



Type of Paper: Original Research

Allele Frequency Distribution and Population Differentiation of Athletic Performance- Associated Polymorphisms (*ACTN3*, *ACE*, *PPARGC1A*) in Southeast Asian Populations: Fixation Index (F_{ST}) and Total Genotype Score (TGS) Analysis Using the 1000 Genomes Project

¹ Bioinformatics Research Group,
Master Program in Statistics,
Universitas Islam Indonesia,
Yogyakarta, Indonesia

² IMERI, Department of Medical
Chemistry, Universitas Indonesia,
Jakarta, Indonesia

* Correspondence: Rohmatul
Fajriyah; rfajriyah@uii.ac.id

Andi Rosilala¹, Rohmatul Fajriyah^{1, *}, and Linda Erlina²

Article history:

Received: March 20, 2026

Revised: April 25, 2026

Accepted: April 30, 2026

How to Cite: Rosilala, A.; Fajriyah, R.;
and Erlina, L. (2026). Allele Frequency
Distribution and Population
Differentiation of Athletic
Performance- Associated
Polymorphisms in Southeast Asian
Populations. JoMaSSH, 2(2), 108–
125. doi: 10.21831/jomassh.
v2i2.2836.



Copyright: © 2026 by the authors. Submitted
for possible open-access publication under the
terms and conditions of the Creative Commons
Attribution (CC BY) license
(<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Background. Athletic performance genomics is dominated by studies conducted in European and African populations, leaving Southeast Asian (SEA) populations systematically underrepresented. The three most extensively studied performance-associated polymorphisms—*ACTN3* rs1815739 (R577X), *ACE* insertion/deletion (proxied by rs4341), and *PPARGC1A* rs8192678 (Gly482Ser), remain poorly characterized in SEA populations at the population-genetics level. While individual loci have been examined in adjacent East Asian cohorts (primarily Japanese, Korean, and Han Chinese), no prior study has jointly reported all three loci alongside F_{ST} and TGS analyses specifically for KHV or CDX populations, to the best of our knowledge. **Objectives.** This study aims to characterise the allele and genotype frequencies of *ACTN3* rs1815739, *ACE* rs4341, and *PPARGC1A* rs8192678 in the Kinh Vietnamese (KHV) and Chinese Dai (CDX) populations, quantify their population differentiation relative to other continental superpopulations, evaluate the applicability of a European-calibrated Total Genotype Score framework in SEA populations, and assess the practical significance of *ACE* allele directionality reversal for endurance genomic scoring in the region. **Methods.** We performed a descriptive bioinformatics analysis of phased genotype data from the 1000 Genomes Project Phase 3 ($n = 2,503$ individuals, 26 populations). Allele frequencies with 95 % Wilson confidence intervals, Hardy–Weinberg Equilibrium (HWE) exact tests, global and pairwise F_{ST} , chi-square tests with Bonferroni correction, and a Total Genotype Score (TGS) framework were applied. **Results.** SEA populations (Kinh Vietnamese [KHV], $n = 99$; Chinese Dai [CDX], $n = 93$) showed intermediate *ACTN3* alt-allele frequencies (KHV: 0.409; CDX: 0.505), high *ACE* rs4341 C- allele frequencies (KHV: 0.672; CDX: 0.656), and intermediate *PPARGC1A* frequencies (KHV: 0.450; CDX: 0.430). Global F_{ST} was highest for *ACTN3* ($F_{ST} = 0.151$, high differentiation). Under the European-calibrated TGS framework SEA scored significantly below EUR reference for TGS Endurance ($\Delta = -16.7$ points for KHV; Mann–Whitney $p = 6.23 \times 10^{-14}$), but not for TGS Power ($p = 0.527$). Reversing *ACE* directionality per the EAS model increased TGS Endurance by +17.2 (KHV) and +15.6 (CDX) points. **Conclusions.** SEA populations show distinct allele frequency profiles from European reference populations, driven by *PPARGC1A* differentiation versus African populations ($F_{ST} = 0.234$) and *ACE* directionality ambiguity. These findings provide, to the best of our knowledge, the first joint population-genetic characterization of these three loci with concurrent F_{ST} and TGS analysis specifically for the KHV and CDX populations, and highlight systematic limitations of European-calibrated TGS frameworks when applied to non-European populations.

Keywords:

F_{ST} ; Performance-Associated Polymorphisms; sports genomics; Total Genotype Score

1. Introduction

The genetic architecture of athletic performance has attracted intense scientific interest since the late 1990s [1]. Genomic studies have now catalogued over 250 variants across more than 140 genes associated with traits influencing athletic capacity [2]. The degree of genetic contribution to athletic performance is estimated at approximately 50%, with heritability estimates for endurance-related phenotypes at 44–68% and power-related phenotypes at 48–56% [3,4].

Among thousands of candidate variants, three polymorphisms have emerged as the most extensively studied: (i) *ACTN3* rs1815739(R577X), encoding the structural protein α -actinin-3 in fast-twitch muscle fibres [5]; (ii) the *ACE* Alu insertion/deletion (rs4646994) [6]; and (iii) *PPARGC1A* rs8192678 (Gly482Ser), affecting PGC-1 α -mediated mitochondrial biogenesis [7]. These three loci collectively exemplify both the promise and the limitations of sports genomics: they are biologically plausible, widely replicated in European cohorts, yet effect sizes are inconsistent and population-dependent [8, 9]. Athletic performance is also closely linked to public health outcomes, national development, and economic productivity: elite sporting achievement inspires youth physical activity, reduces non-communicable disease burden, and generates national prestige [8, 9]. In Southeast Asia and Indonesia specifically, investment in sports science and athlete development has grown substantially in recent decades, yet the genomic evidence base underpinning talent identification and training optimisation remains anchored almost entirely to non-SEA populations—a gap that this study directly addresses.

A critical and largely unaddressed problem in sports genomics is the profound ethnic imbalance of the evidence base. The vast majority of association studies have been conducted in European or, to a lesser extent, African populations [8, 10]. Southeast Asia—home to over 680 million people comprising highly diverse ethnic groups—has been systematically excluded [9]. Athletic performance development has significant socioeconomic and public health relevance in Southeast Asia, where sporting success supports national identity, youth development programmes, and emerging sports medicine infrastructure — making population-specific genomic reference data particularly valuable for evidence-based talent identification and training optimisation in the region.

The most comprehensive meta-analysis to date, screening 4,228 articles and including 54,382 subjects, found no statistically significant association between either *ACE* I/D or *ACTN3* R577X and athletic performance in either power or endurance sports [9]. Critically, 52.6% of all identified variants had been studied in only a single publication, and no eligible study within that meta-analysis reported data specifically from Southeast Asian populations [9]. Existing EAS studies have been conducted predominantly in Japanese, Korean, and Han Chinese cohorts [10, 11], leaving the Kinh Vietnamese and Chinese Dai populations uncharacterised for these three loci in combination.

Miyamoto-Mikami and Fuku [11] demonstrated that European-derived TGS frameworks failed entirely in Japanese athletic cohorts, and that the *ACE* I allele—designated "favourable" for endurance in European studies—was associated with sprint performance in Japanese swimmers, implying a population-specific reversal of effect direction.

The present study is motivated by the following research gaps, which remain unaddressed in the sports genomics literature. First, existing studies on *ACTN3* rs1815739, *ACE* rs4341, and *PPARGC1A* rs8192678 in Asian populations have focused almost exclusively on Japanese, Korean, and Han Chinese cohorts [10, 11], leaving the Kinh Vietnamese (KHV) and Chinese Dai (CDX) populations uncharacterised for these three loci in combination. Second, pairwise *F_{ST}* for these loci between KHV and CDX and other continental superpopulations has not been reported [12]. Third, the behaviour of European-calibrated TGS frameworks applied to KHV and CDX populations has not been evaluated [13, 14], despite evidence of framework failure in adjacent East Asian cohorts [11]. Fourth, the quantitative impact of the *ACE* allele directionality reversal [11] on TGS Endurance scores in KHV and CDX populations has not been assessed.

The present study aims to: (i) characterise allele and genotype frequencies of *ACTN3* rs1815739, *ACE* rs4341, and *PPARGC1A* rs8192678 in the KHV and CDX populations of the 1000 Genomes Project Phase 3; (ii) quantify

population differentiation (F_{ST}) between SEA and other continental superpopulations for these three loci; (iii) evaluate the performance of a European-calibrated TGS framework in SEA populations; and (iv) assess the practical significance of ACE allele directionality reversal for TGS Endurance scores in SEA.

2. Materials and Methods

2.1. *ACTN3* rs1815739 (R577X)

ACTN3 encodes a-actinin-3, a structural protein expressed exclusively in fast-twitch (type II) muscle fibres [5]. The C->T variant (rs1815739) introduces a premature stop codon at position 577, completely abolishing a-actinin-3 protein in XX (T/T) homozygotes [15]. Approximately 18% of the global population carries the XX genotype, but the frequency is substantially higher in East and South Asian populations [5, 16].

Multiple meta-analyses found an enrichment of the R allele in power athletes relative to controls [8, 10, 17]. However, the most comprehensive meta-analysis (n = 14,820 athletes) found no statistically significant association (OR = 1.67, p = 0.995 for power performance) [9].

2.2. *ACE* Insertion/Deletion (rs4646994 / rs4341 proxy)

The *ACE* gene encodes angiotensin-converting enzyme, a central regulator of the renin-angiotensin-aldosterone system (RAAS) [6]. The Alu I/D polymorphism (rs4646994) in intron 16 determines circulating *ACE* levels: the D allele raises *ACE* activity and the I allele lowers it, with lower activity historically linked to endurance performance [18].

The most recent meta-analysis (21 articles, n ≈ 16,000) found no significant association with endurance or power (all ORs, p > 0.49) [9]. Furthermore, Miyamoto-Mikami and Fuku [11] demonstrated that the D allele was favoured among Japanese elite swimmers, the opposite of the European pattern. Because the Alu I/D indel is not represented in biallelic SNV panels, we used rs4341 (G/C, chr17:63488629) as a validated proxy ($r^2 \approx 0.83$) [19].

2.3. *PPARGC1A* rs8192678 (Gly482Ser)

PPARGC1A encodes PGC-1 α , the master regulator of mitochondrial biogenesis and oxidative metabolism [7]. The Gly482Ser substitution reduces PGC-1 α transcriptional activity, impairs mitochondrial density, and has been associated with reduced VO_{2max} response to training [7, 20]. Despite biological plausibility, only 9 studies existed at the time of the 2024 meta-analysis-below the threshold for adequately powered analysis-and no SEA data were available [9].

2.4. Population Genetics: F_{ST} and Global Allele Variation

Wright's fixation index F_{ST} quantifies the proportion of total genetic variation attributable to between-population differences [12]. The 1000 Genomes Project Phase 3 provides the most comprehensive publicly available global reference panel, comprising 2,504 individuals from 26 populations across five continental superpopulations [19]. Previous analyses demonstrated that *ACTN3* R577X exhibits among the highest F_{ST} values of any common variant [15, 21].

2.5. Total Genotype Score (TGS)

The TGS framework aggregates individual genotype scores across a panel of performance-associated SNPs into a single composite index [13, 14]. The formula is:

$$TGS = \frac{100 \times \sum_{j=1}^N S_j}{2N} \quad (1)$$

where N is the number of SNPs and $S_j \in \{0,1,2\}$ denotes unfavourable homozygote, heterozygote, or favourable homozygote, respectively [13]. TGS ranges from 0 (all unfavourable) to 100 (all favourable).

Genotype data were obtained from the 1000 Genomes Project Phase 3 [19], comprising 2,504 phased haplotypes from 26 populations across five continental superpopulations. Population metadata were downloaded

from the IGSF FTP server. SEA populations within this dataset are represented by the Kinh Vietnamese (KHV, $n = 99$) and Chinese Dai from Xishuangbanna (CDX, $n = 93$) [19].

2.6. SNP Selection and VCF Retrieval

Three SNPs were selected based on biological relevance, replication history, and evidence base (Table 1). Phased VCF files were retrieved via tabix-indexed remote access using the Variant Annotation Bioconductor package [22]. After excluding 743 samples present in the high-coverage release but absent from the Phase 3 metadata panel, 2,504 samples with complete population annotations were retained.

Table 1. Metadata for the three target SNPs. Genotype notation: RR=reference homozygote; RX=heterozygote; XX=alternative homozygote.

Gene	rsID	Variant	Trait	Release
ACTN3	rs1815739	R577X (C/T)	Power/Sprint	High Cov.
ACE	rs4341 [†]	Syn. (G/C)	Endurance/VO _{2max}	Biallelic
PPARGC1A	rs8192678	Gly482Ser (C/T)	Mito. biogen.	Biallelic

[†]Proxy for ACEAlu I/D (rs4646994); $r^2 \approx 0.83$ [19].

2.7. Allele Frequency Estimation

For each SNP-population combination the alternative allele frequency p was estimated as [23]:

$$\hat{p} = \frac{n_{RX} + 2n_{XX}}{2n} \quad (2)$$

where n_{RX} , $2n_{XX}$, and n are heterozygote count, alternative homozygote count, and total sample size, respectively.

2.7.1. Wilson Confidence Intervals

Wilson score intervals were used because they outperform the normal approximation for proportions near 0 or 1 [24, 25]:

$$\hat{p}_w = \frac{\hat{p} + \frac{z^2}{2n}}{1 + \frac{z^2}{n}} \quad CI = \hat{p}_w \pm \frac{z \sqrt{\frac{\hat{p}(1 - \hat{p})}{n} + \frac{z^2}{4n^2}}}{1 + \frac{z^2}{n}} \quad (3)$$

with $z = 1.96$ for 95 % intervals.

2.8. Hardy-Weinberg Equilibrium Testing

HWE was assessed using the exact test of Wigginton et al. [26] as implemented in the HardyWeinberg R package [27]. Significance threshold $\alpha = 0.05$; deviating populations were flagged but retained.

2.9. Fixation Index (F_{ST})

Global F_{ST} per SNP was computed via the Nei (1973) formulation [28]:

$$F_{ST} = \frac{H_T - H_S}{H_T} \quad (4)$$

$$H_T = 2\bar{p}(1 - \bar{p}) \quad (5)$$

$$H_S = \frac{1}{K} \sum_{k=1}^K 2p_k(1 - p_k) \quad (6)$$

with $\bar{p}$ the mean allele frequency across all K populations. Interpretation followed Wright [12]: < 0.05 negligible; 0.05-0.15 moderate; 0.15-0.25 high; > 0.25 very high.

2.10. Chi-Square Tests with Bonferroni Correction

Allele frequency differences were tested using χ^2 tests on 2×2 contingency tables [29]. Bonferroni correction ($\alpha_{adj} = 0.05/10 = 0.0050$ per SNP) was supplemented with FDR correction (Benjamini-Hochberg) [30].

2.11. Total Genotype Score (TGS)

TGS was computed per individual using Eq. (1) with three panels: Combined ($N = 3$), Power ($N = 1$: *ACTN3*), Endurance ($N = 2$: *ACE*+*PPARGC1A*). Divergence from EUR reference:

$$\Delta TGS = \overline{TGS}_{pop} - \overline{TGS}_{EUR} \quad (7)$$

Mann–Whitney U tests [31] compared SEA and EUR TGS distributions.

2.12. Sensitivity Analysis: ACE Directionality Reversal

Following Miyamoto-Mikami and Fuku [11], a what-if directional assumption analysis designated *ACE* XX (CC proxy for DD) as favourable under the "EAS model". TGS Endurance was recomputed and A(EAS - EUR model) reported for all 26 populations.

2.13. Software and Reproducibility

All analyses were conducted in R 4.5.2 [32] using tidyverse [33], VariantAnnotation [22], HardyWeinberg [27], hierfstat [34], ggplot2 [35], kableExtra, purrr, and glue. The complete analysis script is available as a reproducible R Markdown document.

3. Results

3.1. Data Summary

After quality control, 2,503 individuals with complete genotype data across all three SNPs were included. SEA populations represent 192 individuals (7.7%) [19].

Table 2. 1000 Genomes Project Phase 3 population summary. **Bold** = SEA populations.

Superpop.	Populations	N
AFR	ACB, ASW, ESN, GWD, LWK, MSL, YRI	661
AMR	CLM, MXL, PEL, PUR	347
SEA	CHB, CHS, JPT, CDX, KHV	504
EUR	CEU, FIN, GBR, IBS, TSI	503
SAS	BEB, GIH, ITU, PJL, STU	489
Total	26 populations	2,504

Table 3. Overall genotype distribution across all 2,504 samples.

Gene (rsID)	n_{RR}	n_{RX}	n_{XX}
<i>ACE</i> (rs4341)	593	1,167	743
<i>ACTN3</i> (rs1815739)	956	1,089	459
<i>PPARGC1A</i> (rs8192678)	1,421	833	249

3.2. Allele and Genotype Frequencies in SEA Populations

Table 4 presents allele and genotype frequencies with 95 % Wilson confidence intervals. For *ACTN3*, the T-allele frequency in KHV (0.409) is substantially lower than CDX (0.505), consistent with intra-EAS heterogeneity [5, 15]. Both SEA populations showed high *ACE* C-allele frequencies (≈ 0.66 - 0.67), considerably exceeding EUR populations (≈ 0.34 - 0.51) [6, 19].

Table 4. Allele and genotype frequencies with 95 % Wilson CI for SEA populations [24].

Gene	Pop	N	n_{RR}	n_{RX}	n_{XX}	f_{RR}	f_{RX}	f_{XX}	Alt AF	CI _{lo}	CI _{hi}
ACE	CDX	93	10	44	39	0.108	0.473	0.419	0.656	0.585	0.720
ACE	KHV	99	13	39	47	0.131	0.394	0.475	0.672	0.604	0.733
ACTN3	CDX	93	28	36	29	0.301	0.387	0.312	0.505	0.434	0.576
ACTN3	KHV	99	33	51	15	0.333	0.515	0.152	0.409	0.343	0.479
PPARGC1A	CDX	93	28	50	15	0.301	0.538	0.161	0.430	0.361	0.502
PPARGC1A	KHV	99	33	43	23	0.333	0.434	0.232	0.450	0.382	0.519

3.3. Hardy-Weinberg Equilibrium

Of 78 SNP-population combinations tested, 75 (96.2 %) were in HWE. Three deviations were detected (**Table 5**). The marginal CDX deviation for *ACTN3* ($p = 0.037$) may reflect population sub-structure within the Xishuangbanna Dai group [19].

Table 5. WE deviations($p < 0.05$) across 78 SNP-population tests **Bold** = SEA deviation.

Gene	Pop	n_{RR}	n_{RX}	n_{XX}	HWE p
<i>ACTN3</i>	CDX	28	36	29	0.037*
<i>ACTN3</i>	GIH	14	61	28	0.046*
<i>PPARGC1A</i>	ESN	89	8	2	0.035*

* $p < 0.05$.

3.4. Population Differentiation (F_{ST})

3.4.1. Global F_{ST}

Global F_{ST} values (**Table 6**) reveal high differentiation for *ACTN3* rs1815739 ($F_{ST} = 0.151$), substantially exceeding the genome-wide average $F_{ST} \approx 0.10$ -0.12 [36]. Moderate differentiation was observed for *PPARGC1A* ($F_{ST} = 0.114$) and *ACE* rs4341 ($F_{ST} = 0.057$).

Table 6. Global F_{ST} per SNP across all 26 populations (Nei 1973 [28]).

Gene (rsID)	H_T	H_S	F_{ST}	Interp.
<i>ACTN3</i> (rs1815739)	0.4803	0.4077	0.151	High
<i>ACE</i> (rs4341)	0.4982	0.4697	0.057	Moderate
<i>PPARGC1A</i> (rs8192678)	0.3904	0.3461	0.114	Moderate

3.4.2. Pairwise F_{ST} : SEA vs. Superpopulations

Near-zero differentiation between SEA and EAS (all ≤ 0.0004) confirms genetic proximity (**Table 7**). The highest pairwise value-*PPARGC1A* SEA vs. AFR ($F_{ST} = 0.234$)-approaches "very high" classification.

Table 7. Pairwise F_{ST} between pooled SEA (KHV+CDX) and each superpopulation [28].

Comparison	ACE	ACTN3	PPARGC1A
SEA vs. AFR	0.044	0.131	0.234
SEA vs. AMR	0.009	0.013	0.035
SEA vs. EAS	0.000	0.000	0.000
SEA vs. EUR	0.046	0.000	0.005
SEA vs. SAS	0.002	0.014	0.020

Bold = maximum value.

3.5. Statistical Comparisons

Chi-square tests (**Table 8**) showed that both SEA populations differed highly significantly from AFR for all three SNPs (all $P_{Bonf} < 0.001$) [29, 30]. Notable intra-SEA heterogeneity emerged: KHV differed significantly from AMR and SAS for *ACTN3* ($P_{Bonf} < 0.001$), while CDX did not.

Table 8. Significant chi-square comparisons after Bonferroni correction [30].

Gene	SEA	vs.	χ^2	P_{Bonf}	Sig.
<i>ACE</i>	CDX	AFR	39.03	<0.001	***
<i>ACE</i>	CDX	EUR	33.55	<0.001	***
<i>ACE</i>	KHV	AFR	46.67	<0.001	***
<i>ACE</i>	KHV	EUR	40.32	<0.001	***
<i>ACTN3</i>	CDX	AFR	182.04	<0.001	***
<i>ACTN3</i>	KHV	AFR	114.77	<0.001	***
<i>ACTN3</i>	KHV	AMR	16.45	<0.001	***
<i>ACTN3</i>	KHV	SAS	21.10	<0.001	***
<i>PPARGC1A</i>	CDX	AFR	286.07	<0.001	***
<i>PPARGC1A</i>	CDX	AMR	21.34	< 0.001	***
<i>PPARGC1A</i>	CDX	SAS	14.19	0.002	**
<i>PPARGC1A</i>	KHV	AFR	317.52	<0.001	***
<i>PPARGC1A</i>	KHV	AMR	27.36	<0.001	***
<i>PPARGC1A</i>	KHV	SAS	19.25	<0.001	***

** $P_{Bonf} < 0.01$; *** $P_{Bonf} < 0.001$.

3.6. Total Genotype Score (TGS) Analysis

3.6.1. TGS Distribution Across Populations

TGS was computed for 2,503 individuals [13]. Under the European-calibrated panel, AFR populations scored highest on both TGS Power (78.7-91.9 %) and TGS Endurance (72.9-79.7%) (**Table 9**), reflecting a fundamental bias in the "favourable" allele designation.

3.6.2. TGS Divergence from EUR Reference

Table 10 quantifies divergence using Eq. (7). KHV showed the largest endurance divergence ($\Delta = -16.7$ points, -27.5%). Mann-Whitney tests [31] confirmed a highly significant TGS Endurance difference between SEA and EUR ($W = 31,292$; $p = 6.23 \times 10^{-14}$), but not for TGS Power ($W = 46,914$; $p = 0.527$).

Table 10. TGS divergence of SEA populations from EUR reference mean and Mann-Whitney test results [13, 31].

Pop	Sup.	TGS Power	TGS End.	Δ Power	Δ End.	Δ End. (%)
CDX	EAS	49.5	45.7	-7.1	-14.9	-24.6%
KHV	EAS	59.1	43.9	+2.6	-16.7	-27.5%
EUR mean	EUR	56.5	60.6	ref	ref	ref

Mann-Whitney - TGS Power: $W = 46,914$; $p = 0.527$ (ns).

Mann-Whitney - TGS Endurance: $W = 31,292$; $p = 6.23 \times 10^{-14}$ (***)

3.6.3. Sensitivity Analysis: ACE Directionality Reversal

Reversing *ACE* directionality per Miyamoto-Mikami and Fuku [11] increased TGS Endurance substantially for all EAS populations (**Table 11**). KHV improved by +17.2 points (43.9 %→61.1 %) and CDX by +15.6 points, both exceeding the EUR reference mean under the EAS model.

Table 11. Sensitivity analysis: TGS Endurance under EUR vs. EAS model (ACE directionality reversal [11]). Selected populations; full results in Supplementary Table S3.

Pop	Sup.	EUR model	EAS model	Δ
CDX	EAS	45.7	61.3	+15.6
KHV	EAS	43.9	61.1	+17.2
CHB	EAS	43.7	68.9	+25.2
CHS	EAS	42.1	62.6	+20.5
CEU	EUR	59.1	55.6	-3.5
FIN	EUR	64.4	55.8	-8.6
YRI	AFR	77.6	68.2	-9.4

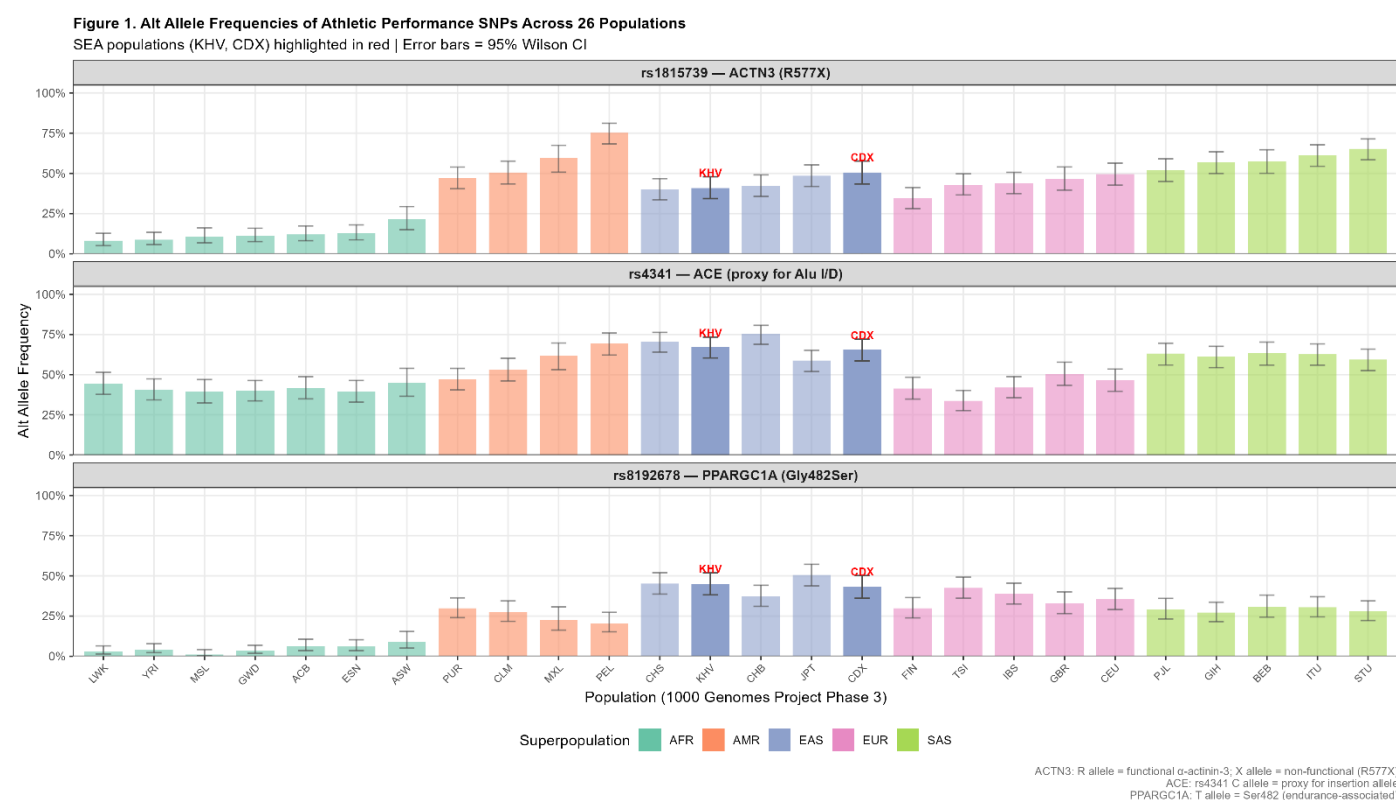
Green=increase;Red=decreaseunderEASmodel.

3.7. Evidence Quality Assessment

Table 12 contextualises the three SNPs within the evidence landscape [9]. Zero studies included SEA population data for any of the three loci prior to this work [9, 37].

Table 12. Evidence quality per SNP based on Psatha et al. [9].

Gene	rsID	NArt.	Meta-anal.	SEA Prev.	This study
ACTN3	rs1815739	29	Yes	0	First report for KHV / CDX jointly with F_{ST} & TGS
ACE	rs4646994	21(+8)	Yes	0	First KHV / CDX estimate (proxy rs4341); prior EAS data from JPN / KOR/ CHN
PPARGC1A	rs8192678	9	No	0	First report for KHV / CDX; no prior SEA data in meta-analysis pool

**Figure 1.** Alternative allele frequencies across 26 populations for three athletic performance-associated SNPs. Error bars = 95 % Wil- son confidence intervals [24]. Red bars = SEA populations (KHV, CDX). Data source: 1000 Genomes Project Phase 3 [19].

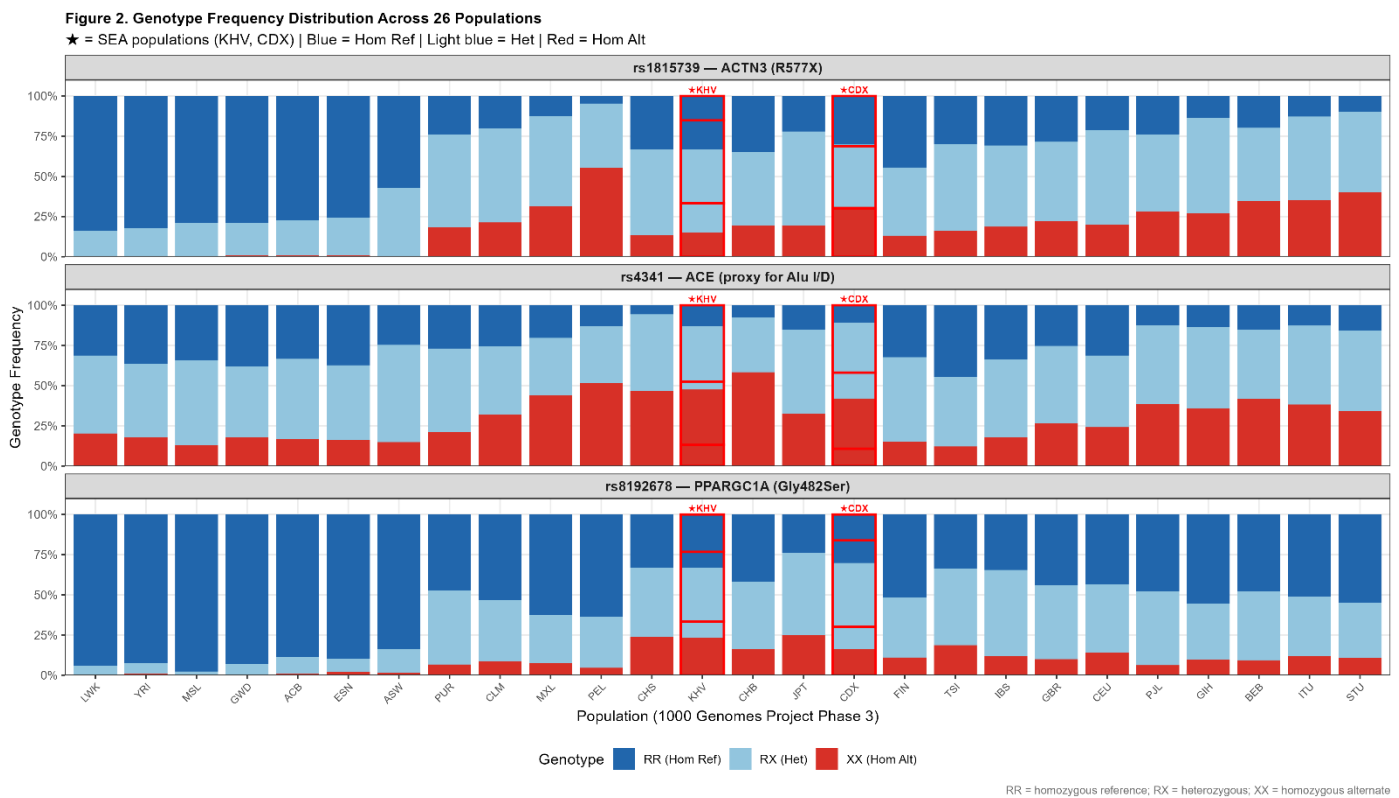


Figure 2. Stacked genotype frequency distribution across 26 populations. RR = reference homozygote; RX = heterozygote; XX = alternative homozygote [19].

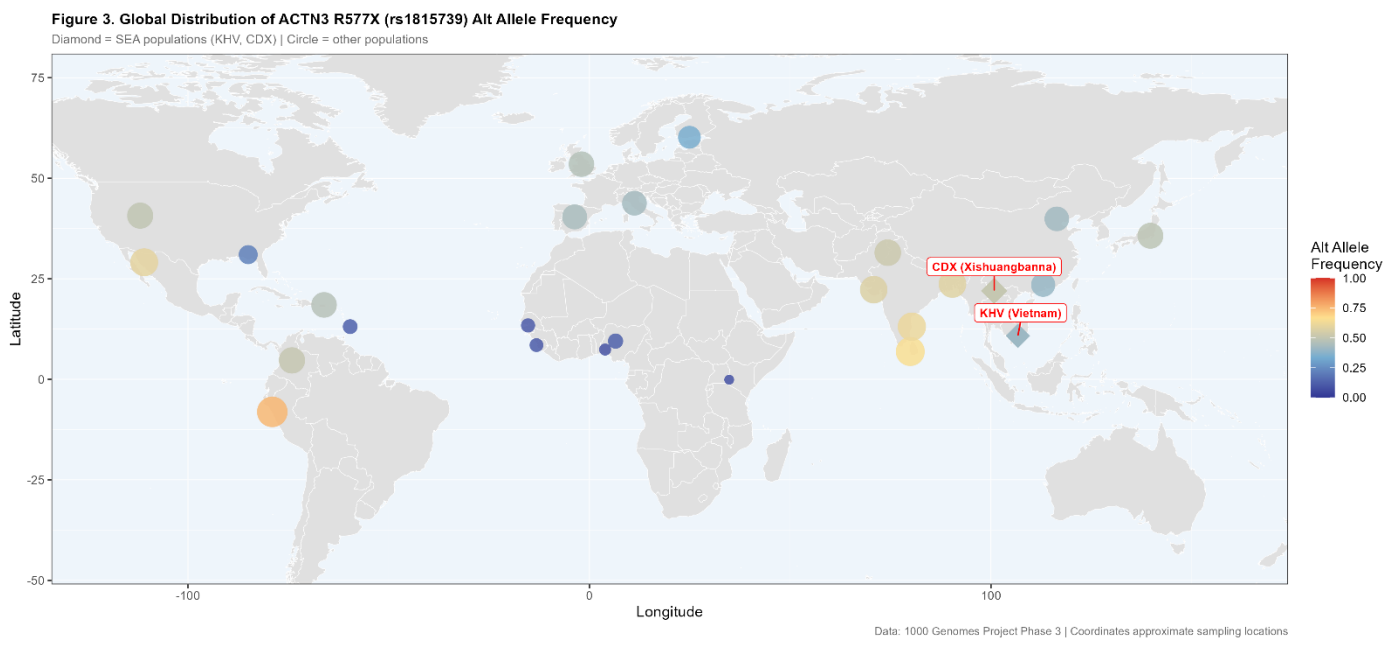


Figure 3. Geographic distribution of ACTN3 rs1815739 alternative allele frequency across 26 populations. Colour scale: blue = low, red = high. Diamond markers = SEA populations [19].

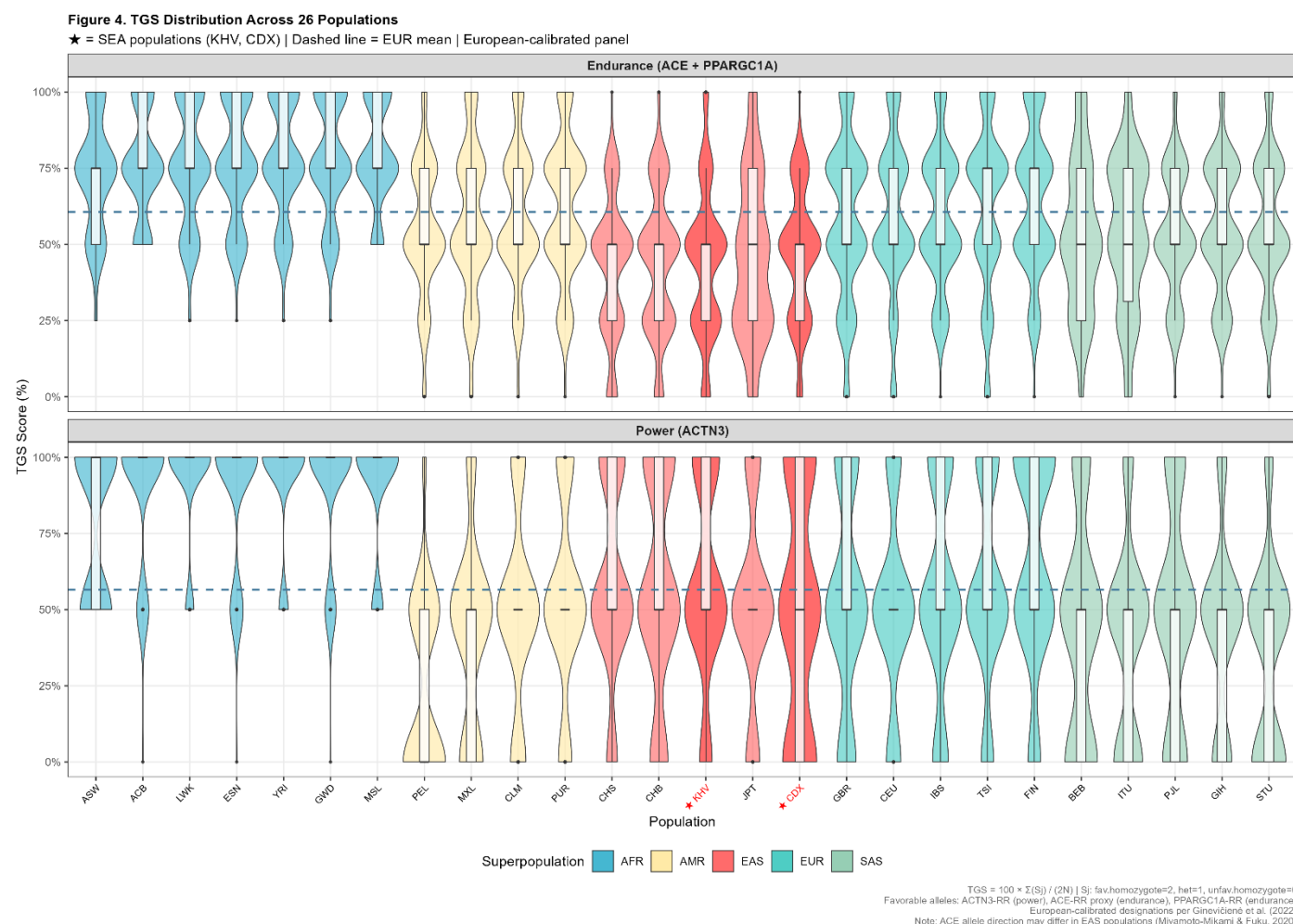


Figure 4. Total Genotype Score (TGS) distributions under the European-calibrated panel [13]. Dashed line = EUR superpopulation mean. Upper panel: TGS Endurance; Lower panel: TGS Power. Red dots = individual SEA samples. Mann-Whitney results in Table 10 [31].

4. Discussion

4.1. SEA Allele Frequencies in Global Context

This study provides, to the best of our knowledge, the first joint population-genetic characterisation of *ACTN3* rs1815739, *ACE* rs4341, and *PPARG1A* rs8192678 specifically for the Kinh Vietnamese (KHV) and Chinese Dai (CDX) populations, with concurrent F_{ST} and TGS analyses. Prior EAS studies of these loci have focused on Japanese, Korean, and Han Chinese cohorts and did not include KHV or CDX. The near-zero pairwise F_{ST} between SEA and the broader EAS superpopulation (all ≤ 0.0004) confirms the genetic proximity expected from phylogeographic analyses [19, 36].

The high *ACE* C-allele frequencies in SEA (≈ 0.66 – 0.67)—substantially exceeding EUR populations (≈ 0.34 – 0.51)—reflect the global EAS pattern and reinforce the need for population-specific directionality assessment [11].

The *PPARG1A* differentiation between SEA and AFR ($F_{ST} = 0.234$) is the most extreme finding of this study, approaching “very high” classification [7]. This level of differentiation suggests that any sports genomics application using an African-population-calibrated *PPARG1A* model in SEA athletes would systematically misassign genetic risk scores [21].

4.2. The European-Calibrated TGS Paradox

Figure 4 and Table 9 expose a fundamental paradox: African populations score highest on both TGS Power and TGS Endurance panels under European-calibrated frameworks [8, 9, 13].

The -27.5% TGS Endurance deficit of KHV versus EUR reference ($\Delta = -16.7$ points, $p = 6.23 \times 10^{-14}$) is entirely a product of the *ACE* directionality assumption. Under the EAS model, both KHV and CDX exceed the EUR reference mean [11]. This result directly extends the observation by Miyamoto-Mikami and Fuku [11] in Japanese swimmers to a broader SEA context.

4.3. HWE Deviations and Population Substructure

The marginal CDX deviation for *ACTN3* ($p = 0.037$) may reflect population substructure within the Xishuangbanna Dai group [19], inbreeding, or the reduced statistical power at $n = 93$. This does not invalidate the frequency estimates but indicates that CDX should be treated with caution [26, 27].

4.4. Implications for Direct-to-Consumer Genetic Testing

The combined evidence strongly cautions against the application of European-calibrated genetic testing panels to SEA individuals. Psatha et al. [9] explicitly concluded that direct-to-consumer genetic testing for athletic performance is “highly risky and even unethical” given current evidence. Without population-specific calibration, TGS-based tools risk systematically misclassifying SEA athletes with high endurance genetic potential [1, 11, 38].

Based on the allele frequency profiles identified in this study, the three polymorphisms offer differential utility for early identification of athletic potential in SEA populations. Athletes carrying the *ACTN3* C-allele (R allele, rs1815739) tend to exhibit greater fast-twitch muscle fibre expression and are more commonly represented among sprint and power athletes, suggesting that genotyping for *ACTN3* rs1815739 may support early identification of power-oriented athletes in SEA. For endurance performance, the *PPARGC1A* rs8192678 G-allele (Gly482) is associated with enhanced mitochondrial biogenesis and VO_2max response to training; however, the high frequency of the T-allele (Ser482) in SEA populations relative to African populations ($F_{ST} = 0.234$) highlights that reference ranges must be population-specific. *ACE* rs4341 C-allele (proxy for the D-allele) is highly prevalent in SEA, but its directionality for endurance versus power is population-dependent and remains ambiguous without SEA-specific athlete cohort data. These findings underscore that no single polymorphism provides definitive athletic phenotype prediction; rather, a multi-locus, SEA-calibrated panel is needed for meaningful early identification of athletic potential.

4.5. Limitations

This study has several limitations. First, rs4341 is an imperfect proxy for the *ACE* Alu I/D polymorphism ($r^2 \approx 0.83$) [19], which introduces measurement uncertainty. Second, major SEA ethnic groups — including Javanese, Filipino, Thai, Malay, and Burmese populations — are absent from the 1000 Genomes Project dataset, limiting the generalisability of findings across the full diversity of Southeast Asia. Third, the dataset contains no phenotypic performance data, meaning no genotype–performance associations can be inferred [9]. Fourth, the sample sizes of $n = 93$ – 99 per population may be insufficient to reliably estimate frequencies of rare genotype combinations [25]. Fifth, TGS as a predictive tool has been criticised in the literature [9, 11, 14] and our TGS analysis is therefore strictly exploratory.

5. Conclusions

This study provides, to the best of our knowledge, the first joint population-genetic characterisation of *ACTN3* rs1815739, *ACE* rs4341, and *PPARGC1A* rs8192678 with concurrent F_{ST} and TGS analyses for the Kinh Vietnamese (KHV) and Chinese Dai (CDX) populations using the 1000 Genomes Project Phase 3 [19]. Existing EAS studies on these loci have been conducted in Japanese, Korean, and Han Chinese cohorts; KHV and CDX have not previously been jointly characterised for these three performances-associated polymorphisms. Future research should focus on expanding Southeast Asian (SEA) genomic reference panels to better represent populations such as Javanese, Malay, Filipino, Thai, and Burmese. In addition, there is a need for properly powered case–control studies involving elite SEA athletes, supported by standardized performance phenotypes to ensure comparability across studies [38]. Calibration of East Asian (EAS)-specific total genotype scores (TGS), particularly for genes such as *ACE*, should also be refined using data derived from EAS athlete cohorts [11]. Furthermore, coordinated multi-centre replication studies for *PPARGC1A* across diverse Asian populations are essential to validate existing findings and improve generalizability [9, 20]. Finally, integrating multi-omics approaches will be crucial to better understand the functional consequences of genetic variation within the context of SEA muscle physiology [13, 37].

Acknowledgments: The authors thank the 1000 Genomes Project Consortium and IGSR for publicly available genotype data and population meta-data [19].

Author Contributions: A.R.: data curation, formal analysis, visualisation, writing. R.F.: conceptualisation, methodology, supervision, writing. L.E.: re-sources, writing (review & editing). All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Data Availability: All genotype data are publicly available at <https://ftp.1000genomes.ebi.ac.uk>.

References

1. Claude Bouchard, Tuomo Rankinen, and James A. Timmons. Genomics and genetics in the biology of adaptation to exercise. *Comprehensive Physiology*, 1(3):1603–1648, 2011. doi: 10.1002/cphy.c100059. <https://pubmed.ncbi.nlm.nih.gov/23733655/>
2. Ekaterina A. Semenova, Elliott C. R. Hall, and Ildus I. Ahmetov. Genes and athletic performance: The 2023 update. *Genes*, 14 (6):1235, 2023. doi: 10.3390/genes14061235. <https://www.mdpi.com/2073-4425/14/6/1235>
3. Hirofumi Zempo, Noriyuki Fuku, Haruka Murakami, Nobuhiro Kikuchi, and Motohiko Miyachi. Heritability of muscle strength: A meta-analysis of twin studies. *Journal of Physical Fitness and Sports Medicine*, 6(2):75–82, 2017. doi: 10.7600/jpfsm.6.75. <https://pubmed.ncbi.nlm.nih.gov/27882617/>
4. Eri Miyamoto-Mikami, Hirofumi Zempo, Noriyuki Fuku, Nobuhiro Kikuchi, Motohiko Miyachi, and Haruka Murakami. Heritability estimates of endurance-related phenotypes: A systematic review and meta-analysis. *Scandinavian Journal of Medicine and Science in Sports*, 28(3):834–845, 2018. doi: 10.1111/sms.12958. <https://pubmed.ncbi.nlm.nih.gov/28801974/>
5. Nan Yang, Daniel G. MacArthur, Jason P. Gulbin, Allan G. Hahn, Alan H. Beggs, Simon Easteal, and Kathryn North. *ACTN3* genotype is associated with human elite athletic performance. *American Journal of Human Genetics*, 73(3):627–631, 2003. doi: 10.1086/377590. <https://pubmed.ncbi.nlm.nih.gov/12879365/>
6. Hugh E. Montgomery, Rosalind Marshall, Harry Hemingway, Simon Myerson, Priscilla Clarkson, Colin Dollery, Martin Hayward, David E. Holliman, Michael Jubb, Michael World, Edward L. Thomas, Ann E. Brynes, Nafeesa Saeed, Michael Barnard, Jimmy D. Bell, Krishna Prasad, Mike Rayson, Philippa J. Talmud, and Steve E. Humphries. Human gene for physical performance. *Nature*, 393:221–222, 1998. doi: 10.1038/30374. <https://pubmed.ncbi.nlm.nih.gov/9607758/>
7. Charlotte Ling, Pernille Poulsen, Erik Carlsson, Martin Ridderstråle, Peter Almgren, Jorgen Wojtaszewski, Henning Beck-Nielsen, Leif Groop, and Allan Vaag. Multiple environmental and genetic factors influence skeletal muscle PGC-1 α and PGC-1 β gene expression in twins. *Journal of Clinical Investigation*, 114 (10):1518–1526, 2004. doi: 10.1172/JCI21889. <https://pubmed.ncbi.nlm.nih.gov/15546003/>
8. Ildus I. Ahmetov and Olga N. Fedotovskaya. Sports genomics: current state of knowledge and future directions. *Cellular and Molecular Exercise Physiology*, 1(1): e1, 2012. doi: 10.7457/cmep.v1i1.e1.
9. Aikaterini Psatha, Zeina N. Al-Mahayri, Christina Mitropoulou, and George P. Patrinos. Meta-analysis of genomic variants in power and endurance sports to decode the impact of genomics on athletic performance and success. *Human Genomics*, 18(1): 47, 2024. doi: 10.1186/s40246-024-00621-9. <https://pubmed.ncbi.nlm.nih.gov/38760851/>
10. Fang Ma, Ying Yang, Xuelong Li, Feng Zhou, Cuicui Gao, Mingli Li, and Longhai Gao. The association of sport performance with *ACE* and *ACTN3* genetic polymorphisms: A systematic review and meta-analysis. *PLOS ONE*, 8(1): e54685, 2013. doi: 10.1371/journal.pone.0054685. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0054685>
11. Eri Miyamoto-Mikami and Noriyuki Fuku. Recent advances in the molecular genetics and sports performance in East Asians. *Journal of Sport and Health Science*, 9(2):106–113, 2020. doi: 10.1016/j.jshs.2019.04.005. <https://pubmed.ncbi.nlm.nih.gov/33308818/>
12. Sewall Wright. Variability within and among natural populations. *Evolution and the Genetics of Populations*, 4, 1978. University of Chicago Press, Chicago.
13. Valentina Ginevičienė, Andrius Pranculis, Kazys Milasius, Vytautas Jaškantas, and Algirdas Utkus. Current state and future directions of sport genomics research for strength and endurance performance. *Biomedicine*, 10(2):298, 2022. doi: 10.3390/biomedicine10020298. <https://www.mdpi.com/2227-9059/10/2/298>
14. Alun G. Williams and Henning Wackerhage. Genetic testing for sports performance, responses to training, and injury risk: Practical and ethical considerations. *Genetic Testing and Molecular Biomarkers*, 21(4):197–199, 2017. doi: 10.1089/gtmb.2017.0011. <https://journals.sagepub.com/doi/full/10.1089/gtmb.2017.29033.gde?cf-mal-redirected=true>
15. Daniel G. MacArthur and Kathryn N. North. A gene for speed? The evolution and function of α -actinin-3. *BioEssays*, 26(7):786–795, 2004. doi: 10.1002/bies.20061. <https://onlinelibrary.wiley.com/doi/10.1002/bies.20061>
16. Aleksandra M. Druzhetskaya, Ildus I. Ahmetov, Irina V. Astratenkova, and Viktor A. Rogozkin. Association of the *ACTN3* R577X polymorphism with power athlete status in Russians. *European Journal of Applied Physiology*, 103:631–634, 2008. doi: 10.1007/s00421-008-0763-1. <https://pubmed.ncbi.nlm.nih.gov/18470530/>
17. Phuntla Tharabenjasin, Noel Pabalan, and Hamdi Jarjanazi. Association of *ACTN3* R577X (rs1815739) polymorphism with elite power sports: A meta-analysis. *PLOS ONE*, 14(1):e0211411, 2019. doi: 10.1371/journal.pone.0211411. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0211411>
18. David R. Woods, Michael World, Mike P. Rayson, Andrew G. Williams, Michael Jubb, James E. Dixon, Yalda Jamshidi, and Steve E. Humphries. Endurance enhancement related to the human angiotensin I-converting enzyme I-D polymorphism is not observed after acute ischaemic forearm exercise. *European Journal of Applied Physiology*, 85:118–121, 2001. doi: 10.1007/s004210100418. <https://pubmed.ncbi.nlm.nih.gov/11513308/>
19. 1000 Genomes Project Consortium, Adam Auton, Gonçalo R. Abecasis, et al. A global reference for human genetic variation. *Nature*, 526:68–74, 2015. doi: 10.1038/nature15393. <https://www.nature.com/articles/nature15393>
20. Zihong He, Yang Hu, Lianyun Feng, Yajun Lu, Guiyu Liu, Yi-Ping Li, Yi Li, and Alejandro Lucia. PGC-1 α gene polymorphisms are associated with elite endurance performance in Chinese national endurance athletes. *International Journal of Sports Physiology and Performance*, 7:306–308, 2012. doi: 10.1123/ijspp.7.3.306.
21. Eugenia Boulygina, Oleg Borisov, Elena Valeeva, Ekaterina Semenova, Elena Kostryukova, Nikolay Kulemin, Dmitry Popov, Edward Generozov, and Ildus Ahmetov. Whole genome sequencing of elite athletes. *Biology of Sport*, 37(3):295–304, 2020. doi: 10.5114/biolsport.2020.96272. <https://pubmed.ncbi.nlm.nih.gov/32879552/>
22. Valerie Obenchain, Michael Lawrence, Vincent Carey, Stephanie Gogarten, Paul Shannon, and Martin Morgan. Variant Annotation: A Bioconductor package for exploration and annotation of genetic variants. *Bioinformatics*, 30(14):2076–2078, 2014. doi: 10.1093/bioinformatics/btu168. <https://pubmed.ncbi.nlm.nih.gov/24681907/>

23. Bruce S. Weir. Genetic data analysis ii. Sinauer Associates, 1996. Sunderland, MA. [Textbook — allele frequency estimation, Chapter 3].
24. Wilson, E. B. (1927). Probable Inference, the Law of Succession, and Statistical Inference. *Journal of the American Statistical Association*, 22(158), 209–212. <https://doi.org/10.1080/01621459.1927.10502953>
25. Lawrence D. Brown, T. Tony Cai, and Anirban DasGupta. Interval estimation for a binomial proportion. *Statistical Science*, 16(2): 101–133, 2001. doi: 10.1214/ss/1009213286. <https://doi.org/10.1214/ss/1009213286>
26. Wigginton, J. E., Cutler, D. J., & Abecasis, G. R. (2005). A note on exact tests of Hardy-Weinberg equilibrium. *American journal of human genetics*, 76(5), 887–893. <https://doi.org/10.1086/429864>
27. Jan Graffelman. Exploring diallelic genetic markers: The Hardy- Weinberg package. *Journal of Statistical Software*, 64(3):1–23, 2015. doi: 10.18637/jss.v064.i03. <https://www.jstatsoft.org/article/view/v064i03>
28. Masatoshi Nei. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences*, 70(12): 3321–3323, 1973. doi: 10.1073/pnas.70.12.3321. <https://doi.org/10.1073/pnas.70.12.3321>
29. Karl Pearson. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine*, 50:157–175, 1900. doi: 10.1080/14786440009463897. https://link.springer.com/chapter/10.1007/978-1-4612-4380-9_2
30. Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple test- ing. *Journal of the Royal Statistical Society B*, 57(1):289–300, 1995. doi: 10.1111/j.2517-6161.1995.tb02031.x.
31. Henry B. Mann and Donald R. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, 18(1):50–60, 1947. <https://doi.org/10.1214/aoms/1177730491>
32. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2024. URL <https://www.R-project.org/>.
33. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the tidyverse. *J Open Source Softw.* 2019;4(43):1686. doi:10.21105/joss.01686. <https://doi.org/10.21105/joss.01686>
34. Goudet J. HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Mol Ecol Notes*. 2005;5(1):184–186. doi:10.1111/j.1471-8286.2004.00828.x. <https://doi.org/10.1111/j.1471-8286.2004.00828.x>
35. Wickham H. ggplot2: elegant graphics for data analysis. 2nd ed. New York: Springer-Verlag; 2016. doi:10.1007/978-3-319-24277-4. <https://doi.org/10.1007/978-3-319-24277-4>
36. Novembre J, Stephens M. Interpreting principal component analyses of spatial population genetic variation. *Nat Genet*. 2008;40(5):646–649. doi:10.1038/ng.139. <https://doi.org/10.1038/ng.139>
37. Juginović A, Kekić A, Aranza I, Biloš V, Armanda M. Next-generation approaches in sports medicine: the role of genetics, omics, and digital health in optimizing athlete performance and longevity—a narrative review. *Life (Basel)*. 2025;15(7):1023. doi:10.3390/life15071023. <https://doi.org/10.3390/life15071023>
38. Varillas-Delgado D, Del Coso J, Gutiérrez-Hellín J, Aguilar-Navarro M, Muñoz A, Maestro A, et al. Genetics and sports performance: the present and future in the identification of talent for sports based on DNA testing. *Eur J Appl Physiol*. 2022;122(8):1811–1830. doi:10.1007/s00421-022-04945-z. <https://doi.org/10.1007/s00421-022-04945-z>

A. Supplementary Table S2: Full Chi-Square Results**Table 13.** Complete chi-square test results: SEA vs. all superpopulations (Bonferroni corrected [30]).

Gene	SEA	vs.	χ^2	<i>Praw</i>	<i>PBonf</i>	Sig.
ACE	CDX	AFR	39.03	<0.001	<0.001	***
ACE	CDX	AMR	4.56	0.033	0.328	ns
ACE	CDX	EAS	0.25	0.618	1.000	ns
ACE	CDX	EUR	33.55	<0.001	<0.001	***
ACE	CDX	SAS	0.93	0.336	1.000	ns
ACE	KHV	AFR	46.67	<0.001	<0.001	***
ACE	KHV	AMR	6.71	0.010	0.096	ns
ACE	KHV	EAS	0.01	0.937	1.000	ns
ACE	KHV	EUR	40.32	<0.001	<0.001	***
ACE	KHV	SAS	1.99	0.159	1.000	ns
ACTN3	CDX	AFR	182.04	<0.001	<0.001	***
ACTN3	CDX	AMR	2.64	0.104	1.000	ns
ACTN3	CDX	EAS	2.43	0.119	1.000	ns
ACTN3	CDX	EUR	3.20	0.074	0.736	ns
ACTN3	CDX	SAS	4.25	0.039	0.393	ns
ACTN3	KHV	AFR	114.77	<0.001	<0.001	***
ACTN3	KHV	AMR	16.45	<0.001	0.001	***
ACTN3	KHV	EAS	0.79	0.373	1.000	ns
ACTN3	KHV	EUR	0.43	0.511	1.000	ns
ACTN3	KHV	SAS	21.10	<0.001	<0.001	***
PPARGC1A	CDX	AFR	286.07	<0.001	<0.001	***
PPARGC1A	CDX	AMR	21.34	<0.001	<0.001	***
PPARGC1A	CDX	EAS	0.10	0.755	1.000	ns
PPARGC1A	CDX	EUR	3.23	0.073	0.725	ns
PPARGC1A	CDX	SAS	14.19	<0.001	0.002	**
PPARGC1A	KHV	AFR	317.52	<0.001	<0.001	***
PPARGC1A	KHV	AMR	27.36	<0.001	<0.001	***
PPARGC1A	KHV	EAS	0.03	0.856	1.000	ns
PPARGC1A	KHV	EUR	5.55	0.019	0.185	ns
PPARGC1A	KHV	SAS	19.25	<0.001	<0.001	***

** $p < 0.01$; *** $p < 0.001$; ns = not significant