Title: Sports genetics moving forward - lessons learned from medical research

Short running title: Sports genetics moving forward

Authors: C. Mikael Mattsson\textsuperscript{1,2}, Matthew Wheeler\textsuperscript{1,3}, Daryl Waggott\textsuperscript{1,3}, Colleen Caleshu\textsuperscript{1,3,4},
Euan A. Ashley\textsuperscript{1,3,5}

Affiliations:

\textsuperscript{1}Division of Cardiovascular Medicine, Department of Medicine, Stanford University, Stanford, CA USA
\textsuperscript{2}Åstrand Laboratory of Work Physiology, The Swedish School of Sport and Health Sciences, Stockholm, Sweden
\textsuperscript{3}Center for Inherited Cardiovascular Disease, Division of Cardiovascular Medicine, Stanford University, Stanford, CA, USA
\textsuperscript{4}Division of Medical Genetics, Department of Pediatrics, Stanford University, Stanford, CA, USA
\textsuperscript{5}Department of Genetics, Stanford University, Stanford, CA USA

Correspondence to:

C. Mikael Mattsson, Stanford University, Falk CVRB, 870 Quarry Rd, Stanford, CA 94305
E-mail: mikaelm@stanford.edu, Phone: +1 650 272-9208

Copyright © 2016 by the American Physiological Society.
ABSTRACT

Sports genetics can take advantage of lessons learned from human disease genetics. By righting past mistakes and increasing scientific rigor, the breadth and depth of knowledge in the field can be magnified. We present an outline of challenges facing sports genetics in the light of experiences from medical research.

Sports performance is complex, resulting from a combination of a wide variety of different traits and attributes. Improving sports genetics will foremost require analyses based on detailed phenotyping. In order to find widely valid, reproducible common variants associated with athletic phenotypes, study sample sizes must be dramatically increased. One paradox is that in order to confirm relevance, replications in specific populations must be undertaken. Family studies of athletes may facilitate the discovery of rare variants with large effects on athletic phenotypes. The complexity of the human genome, combined with the complexity of athletic phenotypes, will require additional metadata and biological validation to identify a comprehensive set of genes involved.

Analysis of personal genetic and multiomic profiles contribute to our conceptualization of precision medicine; the same will be the case in precision sports science. In the refinement of sports genetics it is essential to evaluate similarities and differences between genders and among ethnicities. Sports genetics to date have been hampered by small sample sizes and biased methodology which can lead to erroneous associations and overestimation of effect sizes. Consequently, currently available genetic tests based on these inherently limited data cannot predict athletic performance with any accuracy.

Key words: Sports genetics, individual differences, genetic testing, sports medicine, precision medicine
INTRODUCTION

The first ever sequencing of the human genome was completed in year 2000. It took many years, several research teams from around the world, and the cost was approximately $3 billion US dollars. Since then, we have seen dramatic advances in technology and reduction in cost. By the end of 2015, an estimated 400,000 human genomes will be sequenced at a reagent cost of $1000 per genome. Human genetic studies have led to improved understanding of the genetic variation present in all human genomes and definition of the genetic underpinnings of many rare single-gene disorders. The genetics of most common diseases and traits have been harder to fully explain than was initially expected. Genome-wide association studies (GWAS) have now identified thousands of typically common genetic variants associated with a diverse array of intermediate and disease phenotypes, while resequencing studies (now generally exome or genome sequencing) have expanded the range of genes and variants associated with disease phenotypes. Despite substantial advancements, the heritability explained by current genetic knowledge is in general a fraction of the total heritability for most traits as estimated by twin studies.(8)

Sports genetics would benefit from leveraging lessons learned from human disease genetics, both in order to avoid past mistakes and also to increase and improve the breadth and depth of knowledge in the field. The scope of this review article is to address some of the problems facing sports genetics, and to discuss them in light of experiences from medical research. We will use cardiovascular medicine and endurance sports as specific instructive examples, but the principles discussed are applicable to most domains of disease and sports genetics.
Physiological traits instead of athletic performance

It has been known for more than 30 years that several physiological traits important for sports performance have a substantial genetic component (e.g. cardiorespiratory fitness (10, 17, 46)).

In the early days there were great hopes to find THE sports gene (indeed this remains an attractive concept to the popular press), however, due to the wide range of traits and factors contributing to athletic performance (e.g. physiological, anthropological, psychological, nutritional, environmental, as well as serendipity) such a quest is inevitably bound to fail.

Even the pursuit to find THE gene(s) for performance in a more defined sport setting, such as sprint running or endurance cycling, have slim chances of real success, likely because even these more clearly defined phenotypes are the combination of multiple biological phenotypes, and probably because there are several paths to the same level of athletic performance. Each of these phenotypes and pathways may have a polygenic inheritance underlying, with major environmental interaction determining phenotype. An alternative way to address the situation is to investigate each more narrowly defined piece of the physiological traits involved in sports performance. For endurance sports, pertinent, measurable traits could include maximal oxygen uptake (VO2max), blood volume, capillary density or mitochondrial efficiency in addition to anthropomorphic traits. As an allegory, “performance” is to sport as “heart disease” is to medicine. It should be effective for sports geneticists to investigate more detailed traits that contribute to performance in the same way medicine has studied atherosclerotic heart disease by for example investigating the various heritable traits that contribute to coronary artery disease risk (hypertension, hyperlipidemia, inflammatory response).
Effect sizes

In all individuals, common variants will impact disease risk and athletic potential; delineation of the impact and effect of these variants is and will be the subject of many ongoing and future studies. However, common variants typically have small effects. An athletic genome interpretation could delineate rare variants with negative or positive effects, such as a truncated erythropoietin receptor (20) or a genetic variant leading to the lack of myostatin and subsequent muscular hypertrophy (57) suggesting inborn advantage in endurance or strength sports, or on the other hand mutations associated with underlying cardiomyopathy, which if present could lead to restriction from athletic activity. Although there are many more rare variants of small to no effect, the impact of these variants are difficult or impossible to determine using standard methods and even very large sample sizes. So far, most studies of sports related performance have looked at common variants. To get a more complete profile of a person’s intrinsic fitness level, impactful rare and common alleles associated with response to activity and physiological potential must be integrated. In medicine a commonly used approach is to design a custom genotypic array specific for the area of interest. One example is the Metabochip for genetic studies of metabolic, cardiovascular, and anthropometric traits, analyzing 200,000 SNPs, a number that can be contrasted by the median amount of 6 variants being analyzed in direct-to-consumer tests marketed for sport performance.

Intrinsic, Response and Potential

Many studies have addressed the response to training, e.g. strength, power or endurance (e.g. 11, 39). However, while genetic determinants of response to training, as shown in sedentary
populations, are likely one component of the genetics of athletic ability, there are additional aspects affecting ability that are known to have high heritability, including the intrinsic baseline level and the maximal potential of ability. It is known that there are different genetic variants involved in intrinsic fitness level and response to endurance training (measured as improved VO$_2$max).(12, 13) There are not yet any studies looking at all three levels combined in sports genetics. A corresponding example from medicine is Duchenne (DMD) and Becker (BMD) muscle dystrophy. Even though they both are caused by mutations in the same gene (the dystrophin protein coding gene) the outcome differs. DMD is usually established between ages 2-5, whereas only half of the patients with BMD demonstrate muscle weakness at the age of 10. Patients with DMD typically lose independent ambulation before 12 years of age, but BMD patients typically keep ambulation into their third or fourth decade.(69) Transferred to a sport setting, hypothetically, an athlete with the highest maximal potential might have so even though the intrinsic level is low and the response rate is slow. A person with such a genetic profile will likely have to train for a longer period of time (years) before reaching athletic success. It is empirically well known that some athletes have a quick response to training but rather soon reaches their limit, while other athletes have a slow response but continue to improve their ability over the course of many years, and can reach the very high performance levels. For example, we hypothesize that there will be differences in the impactful genetic variants of athletes reaching extremely high VO$_2$max (maximal potential) at an early age (i.e. as teenagers) as compared to athletes reaching the same value at an older age (i.e. after 30 years old).

Continuum from Health to Disease

An overall hypothesis is that human physiology exists as a spectrum from disease to health. While not valid for many diseases where onset may be in late middle age or older, elite
athletes are often held up as the epitome of health. This thought is supported by findings that elite athletes live longer than the general population, and also that endurance athletes live longer than strength/power athletes.\(^\text{(18, 55)}\) Even though physical exercise is beneficial for overall health, it is not necessarily so that the training of elite athletes is the most beneficial.\(^\text{(42)}\)

A possible way forward is interrogating the tails of distribution of continuous phenotypes, such as VO\(_2\)max, to identify individuals with extreme ‘positive’ phenotypes who may inform prevention, cause and treatment of pathogenic disease states. Studying these extremes will help identify genes and pathways that dually contribute to both disease and extreme health states.\(^\text{(13, 15, 22, 28)}\) The most probable relevance will be by rare variants with large effects (e.g. 20), since the effect of common variants with small effects might be overridden by environmental effect. In a study comparing common disease causing variants elite athletes did not differ from controls.\(^\text{(31)}\) There are a few examples where genetic variation in the (super) healthy have been utilized in the prevention of disease. An incredible example of the success of this hypothesis is work by Hobbs and Cohen in identifying the cholesterol therapeutic target PCSK9. The method of identification involved combining family studies of hypercholesterolemia (high LDL) with population studies at the extreme tails of LDL distribution (low LDL) to identify contributing genes.\(^\text{(1, 19, 33)}\) A similar earlier example resulted from the work done on the role of 5-alpha reductase in enlarged prostates with a group of Caribbean androgen insensitive hermaphrodites with small prostates.\(^\text{(35)}\) Since the salutogenic approach (defined as focusing on factors that support health and well-being, rather than on factors that cause disease) approach is inherent in sports science and sports genetics, but just recently has evoked interest in medical research, collaborations might be fruitful for both entities. In medicine there is a national initiative in the US to advance precision medicine, including through use of genetics at the bedside, in order to tailor therapies to
Exercise could also be one of the precision therapies that, after extensive additional sports genetic research, can be tailored to every individual based on for example genetic profile, training response and optimal dose.

ANALYTICAL AND STATISTICAL CHALLENGES FOR SPORTS GENETICS

Small Sample Size

To date, the genetic study of sports related traits has been limited to sample sizes of tens to several hundreds. The evolution of genome wide association studies in medical research has clearly demonstrated that for common variants (MAF > 5%) with modest effects (OR < 1.2), truly massive sample sizes are required on the orders of tens of thousands. In fact, meta analyses with ~100,000 people are still identifying novel common genetic variants with impact on disease and anthropometric phenotypes. While reduced power for discovery in small sample sizes is generally intuitive, there is also an issue of failure to replicate. These observations are explored at length in seminal papers by Ioannidis including “Why most published research findings are false” and “How to make more published research true”. The failure to replicate is due to an inflation of false positives and an overestimation of effect sizes. False positives can result from technical issues related to statistical assumptions such as inflated variance when analyzing sparse data. For example, a SNP with a minor allele frequency of 1% within a cohort of 200 is predicted by Hardy-Weinberg Equilibrium to have 4 heterozygotes vs 296 common homozygotes. Even a 5% SNP within a cohort of 500 will only have 1 subject with a predicted rare homozygote. In a large association study, sparse genotype classes will co-occur (by chance) with trait outliers or confounders such as population subpopulations, which will bias or even invalidate results. Similarly, exaggerated effect size estimates result in an underestimation of the
required replication sample size and subsequently a failure to replicate. (37) False positives and exaggerated effect sizes are particularly prevalent among top ranked results when a significance threshold is applied. This quantitative conundrum has been explored through the application of the Winner’s curse phenomenon. (71) That is, what is top i.e. ‘wins’ is more likely to be an outlier aspect of the sample and not reflect the general population. In fact, correction methods for results using modest sample sizes (n=500-1500) have shown a dramatic (50-90%) reduction in effect size. (61)

There is obviously a contradiction in the quest for increasing samples size and the fact that the number of athletes at the highest level is very limited. A lesson learned from medical research regarding rare cancers, where the numbers are small so that any lost patient is a loss to science, is that this requires a focused development on infrastructure within hospitals and across borders. Even though the numbers are limited the first step is to formalize the infrastructure around athlete testing and genetics so that it is standardized, integrated and global. That way, as many athletes as possible participates, and we capture multiple phenotypes and covariates. Compared to other genetic studies the numbers will still be low, but in only including the outlier athletes our general hypothesis is that we will be able to find rare variations of large effect. An effort of this type has recently been launched. (54) Another method is to focus on tails of general populations, such as national registries, e.g. UK Biobank with 500,000 participants genotyped. For more general outliers in that type of cohorts our general hypothesis is that common variation of small or modest effect is contributing and be found. Combined, both methods are hypothesized to identify similar genes or pathways.
Generalizability

In order to both confirm relevance and accurately assess effect size, it is imperative that replication in an independent cohort and subsequent populations, is undertaken. For pragmatic reasons this has been limited in the genetics of athletic physiology. Currently, there is an entire industry focused on generalizing genetic findings from effects measured in small underpowered, unvalidated, athlete discovery sets. This is scientifically unsound and ethically dubious, which has recently been addressed in a consensus statement. These data should be treated as hypothesis generating starting points, and not for consumer interpretation or actionability. Applying the stringency and oversight of device and medical regulatory bodies would improve legitimacy in this area.

It’s important to emphasize that genetic association, regardless of how robust, does not infer causality. Causality inference requires specific experimental and analytical methods. Therefore, utility for predicting sport proficiency or injury likelihood has little merit.

Examples of spurious association include indirect associations through linkage disequilibrium and confounder intermediates. Confounders of athletic performance could include variants associated with obesity, depression or ethnicity.

Penetrance

While not specifically penetrance, one analytical challenge with performance is the latent presence of relevant variation in the general population. A subset of the population that does not undergo training for a host of socioeconomic, behavioral and biological reasons will harbor the same genetic variation as trained athletes and thus dilute observed effects with direct comparison with athletes. Athletic performance is a prominent example of Gene x
Environment interaction. In medicine, this interaction is beginning to be studied, for example between SNPs associated with type 2 diabetes and environmental factors (such as pollutants and nutrients). (52) The environmental factors are detected using a model analogous to GWAS called EWAS (environment-wide association study). (51) A similar approach in sports genetics, for example investigating the association between gene variants or genetic profiles and response to different training regimes, would have great potential.

However, from the statistical point of view, in almost all situations, modelling interactions requires larger sample sizes than main effects. (49, 70) This further emphasizes the need for substantially larger samples sizes in sports genetics than the current norm, alternatively large effect sizes.

Heterogeneity

Performance is a qualitative phenotype, not a physiological measurement. As discussed earlier it is advantageous to measure specific physiological traits, such as endophenotypes and subclinical traits, which are more likely to have direct genetic underpinnings. In medical genetics, broad disease spectrum phenotypes, such as cardiovascular disease, are very difficult to uncover associations due to the large number of converging physiological influences. Beyond trait heterogeneity, any single trait is likely to have genetic (different genes) and allelic (different SNPs in the same gene) variability. Analytically modelling the joint effects of multiple genes or allele features without exhaustive multiple testing is an active area of research. (27, 62) Moreover, interpretation of variation in regulatory regions including the potential for epistasis and epigenetics will be an ongoing effort for the community. Leveraging network approaches to define key hubs and communities will be terrifically valuable in making sense of emerging disjoint results in this field. (26, 30, 44). The key point
is that for a single observable trait there are multiple correlated physiological and genetic
underpinnings within an individual and among group of individuals.

Genetic Architecture

Analytically, combining variants across a gene is challenging if the effect of individual
variants are in opposite directions. Generally, in disease cohorts the expectation is a
preponderance of deleterious rare variant burden within genes. If the same genes are involved
in athletic physiology, then that leads to the hypothesis that we are looking for either a paucity
of burden or independent variants that confer an opposite effect. The complexity of the
genetic architecture is further likely to be different among different genes making a single
analytical approach unlikely. Current evolution in burden testing tools which allow dynamic
weighting of variants within genes will be crucial in solving these issues.(43)

DIFFERENT ADAPTATIONS AND PROFILES

Sex differences

The physiological path to athletic performance is in several aspects different among men and
women, and these differences cannot suitably be accounted for by adjusting for body size and
composition. Traits with published sexual dichotomy include metabolism (fat metabolism and
lactate threshold), heat management, work efficiency and VO$_2$max adaptation.(3, 32, 59, 60)
Interesting work by Spina et al. studied exercise intervention in sedentary older men and
women.(60) Both sexes increased VO$_2$max by ~20 %, however each achieved it through a
different mechanism. Men improved primarily through increasing maximal cardiac output due
to enlarged stroke volume. In contrast, women increased oxygen consumption through
improving oxygen extraction by working muscles due to greater capilliarization and numbers
of mitochondria. To date, studies of the physiology of sports performance have almost
exclusively been conducted on men, which underline the importance of studies of both men
and women.

Ethnic differences

Up to present time, the majority of sports genetic studies, especially training studies, have
investigated male Caucasian athletes with a few additional studies on East African,
Jamaican/African American and East Asian (Japanese and Chinese) cohorts. The
anatomically modern human evolved in Africa ~200,000 years ago, and migrated throughout
the world. One branch reached Europe some 40,000 years ago, another branch reached East
Asia about the same time, and South America was inhabited ~10,000 years ago. Over the
years, population bottlenecks associated with migration have led to the evolution of different
genetic profiles. An area with some relevance for endurance performance is the adaptation to
altitude. It has been shown that Andean, Tibetan and Ethiopian highlanders each have
different physiological patterns of adaptation to cope with high altitude. Andean natives living
at high elevation have increased hemoglobin concentration, but Tibetans on the other hand
can tolerate lower oxygen saturation. In those aspects, high altitude-adapted Ethiopians have
similar levels as lowlanders, but instead have superior work efficiency. A relatively large
volume of research have addressed the genetic profiles of these adaptations, and found that
genetic variation in different pathways are responsible in each population. If
differences between ethnicities also holds true for physiological adaptations, such as increase
in VO2max, optimal exercise training would most likely also differ, suggesting a more distinct
athletic individualization. In sports genetics, it is for example already known that the
frequency of the X-allele of the *ACTN3* R577X polymorphism is higher in the Caucasian population, and it has been suggested that this mutation appeared when humans migrated to Eurasia and that it has prevailed because it facilitated adaptation to the new environment. (23)

**Multiomic Profiles**

The field of sports genomics has thus far focused predominantly on genetic variants within the inborn genome. High throughput techniques now allow interrogation of the human multiome, including the epigenome, transcriptome, proteome, metabolome, and inflammasome, both within and across tissues and cell types. Integration of tissue-specific multiomic profiling with genomic and physiologic data will allow a more comprehensive understanding of the biological effects of physical activity and the biological underpinnings of athletic ability.

From a genetics perspective, multiomic profiling of individuals or families with unique athletic attributes may facilitate the discovery or biological validation of genes and/or variants implicated in the genetics of athletic ability. For instance, increase in an aerobic exercise facilitating gene’s expression together with change in its binding or regulatory partners identified through proteomic profiling may confirm that gene’s role as a key actor in the athletic pathway; while hypothesis-directed or agnostic (e.g. network modelling) approaches to multiomic data may identify genes or pathways to consider for variant analysis and interpretation. (5)

An advantage or limitation, depending on context, of much of the multiome is the tissue and temporal specificity of a large portion of the multiomic profile. While blood (and to a more limited extent muscle biopsy derived) multiomic profiles are possible to be derived at multiple time points, even from elite athletes, other pertinent human tissues of import (e.g. heart, lung,
brain) in the processes of training and athletic fitness are not practically available for research from athletes. Autopsy studies of athletes who die from other causes may provide some opportunity, but issues of sample preservation and validity for transcriptomics in particular become limiting. On the other hand, repeated measures of blood or muscle omics from the same individuals can greatly reduce the search space within the genome and allow for focused inquiry into one or a few putative causal pathways that are altered by a given intervention. Recent work has shown the power of this repeated measures method to look at the epigenome in association with endurance exercise training. (45)

Going forward, blood-borne biomarkers of training effects or potential that are the synthesis of a genomic profile may be identified through multiomic profiling. In cardiovascular disease, a few key biomarkers are clinically relevant for a number of distinct disease processes, for example serum troponin for myocardial infarction, myocarditis, or risk of heart failure. Multiomic profiling may help identify a set of biomarkers associated with athletic potential, training effects, or expected response to a particular training regime.

**TRANSLATION & INTERPRETATION CHALLENGES**

One of the ultimate goals of sports genetics research is translation of its findings to improve the training of athletes. In fact, such application is already occurring, with some groups using commercially available genetic tests to guide the training, and even selection, of athletes. (67) Yet such application is dependent on the rigor and robustness of the existing research, which currently faces the many challenges that we have outlined here. Translation of the data to training of athletes is also dependent on how the genetic data is interpreted. The particulars of how to interpret and apply that data arise from the fact that sports performance and its various components are ultimately determined by a complex interplay of multiple genetic variants and
multiple non-genetic factors. This is similar to the genetics of diseases like diabetes and coronary artery disease. A decade of work and thousands of publications have revealed that the genetic basis of such diseases and traits are far more elusive than we had hoped. For example, approximately a hundred locations in the genome have been implicated in coronary artery disease, yet in combination they only explain a quarter of the genetic portion of the etiology and only a tenth of the etiology overall. This inability to explain the majority of the genetic etiology has been seen for the vast majority of diseases and traits. To date the genetics of sports performance have been subject to far less extensive and less rigorous investigation than diseases like coronary artery disease. As such, we understand an even smaller proportion of the genetic etiology and we are not as confident in that understanding. This makes the application and interpretation of current sports genetic tests challenging.

The currently available genetic tests assess a handful of common variants purportedly associated athletic ability (such as ACTN3, ACE, NOS3) or injury (such as COL5A1, COL1A1, MMP3). Given the available data, we can be quite confident in some of these associations (e.g. ACE for endurance and ACTN3 for power events), while most others are somewhat suspect since these associations have not been replicated, and at a minimum need further investigation before any clear conclusion can be drawn. However, interpretation of an athlete’s genetic test results for even the robust associations needs to be done cautiously, taking into account the fact that any one genetic variant only explains a small fraction of performance and that we do not test for the majority of the genetic variants that contribute to performance since they have not yet been discovered. Thus, while there is robust data showing that ACTN3 genotype is associated with potential in power events, such as sprinting, it only predicts a small portion of their overall athletic performance, both in sprinting and in general. Also, bare in mind that not even this most confidently proven gene with a convincing mechanistic reason for its importance is completely excluding, even for the absolutely highest
level of performance in the most specific sport, i.e. 100 m running. Given how few genetic
associations are known and how small their effect size likely is, the ability of current sport
genetic tests to predict the performance of an athlete is so vastly limited that their use is fairly
dubious. In medical genetics the incorporation of genetic counseling has proved to be a
necessary tool in patient care. Genetic counselors are Master’s-level trained healthcare
providers that specialize in evaluating the genetic components of disease and disease risk.
Genetic counselors provide both education and brief psychotherapeutic counseling with the
goals of improved client understand, coping, adaptation, and informed decision making.
Without guidance layman patients cannot well-discrim the difference between the types of
mutations, for example between a variant of unknown significance (VUS) and a likely benign
mutation, much less understand the significance of either a truncating or frameshift variant.
Oftentimes patients may misinterpret the effect size or miscalculate the heritability of a
mutation, or not consider the genetic implications for past and future generations. Integration
of genetic counseling practices into routine care removes these burdens and confusions from
the shoulders of the patient, and instead repurposes the raw genetic data into an impactful
learning opportunity for the patient. Both medical genetic tests and sports-related
genetics involve careful interpretation and application. If sports genetic testing reaches its
promise then, much like medical genetic testing, it could have potential significant effects on
an individual’s life trajectory. As such, genetic counseling should be provided alongside any
sports-related genetic testing.

FUTURE DIRECTIONS

As the body of knowledge in sports genetics expands and refines, the ability to apply it to
athletic training in a valid and meaningful way will increase. Much work is being done to
develop valid methodology in the application of genetic knowledge to the prediction of multifactorial, polygenic diseases and traits.(29) In order to get a comprehensive picture large studies are needed to find common variants, in combination with family studies to discover rare variants. Once genetic associations and their odds ratios have been identified in a robust and valid manner, a genetic score can be developed by building a predictive model that includes each genotype, weighted by its effect size. The model then needs to be validated in a prospective cohort that includes both genotyping and detailed phenotyping information. The ability of the genetic score to predict the trait better than existing non-genetic factors needs to be assessed. Such work in coronary artery disease has found that while a genetic risk score is independently statistically significantly better at predicting disease than a clinical risk score, but the actual difference in predictive power is small and the suggestion is to use a combination of the two..(29) If valid genetic associations can be found for sports performance similar work could be done, but as stated earlier, the ultimate utility of a genetic test that predicts sports performance will depend on whether enough reliable and robust genetic associations can be found to explain a sufficiently large portion of the overall variability, which at the moment is very far from the case. In analogue with the example of combined risk scores, a more relevant and practicable way of utilizing a sports genetic score might be in determining what type of training that will give the best response for each individual, as suggested in the segment about gene x environment interaction.

CONCLUSION

The scope of this article was to highlight some of the challenges facing sports genetics moving forward. As described above, some of the most important aspects are:
Since sports performance is very complex and a result of a combination of many different traits and aspects it is necessary to base analyses and conclusions of genetic profiles on more detailed phenotyping.

In order to find valid common variants sample sizes must increase dramatically. To identify rare variants with large effects, ethnicity specific and family studies are needed.

Personalized genetic and multiomic profiles are prominent goals of precision sports science. In order to confirm relevance and effect sizes, it is imperative that replications in specific populations are undertaken. Gender and ethnicity specific analyses will be required to refine sports genetics.

Sports genetics to date has been hampered by small sample sizes and biased methodology which can lead to erroneous associations and overestimation of effect sizes. As a consequence, commercial genetic tests currently available cannot predict performance with any accuracy. Collaboration and collation of larger cohorts, together with improved study design are required to overcome these limitations.
REFERENCES


