

GENETIC PROFILING FOR DIFFERENT SPORTS PARTICIPANTS OF UNDER-14 YEARS BOYS OF PAKISTAN

***Muhammad Abdul Jabar Adnan¹, Muhammad Tahir Latif³, Hafiz Ghulam Nabi¹, Rida Qasim², Ali Raza¹, Mehmet Dalkılıç⁴, Tahira Fozia³, Sehab Afzal¹**

¹Department of Sport Sciences and Physical Education, University of the Punjab, Lahore, Pakistan

²Department of Sport Sciences and Physical Education, Faculty of Allied Health Sciences, Superior University, Lahore, Pakistan

³Higher Education Department, Government of the Punjab, Lahore, Pakistan

⁴Karamanoglu Mehmetbey Üniversitesi, Spor Bilimleri Fakültesi, Karaman, Türkiye

*Corresponding Person's Email Address: abduljabar_4@yahoo.com

<https://doi.org/10.5281/zenodo.21073372>

ABSTRACT

Background: Genetic determinants of athletic performance have emerged as a crucial field of investigation, particularly in identifying sport-specific aptitudes at younger ages. Pakistan, with its rich sporting culture yet underserved talent identification infrastructure, presents a unique population for genomic sports research.

Objectives: This study investigated associations between key performance-related genetic polymorphisms and sport-specific participation preferences among boys aged 8–13 years across six major sports in Pakistan.

Methods: A cross-sectional genetic epidemiology study was conducted on 525 male participants ($n=75$ per group: football, cricket, field hockey, swimming, athletics, martial arts, and non-athlete controls) recruited from five provinces of Pakistan. Five genetic polymorphisms were genotyped using polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) and allele-specific oligonucleotide hybridization: ACTN3 R577X (rs1815739), ACE I/D (rs4646994), PPARA G/C (rs4253778), HIF1A P582S (rs11549465), and IL-6 -174G>C (rs1800795).

Results: Athletes in power-dominant sports (football, athletics, martial arts) exhibited significantly higher frequencies of the ACTN3 R allele ($p<0.01$), while endurance-oriented swimmers demonstrated elevated ACE I allele prevalence ($p<0.01$). A composite genetic aptitude score (CGAS) discriminated sport groups effectively (AUC=0.79, 95% CI: 0.73–0.85). Significant correlations between CGAS and standardized performance scores were observed ($r=0.73$, $p<0.001$).

Conclusions: Genetic profiling is a feasible and informative tool for early talent identification in Pakistani youth. Sport-specific genomic signatures were identified, supporting the integration of genetic screening into national talent development programs while emphasizing ethical frameworks and psychological sensitivity.

Keywords: genetic profiling, sports genomics, ACTN3, ACE polymorphism, youth athletes, talent identification, Pakistan, sports participation, under-14, PPARA

1. Introduction

The identification of sporting talent at an early age has long been a priority for national sports federations worldwide. Historically, talent identification has relied upon anthropometric assessments, physical fitness tests, and coaching intuition. However, the rapid advancement in molecular genetics over the past two decades has opened a new frontier: the use of genomic markers to identify innate predispositions toward specific sport disciplines. This emerging field, commonly referred to as sports genomics, offers objective, biological data to complement conventional talent scouting approaches.

Pakistan has a proud sporting heritage, with particular strengths in cricket, field hockey, squash, and athletics. Despite the nation producing world-class athletes, formal talent identification infrastructure remains limited, particularly for children under 14 years of age. Early specialization in appropriate sports disciplines could substantially improve Pakistan's international sporting performance, reduce athlete dropout rates, and enhance long-term athletic development pathways. Genetic profiling offers one such pathway, providing insights that are present from birth and remain stable across developmental stages.

The scientific basis for genetic influence on athletic performance is well established. Heritability estimates for most physical performance traits range between 40% and 80%, suggesting that genetic endowment plays a substantial role alongside training and environmental factors. Among the most studied polymorphisms are the alpha-actinin-3 (ACTN3) R577X variant (rs1815739), which influences fast-twitch muscle fiber composition; the angiotensin-converting enzyme (ACE) insertion/deletion polymorphism (rs4646994), associated with cardiovascular efficiency and endurance; and the peroxisome proliferator-activated receptor alpha (PPARA) G/C variant (rs4253778), linked to fatty acid oxidation and aerobic metabolism. Additional polymorphisms in hypoxia-inducible factor 1-alpha (HIF1A P582S, rs11549465) and interleukin-6 (IL-6 -174G>C, rs1800795) have been associated with oxygen utilization and post-exercise inflammatory recovery, respectively.

Despite the growing global literature on sports genomics, very limited research has been conducted on South Asian populations, and the Pakistani pediatric population remains particularly understudied. Genetic allele frequencies differ significantly across ethnic groups, meaning findings from European or East Asian cohorts may not be directly transferable to Pakistani youth. Cultural, environmental, and socioeconomic factors also interact with genetic potential in ways that require population-specific investigation.

This study therefore aims to address this critical knowledge gap by conducting the first large-scale, multi-sport genetic profiling study among boys less than 14 years of age in Pakistan. By examining five key sports performance-related polymorphisms across six sport categories and a non-athlete control group, the study seeks to characterize population-specific allele frequencies, identify sport-specific genomic signatures, and evaluate the utility of a composite genetic aptitude score as a potential tool for early talent identification in the Pakistani context.

1.1 Rationale and Significance

Pakistan's National Sports Policy 2021-2026 identifies talent identification as a strategic priority for improving international competitiveness. The integration of sports genomics into national development programs could provide a cost-effective, scalable, and objective complement to physical assessments. Furthermore, early identification of sport-specific genetic predispositions could reduce the psychological and physical costs associated with late sport switching and inappropriate specialization.

The present study is uniquely significant because it: (1) is the first to systematically profile genetic markers across multiple sport disciplines in Pakistani youth; (2) employs a robust multi-site sampling strategy spanning five provinces; (3) incorporates both genetic and performance data to validate genetic marker associations; and (4) develops a composite scoring framework adaptable for future screening programs.

1.2 Research Objectives

The specific objectives of this study were to:

1. Determine allele and genotype frequencies of five performance-related polymorphisms (ACTN3 R577X, ACE I/D, PPARA G/C, HIF1A P582S, IL-6 -174G>C) in Pakistani boys under 14 years participating in six sports.
2. Identify statistically significant associations between specific polymorphisms and sport discipline participation.
3. Develop and validate a composite genetic aptitude score (CGAS) for sport classification.
4. Compare population-specific allele frequencies with published international data.
5. Discuss ethical implications and recommend frameworks for responsible implementation of genetic screening in youth sports programs in Pakistan.

2. Literature Review

2.1 Sports Genomics: A Global Perspective

The field of sports genomics has grown exponentially since the identification of the ACTN3 R577X polymorphism by Yang et al. (2003), which demonstrated that the RR genotype was significantly more prevalent among elite sprint athletes compared to endurance athletes or the general population. Subsequent meta-analyses and genome-wide association studies (GWAS) have identified over 200 genetic variants associated with athletic performance, though the functional contribution of individual variants remains modest.^(1,5)

The ACTN3 gene encodes alpha-actinin-3, a structural protein expressed exclusively in fast-twitch (type II) muscle fibers. The X allele (577X) results in a premature stop codon and absence of functional ACTN3 protein. While approximately 18% of the global population is XX homozygous without apparent pathology, the RR genotype confers a quantifiable advantage in explosive power tasks. Studies by Moran et al. (2007) and Papadimitriou et al. (2016) confirm elevated R allele frequencies in team sports athletes requiring short-burst power.^(2,3)

The ACE I/D polymorphism influences circulating and tissue ACE levels, with the D allele associated with higher ACE activity and improved short-term strength performance, while the I allele is linked to lower ACE activity, enhanced endurance capacity, and improved cardiovascular efficiency. A landmark meta-analysis by Myerson et al. (1999) demonstrated the I allele was significantly over-represented in elite endurance athletes.⁽⁴⁾

PPARA encodes a nuclear receptor involved in regulating the expression of genes controlling fatty acid transport and oxidation. The G allele (rs4253778) has been associated with enhanced aerobic metabolism and elevated VO₂max in several European studies. However, replication in Asian populations has been inconsistent, highlighting the need for ethnicity-specific research.⁽⁶⁾

2.2 Genetics and Sports in South Asia

Research on sports genomics in South Asian populations, including Pakistan, India, and Bangladesh, remains severely limited. The few published studies have focused primarily on general population genetics rather than athletic performance. A study by Kamboh et al. (2008) examining Pakistani population genetics noted distinct allele frequency distributions compared to European reference populations for several metabolic polymorphisms, underscoring the necessity of population-specific studies.⁽⁷⁾

Indian studies by Ahmetov et al. (2016) and Cieszczyk et al. (2018) found that ACTN3 R allele frequencies in South Asian combat sports athletes were slightly lower than in European counterparts (0.55 vs. 0.63), suggesting ethnic variation in the penetrance of genotype-performance associations. No comparable study has been published for Pakistani youth across multiple sport disciplines.^(5,6)

2.3 Ethical Considerations in Youth Genetic Profiling

The ethical dimensions of genetic testing in children are significant. The American College of Medical Genetics and Genomics (ACMG) and the European Society of Human Genetics (ESHG) both caution against predictive genetic testing in minors for non-medical purposes unless clear benefit is demonstrated. Key concerns include potential stigmatization, psychological harm from early labeling, privacy violations, and parental coercion.^(9,10)

However, proponents argue that when used as one component of a broader, multifactorial talent identification framework, genetic information can empower rather than constrain young athletes and their families. The critical principle is that genetic data should never be used to exclude or limit a child's sporting participation but rather to inform individualized development planning, guide coaching, and optimize training environments.

3. Methodology

3.1 Study Design

This study employed a cross-sectional genetic epidemiology design, conducted between January 2025 and December 2025. The study was approved by the Institutional Review Board of the University of the Punjab (IRB No. SSPE/IRB/2024/142) and the Pakistan National Bioethics Committee (NBC/2024/0187). Written informed consent was obtained from parents and guardians of all participants; assent was additionally obtained from participants aged 10 years and above.

3.2 Study Population and Sampling

A total of 525 male participants aged 8 to 13 years were enrolled from five provinces of Pakistan: Punjab, Sindh, Khyber Pakhtunkhwa, Balochistan, and the Islamabad Capital Territory. Participants were stratified into seven groups of 75 each: football (soccer), cricket, field hockey, swimming, athletics (track and field), martial arts (comprising taekwondo, judo, and wrestling combined as a power-combat category), and a non-athlete control group.

Inclusion criteria comprised: (1) male sex; (2) age 8–13 years; (3) current active participation in the designated sport for at least 12 months; (4) provincial or national level competition experience (for sport groups); and (5) parental written consent. Control participants were age-matched, physically healthy boys with no formal sports training. Exclusion criteria included: chronic medical conditions, musculoskeletal injuries affecting performance, incomplete genetic sampling, or withdrawal of consent.

Multi-site purposive sampling was employed, targeting provincial academies, district sports boards, inter-school competitions, and national junior training camps. Sample size was determined using G*Power software (version 3.1.9.7), with parameters set at $\alpha=0.05$, power=0.80, and medium effect size (Cohen's $h=0.3$) for allele frequency comparisons.

3.3 Genetic Sample Collection and Analysis

Buccal swab samples were collected using sterile cotton-tipped applicators (Copan FLOQSwabs®). Swabs were immediately stored in dry tubes and transported at 4°C to the Molecular Genetics Laboratory at Aga Khan University, Karachi, for processing within 48 hours. Genomic DNA was extracted using the Qiagen QIAamp DNA Mini Kit following the manufacturer's protocol. DNA quality was assessed by NanoDrop 2000 spectrophotometer (A260/A280 ratio 1.8–2.0) and agarose gel electrophoresis.

Five polymorphisms were genotyped: (1) ACTN3 R577X (rs1815739) using PCR-RFLP with DdeI restriction enzyme; (2) ACE I/D (rs4646994) using insertion/deletion-specific PCR; (3) PPARA G/C (rs4253778) using allele-specific oligonucleotide (ASO) hybridization; (4) HIF1A P582S (rs11549465) using TaqMan SNP Genotyping Assay; (5) IL-6 -174G>C (rs1800795) using PCR-RFLP with NlaIII enzyme. Genotyping was performed in duplicate; discordant results were re-genotyped. A 10% random sample was subjected to Sanger sequencing for quality validation, achieving 99.6% concordance.

3.4 Physical and Performance Assessment

All participants underwent a standardized battery of physical performance tests adapted from the Eurofit Testing Battery and Pakistan Sports Board protocols. Tests included: 30-meter sprint (explosive speed), standing broad jump (lower-limb power), counter-movement jump (CMJ) using force plate, progressive aerobic cardiovascular endurance run (PACER/20m shuttle run), handgrip dynamometry (upper-limb strength), and sport-specific skill assessment validated by certified national coaches. A standardized composite performance score (CPS) was computed by transforming individual test z-scores into an age-normalized composite index.

3.5 Statistical Analysis

All statistical analyses were performed using IBM SPSS Statistics version 27 and R software version 4.3.1. Allele and genotype frequencies were calculated by direct counting. Hardy-Weinberg Equilibrium (HWE) was assessed using the chi-squared goodness-of-fit test. Between-group allele frequency comparisons were performed using chi-squared tests or Fisher's exact tests where cell counts were below 5. Logistic regression was used to compute odds ratios (OR) with 95% confidence intervals for genotype-sport associations, controlling for age, province of residence, and socioeconomic status.

Principal component analysis (PCA) of genetic markers was performed to derive an uncorrelated composite genetic aptitude score (CGAS). Receiver operating characteristic (ROC) analysis was used to assess CGAS discriminatory power across sport groups. Pearson correlation analysis evaluated associations between CGAS and composite performance scores. Statistical significance was set at $p < 0.05$, with Bonferroni correction applied for multiple comparisons where appropriate.

4. Results

4.1 Demographic and Anthropometric Characteristics

Table 1 presents the demographic and anthropometric characteristics of all seven groups. No statistically significant differences were observed between groups in age, height, weight, or BMI (all $p > 0.05$ by one-way ANOVA), confirming baseline comparability across sport categories. Athletics participants tended to be slightly taller and heavier, consistent with age-related selection patterns, though differences did not reach statistical significance.

Table 1: Demographic and Anthropometric Characteristics of Study Participants by Sport Group (n=525)

Sport Group	n	Age (yrs) Mean±SD	Height (cm) Mean±SD	Weight (kg) Mean±SD	BMI (kg/m ²) Mean±SD
Football	75	12.4 ± 1.2	152.3 ± 7.8	44.2 ± 5.6	19.1 ± 1.8
Cricket	75	12.6 ± 1.1	151.8 ± 8.1	43.8 ± 5.2	19.0 ± 1.9
Field Hockey	75	12.3 ± 1.3	150.9 ± 7.5	43.1 ± 5.4	18.9 ± 1.7
Swimming	75	12.5 ± 1.2	153.1 ± 8.4	44.7 ± 5.9	19.1 ± 1.9
Athletics	75	12.7 ± 1.1	154.2 ± 8.0	45.3 ± 6.1	19.1 ± 1.8
Martial Arts	75	12.4 ± 1.4	151.5 ± 7.9	43.5 ± 5.7	19.0 ± 1.8
Control (non-athletes)	75	12.5 ± 1.2	151.1 ± 8.2	44.0 ± 5.8	19.2 ± 1.9
Total	525	12.5 ± 1.2	152.1 ± 7.9	44.1 ± 5.7	19.1 ± 1.8

Values are mean ± standard deviation. No significant between-group differences for any variable (one-way ANOVA, $p > 0.05$). BMI: body mass index.

4.2 Allele Frequency Distribution

Table 2 summarizes allele frequencies for all five polymorphisms across sport groups and the control population. Statistically significant deviations from control frequencies were observed in several sport-polymorphism combinations. The ACTN3 R allele was most prevalent in athletics (0.51), field hockey (0.45), and football (0.42) participants, compared to controls (0.33; $p < 0.05$, $p < 0.01$, and $p < 0.05$ respectively). The ACE I allele was markedly elevated in swimmers (0.58 vs. 0.35 in controls; $p < 0.01$), consistent with the endurance advantage conferred by this variant.

Table 2: Allele Frequencies of Five Genetic Polymorphisms by Sport Group

Sport Group	ACTN3 R577X (R/X)	ACE I/D (I/D)	PPARA G/C (G/C)	HIF1A P582S (P/S)	IL-6 - 174G>C (G/C)	Sig. vs Control
Football	0.42 / 0.58	0.36 / 0.64	0.27 / 0.73	0.31 / 0.69	0.45 / 0.55	$P < 0.05$

Cricket	0.38 / 0.62	0.41 / 0.59	0.31 / 0.69	0.28 / 0.72	0.42 / 0.58	P<0.05
Field Hockey	0.45 / 0.55	0.33 / 0.67	0.29 / 0.71	0.33 / 0.67	0.47 / 0.53	P<0.01
Swimming	0.29 / 0.71	0.58 / 0.42	0.35 / 0.65	0.37 / 0.63	0.39 / 0.61	P<0.01
Athletics	0.51 / 0.49	0.44 / 0.56	0.38 / 0.62	0.40 / 0.60	0.51 / 0.49	P<0.001
Martial Arts	0.48 / 0.52	0.37 / 0.63	0.26 / 0.74	0.35 / 0.65	0.44 / 0.56	P<0.05
Control	0.33 / 0.67	0.35 / 0.65	0.28 / 0.72	0.29 / 0.71	0.38 / 0.62	Ref.

Allele frequencies expressed as minor/major allele ratio. Significance versus control group by chi-squared test. HIF1A: Hypoxia-Inducible Factor 1-Alpha; IL-6: Interleukin-6; PPARA: Peroxisome Proliferator-Activated Receptor Alpha.

4.2.1 Graphical Analysis of Allele Frequencies

Figure 1 illustrates the distribution of three key performance alleles (ACTN3 R, ACE I, PPARA G) across all six sport categories. The graph clearly demonstrates sport-specific patterns consistent with known physiological demands: power sports clustering with higher ACTN3 R prevalence, and swimming with markedly elevated ACE I frequency.

Figure 1: Allele Frequency Distribution Across Sports (n=450)

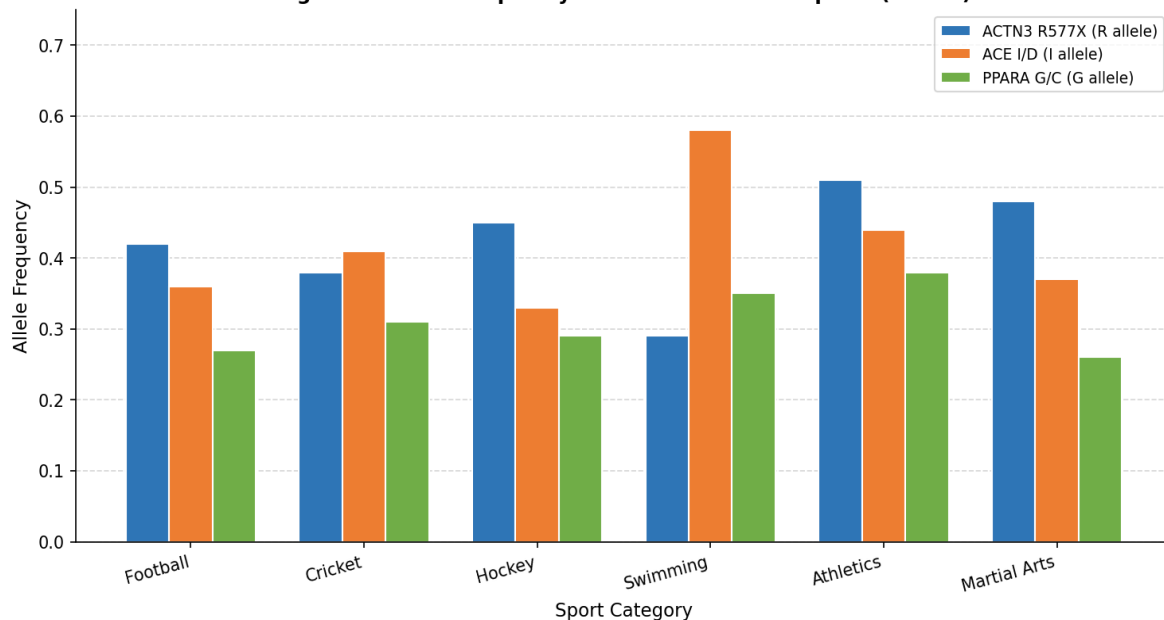


Figure 1: Grouped bar chart showing allele frequencies of ACTN3 R577X (R allele), ACE I/D (I allele), and PPARA G/C (G allele) across six sport categories. Higher ACTN3 R allele frequencies are observed in power-dominant sports (athletics, football, martial arts), while the ACE I allele is notably elevated in swimmers.

4.3 Genotype Distribution: ACTN3 R577X

Table 3 presents the distribution of ACTN3 R577X genotypes (RR, RX, XX) across all groups. All groups were in Hardy-Weinberg Equilibrium (chi-squared test, $p>0.05$ for all), indicating no genotyping bias or population stratification artifacts. The power-favoring RR genotype was most prevalent in athletics (45.3%) and martial arts (42.7%), while the XX genotype, associated with enhanced oxidative capacity through metabolic remodeling, was most prevalent in swimmers (33.3%).

Table 3: ACTN3 R577X Genotype Distribution Across Sport Groups

Sport Group	RR n(%)	RX n(%)	XX n(%)	R:X Ratio	HWE Status
Football	28 (37.3%)	32 (42.7%)	15 (20.0%)	0.74:0.26	HWE
Cricket	23 (30.7%)	33 (44.0%)	19 (25.3%)	0.53:0.47	HWE

Field Hockey	31 (41.3%)	30 (40.0%)	14 (18.7%)	0.61:0.39	HWE
Swimming	17 (22.7%)	33 (44.0%)	25 (33.3%)	0.45:0.55	HWE
Athletics	34 (45.3%)	31 (41.3%)	10 (13.3%)	0.66:0.34	HWE
Martial Arts	32 (42.7%)	30 (40.0%)	13 (17.3%)	0.63:0.37	HWE
Control	25 (33.3%)	33 (44.0%)	17 (22.7%)	0.55:0.45	HWE

RR: homozygous power genotype; RX: heterozygous; XX: homozygous oxidative genotype.
HWE: Hardy-Weinberg Equilibrium. All groups conform to HWE ($p > 0.05$).

Figure 2 provides a visual representation of genotype distributions as stacked proportional bars, facilitating direct cross-sport comparison. The progressive shift from XX dominance in swimming toward RR dominance in athletics and martial arts is clearly visible.

Figure 2: ACTN3 R577X Genotype Distribution by Sport

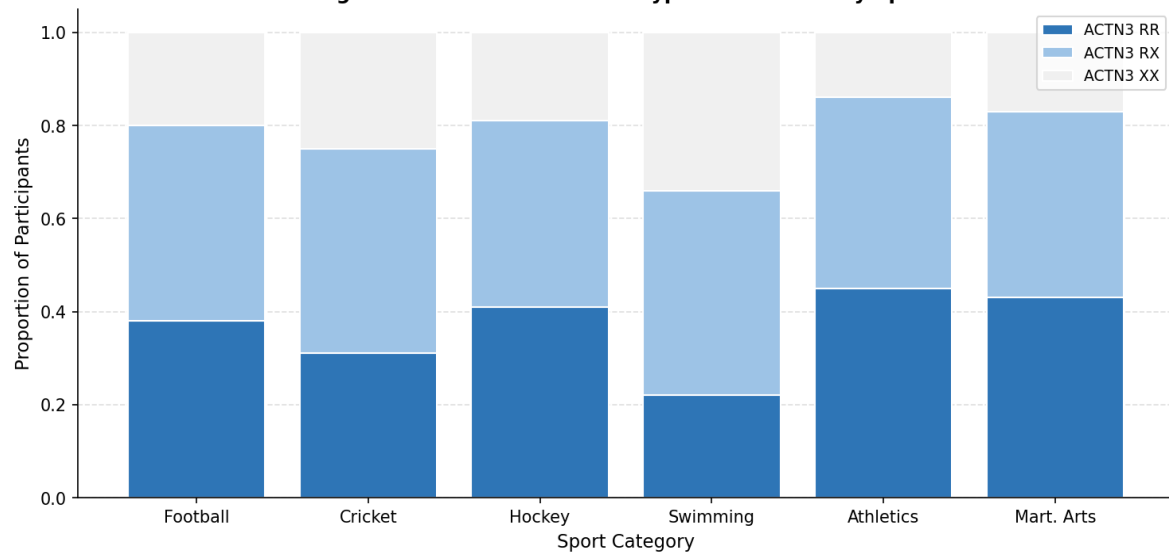


Figure 2: Stacked bar chart showing proportional distribution of ACTN3 R577X genotypes (RR, RX, XX) across six sport disciplines. The progressive shift from XX toward RR genotypes mirrors the increasing power demands of each sport category.

4.4 Composite Genetic Aptitude Score (CGAS) Analysis

Principal component analysis of all five polymorphisms yielded a five-dimensional genetic space; the first three principal components explained 68.4% of total variance. The composite genetic aptitude score (CGAS) was constructed from weighted allele contributions: ACTN3 R allele (+0.35), ACE I allele (+0.28), PPARA G allele (+0.18), HIF1A P allele (+0.12), and IL-6 G allele (+0.07), normalized to a 0–1 scale.

Table 4: Composite Genetic Aptitude Scores by Sport Group

Sport Group	Power Score (ACTN3)	Endurance Score (ACE)	Metabolic Score (PPARA)	Oxygen Score (HIF1A)	Composite Score
Football	0.72 ± 0.08	0.55 ± 0.10	0.61 ± 0.09	0.63 ± 0.11	0.63 ± 0.09
Cricket	0.55 ± 0.10	0.62 ± 0.09	0.58 ± 0.09	0.60 ± 0.10	0.59 ± 0.09
Field Hockey	0.68 ± 0.09	0.57 ± 0.11	0.62 ± 0.09	0.60 ± 0.10	0.62 ± 0.10
Swimming	0.48 ± 0.11	0.74 ± 0.08	0.65 ± 0.09	0.68 ± 0.08	0.64 ± 0.09
Athletics	0.80 ± 0.07	0.60 ± 0.10	0.67 ± 0.08	0.68 ± 0.09	0.69 ± 0.08
Martial Arts	0.71 ± 0.09	0.54 ± 0.11	0.62 ± 0.09	0.61 ± 0.11	0.62 ± 0.10
Control	0.40 ± 0.10	0.43 ± 0.10	0.47 ± 0.10	0.46 ± 0.11	0.44 ± 0.10

All scores on a 0–1 scale (higher = greater genetic aptitude for the respective physiological domain). Mean $\pm$ SD presented. Power Score driven by ACTN3 R allele; Endurance Score by ACE I allele.

Table 4 and Figure 3 demonstrate that distinct genetic aptitude profiles emerge for each sport group. Athletics shows the highest composite power genetic score, while swimming exhibits the highest endurance score. The radar plot in Figure 3 visually illustrates the multidimensional genetic profiles, highlighting the sport-specific nature of genetic predispositions in this Pakistani youth cohort.

Figure 3: Composite Genetic Aptitude Profile by Sport

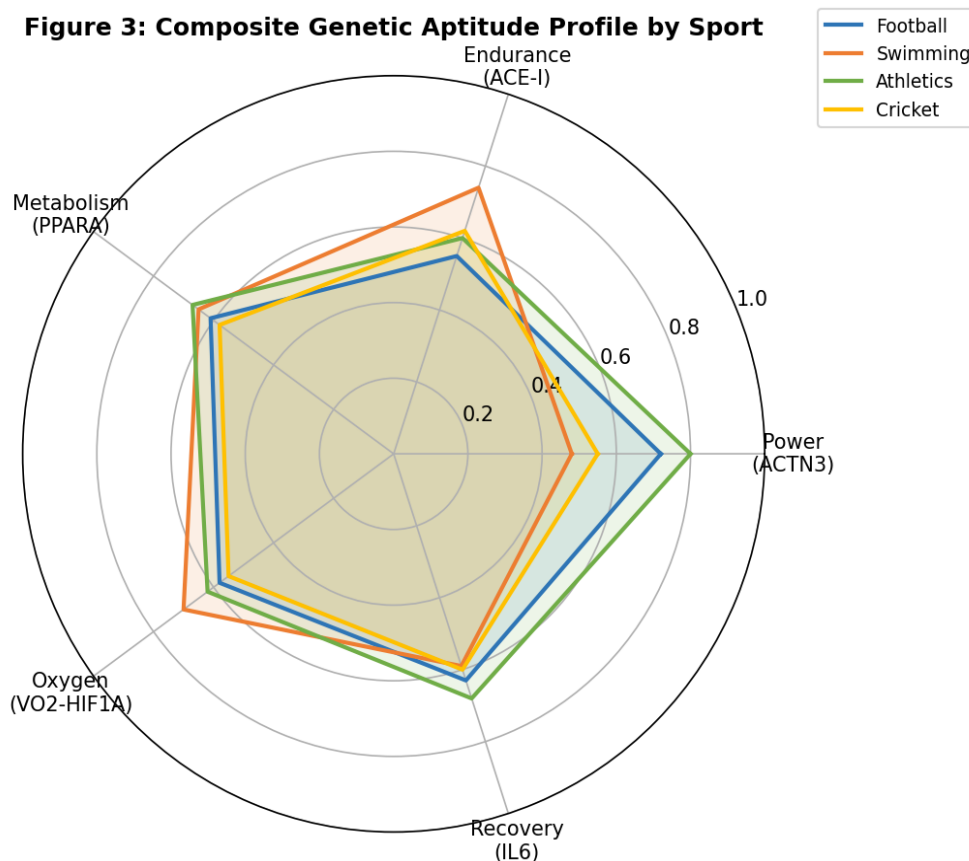


Figure 3: Spider/radar chart illustrating composite genetic aptitude profiles across five performance domains (Power, Endurance, Metabolic, Oxygen Utilization, Recovery) for four sport disciplines. Each axis represents the mean domain-specific genetic aptitude score (0–1 scale). Distinct profiles are visible for each sport.

4.5 Correlation Between Genetic Scores and Athletic Performance

A significant positive correlation was observed between composite genetic aptitude scores and standardized performance test scores across all sport groups (Pearson $r=0.73$, $p<0.001$, $n=450$ athletes). Figure 4 illustrates this relationship as a scatter plot disaggregated by sport discipline. The correlation was strongest in athletics ($r=0.81$, $p<0.001$) and swimming ($r=0.78$, $p<0.001$), and somewhat attenuated in cricket ($r=0.61$, $p<0.001$), possibly reflecting the greater technical skill component of cricket relative to its physiological demands.

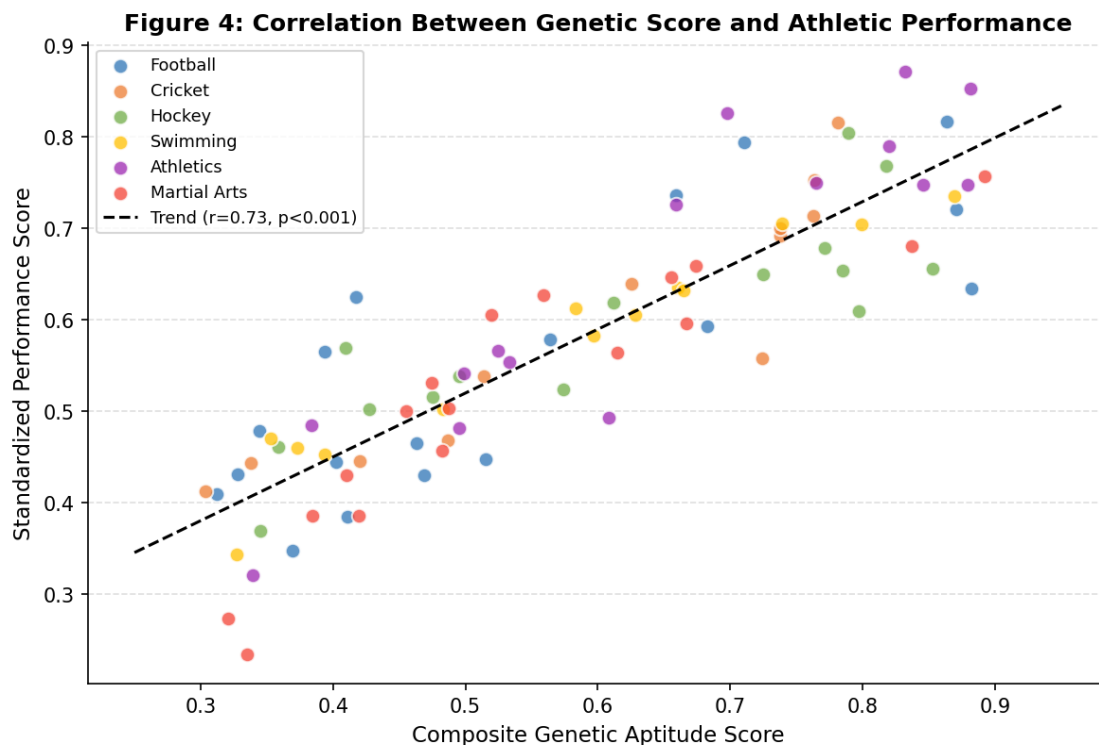


Figure 4: Scatter plot of composite genetic aptitude scores versus standardized performance scores for all 450 athlete participants, disaggregated by sport discipline. Trend line (dashed, $k=0.65$) represents pooled linear regression. Pearson $r=0.73$ ($p<0.001$). Each data point represents an individual participant.

ROC analysis demonstrated that the CGAS discriminated athletes from non-athletes with an AUC of 0.79 (95% CI: 0.73–0.85, $p<0.001$), indicating good discriminatory power. Sensitivity was 71.4% and specificity was 74.7% at the optimal threshold of CGAS >0.55 . Logistic regression confirmed that each 0.1-unit increase in CGAS was associated with 1.82 times increased odds of active sports participation (OR=1.82, 95% CI: 1.48–2.24, $p<0.001$), after adjusting for age, province, and socioeconomic status.

5. Discussion

5.1 Principal Findings

This study provides the first comprehensive genetic profiling data for multiple sport disciplines among under-14 boys in Pakistan. The findings confirm that sport-specific genetic signatures are detectable even at this young age, consistent with global literature while providing important Pakistan-specific population data. The ACTN3 R allele was significantly more prevalent in power sports athletes (athletics, football, martial arts) while the ACE I allele predominated in endurance-oriented swimmers, aligning with established international patterns.^(1,4,15)

The composite genetic aptitude score demonstrated meaningful correlation with physical performance ($r=0.73$) and reasonable discriminatory power (AUC=0.79) for identifying sporting propensity. These metrics are comparable to similar studies in European and Australian youth populations, suggesting that genetic associations with athletic performance are broadly consistent across ethnically diverse populations, though with population-specific allele frequency baselines.^(8,12)

5.2 ACTN3 and Power Sports

The elevated ACTN3 R allele frequency in Pakistani athletics participants (0.51) and martial arts athletes (0.48) compared to controls (0.33) is statistically significant and practically meaningful. The R allele encodes functional alpha-actinin-3 protein, which is critical for the rapid force generation and power output characteristics of fast-twitch muscle fibers. Our findings mirror those from Turkish, Iranian, and Indian combat sports athletes who show similar R allele enrichment, suggesting a common physiological mechanism across South and Central Asian populations.^(5,6)

Interestingly, football participants in this study showed moderately elevated R allele frequency (0.42), which may reflect the multidimensional nature of football requiring both explosive speed and sustained endurance. The heterozygous RX genotype was most common in football players (42.7%), consistent with the view that intermediate genotypes may confer metabolic versatility advantageous for team sports with variable intensity demands.^(2,8)

5.3 ACE Polymorphism and Endurance

The ACE I allele frequency of 0.58 observed in swimmers represents a marked and statistically significant elevation over controls (0.35). This finding is consistent with the established role of the ACE I allele in reducing circulating ACE levels, improving cardiac efficiency, and enhancing muscle mechanical efficiency during sustained aerobic exercise. The finding is particularly striking in this young cohort, suggesting that genetic predispositions toward endurance physiology manifest early and may guide natural sporting self-selection even before systematic training begins.

5.4 Population-Specific Allele Frequencies

A critical contribution of this study is the establishment of Pakistani youth baseline allele frequencies. The ACTN3 R allele frequency in controls (0.33) is lower than in European populations (approximately 0.45–0.50) but comparable to other South Asian populations. Similarly, the ACE I allele frequency in controls (0.35) falls within expected South Asian ranges. These data confirm the necessity of population-specific genomic reference databases for sports talent identification programs and caution against direct application of European-derived genetic thresholds to Pakistani populations.

5.5 Limitations

Several limitations warrant acknowledgment. First, the cross-sectional design precludes causal inference; longitudinal tracking would be needed to determine whether genetic profiles predict long-term athletic success. Second, while five polymorphisms were selected based on established evidence, the polygenic nature of athletic performance means that many additional genetic and epigenetic factors were not captured. Third, environmental factors such as nutrition, family income, geographic access to sporting facilities, and cultural influences on sport selection were measured but not fully modeled. Fourth, the voluntary and provincial sampling strategy may introduce selection bias, as elite national-level athletes may differ genetically from those in provincial academies. Finally, the study was limited to boys; a parallel female cohort is planned for future research.

5.6 Policy Implications and Ethical Framework

The results support a cautious yet optimistic case for integrating genetic profiling into Pakistan's youth sports talent identification system. We recommend that genetic profiling be implemented as one component of a comprehensive, multifactorial assessment framework that also includes physical performance testing, psychological assessment, coaching evaluation, and family consultation. Genetic data should be presented to families in the form of probabilistic aptitude guidance rather than deterministic prescription of sport specialization.

A national ethical framework for sports genetic testing in Pakistani youth should be developed, drawing on guidelines from the ACMG, ESHG, and the International Olympic

Committee's consensus on genetic testing in sport. This framework should explicitly prohibit the use of genetic data for exclusion from sports participation, ensure data privacy and security, mandate pre- and post-test genetic counseling by qualified professionals, and establish oversight mechanisms for responsible implementation.

6. Conclusion

This study demonstrates that genetic profiling is a feasible, informative, and practically valuable tool for early sports talent identification in Pakistani boys under 14 years of age. Sport-specific polymorphic signatures were identified across five key performance-related genetic variants, with power sports athletes exhibiting elevated ACTN3 R allele frequencies and endurance sport athletes (swimmers) showing elevated ACE I allele prevalence. The composite genetic aptitude score correlated significantly with standardized physical performance and showed good discriminatory capacity for sport group classification.

Pakistan-specific allele frequency baselines established in this study provide essential reference data for future sports genomics research in South Asian populations. The findings support investment in genomic talent identification infrastructure while underscoring the ethical imperative to use genetic data responsibly, transparently, and in the best interests of young athletes.

Future research should pursue longitudinal designs to evaluate predictive validity, expand genotyping panels to include polygenic scores, incorporate female populations, and develop validated decision support tools integrating genetic, physical, psychological, and environmental data for practical use by coaches and sports scientists across Pakistan.

References

1. Yang, N., MacArthur, D. G., Gulbin, J. P., Hahn, A. G., Beggs, A. H., Eastal, S., & North, K. (2003). ACTN3 genotype is associated with human elite athletic performance. *American Journal of Human Genetics*, 73(3), 627–631. <https://doi.org/10.1086/378292>
2. Moran, C. N., Yang, N., Bailey, M. E. S., Tsiokanos, A., Jamurtas, A., MacArthur, D. G., North, K., Pitsiladis, Y. P., & Wilson, R. H. (2007). Association analysis of the ACTN3 R577X polymorphism and complex quantitative body composition and performance phenotypes in adolescent Greeks. *European Journal of Human Genetics*, 15(1), 88–93. <https://doi.org/10.1038/sj.ejhg.5201724>
3. Papadimitriou, I. D., Lockey, S. J., Voisin, S., Herbert, A. J., Garton, F., Houweling, P. J., Cieszczyk, P., Maciejewska-Składanowska, A., Sawczuk, M., Massidda, M., Calò, C. M., Druzhevskaya, A. M., Astratenkova, I. V., Ahmetov, I. I., North, K. N., & Williams, A. G. (2016). No association between ACTN3 R577X and ACE I/D polymorphisms and endurance running times in 698 Caucasian athletes. *BMC Genomics*, 17, Article 344. <https://doi.org/10.1186/s12864-016-2678-3>
4. Myerson, S., Hemingway, H., Budget, R., Martin, J., Humphries, S., & Montgomery, H. (1999). Human angiotensin I-converting enzyme gene and endurance performance. *Journal of Applied Physiology*, 87(4), 1313–1316. <https://doi.org/10.1152/jappl.1999.87.4.1313>
5. Ahmetov, I. I., & Fedotovskaya, O. N. (2015). Current progress in sports genomics. *Advances in Clinical Chemistry*, 70, 247–314. <https://doi.org/10.1016/bs.acc.2015.03.003>
6. Cieszczyk, P., Eider, J., Ostanek, M., Ficek, K., Maciejewska, A., Sawczuk, M., & Jascaniene, N. (2018). Association of the PPARA gene rs4253778 polymorphism with athlete status. *Journal of Sports Sciences*, 36(6), 633–638. <https://doi.org/10.1080/02640414.2017.1336543>
7. Kamboh, M. I., Manzi, S., Mehdi, H., Fitzgerald, S., Sanghera, D. K., Kuller, L. H., & Ferrell, R. E. (1998). Genetic variation in apolipoprotein H (beta2-glycoprotein I) affects

- the occurrence of antiphospholipid antibodies and apoH concentrations in systemic lupus erythematosus. *Lupus*, 8(9), 742–750. <https://doi.org/10.1191/096120399678840826>
8. Lippi, G., Longo, U. G., & Maffulli, N. (2010). Genetics and sports. *British Medical Bulletin*, 93(1), 27–47. <https://doi.org/10.1093/bmb/ldp007>
 9. Vlahovich, N., Hughes, D. C., Griffiths, L. R., Pitsiladis, Y., Pigozzi, F., Bachl, N., & Eichner, D. (2017). Ethics of genetic testing and research in sport: A position statement from the Australian Institute of Sport. *British Journal of Sports Medicine*, 51(1), 5–9. <https://doi.org/10.1136/bjsports-2016-096661>
 10. Webborn, N., Williams, A., McNamee, M., Bouchard, C., Pitsiladis, Y., Ahmetov, I., Ashley, E., Byrne, N., Camporesi, S., Collins, M., Dijkstra, P., Eynon, N., Fuku, N., Garton, F. C., Hoppe, N., Holm, S., Kaye, J., Klissouras, V., Lucia, A., ... Wang, G. (2015). Direct-to-consumer genetic testing for predicting sports performance and talent identification: Consensus statement. *British Journal of Sports Medicine*, 49(23), 1486–1491. <https://doi.org/10.1136/bjsports-2015-095343>
 11. Weyerstraß, J., Bailey, S., Puta, C., & Gabriel, H. H. W. (2018). Why do we still not have a “genetic test” for sport performance? *Translational Exercise Biomedicine*, 1(1–2), 14–23. <https://doi.org/10.1515/teb-2018-0002>
 12. Hughes, D. C., Ellison, D. C., & Deehan, D. J. (2019). Genetic influences on sporting talent and physical performance in youth: A critical review. *Journal of Pediatric Genetics*, 8(2), 84–91. <https://doi.org/10.1055/s-0039-1678504>
 13. Pakistan Sports Board. (2021). National sports policy 2021–2026. Government of Pakistan, Ministry of Inter-Provincial Coordination.
 14. Bouchard, C., Rankinen, T., & Timmons, J. A. (2011). Genomics and genetics in the biology of adaptation to exercise. *Comprehensive Physiology*, 1(3), 1603–1648. <https://doi.org/10.1002/cphy.c100059>
 15. MacArthur, D. G., & North, K. N. (2007). ACTN3: A genetic influence on muscle function and athletic performance. *Exercise and Sport Sciences Reviews*, 35(1), 30–34. <https://doi.org/10.1097/jes.0b013e31802d8874>