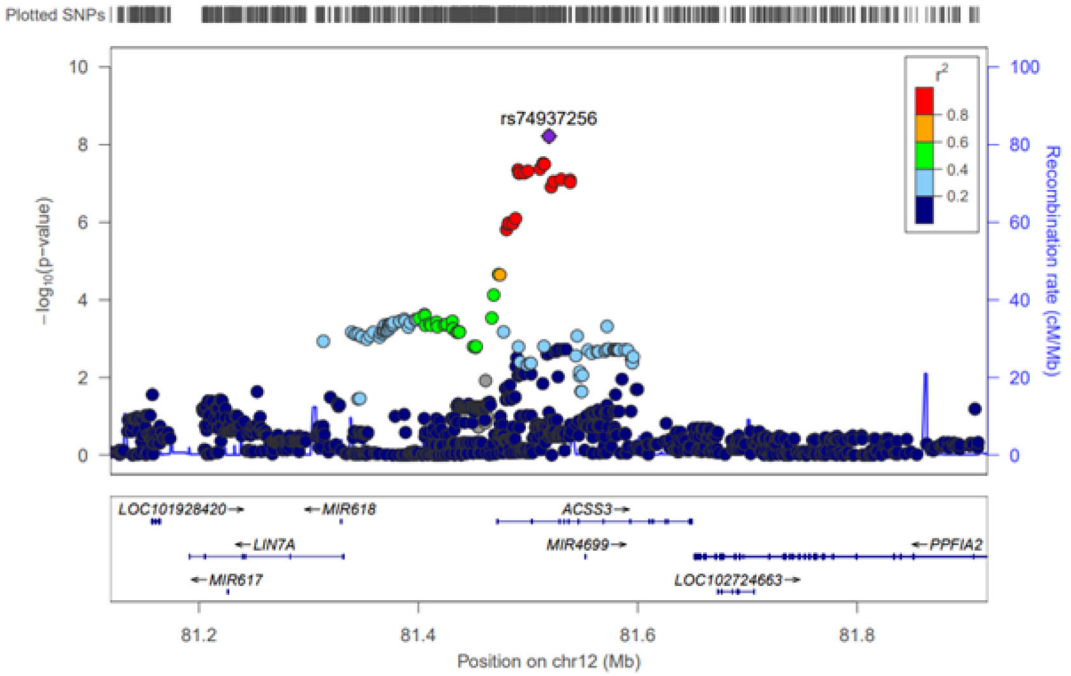

FIGURE 4 Regional plot of GWAS for PA. r^2 with rs74937256 are represented in five colors. The most highly correlated SNPs are colored in red and orange, green, blue, and navy in decreasing order.

3.3 | Genome-wide association studies and heritability estimation

GWASs with discovery data were conducted for inverse-normalized PA. QQ and Manhattan plots in Figure 3A, B also show that the GWASs were statistically valid, with a genomic inflation factor of 0.98. One significant genome-wide region was identified. The results for the most significant SNP, rs74937256, are summarized in Table 2 ($P = 6.06 \times 10^{-9}$). The MAF of rs74937256 was 0.208, which is similar to that of 0.211 from the EAS population in the gnomAD database[35] and 66,407 subjects of the Korean reference panel (KRP). The coefficient of rs74937256 (0.110) implies that subjects with minor alleles tended to spend more time performing PA than the others. The regional plot in Figure 4 shows that rs74937256 is the most significant locus, located in the *ACSS3* gene region with a heritability estimate of 0.034 (SE = 0.032, $P = 0.129$).

For the generalization analyses, we considered 580 SNPs located in *ACSS3*. The same phenotype was not available; alternatively, the binary trait, measuring whether subjects displayed a level above that of the recommended moderate or vigorous level of PA, was considered. The genome-wide significant SNP rs74937256 was not generalized ($P = 0.882$). MAFs of rs74937256 were 0.060, 0.065, and 0.208 for the UKB, gnomAD EUR (non-Finnish), and KoGES, respectively. Such differences were indicative of the influence of geographic origin on linkage disequilibrium blocks between Korea and Europe. We found 580 SNPs while considering SNPs in the same gene. Table 3 shows that rs2525840 satisfied the Bonferroni-corrected significance level $\alpha = 8.62 \times 10^{-5}$ ($P = 3.77 \times 10^{-5}$), with MAFs of 0.322, 0.427, and 0.448 in the UKB, gnomAD EUR (non-Finnish), and KoGES cohorts, respectively.

3.4 | Gene-based analyses

Gene-based analyses were conducted using the discovery dataset (Figure 5A, B). The Manhattan plot in Figure 5**(b)** shows that *CSGALNACT1* satisfied the Bonferroni-corrected significance level $\alpha = 3.27 \times 10^{-6}$ ($P = 3.10 \times 10^{-6}$). Table 4 presents the results of the ten most significant genes identified from the gene-based analyses. Associations with *CSGALNACT1* were significant after adjusting for multiple testing with Bonferroni correction. *CSGALNACT1* is associated with high-density lipoprotein cholesterol (HDL-C) and triglyceride levels[36] and participants in the development of osteoarthritis,[37] which is associated with obesity. Though the remaining genes in Table 4 were not significant at Bonferroni-corrected significance level, they reportedly exhibit associations with indicators of PA-related phenotypes, including abdominal obesity.[38–41] However, *CSGALNACT1* did not achieve the significance level in the generalization dataset (Table 4).

3.5 | Differentially expressed gene analyses

We conducted eQTL analyses of the genome-wide significant SNPs using the GTEx portal. As rs74937256 was not available, rs57018719,

TABLE 2 Summary statistics of GWAS

GENE	CHR	BP	SNP	REF	ALT	MAF			HWE	MR	N	BETA	STAT	P
						KoGES	gnomAD	KRP						
ACSS3	12	81,519,020	rs74937256	A	C	0.208	0.214	0.211	0.590	0.017	8109	0.110	5.821	6.06×10^{-9}

Abbreviations: ALT, alternative allele; REF, Reference allele; MAF gnomAD, MAF of EAS in gnomAD database; MAF KRP, MAF of 66,407 Korean reference panel; HWE, Hardy–Weinberg Equilibrium; MR, SNP missing rate; N, the number of subjects included in GWAS; STAT, T-statistics.

which was also genome-wide significant and closely located to rs74937256, was considered. Rs57018719 presented the highest R^2(0.938) and D′(0.999) with rs74937256. Table 5a shows that *ACSS3* expression was significantly associated with rs57018719 in skeletal muscle tissue ($P = 7 \times 10^{-8}$) and cultured fibroblasts ($P = 5 \times 10^{-5}$). mRNA expression results in skeletal muscle tissue (Table 5 b) showed that the predicted mRNA expression of *ACSS3* was significantly associated with PA ($P = 10^{-4}$). While conducting DEG analyses using the publicly available dataset GSE1786, *ACSS3* was found to be significantly differentially expressed (Table 6).

3.6 | Mediation analyses

Mediation analyses were conducted using the KoGES dataset to identify the mediation effect of rs74937256 and the predicted *ACSS3* expression level using PrediXcan. Although the average direct effect of PA on cerebrovascular disease was not significant in rs74937256 (with effect size -2.42×10^{-5}; $P = 0.988$) nor was the predicted *ACSS3* expression level (effect size 0.005; $P = 0.558$), multiple studies have established a theoretical basis for this relationship.[8,10,11,42] The average mediation effects of rs74937256 and the predicted *ACSS3* expression level were -6.77×10^{-4} ($P = 0.044$) and 9.46×10^{-4} ($P = 0.008$), respectively. Our results also indicate that genetic factors in PA affect cerebrovascular disease.

4 | DISCUSSION

Multiple studies have investigated the genetic components associated with PA. However, GWASs of PA require careful interpretation as the significantly associated loci vary widely based the definition of PA as well as the genetic, ancestral backgrounds of subjects. For example, in a study investigating PA defined as sports performance, more than 155 genetic markers were reported as positively associated with non-EAS athletic status,[43–50] most of which were not identified in the studies of PA defined as exercise in EAS. In parallel with the reported associations between PA and brain-related phenotypes, we found potential mediation effects of genetic factors in the association between PA and cerebrovascular disease.

In this study, we defined the daily sum of bodily movement produced by skeletal muscles that results in energy expenditure as PA with the aim of elucidating the genetic effects of PA in EAS non-athletes. To this end, we conducted a GWAS of PA and found that intronic rs74937256 in *ACSS3* was significantly associated with PA, followed by the generalization analyses on the variants in *ACSS3*. Combined with the fact that the alternative variant rs57018719 affected PA in the GWAS (BETA = −0.105) and eQTL (BETA = -0.38), it can be inferred that rs74937256 partially regulates the expression of *ACSS3*. *ACSS3*, a member of the acyl-coenzyme A synthetase (*ACSS*) family, catalyzes fatty acid metabolism through energy storage and metabolic functions. *ACSS3* degrades ketone bodies, leading to energy production.[51–53] In summary, rs74937256, followed by *ACSS3* expression, partially regulates the activity of ketone body synthesis, and consequently, the level

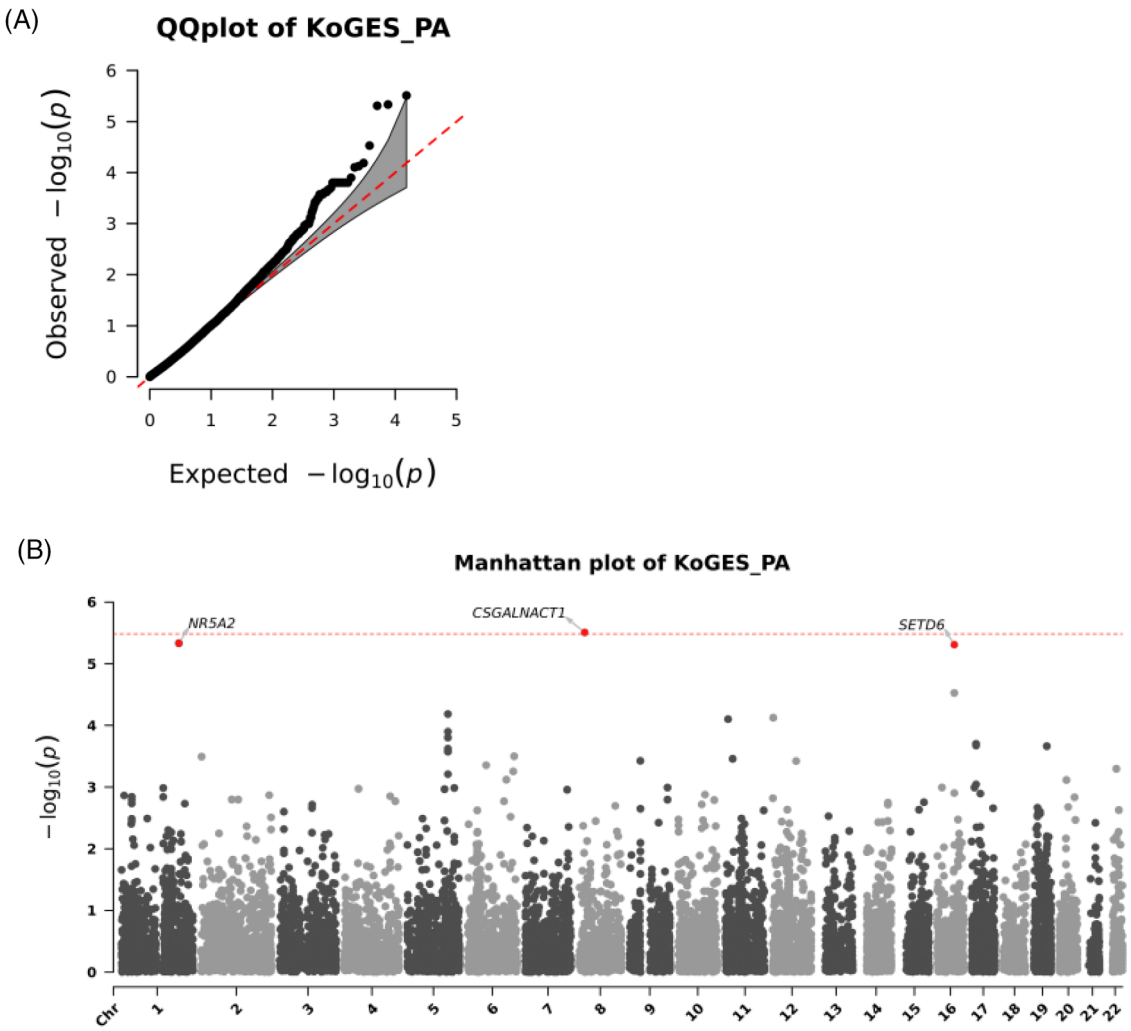

FIGURE 5 QQ and Manhattan plots of gene-based analyses. (A) QQ plot constructed using the summary statistics of the gene-based analyses. (B) Manhattan plot constructed using the summary statistics of the gene-based analyses.

of PA. This response allows skeletal muscles to use ketone bodies as the primary energy source,[54–58] which may lead to the association between rs74937256 and PA.

Although we present evidence of genetic factors in PA, the present study has certain limitations. First, genome-wide significant SNPs from EAS were not generalized from the UKB data. Distinct genetic architectures among different genetic groups were reflected in the MAF patterns and linkage disequilibrium block, and this inconsistency may account for the unsuccessful generalization. Alternatively, different SNPs located in *ACSS3* were generalized. Further generalization may increase the strength of our findings. Secondly, inconsistencies exist in PA definitions and the geographical, genetic background between the discovery/generalization dataset, PrediXcan, and GEO database. We attempted to use the training weight data from EAS for PrediXcan; however, this was not available for skeletal muscle tissue. Hence, our results from PrediXcan may be biased due to genetic and ancestral differences. In addition, the gene expression dataset from GEO contains subjects from EUR ancestry and estimated effect sizes may differ for EAS. Further studies are required to determine the effects related to trans-ancestral genomic architecture. Third, our outcome was based on a self-report questionnaire prone to several biases (e.g., recall bias). Lastly, although PA is reportedly affected by genetic and environmental factors, this study only considered genetic factors. Therefore, further research is needed to investigate the mechanisms underlying the gene-by-environmental interactions associated with PA.

5 | CONCLUSION

The *ACSS3* variant rs74937256 influences energy supply to muscles during conditions of energy exhaustion. In the present study, probable evidence of association was observed between rs74937256 and PA. We identified the genome-wide association of *ACSS3*, which is primarily involved in energy mechanisms characterized by ketone body synthesis. The biological functions of *ACSS3* illustrate the importance of *ACSS3* in PA, these results may increase our understanding of the biological aspects of PA. Additionally, given that PA has associations with chronic diseases or metabolic syndrome, our mediation

TABLE 3 Summary statistics of generalization analyses

CHR	BP	SNP	REF	ALT	MAF			HWE	MR	N	OR	STAT	P	INFO
					UKB	gnomAD	KoGES							
12	81,519,020	rs74937256	A	C	0.060	0.065	0.208	0.753	0.002	316,831	1.001	0.149	0.882	0.995
12	81,531,743	rs2525840	C	T	0.322	0.427	0.448	5.47×10^{-2}	0.008	314,026	0.978	−4.121	3.77×10^{-5}	0.995

Abbreviations: INFO, Information Quality; MAF gnomAD, MAF of European (non-Finnish) in gnomAD database.

TABLE 4 Top 10 significant genes identified using gene-based analyses

CHR	GENE	STAT	P for EAS	P for UKB
8	*CSGALNACT1*	4.520	3.10×10^{-6}	0.92
1	*NR5A2*	4.433	4.64×10^{-6}	0.15
16	*SETD6*	4.421	4.90×10^{-6}	0.86
16	*CNOT1*	4.015	2.97×10^{-5}	0.70
5	*PCDHGA1*	3.826	6.53×10^{-5}	0.01
12	*CACNA2D4*	3.792	7.49×10^{-5}	0.84
11	*TRIM66*	3.778	7.91×10^{-5}	0.43
5	*PCDHGA2*	3.660	1.26×10^{-4}	0.01
5	*PCDHGA3*	3.604	1.57×10^{-4}	0.01
5	*PCDHGB1*	3.604	1.57×10^{-4}	0.01

TABLE 5 Differentially expressed gene (DEG) analyses using eQTL and PrediXcan

GENE	SNP	BETA	P	TISSUE
(a) eQTL for rs57018719				
ACSS3	rs57018719	−0.38	7×10^{-8}	Muscle-skeletal
ACSS3	rs57018719	−0.19	7×10^{-5}	Cultured fibroblasts
(b) PrediXcan for *ACSS3*				
GENE	BETA	P	TISSUE	
ACSS3	−0.012	10^{-4}	Muscle-Skeletal	

Note: Top SNP rs74937256 was not available in GTEx portal, we substituted it to rs57018719 which is genome-wide significant in KoGES and the most highly correlated with rs74937256 ($r^2 = 0.938$ and $D' = 0.999$).

TABLE 6 Differentially expressed gene (DEG) analyses using LIMMA and BALLI

GENE	BETA	LIMMA P	BALLI P
ACSS3	18.56	0.042	0.043

Abbreviation: BETA, Beta coefficient from LIMMA.

analyses findings can be applied to determine genetic correlations between PA and other diseases, and to develop personalized recommendations regarding PA. Collectively, these findings broaden our understanding of the biology of, and genetic disposition for, PA.

AUTHOR CONTRIBUTIONS

Jinyeon Jo contributed to the design, analysis of the results and writing of the manuscript. Sungho Won supervised overall direction and planning. Youngkyu Song aided in interpreting the results. Dankyu Yoon and Chung Gun Lee contributed to the discussions. All authors discussed the results and commented to the manuscript.

ACKNOWLEDGMENTS

Ms. Jo and Prof. Won had full access to all of the data in the study and take responsibility for the integrity of the data and the

accuracy of the data analysis. The researcher claims no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from Korea Disease Control and Prevention Agency and UK Biobank. Restrictions apply to the availability of these data, which were used under license for this study. KoGES data are available at https://nih.go.kr/contents.es?mid=a50401010100 with the permission of Korea Disease Control and Prevention Agency. UKB data are available at https://www.ukbiobank.ac.uk with the permission of UK Biobank. The data that support the findings of this study are available from the corresponding author, Jinyeon Jo, upon reasonable request.

ORCID

Jinyeon Jo https://orcid.org/0000-0002-4192-1725
Chung Gun Lee https://orcid.org/0000-0002-3314-3513

REFERENCES

1. Bouchard C, Blair SN, Haskell WLC. *Physical Activity and Health*. second ed. Human Kinetics; 2007.
2. Kokkinos P, Myers J, Doumas M, et al. Exercise capacity and all-cause mortality in prehypertensive men. *Am J Hypertens*. 2009;22(7):735-741.
3. Kokkinos PF, Holland JC, Narayan P, Colleran JA, Dotson CO, Papademetriou V. Miles run per week and high-density lipoprotein cholesterol levels in healthy, middle-aged men: a dose-response relationship. *Arch Intern Med*. 1995;155(4):415-420.
4. Kokkinos P. Physical activity, health benefits, and mortality risk. *Int Sch Res Not* 2012.
5. Helmrich SP, Ragland DR, Leung RW, Paffenbarger RS Jr. Physical activity and reduced occurrence of non-insulin-dependent diabetes mellitus. *N Engl J Med*. 1991;325(3):147-152.
6. Blair SN, Kohl HW, Paffenbarger RS, Clark DG, Cooper KH, Gibbons LW. Physical fitness and all-cause mortality: a prospective study of healthy men and women. *JAMA*. 1989;262(17):2395-2401.
7. Carnethon MR. Physical activity and cardiovascular disease: how much is enough? *Am J Lifestyle Med*. 2009;3(1 Suppl):44 s-49 s.
8. Huerta JM, Chirlaque MD, Tormo MJ, et al. Physical activity and risk of cerebrovascular disease in the European prospective investigation into cancer and nutrition-Spain study. *Stroke*. 2013;44(1):111-118.
9. Doherty A, Smith-Byrne K, Ferreira T, et al. GWAS identifies 14 loci for device-measured physical activity and sleep duration. *Nat Commun*. 2018;9(1):5257.
10. Fornaro M, Solmi M, Veronese N, et al. The burden of mood-disorder/cerebrovascular disease comorbidity: essential neurobiology, psychopharmacology, and physical activity interventions. *Int Rev Psychiatry*. 2017;29(5):425-435.
11. Schmidt W, Endres M, Dimeo F, Jungehulsing GJ. Train the vessel, gain the brain: physical activity and vessel function and the impact on stroke prevention and outcome in cerebrovascular disease. *Cerebrovasc Dis*. 2013;35(4):303-312.
12. Karvinen S, Waller K, Silvennoinen M, et al. Physical activity in adulthood: genes and mortality. *Sci Rep*. 2015;5(1):18259.
13. Hara M, Hachiya T, Sutoh Y, et al. Genomewide association study of leisure-time exercise behavior in Japanese adults. *Med Sci Sports Exerc*. 2018;50(12):2433-2441.
14. Klimentidis YC, Raichlen DA, Bea J, et al. Genome-wide association study of habitual physical activity in over 377,000 UK biobank participants identifies multiple variants including CADM2 and APOE. *Int J Obes (Lond)*. 2018;42(6):1161-1176.
15. Lee CG, Moon H, Park S. The effects of dopamine receptor genes on the trajectories of sport participation from adolescence through young adulthood. *Ann Hum Biol*. 2020;47(3):256-262.
16. de Geus EJC. A genetic perspective on the association between exercise and mental health in the era of genome-wide association studies. *Ment Health Phys Act*. 2021;20:100378.
17. Cho YS, Go MJ, Kim YJ, et al. A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nat Genet*. 2009;41(5):527-534.
18. Kim Y, Han BG, Ko GESg. Cohort profile: the Korean genome and epidemiology study (KoGES) Consortium. *Int J Epidemiol*. 2017;46(2):e20.
19. Bycroft C, Freeman C, Petkova D, et al. The UK biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203-209.
20. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 2009;5(6):e1000529.
21. Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74.
22. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet*. 2012;44(8):955-959.
23. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-575.
24. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011;88(1):76-82.
25. Song YE, Lee S, Park K, Elston RC, Yang H-J, Won S. ONETOOL for the analysis of family-based big data. *Bioinformatics*. 2018;34(16):2851-2853.
26. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.
27. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010;26(18):2336-2337.
28. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun*. 2017;8(1):1826.
29. Consortium GT. The genotype-tissue expression (GTEx) project. *Nat Genet*. 2013;45(6):580-585.
30. Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.
31. Park K, An J, Gim J, et al. BALLI: Bartlett-adjusted likelihood-based linear model approach for identifying differentially expressed genes with RNA-seq data. *BMC Genom*. 2019;20(1):540.
32. Gamazon ER, Wheeler HE, Shah KP, et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet*. 2015;47(9):1091-1098.
33. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res*. 2012;41(D1):D991-D995.
34. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. Mediation: R package for causal mediation analysis. *J Stat Softw*. 2014;59(5):1-38.
35. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434-443.
36. Richardson TG, Sanderson E, Palmer TM, et al. Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease: a multivariable Mendelian randomisation analysis. *PLoS Med*. 2020;17(3):e1003062.
37. Kogelman LJA, Zhernakova DV, Westra H-J, et al. An integrative systems genetics approach reveals potential causal genes and pathways related to obesity. *Genome Med*. 2015;7(1):105.

38. Pulit SL, Stoneman C, Morris AP, et al. Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum Mol Genet*. 2019;28(1):166-174.
39. Genovesi S, Zaccaria D, Rossi E, Valsecchi MG, Stella A, Stramba-Badiale M. Effects of exercise training on heart rate and QT interval in healthy young individuals: are there gender differences? *Europace*. 2007;9(1):55-60.
40. Astle WJ, Elding H, Jiang T, et al. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell*. 2016;167(5):1415-1429.
41. Graff M, Scott RA, Justice AE, et al. Genome-wide physical activity interactions in adiposity - a meta-analysis of 200,452 adults. *PLoS Genet*. 2017;13(4):e1006528.
42. Shrout P, Bolger N. Mediation in experimental and nonexperimental studies: new procedures and recommendations. *Psychol Methods*. 2002;7:422-445.
43. Bray MS, Hagberg JM, Pérusse 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):35-73.
44. Buxens A, Ruiz JR, Arteta D, et al. Can we predict top-level sports performance in power vs endurance events? A genetic approach. *Scand J Med Sci Sports*. 2011;21(4):570-579.
45. Ben-Zaken S, Meckel Y, Nemet D, Eliakim A. Genetic score of power-speed and endurance track and field athletes. *Scand J Med Sci Sports*. 2015;25(2):166-174.
46. Weyerstraß J, Stewart K, Wesselius A, Zeegers M. Nine genetic polymorphisms associated with power athlete status - a meta-analysis. *J Sci Med Sport*. 2018;21(2):213-220.
47. Grealy R, Herruer J, Smith CL, Hiller D, Haseler LJ, Griffiths LR. Evaluation of a 7-gene genetic profile for athletic endurance phenotype in ironman championship triathletes. *PLoS One*. 2015;10(12):e0145171.
48. Guilherme J, Bertuzzi R, Lima-Silva AE, Pereira ADC, Lancha Junior AH. Analysis of sports-relevant polymorphisms in a large Brazilian cohort of top-level athletes. *Ann Hum Genet*. 2018;82(5):254-264.
49. Egorova ES, Borisova AV, Mustafina LJ, et al. The polygenic profile of Russian football players. *J Sports Sci*. 2014;32(13):1286-1293.
50. Ahmetov II, Egorova ES, Gabdrakhmanova LJ, Fedotovskaya ON. Genes and athletic performance: An update. *Med Sport Sci*. 2016;61:41-54.
51. Watkins PA, Maiguel D, Jia Z, Pevsner J. Evidence for 26 distinct acyl-coenzyme a synthetase genes in the human genomes. *J Lipid Res*. 2007;48(12):2736-2750.
52. Williamson DH, Thornton PS. 45 - ketone body production and metabolism in the fetus and neonate. In: Polin RA, Fox WW, Abman SH, eds. *Fetal and Neonatal Physiology*. Fourth ed. W.B. Saunders; 2011:487-497.
53. Laffel L. Ketone bodies: a review of physiology, pathophysiology and application of monitoring to diabetes. *Diabetes Metab Res Rev*. 1999;15(6):412-426.
54. Sass JO. Inborn errors of ketogenesis and ketone body utilization. *J Inherit Metab Dis*. 2012;35(1):23-28.
55. Caballero BTLCFPM. *Encyclopedia of Food Sciences and Nutrition*. Academic; 2003.
56. Timmins IR, Zaccardi F, Nelson CP, Franks PW, Yates T, Dudbridge F. Genome-wide association study of self-reported walking pace suggests beneficial effects of brisk walking on health and survival. *Commun Biol*. 2020;3(1):634.
57. Evans M, Cogan KE, Egan B. Metabolism of ketone bodies during exercise and training: physiological basis for exogenous supplementation. *J Physiol*. 2017;595(9):2857-2871.
58. Pinckaers PJM, Churchward-Venne TA, Bailey D, van Loon LJC. Ketone bodies and exercise performance: the next magic bullet or merely hype? *Sports Med*. 2017;47(3):383-391.

How to cite this article: Jo J, Song Y, Yoon D, Lee CG, Won S. Genome-wide assessment reveals a significant association between *ACSS3* and physical activity. *Genes, Brain and Behavior*. 2023;22(1):e12834. doi:10.1111/gbb.12834