

# **Genomics of Elite Sporting Performance: What little we know and necessary advances**

Yannis Pitsiladis<sup>1</sup> PhD, Guan Wang<sup>1</sup> MRes, Bernd Wolfarth<sup>2</sup> MD, Robert Scott<sup>3</sup> PhD, Noriyuki Fuku<sup>4</sup> PhD, Eri Mikami<sup>4</sup> MRes, Zihong HE<sup>5</sup> PhD, Carmen Fiuza-Luces<sup>6</sup> PhD, Nir Eynon<sup>7</sup> PhD, Alejandro Lucia<sup>6</sup> PhD

<sup>1</sup>College of Medicine, Veterinary and Life Sciences, Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, United Kingdom.

<sup>2</sup>Department of Preventive and Rehabilitative Sports Medicine, Technical University Munich, Munich, Germany.

<sup>3</sup>MRC Epidemiology Unit, Institute of Metabolic Science, Cambridge, United Kingdom

<sup>4</sup>Department of Genomics for Longevity and Health, Tokyo Metropolitan Institute of Gerontology, Tokyo, Japan.

<sup>5</sup>Biology Center, China Institute of Sport Science, Beijing, China.

<sup>6</sup>School of Doctorate Studies and Research, European University of Madrid, Madrid, Spain.

<sup>7</sup>College of Sport and Exercise Science, and Institute of Sport, Exercise and Active Living (ISEAL), Victoria University, Melbourne, Australia.

## **Running title: Genes and elite sporting performance**

Addresses for correspondence:

Yannis Pitsiladis PhD  
College of Medicine, Veterinary and Life Sciences  
Institute of Cardiovascular and Medical Sciences  
University of Glasgow  
Glasgow, G12 8QQ  
Scotland  
Phone: +44-141-330-3858

Fax: +44-141-330-3858

E-mail: [Yannis.Pitsiladis@Glasgow.ac.uk](mailto:Yannis.Pitsiladis@Glasgow.ac.uk)

## **Abstract**

Numerous reports of genetic associations with performance-related phenotypes have been published over the past three decades but there has been limited progress in discovering and characterising the genetic contribution to elite/world-class performance, mainly due to few coordinated research efforts involving major funding initiatives/consortia and the use primarily of the candidate gene analysis approach. It is timely that exercise genomics research has moved into an era utilising well-phenotyped, large cohorts and genome-wide technologies: approaches that have begun to elucidate the genetic basis of other complex traits/diseases. This review summarizes the most recent and significant findings from sports genetics and explores future trends and possibilities.

**Key words: Sports Genetics, Polymorphisms, Candidate Gene Approach, Genome-Wide Association Study**

## Introduction

Despite numerous attempts in recent years to discover genetic variants associated with elite athletic performance and more specifically elite/world-class athletic status, there has been limited progress due to few coordinated research efforts involving major funding initiatives/consortia and the reliance on candidate gene analyses, involving a small number of single nucleotide polymorphisms (SNPs) and structural variants (e.g., the commonly studied insertion/deletion polymorphisms). Nevertheless, over 200 SNPs associated with physical-performance traits, and over 20 SNPs associated with elite athletic status, have been reported in the literature and have been summarized on a yearly basis in the “The human gene map for performance and health-related fitness phenotypes” until 2009 [1]. Due to the massive increase in related papers, the authors changed the format now summarizing only the key findings of each year in “Advances in Exercise, Fitness, and Performance Genomics” [2-5] series. However, most reported associations are reported in studies with small sample sizes, without robust replication, and therefore most likely type 1 errors. It is widely acknowledged that there will be many genes involved in physical performance phenotypes and hence it is timely that genetic research has moved to the genomics era, i.e., the simultaneous testing of multiple genes is now possible. New approaches involving large well-funded consortia and utilising well-phenotyped large cohorts and genome-wide technologies will be necessary for meaningful progress to be made. This review summarizes most recent and significant findings from sports genetics and explores future trends and possibilities.

## Study Designs, Strategies and Methodologies

### *Twin and family studies*

Similar to other areas of research, family or twin studies were initially the main focus investigating the genetic basis of human performance. The first studies in the 1970s assessed indirectly the genetic basis of human performance using twin models and comparing the intra-pair variation between monozygotic (MZ) and dizygotic (DZ) twins; a concept referred to as the heritability estimate ( $h^2$ ), which reflects the population variance in a trait attributable to genetic factors (assuming a simple additive model of heredity) and calculated by dividing the difference of the variance between DZ and MZ twins by the variance of DZ twins. When this approach was applied to maximal oxygen consumption ( $\dot{V}O_2 \text{ max}$ ) using 25 pairs of

twins (15 MZ and 10 DZ preadolescent boys), high heritability estimates were reported (e.g.,  $h^2 = 93.4\%$ ).[6] In other words, genetics could explain as much as 93.4% of the phenotypic variation in  $\dot{V}O_2$  max. Similarly, high heritability estimates were reported in 15 MZ and 16 DZ twins of both genders ( $h^2 = 96.5\%$ )[7] for the variation in skeletal muscle fibre composition. Genetic influences on other performance-related attributes such as body composition and motor activities (e.g., walking, running, throwing, balancing) as well as training-induced improvements in  $\dot{V}O_2$  max were also reported using similar methods.[8-12] A more contemporary view based on results from the HERITAGE Family Study is that genetic factors can explain ~50% of  $\dot{V}O_2$  max when adjusted for age, body mass and body composition as covariates.[13] Notably, in a study of 4488 adult British female twins, the heritability of athlete status was estimated at 66%.[14] Given these exceptionally high heritability estimates, this concept has received considerable criticism with arguments that the high heritability estimates are due to low twin numbers and the near identical social environment of the studied twins.[15,16] Recent studies also report high heritability estimates for neuromuscular performance and body composition.[17-19]

Despite the suggestion of significant heritable components for a range of performance-related traits, family-based studies do not offer insight into the specific genetic variation underlying these heritable components. The limitations and criticisms of the early indirect methods required the focus to be shifted to the continuously developing molecular-based laboratory methods to test directly the interaction between genetic and environmental factors, not only in family or twin studies but also in populations of interest for complex traits; for example, elite athletic status, where inter-individual phenotypic variations are caused by a heterogeneous, polygenic model with multiple gene variants involved and interacting with environmental factors.[1,20,21]

#### *Genome-Wide Association studies: hypothesis-free approach*

A number of different methodological approaches within the field of genetic epidemiology have been utilized to unravel the genetic basis of elite human performance. Due to the development of more advanced gene discovery techniques, genetic studies are no longer restricted to family/twin studies but expanded to include the assessment of genetic variants (i.e., mostly SNPs) within a population of interest. Population-based case-control studies are extensively being used at present, and can be further differentiated into hypothesis-free

(sometimes referred to as a “fishing trip”, i.e., no assumptions made about the genomic location of associated variants) and the more commonly used hypothesis-driven approaches, where the search for association may be restricted to particular genes of interest.

Advances in molecular technologies enabled researchers to apply genome-wide association (GWAS) approaches to the field. GWAS examine the ‘association of genetic variation with outcomes or phenotypes of interest by analysing 100,000 to several millions of SNPs across the entire genome without any previous hypotheses about potential mechanisms’. GWAS has been successful in identifying novel genetic variants for age-related macular degeneration,[22] Type 2 Diabetes Mellitus (T2DM),[23] the interleukin 23 (IL-23) pathway in Crohn's disease,[24] and obesity-related traits.[25] This promising approach is not without important limitations. For example, human height is a highly heritable quantitative trait (up to 90% of population variance)[26-30] as well as stable and easy to measure, although one of the largest studies to date (n = 183,727) identified at least 180 loci associated with adult height, but together these explained only 10% of the variation in height. This is in large part due to the small effect size of most of these genetic variants. Furthermore, genetic variants associated with most complex diseases do not show predictive utility.[31] Thus, much of the heritability of complex traits remains missing,[32] and numerous explanations have been proposed to account for this missing heritability. There have been suggestions that common variants do explain up to 45% of the variance in height,[33] but the small effect size of these variants may render these variants undetectable by common study designs.[32] Despite several limitations, these studies confirm that GWAS is able to detect many loci that implicate biologically related genes and pathways.[34] The occurrence of rare variants which are not captured by GWAS may partly explain this limited success in determining the genomics of adult height. Following GWAS, additional approaches, such as fine mapping and sequencing, may be used to find common SNPs with larger effect sizes than GWAS tagging SNPs or identify rarer variants across GWAS loci. The hypothesis-free GWAS design is the most popular of the current widespread approaches as it allows to 1) detect smaller gene effects by narrowing down the genomic target region precisely with new chips; 2) maximize the amount of variation captured per SNP with the fixed set of markers; and 3) reduce genotyping costs, which make this approach attractive.[35] Other factors relating to the sample population (family history of traits, ethnically homogeneous populations, sample size of at least several thousands) and differences in statistical approaches (e.g., the conservative Bonferroni correction for multiple testing vs. the non-conservative false discovery rate correction or the

application of permutation testing approaches) need to be carefully considered to ensure successful application of the GWAS approach.[36] GWAS of elite human athletic performance are ongoing,[37-40] but no published papers to date.

#### *Candidate gene analysis: hypothesis-driven approach*

The most extensively used candidate gene association study approach requires a prior hypothesis that particular genes of interest contain variants that may be associated with a trait or disease. Typically, variants in a gene or genes of interest are genotyped in cases and controls, or to investigate these associations with a quantitative trait. This approach is effective in detecting genetic variants with small or modest influence on common disease or complex traits. Functional SNPs with tag SNPs (by use of linkage disequilibrium) which would cover the entire candidate gene have been used in many candidate gene association studies.[35] However, in this approach, candidate genes ought to be selected if there is good evidence that 1) the proposed candidate gene is biologically relevant to main phenotype/complex trait of interest (e.g., physical performance/aerobic capacity, adiposity); 2) the variants of the candidate gene influences the overall function of the gene (e.g., variation in physiological angiotensin converting enzyme (ACE) activity levels are linked to polymorphisms in the *ACE* gene -see section on ‘Genes and polymorphisms with reasonable replication’); and 3) the polymorphisms of the selected candidate gene are frequent enough in the population to allow meaningful statistical analysis (e.g., typical allele frequencies for the I and D allele of the *ACE* I/D polymorphism in a European population are ~43% and 57% respectively). When these criteria are not fulfilled and candidate genes are selected based primarily on the interest of the research group, this approach generates conflicting results with low statistical power and difficulty to be replicated in other populations, and thus low validity.[41] In addition, the candidate gene approach has largely been implemented in studies with small subject numbers and often without robust replication, and perhaps partly attributable to publication bias.

#### *Major study cohorts*

The genotyping of athletes of the highest performance caliber such as world record holders, world champions and Olympians is desirable and may circumvent the need for very large athlete cohorts in order to discover performance-associated polymorphisms. The number of large genetic cohorts of world-class athletes from a variety of countries and sports with

extensive physical performance phenotypes is limited. The following are the most significant elite athlete cohorts based on current publication outcomes:

*Genathlete Study:* In the Genathlete study, a classical case-control study, endurance athletes with high  $\dot{V}O_2$  max were compared to control participants with a low to average  $\dot{V}O_2$  max.[42-44] Data were analysed using the candidate gene approach, which allowed the distribution of particular genetic variants with respect to the phenotype  $\dot{V}O_2$  max in both groups to be assessed. As this type of assessment requires large subject numbers, this study was designed as a multi-centre study. To exclude influences due to regionally different distribution of genetic variants, particular attention was paid to a comparable distribution of the regional origin of participants. Currently, this cohort involves more than 600 participants (~300 athletes and ~300 controls) and therefore constitutes one of the largest matched case-control studies in this field.[43,44]

*Elite Russian athlete cohort:* One of the largest studies of elite athletes involves elite Russian athletes from mixed athletic disciplines.[45-47] In the most recent study, 998 male and 425 female Russian athletes of regional or national competitive standard were recruited from 24 different sports.[47] Athletes were stratified into 5 groups according to event duration (very long-, long- and middle-endurance), mixed ‘anaerobic/aerobic’ activity group and power group (predominantly anaerobic energy production).

*Elite east African athlete cohorts:* The phenomenal success of athletes from Ethiopia and Kenya in endurance running events is well recognized. Middle- and long-distance runners from Ethiopia and Kenya hold over 90% of both all-time world-records and current top-10 positions in world event rankings.[48] Moreover, these successful athletes come from localized ethnic subgroups within their respective countries.[49,50] In order to investigate the east African running phenomenon, a first study[50-52] involved 76 endurance runners from the Ethiopian junior- and senior-level national athletics teams (12 female, 64 male), 315 controls from the general Ethiopian population (34 female, 281 male), 93 controls from the *Arsi* region of Ethiopia (13 female, 80 male), and 38 sprint and power event athletes from the Ethiopian national athletics team (20 female, 18 male). A similar approach was conducted in a study[53] with 291 elite Kenyan endurance athletes (232 male) and 85 control subjects (40



male). 70 of the athletes (59 male) had competed internationally representing Kenya and achieved remarkable success.

*Elite Jamaican and USA sprint cohorts:* These cohorts are comprised of elite Jamaican and African-American athletes representing the highest level of sprinting performance and geographically matched controls. In the Jamaican cohort, 116 athletes (male = 60, female = 56) and 311 control subjects (throughout the whole island; male = 156, female = 155) were recruited.[54] 71 and 35 athletes had participated in 100-200 m and 400 m sprint events, respectively; and 10 athletes were involved in the jump and throw events. These athletes can be further classified into national (n = 28) and international athletes (n = 88) who were competitive at the national level in Jamaica and the Caribbean or at major international competitions for Jamaica. Among the 88 international athletes, 46 had won medals at major international events or held world records in sprinting. In the African-American cohort, samples from 114 elite sprint athletes (male = 62, female = 52) and 191 controls (throughout the United States; male = 72, female = 119) were collected.[54] Among these athletes, 48, 42 and 24 athletes participated 100-200 m, 400 m, and jump and throw events, respectively. Athletes can be subdivided into 28 national and 86 international athletes; 35 of these athletes had won medals at international games and/or broken sprint world records.

*Elite Australian athlete cohort:* Australia has provided valuable genetic information on elite sprinters and endurance performers. The cohort comprises of 429 elite athletes from 14 different sports and 436 unrelated controls. A subgroup of 107 and 194 subjects were classified as elite sprinters and endurance runners respectively.[55] This cohort was studied to postulate, for the first time, the *ACTN3* gene as a strong candidate to influence elite athletic performance (to be discussed in ‘Genes and polymorphisms with reasonable replication’ section).

*Elite Japanese athlete cohort:* Japanese athletes are successful in international competitions such as Olympics, especially in endurance-oriented events such as Marathon and swimming events. This cohort is comprised of 717 elite Japanese athletes and 814 controls. Athletes are either national (participants in national competitions) or international athletes (participants in Olympic Games, World and Asian Championships), including several medallists at these international games and world record holders. This Japanese athlete cohort comprises 381 track & field athletes, 166 swimmers, and 170 Olympians from various sports. This cohort

was initially established in order to identify both nuclear and mitochondrial DNA polymorphisms/haplogroups associated with elite Japanese athlete status and performance-related traits.[56-61]

*Elite European and Asian swim cohort:* Two elite swim cohorts, comprising Caucasian and East Asian swimmers, respectively have been established. The Caucasian cohort comprised of 200 elite Caucasian swimmers from European, Commonwealth, American and Russian sub-cohorts. Swimmers were categorized as short and middle distance ( $\leq 400$  m,  $n = 130$ ) or long distance swimmers ( $> 400$  m,  $n = 70$ ). Caucasian swimmers were all highly competitive and of world-class status having represented their countries in international competitions. Caucasian controls were drawn from a previous published report.[62] Elite Japanese ( $n = 158$ ) and Taiwanese ( $n = 168$ ) swimmers were recruited and classified as short distance ( $< 200$  m,  $n = 166$ ) and middle distance (200 – 400 m,  $n = 160$ ), and none of these Asian swimmers competed at a distance greater than 400 m. East Asian swimmers were world-class having also competed in international competitions such as the Olympics, World Championships and Asian Games, or were competitive in national competitions. Controls were pooled from general Japanese ( $n = 649$ ) and Taiwanese ( $n = 603$ ) populations, respectively and were healthy adults of both sexes and not professionally connected with athletics/sport. Two candidate genes (Angiotensin Converting Enzyme &  $\alpha$ -Actinin-3; see next section) have been studied to date in this cohort.[63]

*Spanish cohort:* Extensive researches have been conducted in Spanish male athletes, with the most representative cohort comprising endurance world-class athletes ( $n = 100$ , including 50 Olympic-class endurance runners and 50 professional cyclists (most of whom are *Tour de France* finishers, including stage winners)[64,65]) and world-class rowers ( $n = 54$ , lightweight category, most of whom are medallists in world championships).[66] This cohort also includes the majority of all-time best Spanish judo male athletes ( $n = 108$ ),[67] elite swimmers ( $n = 88$ )[68] and track & field elite power athletes ( $n = 53$ ).[69]

*Elite Israeli cohort:* This cohort is comprised of 74 endurance-, and 81 sprint-/power- male athletes, who are current and former track and field athletes, as well as 240 matched controls. Athletes were carefully selected and included only if their main events were the 10,000 m run or the marathon (endurance group); and only if their main events were the 100-200 m dash and long-jump (sprint/power group). According to their personal best, athletes were further

divided into 2 subgroups: the elite-level (those who had represented Israel in track and field world championships or in the Olympic Games), and the national-level.

*Chinese cohort:* He et al. have recently conducted research in a Chinese cohort (of *Han* origin) made up of the best endurance runners (from 5,000m to marathon) of this country (current total n=241, 118 male).[70,71] An important novelty of research on this Chinese cohort is results' replication in a different (Caucasian) group of athletes and especially analysis of the functionality of the SNPs been found to be associated with elite endurance status (using dual-luciferase reporter assay). [70]

### **Genes and polymorphisms with reasonable replication**

*Angiotensin Converting Enzyme (ACE) and the Renin-Angiotensin-Aldosterone-System (RAAS).* One of the most widely-studied candidate genes for athletic performance is the Angiotensin Converting Enzyme (*ACE*) gene. *ACE* is a peptidase known to regulate blood pressure by catalyzing the conversion of Angiotensin I to the vasoconstrictor Angiotensin II and also degrading the vasodilator bradykinin.[72] Inter-individual variation in physiological *ACE* activity levels has been linked to polymorphisms in the *ACE* gene. Notably, *ACE* Insertion-Deletion (*I/D*) (rs4340), in which 'I' refers to the presence (insertion) and 'D' to the absence (deletion) of a 287 bp sequence in an *Alu* sequence of intron 16 in the *ACE* gene at chromosome location 17q23, can account for up to 47% of *ACE* activity variance in subjects (i.e., Caucasians and Asians, not Africans) with an additive effect across the II, ID, and DD genotypes.[72] Regarding physical performance, *ACE* *I/D* genotypes have been associated with a wide range of phenotypes. *ACE* II subjects show significantly higher muscle efficiency gains from training than DD individuals,[73] as well as greater improvements in running economy, or ability to sustain sub-maximal pace with lower oxygen consumption.[74] Additionally, the I-allele associates with muscular endurance gains from training,[75] which may relate to higher type 1 (slow-twitch) muscle fibre preponderance in *ACE* II subjects.[76] In terms of overall sporting ability, the I-allele has been associated with superior performance in British mountaineers,[75] South-African triathletes,[77] British distance runners,[78] and Australian rowers.[79] In contrast, the D-allele has been associated with success in power-oriented sports such as short-distance swimming[63,80] and sprinting.[78] Notably, a recent study[63] reported that *ACE* D-allele was associated with short and middle distance swimmer status in Caucasian swimmers, whereas *ACE* I-allele was found to be overrepresented in East

Asian short distance swimmers. Although this finding might be explained by different risk alleles being responsible for the associations in swimmers of different ethnicities, it requires to be further confirmed in future studies. Other inconsistencies in the literature regarding *ACE* findings also exist. For instance, the D-allele has been found to be both positively[81] and negatively[82] associated with  $\dot{V}O_2$  max. A study of 230 elite Jamaican and American sprinters found no association of either allele with sprint athlete status.[54] A cohort of 192 athletes of mixed Caucasian nationalities and endurance sporting disciplines did not exhibit I-allele frequencies that were significantly different from geographically-matched controls, nor did I-allele frequency associate with  $\dot{V}O_2$  max in these athletes.[43] Several other studies involving Caucasian populations also found no association between the I-allele and elite physical performance,[45,62,83] and an opposite finding (i.e. D-allele associated with endurance performance) in Israeli[84] and Korean[85] cohorts. Despite many inconsistencies in replication, the *ACE* gene remains to be a candidate possibly influencing elite performance.

*$\alpha$ -Actinin-3 (ACTN3)*. *Alpha-actinin-3* is an actin-binding protein and a key component of the sarcomeric Z-line in skeletal muscle. Expression of *ACTN3* (at 11q13.1) is limited to type II (i.e. fast, mostly glycolytic) muscle fibres which can generate more force at high velocity. Homozygosity for the common nonsense polymorphism R577X (rs1815739) in the  *$\alpha$ -Actinin-3 (ACTN3)* gene results in deficiency of actinin-3 in a large proportion of the global population.[86] Yang et al.[87] examined R577X genotype frequencies in three African populations (Kenya, Nigeria and Ethiopia) in comparison with non-African populations (Europe, Asia, Australia). Extremely low R577XX genotype frequencies were observed in Kenyan and Nigerian athletes versus controls (1% vs. 1% and 0% vs. 0% respectively) as well as much lower than in any other non-African populations, i.e., frequency of the R577XX genotype of: 18% in Australian Caucasians, 10% in Aboriginal Australians, 18% in Spanish Caucasians, and 25% in Japanese). These results also implied that the *ACTN3* deficiency was not a major influence on performance in African athletes. This polymorphism does not appear to result in pathology, although it could alter muscle function.[88-92] Furthermore, a strong association has been reported between the *ACTN3* R577X polymorphism and elite athletic performance in Caucasian populations.[55,93-98]

The 577XX genotype was found at a lower frequency in elite Australian sprint/power athletes relative to controls,[55] and this finding replicated in Finnish,[96] Greek,[97] Israelis,[99] and Russian athletes.[94] Particularly, in a study of 429 elite white athletes from 14 different sporting disciplines and 436 controls, the sprint athlete group showed a higher frequency of the 577RR genotype (50%) and a lower frequency of the 577RX genotype (45%), compared with controls (30% and 52%, respectively), while the elite endurance athletes displayed a higher frequency of the 577XX genotype (24%) than controls (18%);[55] however, the sample sizes in the truly elite subgroups are very small therefore, any conclusions drawn from them are prone to high risk of type I error and should be treated with caution. Interestingly, MacArthur et al.[100] developed an exciting *Actn3* knockout (KO) mouse model in order to investigate the mechanisms underlying *ACTN3* deficiency. These authors found the KO mice had similar muscle fibre proportions as the wild type but reduced muscle mass that appeared to be accounted for by the reduced fibre diameter of the fast-twitch muscle observed in KO mice.[100] In addition to alterations in muscle fibre size, increased activity of muscle aerobic enzymes, longer muscle contracting time and shorter recovery period from fatigue were attributed to the characteristics of the *Actn3* KO mouse. Thus, the phenotypes of the *Actn3* KO mouse mimic the gene association studies performed in humans and provide a plausible explanation for the reduced sprint/power capacity and improved endurance performance in humans with the *ACTN3* 577XX genotype.

### **Implications of current genetic findings in association with elite athlete status**

World-class athletic performance is a complex multi-factorial phenotype, and it is acknowledged that to become an elite athlete, a synergy of physiological, behavioural and other environmental factors is required. It is commonly perceived that genetic endowment is one of the arbiters of elite athletic performance; a belief perhaps augmented by the striking geographical variation in athletic success.[49,50,101] As reviewed above, a number of genes have been found to associate with elite performance. These studies have employed primarily the candidate gene approach to identify those genes which are associated with elite performance, or associated with variation in performance-related traits. As such, a number of genes have been found to associate with elite performance, although generally with small effect sizes and heavily prone to type I statistic error. The number of candidate genetic variants that can potentially explain elite athletic status will be much higher than that examined by numerous biotechnology companies such as Sports X factor (e.g., 7 genes:

www.sportsxfactor.com). While genetic testing will likely become a part of talent ID programmes in the future, current genetic testing is of **zero** predictive capacity. Whether new approaches such as GWAS will significantly improve prediction outcomes for athletes is unknown.

## **The Future**

Most of the knowledge in sports genetics (including most of the information presented in this review) has been generated primarily using classical/“old fashioned” genetic methods such as candidate gene analysis and almost exclusively applied to cohorts with small sample sizes (usually  $n \leq 300$ ) and rather unsophisticated multifactorial phenotypes (e.g., aerobic capacity, athlete status). The data generated therefore from these studies and reviewed here need to be examined in light of the view held by most “hard core” geneticists that a study of any complex phenotype in humans is futile unless a cohort size of between 20,000-100,000 is used and therefore possessing sufficient statistical power for meaningful analysis and interpretation. If one accepts this view (not currently held by the authors of this review) then all studies reviewed here should be ignored. While somewhat extreme, an intermediate view (currently held by the authors of this review) is that, perhaps beside the *ACTN3* R577X and possibly *ACE* I/D, the vast majority of the candidate genes for sporting performance discovered to date, are not the key candidates seriously implicated in the phenotypes of interest. Priority should therefore be given to recruiting sufficiently large study cohorts with adequately measured phenotypes to increase statistical power. Some of the elite athlete cohorts described in this review may suffice and collectively, these cohorts could be used for replication purposes.

As stated previously, it is accepted that there will be many genes involved in sporting performance and hence it is timely that genetic research has moved to the genomics era. New approaches and technologies will no doubt be increasingly applied to searching the whole human genome instead of studying single genes or indeed SNPs as the cost of using such whole-genome methods becomes more affordable. Particularly, the cost of large-scale sequencing has dramatically dropped, from the first complete human genome costing \$3 billion to sequence in 2000 to \$1,000 per genome as promised by the company Ion Torrent (a division of Life Technologies) using its new Ion Proton sequencer in 2012.[\[102\]](#) At present, no matter the success or failure of the GWAS approach, this approach is certainly providing,

and will continue to provide, the insight into genetic architecture and the molecular basis underlying human diseases and complex traits. The first genome-wide association study in age-related macular degeneration (AMD) revealed an intronic and common variant significantly related to AMD by comparing 96 cases to 50 controls and consequently a functional polymorphism in Complement Factor H (*CFH*) gene was identified by resequencing,[22] suggesting that it is not unreasonable to expect and detect variants with large effects in a small study; however, large cohorts (i.e., ranging from several thousands to 20,000-100,000) will be routinely studied by GWAS and will provide good resources for all scientific fields including genomics of the world-class athlete.[103] This development will require a move away from the traditional way of researching in exercise science/medicine (i.e., predominantly single-laboratory studies) to large well-funded collaborations/consortia with leading industry partners and therefore substantial statistical/technological power and knowhow. Only with such resources can the most strongly-acting genes be identified with confidence, enabling gene x gene interactions being revealed and gene x environment interactions being studied more accurately. One first example in the area of sports performance has recently been successfully piloted.[104] The overall purpose was to identify new SNPs that confer susceptibility to sprint and endurance performance by use of world-class athletes as subjects. GWAS was initially performed using Illumina<sup>®</sup> HumanOmni-Quad BeadChip (> 1,000,000 SNPs/sample) or HumanOmniExpress BeadChip (> 700,000 SNPs/sample) in 95 sprinters and 102 controls from Jamaica. After removing individuals and markers failing quality control, remaining SNPs were taken forward for association analysis. Genotype frequencies were compared for 88 Jamaican sprinters and 87 Jamaican controls using logistic regression (corrected for population stratification), assuming an additive model. Seventeen SNPs crossed a predetermined significance threshold of  $5 \times 10^{-5}$ .[40] Further validation of these signals in independent cohorts is underway, and the replicated SNPs will be taken forward for fine-mapping and functional studies to uncover the underlying biological mechanisms. Further analyses using other cohorts such as the ones described in this review will also provide opportunity for verifying these GWAS findings across different ethnicities.

## References

1. Bray MS, Hagberg JM, PÉRusse L, et al. The Human Gene Map for Performance and Health-Related Fitness Phenotypes: the 2006-2007 update. *Medicine and science in sports and exercise*. 2009;41:35-73.
2. Roth SM, Rankinen T, Hagberg JM, et al. Advances in exercise, fitness, and performance genomics in 2011. *Medicine and science in sports and exercise*. May 2012;44(5):809-17.
3. Rankinen T, Roth SM, Bray MS, et al. Advances in exercise, fitness, and performance genomics. *Medicine and science in sports and exercise*. 2010;42:835-46.
4. Hagberg JM, Rankinen T, Loos RJ, et al. Advances in exercise, fitness, and performance genomics in 2010. *Medicine and science in sports and exercise*. 2011;43:743-52.
5. Pérusse L, Rankinen T, Hagberg JM, et al. Advances in Exercise, Fitness, and Performance Genomics in 2012. *Medicine and science in sports and exercise*. 2013 Epub ahead of print.
6. Klissouras V. Heritability of adaptive variation. *J Appl Physiol*. 1971;31:338-44.
7. Komi P, Viitasalo J, Havu M, et al. Skeletal muscle fibres and muscle enzyme activities in monozygous and dizygous twins of both sexes. *Acta Physiol Scand*. 1977;100:385 -92.
8. Weber G, Kartodihardjo W, Klissouras V. Growth and physical training with reference to heredity. *J Appl Physiol*. 1976;40:211 -5.
9. Kovar R. Somatotypes of twins. *Acta Univ Carol Gymn*. 1977;13:49 -59.
10. Sklad M. Skeletal maturation in monozygotic and dizygotic twins. *J Hum Evol*. 1977;6:145-9.



11. Williams LRT, Gross JB. Heritability of motor skill. *Acta Genet Med.* 1980;29:127 - 36.
12. Orvanova E. Body build, heredity and sport achievements. In Genetics of Psychomotor Traits in Man. *International Society of Sport Genetics and Somatology, Warsaw.* 1984:111-23.
13. Bouchard C, An P, Rice T, et al. Familial aggregation of VO<sub>2</sub> max response to exercise training: results from the HERITAGE Family Study. *J Appl Physiol.* 1999;87:1003-8.
14. De Moor MH, Spector TD, Cherkas LF, et al. Genome-wide linkage scan for athlete status in 700 British female DZ twin pairs. *Twin Res Hum Genet.* 2007;10(6):812-20.
15. Bouchard C, Lesage R, Lortie G, et al. Aerobic performance in brothers, dizygotic and monozygotic twins *Med Sci Sports Exerc.* 1986;18:639-46.
16. Fagard R, Bielen E, Amery A. Heritability of aerobic power and anaerobic energy generation during exercise. *J Appl Physiol.* 1991;70:357-62.
17. Missitzi J, Geladas N, Klissouras V. Heritability in neuromuscular coordination: implications for motor control strategies. *Med Sci Sports Exerc.* 2004;36:233-40.
18. Peeters M, Thomis M, Loos R, et al. Heritability of somatotype components: a multivariate analysis. *Int J Obes (Lond).* 2007;31:1295-301.
19. Missitzi J, Geladas N, Klissouras V. Genetic variation of maximal velocity and EMG activity. *Int J Sports Med.* 2008;29:177-81.
20. Bouchard C, Simoneau J, Lortie G, et al. Genetic effects in human skeletal muscle fiber type distribution and enzyme activities. *Can J Physiol Pharmacol.* 1986;64:1245-51.
21. Ahmetov I, Rogozkin V. Genes, athlete status and training -- An overview. *Med Sport Sci.* 2009;54:43-71.

22. Klein RJ, Zeiss C, Chew EY, et al. Complement factor H polymorphism in age-related macular degeneration. *Science*. 2005;308(5720):385-9.
23. McCarthy MI, Zeggini E. Genome-wide association studies in type 2 diabetes. *Curr Diab Rep*. 2009;9(2):164-71.
24. Massey D, Parkes M. Genome-wide association scanning highlights two autophagy genes, ATG16L1 and IRGM, as being significantly associated with Crohn's disease. *Autophagy*. 2007;3:649-51.
25. Frayling TM, Timpson NJ, Weedon MN, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science*. May 11 2007;316(5826):889-94.
26. Preece MA. The genetic contribution to stature. *Horm Res*. 1996;45 Suppl2:56-8.
27. Silventoinen K, Kaprio J, Lahelma E, et al. Relative effect of genetic and environmental factors on body height: differences across birth cohorts among Finnish men and women. *Am J Public Health*. 2000;90:627-30.
28. Silventoinen K, Sammalisto S, Perola M, et al. Heritability of adult body height: A comparative study of twin cohorts in eight countries. *Twin Research*. 2003;6:399-408.
29. Macgregor S, Cornes B, Martin N, et al. Bias, precision and heritability of self-reported and clinically measured height in Australian twins. *Human Genetics*. 2006;120:571-80.
30. Perola M, Sammalisto S, Hiekkalinna T, et al. Combined genome scans for body stature in 6,602 European twins: Evidence for common Caucasian loci. *PLoS Genetics*. 2007;3:e97.
31. Talmud PJ, Hingorani AD, Cooper JA, et al. Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study. *BMJ*. 2010;340:b4838.

32. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461(7265):747-53.
33. Yang J, Benyamin B, McEvoy BP, et al. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet*. 2010;42(7):565-9.
34. Allen HL, Estrada K, Guillaume L, et al Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*. 2010;467:832-8.
35. Yang W, Kelly T, He J. Genetic epidemiology of obesity. *Epidemiol Rev*. 2007;29:49-61.
36. McCarthy MI, Abecasis GR, Cardon LR, et al Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet*. 2008;9(5):356-69.
37. Wang G, Fuku N, Padmanabhan P, et al. Genome-Wide Association Study (GWAS) of elite sprinters of West African Ancestry [abstract]. In: XXXII World Congress of Sports Medicine; 27 - 30 Sep 2012; Rome, Italy.
38. Wang G, Fuku N, Padmanabhan P, et al. Genome-wide association study approach in elite Jamaican athletes to identify common genetic variations associated with world class athlete status [abstract]. In: Glasgow Polyomics Launch Symposium; 14 Sep 2012; Glasgow, UK.
39. Wang G, Fuku N, Padmanabhan P, et al. A genome wide association study of elite Jamaican athletes identifies multiple putative single nucleotide polymorphisms associated with world class athlete status [abstract]. In: The West of Scotland Health and Ethnicity Network conference; 20 Aug 2012; Glasgow, UK.
40. Wang G, Fuku N, Padmanabhan P, et al. Genome-wide association study of world class athlete status using elite Jamaican sprinters [abstract]. In: the Malawi-Liverpool-Wellcome Trust Clinical Research Programme; 21 Mar 2013; Glasgow, UK.

41. Cowley AWJr. The genetic dissection of essential hypertension. *Genetics*. 2006;7:829-40.
42. Rivera MA, Dionne FT, Wolfarth B, et al. Muscle-specific creatine kinase gene polymorphisms in elite endurance athletes and sedentary controls. *Medicine and science in sports and exercise*. 1997;29:1444-7.
43. Rankinen T, Wolfarth B, Simoneau JA, et al. No association between the angiotensin-converting enzyme ID polymorphism and elite endurance athlete status. *J Appl Physiol*. 2000;88:1571-5.
44. Wolfarth B, Rankinen T, Mühlbauer S, et al. Endothelial nitric oxide synthase gene polymorphism and elite endurance athlete status: the Genathlete study. *Scand J Med Sci Sports*. 2008;18:485-90.
45. Nazarov IB, Woods DR, Montgomery HE, et al. The angiotensin converting enzyme I/D polymorphism in Russian athletes. *Eur J Hum Genet*. 2001;9:797-801.
46. Ahmetov II, Mozhayskaya IA, Flavell DM, et al. PPAR $\alpha$  gene variation and physical performance in Russian athletes. *Eur J Appl Physiol*. 2006;97:103-8.
47. Ahmetov II, Williams AG, Popov DV, et al. The combined impact of metabolic gene polymorphisms on elite endurance athlete status and related phenotypes. *Hum Genet*. 2009;126:751-61.
48. Ash GI, Scott RA, Deason M, et al. No association between ACE gene variation and endurance athlete status in Ethiopians. In Press. *Medicine and science in sports and exercise*. 2010.
49. Onywera V, Scott R, Boit M, et al. Demographic characteristics of elite Kenyan endurance runners. *J Sports Sci*. 2006;24:415-22.

50. Scott RA, Georgiades E, Wilson RH, et al. Demographic Characteristics of Elite Ethiopian Endurance Runners. *Medicine and science in sports and exercise*. 2003;35:1727-32.
51. Moran CN, Scott RA, Adams SM. Y Chromosome Haplogroups of Elite Ethiopian Endurance Runners. *Hum Genet*. 2004;115:492-7.
52. Scott RA, Wilson RH, Goodwin WH, et al. Mitochondrial DNA Lineages of Elite Ethiopian Athletes. *Comp Biochem Physiol B Biochem Mol Biol*. 2005;140:497-503.
53. Scott RA, Moran CN, Wilson RH, et al. No Association Between Angiotensin Converting Enzyme (ACE) Gene Variation and Endurance Athlete Status in Kenyans. *Comp Biochem Physiol A Mol Integr Physiol*. 2005;141:169-75.
54. Scott RA, Irving R, Irwin L, et al. ACTN3 and ACE genotypes in elite Jamaican and US sprinters. *Medicine and science in sports and exercise*. 2010;42(1):107-12.
55. Yang N, MacArthur DG, Gulbin JP, et al. ACTN3 genotype is associated with human elite athletic performance. *Am J Hum Genet*. Sep 2003;73(3):627-31.
56. Fuku N, Mikami E, Tanaka M. Association of mitochondrial DNA polymorphisms and/or haplogroups with elite Japanese athlete status. *The Journal of Physical Fitness and Sports Medicine*. 2013;2(1):17-29.
57. Fuku N, Mori S, Murakami H, et al. Association of 29C>T polymorphism in the transforming growth factor-beta1 gene with lean body mass in community-dwelling Japanese population. *Geriatr Gerontol Int*. Apr 2012;12(2):292-7.
58. Fuku N, Murakami H, Iemitsu M, et al. Mitochondrial macrohaplogroup associated with muscle power in healthy adults. *Int J Sports Med*. May 2012;33(5):410-4.
59. Mikami E, Fuku N, Takahashi H, et al. Mitochondrial haplogroups associated with elite Japanese athlete status. *Br J Sports Med*. 2011(45):1179-83.

60. Mikami E, Fuku N, Takahashi H, et al. Polymorphisms in the control region of mitochondrial DNA associated with elite Japanese athlete status. *Scand J Med Sci Sports*. [in press].
61. Saito D, Fuku N, Mikami E, et al. The ACTN3 R577X nonsense allele is under-represented in elite-level Japanese endurance runners. *Jpn J Phys Fitness Sports Med*. 2011;60(4):443-51.
62. Woods D, Hickman M, Jamshid Y, et al. Elite swimmers and the D allele of the ACE I/D polymorphism. *Hum Genet*. 2001;108:230-2.
63. Wang G, Mikami E, Chiu L, et al. Association analysis of ACE and ACTN3 in Elite Caucasian and East Asian Swimmers. *Medicine and science in sports and exercise*. [in press].
64. 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*. Aug 2011;21(4):570-9.
65. Gomez-Gallego F, Ruiz JR, Buxens A, et al. Are elite endurance athletes genetically predisposed to lower disease risk? *Physiol Genomics*. Mar 3 2010;41(1):82-90.
66. Santiago C, Ruiz JR, Muniesa CA, et al. Does the polygenic profile determine the potential for becoming a world-class athlete? Insights from the sport of rowing. *Scand J Med Sci Sports*. Feb 2010;20(1):e188-94.
67. Rodriguez-Romo G, Yvert T, de Diego A, et al. ACTN3 R577X Polymorphism is not Associated with Elite Judo Athletic Status. *International journal of sports physiology and performance*. Jan 23 2013.
68. Ruiz JR, Santiago C, Yvert T, et al. ACTN3 genotype in Spanish elite swimmers: No "heterozygous advantage". *Scand J Med Sci Sports*. Jan 14 2013.

69. Ruiz JR, Arteta D, Buxens A, et al. Can we identify a power-oriented polygenic profile? *J Appl Physiol*. Mar 2010;108(3):561-6.
70. He ZH, Hu Y, Li YC, et al. Are calcineurin genes associated with athletic status? A function, replication study. *Medicine and science in sports and exercise*. Aug 2011;43(8):1433-40.
71. Yvert T, He ZH, Santiago C, et al. Acyl coenzyme A synthetase long-chain 1 (ACSL1) gene polymorphism (rs6552828) and elite endurance athletic status: a replication study. *PloS one*. 2012;7(7):e41268.
72. Rigat B, Hubert C, Alhenc-Gelas F, et al. An insertion/deletion polymorphism in the angiotensin I-converting enzyme gene accounting for half the variance of serum enzyme levels. *J Clin Invest*. Oct 1990;86(4):1343-6.
73. Williams AG, Rayson MP, Jubb M, et al. The ACE gene and muscle performance. *Nature*. 2000;403:614.
74. Woods DR, World M, Rayson MP, et al. Endurance enhancement related to the human angiotensin I-converting enzyme I-D polymorphism is not due to differences in the cardiorespiratory response to training. *Eur J Appl Physiol*. 2002;86:240-4.
75. Montgomery HE, Marshall R, Hemingway H, et al. Human gene for physical performance. *Nature*. 1998;393:221-2.
76. Zhang B, Tanka H, Shono N, et al. The I allele of the angiotensin-converting enzyme gene is associated with an increased percentage of slow-twitch type I fibers in human skeletal muscle. *Clin Genet*. 2003;63:139-44.
77. Collins M, Xenophontos SL, Cariolou MA, et al. The ACE gene and endurance performance during the South African Ironman Triathlons *Medicine and science in sports and exercise*. 2004;36:1314-20.

78. Myerson S, Hemingway H, Budget R, et al. Human angiotensin I-converting enzyme gene and endurance performance. *J Appl Physiol.* 1999;87:1313-6.
79. Gayagay G, Yu B, Hambly B. Elite endurance athletes and the ACE I allele- the role of genes in athletic performance. . *Hum Genet.* 1998;103:48-50.
80. Woods DR, Hickman M, Jamshidi Y, et al. Elite swimmers and D allele of the ACE I/D polymorphism. *Hum Genet.* 2001;108(3):230-2.
81. Zhao B, Moochhala SM, Tham S, et al. Relationship between angiotensin-converting enzyme ID polymorphism and VO<sub>2</sub>max of Chinese males. *Life Sci.* 2003;73:2625-30.
82. Hagberg JM, Ferrell RE, McCole SD, et al. VO<sub>2</sub> max is associated with ACE genotype in postmenopausal women. *J Appl Physiol.* 1998;85:1842-6.
83. Kalson NS, Thompson J, Davies AJ, et al. The effect of angiotensin-converting enzyme genotype on acute mountain sickness and summit success in trekkers attempting the summit of Mt. Kilimanjaro (5,895 m). *Eur J Appl Physiol.* 2009;105:373-9.
84. Amir O, Amir R, Yamin C, et al. The ACE deletion allele is associated with Israeli elite endurance athletes. *Exp Physiol.* Sep 2007;92(5):881-6.
85. Kim CH, Cho JY, Jeon JY, et al. ACE DD genotype is unfavorable to Korean short-term muscle power athletes. *Int J Sports Med.* Jan 2010;31(1):65-71.
86. North KN, Yang N, Wattanasirichaigoon D, et al. A common nonsense mutation results in alphaactinin-3 deficiency in the general population. *Nat Genet.* 1999;21:353-4.
87. Yang N, MacArthur DG, Wolde B, et al. The ACTN3 R577X polymorphism in East and West African athletes. *Medicine and science in sports and exercise.* 2007;39:1985-8.



88. Clarkson PM, Devaney JM, Gordish-Dressman H, et al. ACTN3 genotype is associated with increases in muscle strength in response to resistance training in women. *J Appl Physiol.* 2005;99:154-63.
89. Clarkson PM, Hoffman EP, Zambraski E, et al. ACTN3 and MLCK genotype associations with exertional muscle damage. *J Appl Physiol.* 2005;99:564-9.
90. Delmonico MJ, Kostek MC, Doldo NA, et al. Alpha-actinin-3 (ACTN3) R577X polymorphism influences knee extensor peak power response to strength training in older men and women. *J Gerontol A Biol Sci Med Sci.* 2007;62:206-12.
91. Moran CN, Yang N, Bailey ME, et al. Association analysis of the ACTN3 R577X polymorphism and complex quantitative body composition and performance phenotypes in adolescent Greeks. *Eur J Hum Genet.* 2007;15:88-93.
92. Walsh S, Liu D, Metter EJ, et al. ACTN3 genotype is associated with muscle phenotypes in women across the adult age span. *J Appl Physiol.* 2008;105:1486-91.
93. Ahmetov, II, Druzhevskaya AM, Astratenkova IV, et al. The ACTN3 R577X polymorphism in Russian endurance athletes. *Br J Sports Med.* Jul 2010;44(9):649-52.
94. Druzhevskaya AM, Ahmetov, II, Astratenkova IV, et al. Association of the ACTN3 R577X polymorphism with power athlete status in Russians. *Eur J Appl Physiol.* Aug 2008;103(6):631-4.
95. Eynon N, Ruiz JR, Femia P, et al. The ACTN3 R577X polymorphism across three groups of elite male European athletes. *PloS one.* 2012;7(8):e43132.
96. Niemi AK, Majamaa K. Mitochondrial DNA and ACTN3 genotypes in Finnish elite endurance and sprint athletes. *Eur J Hum Genet.* Aug 2005;13(8):965-9.
97. Papadimitriou ID, Papadopoulos C, Kouvatsi A, et al. The ACTN3 gene in elite Greek track and field athletes. *Int J Sports Med.* Apr 2008;29(4):352-5.

98. Roth SM, Walsh S, Liu D, et al. The ACTN3 R577X nonsense allele is under-represented in elite-level strength athletes. *Eur J Hum Genet.* Mar 2008;16(3):391-4.
99. Eynon N, Duarte JA, Oliveira J, et al. ACTN3 R577X polymorphism and Israeli top-level athletes. *Int J Sports Med.* Sep 2009;30(9):695-8.
100. MacArthur DG, Seto JT, Chan S, et al. An Actn3 knockout mouse provides mechanistic insights into the association between  $\alpha$ -actinin-3 deficiency and human athletic performance. *Hum Mol Genet.* 2008;17:1076-86.
101. Scott RA, Pitsiladis YP. Genotypes and Distance Running: Clues from Africa. *Sports Medicine.* 2007;37:1-4.
102. Ion Proton System. <http://www.invitrogen.com/site/us/en/home/Products-and-Services/Applications/Sequencing/Semiconductor-Sequencing/proton.html> Jonathan. Accessed Accessed March 1, 2013.
103. Telegraph completely mangles debate over value of genetic research. [http://scienceblogs.com/geneticfuture/2009/04/telegraph\\_completely\\_mangles\\_d.php](http://scienceblogs.com/geneticfuture/2009/04/telegraph_completely_mangles_d.php). Accessed Accessed Mar 04, 2013.
104. Fuku N, Scott RA, Mikami E, et al. Analysis of multiple performance-associated genetic polymorphisms in sprint and endurance running world record holders [abstract]. In: The American College of Sports Medicine Annual Meeting; Jun 2010; USA.