Genomics of elite sporting performance: what little we know and necessary advances

This is the Accepted version of the following publication


The publisher's official version can be found at

Note that access to this version may require subscription.

Downloaded from VU Research Repository  https://vuir.vu.edu.au/23821/
Genomics of Elite Sporting Performance: What little we know and necessary advances

Yannis Pitsiladis¹ PhD, Guan Wang¹ MRes, Bernd Wolfarth² MD, Robert Scott³ PhD, Noriyuki Fuku⁴ PhD, Eri Mikami⁴ MRes, Zihong HE⁵ PhD, Carmen Fiuza-Luces⁶ PhD, Nir Eynon⁷ PhD, Alejando Lucia⁶ PhD

¹College of Medicine, Veterinary and Life Sciences, Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, United Kingdom.
²Department of Preventive and Rehabilitative Sports Medicine, Technical University Munich, Munich, Germany.
³MRC Epidemiology Unit, Institute of Metabolic Science, Cambridge, United Kingdom
⁴Department of Genomics for Longevity and Health, Tokyo Metropolitan Institute of Gerontology, Tokyo, Japan.
⁵Biology Center, China Institute of Sport Science, Beijing, China.
⁶School of Doctorate Studies and Research, European University of Madrid, Madrid, Spain.
⁷College 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
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 ($\dot{V}O_2$ 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 \( VO_2 \text{ max} \) were compared to control participants with a low to average \( VO_2 \text{ max} \).[42-44] Data were analysed using the candidate gene approach, which allowed the distribution of particular genetic variants with respect to the phenotype \( VO_2 \text{ max} \) 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. 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. 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. 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 (≤ 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 & α-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). 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). 

**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. 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. 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, as well as greater improvements in running economy, or ability to sustain sub-maximal pace with lower oxygen consumption. Additionally, the I-allele associates with muscular endurance gains from training, which may relate to higher type 1 (slow-twitch) muscle fibre preponderance in ACE II subjects. In terms of overall sporting ability, the I-allele has been associated with superior performance in British mountaineers, South-African triathletes, British distance runners, and Australian rowers. In contrast, the D-allele has been associated with success in power-oriented sports such as short-distance swimming and sprinting. Notably, a recent study 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 VO₂ 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 VO₂ 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.

α-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 α-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 577XX 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 577XX 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:
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 ≤ 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® HumanOmni1-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 5x10^{-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


