

Genetic limitations to athletic performance

Dr Colin N Moran¹ and Dr Guan Wang²

Dr Colin N Moran

¹ Physiology, Exercise and Nutrition Research Group, Faculty of Health Sciences and Sport, University of Stirling, Scotland, UK

<https://orcid.org/0000-0001-6226-8131>

Dr Guan Wang

² Collaborating Centre of Sports Medicine, University of Brighton, Eastbourne, UK

<https://orcid.org/0000-0002-7432-7933>

Abstract

Elite performance is a complex phenotype. It requires athletes to train hard, train well and look after their bodies through appropriate rest and nutrition. However, it also has a well-established genetic component that is larger than most people would estimate. In this chapter we will explore whether genetic variation could limit an individual's athletic ability. We will consider the determinants of performance; both nature and nurture. We will explore the extent of genetic variation and how it can impact on physiological traits; as well as the issue of responders versus non-responders. We will discuss in detail some of the best known common and rare genetic variants that influence performance as well as the issues around, and strategies being employed in, discovering more. Lastly we will consider the potential for gene doping and for athletes to try to modify their performance through altering their DNA. Much about the genetics of sporting performance remains unknown. Athletes should not feel limited by any genetic information they currently have. In fact they should want to know more.

Genetic limitations to athletic performance

How do we define athletic performance?

Athletic performance ultimately is defined by success in winning medals and trophies. However, across sports there is no single method to achieve these goals. Some sports require power and strength, such as rugby or sprinting. Others, like cross-country skiing or marathon running, require endurance. For some, like the 100 m sprint, athletes must run in straight lines. For others, like tennis or basketball, they must twist and turn. Each of these places different demands on an athlete's physiology; and, we are still only scratching the surface of the potential variety and potential for limitation.

Elite athletic performance is a complex phenomenon. In the biological context, human elite sporting performance is a phenotype, which reflects a collection of abilities required to achieve peak physiological performance. These include the ability to utilise and transport oxygen, the ability to adapt to training, and cognitive ability; to mention a few. Additionally, the lines between the necessary aspects of physiology are not easily determined. Some, in fact most, sports require a mixture of these characteristics; albeit to widely varying degrees. Whilst a 100 m sprinter will mostly benefit from strength and power, athletes playing hockey, tennis or football would all be likely to benefit from strength, power and endurance. Even within one sport, there is no fixed path to reach the top level of performance and achieve success. Athletes come in different shapes and sizes and excel in different aspects of their physiology (see Figure 1) despite having the same target.



Figure 1. Images showing differences in physiology within one sport. Lionel Messi (1.7 m) and Cristiano Ronaldo (1.87 m), both renowned international footballers with markedly different statures. (Images obtained from (16) and (3); cropped, resized and decoloured).

What determines athletic performance?

The various aspects of physiology required in a competitive sport help define whether and how an athlete performs remarkable athletic feats. Despite the variety of paths to success, all top athletes display elite athleticism, train hard and manage their bodies to achieve elite performance. Typically, they primarily attribute their success to hard work rather than the natural talents and attributes they were born with. Others consider the converse to be true leading to the nature versus nurture debate. However, as David Epstein writes in *The Sports Gene: Talent, Practice and the Truth about Success*, elite athleticism is a result of both innate hardware (innate ability) and learned software (e.g. training and practice); the greatness of athletes is always characterised by both their genes and their training environments (37). Neither nature nor nurture is absolute for sports expertise. Nature and nurture are intertwined in predisposition to elite athleticism, or any complex trait. Sports geneticists seek to understand where the balance lies and how the two interact.

The complexity of nurture

Nurture plays a crucial role in developing or limiting athletic success. The characteristics of the environment in which an athlete develops and trains are the only part of the equation that we (at least) perceive that we have control over: where to live; which sports to focus on; how often to train; what nutritional advice to follow; etc. Socioeconomics also plays a prominent role. On one hand, low socioeconomic status may motivate athletes to perform and maintain their best condition to win competitions, to improve quality of life, and improve social status. On the other hand, socioeconomics influences the opportunities and resources, such as facilities or equipment, which are available for training of athletes and are also vital for discovering and shaping future Olympians.

Intensity, frequency and type of training are important environmental components that are under an individual's control. However, the magnitude of an individual's response to training is also highly variable. In other words, we don't all respond in the same way or to the same extent. Not just because of varied training loads, durations and patterns; but, as we will explore later, this too is subject to genetic influence (78). Whilst individualised training, according to genetic makeup, would be an appealing aspect to develop, the challenge lies in decomposing the complexity of genetics (nature) that will be further illustrated in the sections below.

The complexity of nature

In the current context, nature refers to the genetic blueprints that define who we are. These blueprints are known as a genome and contain all the necessary instructions for building a human. They are the sum of all the genes in a cell or an organism. However, each one of us is unique; both in obvious ways such as hair colour or eye colour and in more complex subtle ways such as disease risks or sporting potential. Therefore, the instructions that were used to make each of us are also unique: similar because we are all humans, but different because we are all individuals; and they predispose us to be good, or not good, at different things.

Our genomes are made of four DNA letters: A, C, T and G. The sequence of letters in the human genome was first read, or sequenced, in full during the Human Genome Project (HGP) between 1988 and 2003. The HGP was a concerted international effort across 6 nations outlining for the first time a complete human genetic blueprint creating a reference sequence and revealing approximately 20,500 human genes (15). A series of achievements from the HGP include characterisation of 99% of the genic regions of the human genome with a high accuracy (99.99%) and identification of 3.7 million human common genetic variations (although subsequently many more have been found) (15). The outcomes of the HGP had huge impacts throughout human biology and for our understanding of evolution, development and function of human cells, medicine and physiology. The HGP paved the way for decoding the human genetic instruction book and understanding the differences between us as individuals.

Following on from the success of the HGP, the 1000 Genomes Project (2008–2015) sought to sequence multiple human genomes to understand the extent of variation. It used 2,504 individuals from ancestrally diverse populations and created a global reference for common human genetic variation (25). It catalogued a variety of different types of variation in our genomes: 84.7 million single nucleotide polymorphisms (SNPs; i.e. alternative letters in the sequence) with a frequency $\geq 1\%$ across ancestries; 3.6 million short insertions/deletions (indels; i.e. additional or missing letters in the sequence); and, 60,000 structural variants (≥ 50 bp; i.e. large insertion/deletions or sections in unexpected orientations or locations) (25). Each person only carries a subset of these variants; but this library of possibilities allows each of us to be a unique individual. Typically, an individual human genome differs from the reference human genome at 4.1-5.0 million variant sites (25). These genetic variations contribute to the different physical characteristics among individuals, disease susceptibility within and among populations, and differences in almost any human trait. They predispose us to our individual sporting abilities (or lack of) making major contributions to how athletes achieve the highest performance calibre in their specialised sports (for example, by modulating training adaptation or protecting athletes from injury).

The 1000 Genomes Project is a cornerstone of studying the functional impact of human genetic variation and disease association. The 1000 Genomes samples and data remain

valuable open resources (12) providing samples for other large genomic projects such as GEUVADIS (13) and ENCODE (10). Building on this, several nations have launched population-wide biobank projects, such as the pioneer Iceland deCODE project (founded in 1996) (6), the UK Biobank (2006–2010) (31) and the Auriia Biobank in Finland (2012–ongoing) (2). These biobank projects aim to collect human genetic and clinical data on a population scale to improve disease pre-diagnoses and treatments for individuals. Significant expansions of these efforts are also already underway. For example, the 100,000 Genomes Project (England) will grow to sequence one million genomes through the United Kingdom's National Health Service (NHS) centres and the UK biobank, with a further plan announced for sequencing 5 million genomes in the UK within 5 years by the UK Health and Social Care Secretary Matt Hancock in October 2018 (17). Despite these biobank projects focusing on improving patient care, they also facilitate a better understanding of general populations and the interaction between an individual's unique genetic makeup and their environment including how traits related to sporting performance are influenced positively and negatively by genetics.

Simple versus complex genetic traits

Traditionally, in high schools we are taught that genetic traits are simple and follow Mendelian rules of inheritance. This implies that there is a gene for each trait with simple alternative versions. For example, the often-used example of the eye colour gene with blue and brown versions. Some traits are relatively simple. Rare disorders such as cystic fibrosis, sickle-cell anaemia, and Huntington's disease are the result of changes in single genes that can cause significant disease. However, the situation is rarely so simple. Even eye colour in reality involves differing versions of multiple genes (82). Height, a relatively simple aspect of physiology, is estimated to be determined by common variants in at least 700 genes – each with a small effect of only a few millimetres (85) – and rare variants in at least 83 genes with slightly larger effects of up to 2 centimetres (49). Other traits are also determined by multiple genes with small effects. For example, ~900 genes are involved determining risk of hypertension (38). In fact, the vast majority of traits are complex and the result of the combined effects of many genes.

However, genetics does not fully explain the differences in our heights or disease risks on its own. Heritability studies estimate that ~80% of the differences between us in height and ~30-50% of the differences between us in hypertension risk can be explained by heritable (genetic) factors. Consequently, the remaining 20% or 50-70% respectively must be explained by non-genetic factors i.e. environmental differences between us, such as diet and lifestyle. Traits involving the interaction of multiple genetic variants and multiple environment variables are known as complex, polygenic or multifactorial traits.

Elite sporting performance is an example of such a complex trait. A twin study investigating the heritability of elite athlete status found that 66% of the differences between us could

be explained by genetic differences (35). This is consistent with the well-known quote from the renowned exercise physiologist Per-Olof Åstrand “anyone interested in winning Olympic gold medals must select his or her parents very carefully.” However, the result means that genetics and the environment are simultaneously crucial to understanding elite sports performance; not exclusively one or the other. Whilst 66% of the differences between us can be explained by our biological inheritance from our parents, 34% cannot be explained this way and can be influenced by the choices that we make and things over which we have control. These will include training, diet, opportunity, etc. To understand fully the genetic determinants and limits of sporting performance, it is vital that we capture all the genetic effects, large or small, and how genes interact with the environment and choices that we make in the development of performance.

How do we investigate complex performance traits?

Complex traits, almost by definition are difficult to study. The fact that a large proportion of the differences among individuals are the result of environmental factors and choices made by individuals creates a lot of noise in experiments. In part, this problem is addressed by developing larger and larger studies. For example, recent analyses of hypertension have involved >1 million people (38). However, this approach would not be possible when studying elite athletes. Elite athletes are by definition rare. The problem is further compounded by the variety of paths to successful elite performance. For example, an elite 10,000 metre runner may win medals by stretching the field out from the start of the race or by bursting from the pack in a sprint finish over the final 100 metres. Each strategy relying on different aspects of physiology and therefore variation in different genes. Consequently, studies into the genetic determinants of sporting performance typically focus on measureable aspects of physiology known to be important for performance, such as $\dot{V}O_2$ max for endurance or muscle fibre type for power, rather than directly on performance itself.

Heritable aspects of endurance performance

The Hhealth, Rrisk factors, exercise Ttraining And Genetics (HERITAGE) Family Study was kicked started with funding from the US government in 1992. The primary aim was to identify the role of genes in the cardiovascular and metabolic responses to regular endurance exercise (14). The was led by Professor Claude Bouchard and was a multicentre study across 5 institutes. It spanned 12 years and collected a range of physiological data including, but not limited to, blood pressure, blood lactate, glucose, plasma lipids and lipoproteins, cardiac output and $\dot{V}O_2$ max in two-generations of Caucasian and African-American families who participated in a 20-week standardised stationary cycle ergometer programme (14). This allowed investigation of the genetic component of these traits both at baseline and in response to exercise training.

At entry to the study, participants' $\dot{V}O_2$ max scores ranged from $\sim 1750 \text{ ml}\cdot\text{min}^{-1}$ to $3500 \text{ ml}\cdot\text{min}^{-1}$. Since the participants were family members, some more closely related than others, it was possible to see if the degree of relatedness associated with differences in $\dot{V}O_2$ max. Approximately 51% of the differences between individuals' $\dot{V}O_2$ max scores could be explained by heritable factors (28). Following the exercise training protocol there was, as expected, an average increase in $\dot{V}O_2$ max of $\sim 16\%$. However, there were also considerable individual differences in response to training. Roughly 5% of participants had little or no improvement (i.e. $<5\%$ increase) whilst $\sim 5\%$ improved dramatically (i.e. $>40\%$ improvement); although every size of response in between was also observed (14). This could not be accounted for by age, sex, initial fitness, or ethnicity. Importantly, there was more variance (around 2.5 times) between families than within families – i.e. two individuals were more likely to have a similar $\dot{V}O_2$ max if they were from the same family than if they were from different families – indicating a genetic component to $\dot{V}O_2$ max trainability (27). The heritability estimate reached 47% and could not be explained by baseline variables including baseline $\dot{V}O_2$ max. Thus, genetics is equally important to sedentary $\dot{V}O_2$ max (or $\dot{V}O_2$ max response to training) as environmental factors that are widely accepted to be important.

What makes a champion?

This study additionally suggests that the genes responsible for baseline sedentary $\dot{V}O_2$ max are different from the genes responsible for $\dot{V}O_2$ max response to training (27). Therefore, athletes with the highest $\dot{V}O_2$ max are likely to have a set of genetic variants giving them naturally high $\dot{V}O_2$ max when sedentary and have another set of genetic variants giving them a high $\dot{V}O_2$ max response to training. Additionally, of course they must do the appropriate training and look after their diet and general health. Individuals with the best genetic profiles will not become World Champion athletes if they spend their days sitting on their couch. Equally, individuals with the highest levels of performance may not be the most genetically gifted. See Figure 2 adapted from ref (78).

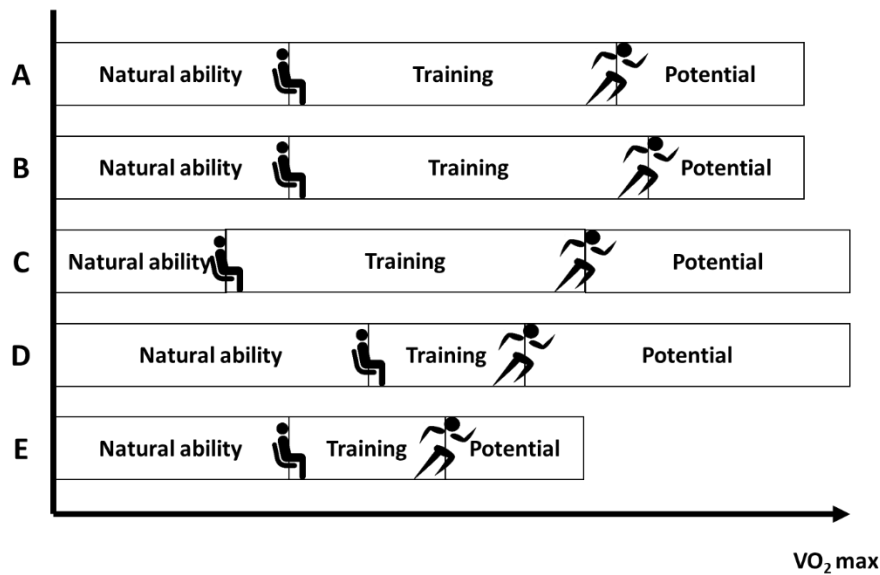


Figure 2. Figure showing the importance of genetics in natural talent, genetics in trainability and training. Athletes A and B have similar natural abilities (baseline genetics) and similar potential (trainability genetics) but athlete B trains harder realising more of their potential. Athlete C trains equally hard to athlete B but has less natural ability putting them behind. However, athlete C actually has a higher genetic potential were they willing to train even harder. Athlete D has an extremely high natural ability and similar potential to athlete C but spends most of their time being sedentary. Athlete E has an average natural ability but much more limited potential than the other athletes.

Are there really non-responders to exercise?

As well as quantifying the genetic contribution to training response, the HERITAGE study also suggests that some individual's $\dot{V}O_2$ max does not respond to aerobic exercise training. Genetics is often incorrectly thought of as being both deterministic and categorical: e.g. athletes have the gene for strength, or the gene for speed, whilst the rest of us do not – or that there are groups of extreme people who are responders (or non-responders) to exercise. A study by Montero and Lundby in 2017 (51) reported that non-responders of $\dot{V}O_2$ max can respond given a sufficient exposure to exercise. In this study, 78 healthy young men participated in a successive 6-week endurance training programme in 5 groups with differing training volumes (i.e. 1 - 5 x 60 mins training sessions per week, respectively), with an average training intensity equivalent to 65% of maximal power output (W_{max}) for 60 mins. The non-responders were classified given a technical error of 3.96% for W_{max} based on the baseline measurements in all participants. They were then subjected to a further 6-week training (identical to the first training protocols) plus 2 additional exercise sessions per week. The authors concluded that the non-responsiveness is diminished and eventually removed from all training groups with the increased training, and demonstrated that total haemoglobin mass is a primary determinant of $\dot{V}O_2$ max. The authors pointed out that the

technical error (%) for $\dot{V}O_2$ max was not calculated due to the lack of multiple $\dot{V}O_2$ max measurements at baseline, and physiological responses in body composition changes were not measured in the study. Other criticisms that this work has received include the low W max error used, failure of considering the W max error both at baseline and post training, and potential recruitment bias by Phillips et al. (60). The latter group of authors (60) studied $\dot{V}O_2$ max, blood pressure and HOMR-IR responses to a 6-week high-intensity interval training (3 sessions per week at workloads equivalent to $\sim 100\%$ or $\sim 125\%$ $\dot{V}O_2$ max) in 189 sedentary women and men (including 13 non-exercise participants) with impaired glucose intolerance and/or a body mass index >27 kg/m². They observed a comparable non-responder rate of $\sim 15\text{-}20\%$ for $\dot{V}O_2$ max to other large and robust exercise training studies, re-stating the heterogeneity of responses to exercise training (60). However, it is important to remember that complex genetics is not categorical. As clearly shown in the figures of Bouchard et al (28), there is not a group of responders and a group of non-responders; there is a continuum. It would be more appropriate to consider the degree of responsiveness, removing much of the conflict between the above studies. Those requiring extra training sessions in the Montero and Lundby study could be described as being less responsive to training rather than non-responders.

Which genes determine $\dot{V}O_2$ max?

The HERITAGE and other studies established the magnitude of the importance of heritable factors in determining $\dot{V}O_2$ max. A best guess, based on other complex phenotypes, is that as many as 1000 genes may be involved, each containing common genetic variants with small effects on $\dot{V}O_2$ max. However, the identity of those genes and variants remains largely unknown. Initial efforts took a candidate gene approach. This approach relies on existing knowledge of the underlying physiology of $\dot{V}O_2$ max and investigates variation in genes related to these physiological processes or structures.

Physiologically, $\dot{V}O_2$ max is largely determined by the capacity of the heart and oxygen transportation / delivery systems (73) as well as the muscles' ability to perform aerobic respiration. However, many studies have used relatively small samples or have used differing training protocols leading to results that are often irreproducible (73). Two of the most robustly reproduced genes associated with sporting performance are angiotensin I converting enzyme (*ACE*) and alpha-actinin-3 (*ACTN3*) both with a link to endurance performance.

ACE is part of the renin angiotensin system and is involved in blood pressure regulation and fluid-electrolyte balance. This gene contains a common ($\sim 40\%$ minor allele frequency; MAF) well studied 287 bp *A/u* insertion / deletion polymorphism in intron 16 (Ensembl Variant rs1799752). The insertion (I)-allele is associated with lower levels of circulating (68) and tissue (33) *ACE* activity; whilst the deletion (D)-allele is associated with higher levels of circulating and tissue *ACE* activity (61). The *ACE* enzyme is a dipeptidase catalysing the

conversion of inactive angiotensin I to active angiotensin II. Angiotensin II is a potent vasopressor and aldosterone stimulating peptide. Consequently, the *ACE* I/D polymorphism has potential to alter blood pressure control and fluid electrolyte balance depending on which version of the gene individuals carry. Control of blood flow to working muscles and fluid electrolyte balance are crucial for sporting activity making *ACE* I/D an excellent candidate polymorphism for sporting performance.

In 1998, the *ACE* I/D polymorphism was the first genetic variation to be associated with performance (52). Since then, the *ACE* I-allele, or II homozygote, has been repeatedly associated with aerobic performance; whilst the D-allele, or DD homozygote, has been repeatedly associated with strength and power performance. Some studies have failed to find these associations; although, evidence from a recent meta-analysis suggests that these associations are genuine with II individuals being 1.35 (95%CI 1.17–1.55) times more likely to be endurance athletes and DD individuals being 1.21 (95%CI 1.03–1.42) times more likely to be strength athletes (46). Nonetheless, the exact molecular mechanisms by which the I/D polymorphism influences both $\dot{V}O_2$ max and strength remain elusive.

ACTN3 is perhaps the best-known sports performance gene. It is often referred to as the sprinting gene. It is a muscle structural protein primarily expressed in type II (fast) skeletal muscle fibres. There, it binds to the actin thin filaments anchoring them at the Z discs between the sarcomeres where it is crucial for muscle function and contraction. However, it contains an unusual nonsense polymorphism at amino acid 577 (R577X; Ensembl Variant rs1815739; (56)). Unusual in that it is both well tolerated and common in human populations (50). Any variant that results in the absence of a structural protein ought to have a dramatic effect on phenotype, be strongly selected against and therefore rare in human populations. However, the effects of this change in *ACTN3* are tolerated far better than would be predicted and the underlying polymorphism far more common (~40% globally) than would be expected. This tolerance appears to be due to overlapping expression patterns and functional redundancy with the related protein *ACTN2*. *ACTN2* can carry out the essential functions of *ACTN3* meaning that *ACTN3*'s absence is not so damaging (50). Although, given the *ACTN3* R577X association with sporting performance, *ACTN2* clearly cannot carry out all of *ACTN3*'s functions equally well.

ACTN3 was first identified as an elite performance gene in 2003 in a cohort of elite Australian athletes (86). The authors compared the frequency of the RR, RX and XX genotypes and the R and X alleles in 107 power athletes, 194 endurance athlete to 436 controls (see Figure 3). They highlighted an increase in R-alleles in the sprint athletes and a concomitant decrease in X-alleles in the sprint athletes. Whilst the converse was true in the endurance athletes. They suggested that the R-allele was of benefit to elite sprinters, whilst the X-allele was of benefit to elite endurance athletes.

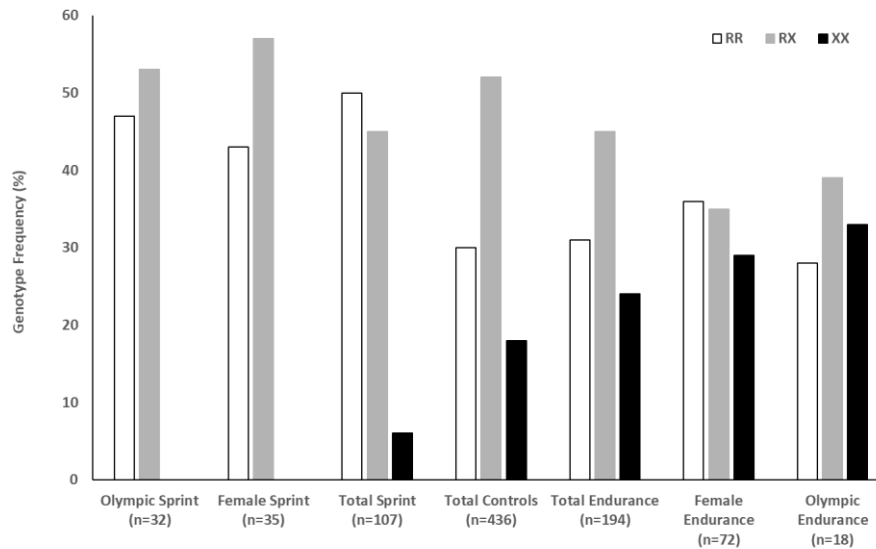


Figure 3. Graph showing the *ACTN3* R577X genotype frequencies in power athletes, endurance athletes and controls. Redrawn from data in Yang *et al* (2003) (86).

Since then the R-allele has been very robustly associated with sprinting ability in both elite (e.g. (41)) and the general populations (53). Although the effect size is small, explaining ~2.3% of the variance in 40 m sprint ability of the adolescent males. The X-allele has also less clearly been associated with endurance ability (23, 86). However, recent work in animal models, and replicated in humans, has identified a plausible mechanism for the X-allele to improve endurance performance through a shift towards slow myogenic programming (76). The association with endurance may become more reproducible as we understand the mechanism of action and therefore test associations in populations with the most appropriate training backgrounds.

Rare genetic variants that influence $\dot{V}O_2$ max

ACE I/D and *ACTN3* R577X are both common genetic variants with MAFs close to 40%. It is likely that some of the heritability of sporting performance will be the result of rarer genetic variants. These are harder to identify and study simply because they are rarer meaning that studies to identify and investigate them need to be much larger to gather enough carriers together.

An example of a rare variant with a beneficial effect on sporting performance comes from the famous Finnish cross-country skier, Eero Mäntyranta, and his family. Mäntyranta was a phenomenal athlete. He competed in four Winter Olympics (1960-1972), winning seven medals (3 golds, 2 silver and 2 bronze) as well as five World Championships medals (2 gold, 2 silver and 1 bronze; 1962 and 1966) (7, 8). He won some races by unsurpassed margins. However, he was also known to have a high haematocrit and associated high $\dot{V}O_2$ max, leading to accusations of doping that he could not shake. But doping was not the source of

his advantage. After his career had finished, scientists studied samples of his bone marrow and DNA to try to understand why he had such a high blood cell count (34). They were astonished to see his bone marrow producing red blood cells without any stimulation.

Normally, bone marrow produces red blood cells only when stimulated by a hormone called erythropoietin (EPO). To most sports fans, EPO is synonymous with doping. However, it is a naturally occurring hormone produced by the kidneys when oxygen levels are low. Once in the circulation, EPO binds to the erythropoietin receptor (EPOR) on bone marrow cells stimulating the production of new red blood cells. These red blood cells increase the oxygen carrying capacity of the blood counterbalancing the low oxygen levels. It is only doping when athletes inject themselves with additional EPO to stimulate red blood cell production artificially.

The investigation of Mäntyranta's bone marrow identified a rare genetic variant (Ensembl Variant rs121917830) in his *EPOR* (34). This variant meant that the receptor constantly signalled the presence of EPO, even when there was not any there, producing a haematocrit up to 50% higher than normal. In the general population this variant has a MAF of <0.001% making it incredibly rare (40). Although, the study including almost 100 members of Mäntyranta's extended family found it to be at ~30% in his close family members giving many of them higher than normal haematocrit. These extra red blood cells gave Mäntyranta a significant natural advantage in endurance events. However, a permanently elevated haematocrit comes at a cost. Whilst, it allows the blood to carry more oxygen, it also thickens the blood potentially increasing the risk of heart attack and stroke. This condition is known as polycythaemia (also known as erythrocytosis) (18).

Rare and common genetic variants that influence injury risk

Injury risk is a major concern for top athletes. Injuries can restrict training schedules or participation in major sporting events. Soft tissue injuries ranging from minor to severe, such as twisted ankles or anterior cruciate ligament (ACL) injury, are relatively common in some sports. Collagen is one of the most abundant and crucial proteins in the human body. It is a main component in the structure and support of our soft connective tissues. Collagen malfunction can manifest itself in rare genetic disorders, such as osteogenesis imperfecta, chondrodysplasias, or Ehlers-Danlos syndrome, depending on the types of collagens involved (32). However, collagen variants can also have more benign but nonetheless significant effects on athletes by influencing injury risk. Common genetic variants (e.g. rs12722, T-allele MAF=35% and rs13946 C-allele MAF=25% estimated from 1000 Genomes Project populations; see (4) and (5)) in the *COL5A1* gene, which encodes the type V collagen, are the most studied genetic loci related to tendon and ligament injuries (24, 75). A recent systematic review showed that carriers of the TT genotype are 1.58 (95%CI 1.33–1.89) times (a combined effect across multiple studies) more likely to suffer soft tissue injuries such as tennis elbow, ACL rupture and Achilles tendon pathology (45). This finding was

further supported by a subsequent meta-analysis illustrating the protective role of the CC genotype of rs12722 and rs13946 to tendon-ligament injuries (57). The type V collagen molecules align themselves alongside the type I collagen, regulate the diameter of these fibrils and modulate assembly of other collagen types in several tissues (26, 81). Rare mutations in *COL5A1* also associate with the Ehlers-Danlos syndrome, characterised by joint hypermobility (47, 48). Links between rarer variants of the *COL5A1* or other collagen genes and tissue injuries or endurance performance in athletes warranting further investigation are nevertheless interesting.

Beyond candidate gene studies of $\dot{V}O_2$ max

An issue with a candidate gene approach is that it can only ever find associations in genes and process already known to be involved. It cannot find new and unexpected relationships with $\dot{V}O_2$ max. To avoid this pitfall, variants related to any phenotype can be identified using a hypothesis free approach known as a genome wide association study (GWAS). GWAS test variation at nearly all known sites in the genome and compare genotype frequencies in athletes and controls, or average phenotypic scores across genotypes in a very similar manner to candidate gene studies. However, GWAS test all known variants regardless of whether they are from pathways known to be involved in the underlying physiology. This allows them to identify unexpected pathways involved in sporting performance.

An inherent requirement of GWAS is that they must correct for the large number of statistical tests performed. This is done by lowering the threshold at which significance is accepted from the more familiar 0.05 of many candidate gene studies to 5×10^{-8} which is known as genome wide significance. This reduces the number of false positives, but has the unfortunate consequence that GWAS need very large numbers of participants to achieve such low p-values given that the variants mostly have small effects. Gathering a large number of high-level elite athletes is rather difficult as only a small fraction of the athletes reach the necessary performance calibre to be considered elite. GWAS are widely used in the study of complex phenotypes; although the majority of the research focuses on health conditions such as diabetes or cardiovascular disease (see the GWAS Catalog (54)). For such common conditions, large numbers of affected and unaffected individuals often with extensive physiological measurements are relatively easy to identify. Whilst an initial GWAS conducted in 2005 on age-related macular degeneration had only a few hundred participants (42), modern GWAS have hundreds of thousands of participants (e.g. a 5-year GWAS review (79) published in 2012 and a recent meta-analysis of GWAS of height and body mass index in ~700000 individuals (87)) increasing their power to detect variants associated with the phenotype under investigation. Typically, a GWAS will identify one genetic variant at genome wide significance for every ~1000 participants. This presents an additional obstacle for the study of elite sporting performance. Elite athletes are, by definition, rare making it difficult to achieve the necessary numbers to make this approach viable.

Consequently, GWAS with sufficient power to detect genetic variants predisposing to elite athletic performance are scarce. Some studies have used this approach to identify genes associated with elite endurance performance (20, 63). Additionally, a recent report combines GWAS and metabolic profiling in 490 endurance athletes (who were also tested negative for doping) (22). However, despite their value there are also caveats to these studies, not just the relatively small number of samples analysed, but also problematic categorisation of elite athletes, the use of customised GWAS arrays containing a limited number of genetic variations, lack of stringent validation and replication studies, and/or multiple testing issue. Attempts are being made to unravel variants associated with elite sprint performance in three ethnic populations of World-class athletes in the hope that this may circumvent the need for a very large number of participants required for conventional GWAS. Genotype imputation and meta-analysis of the three ethnic GWASs, and replication of the top finding were then followed to maximise the power of GWAS to identify and verify putative common genetic variations (MAF >5%) with modest effect (effect size >2; unpublished data, Guan Wang et al). Other studies have attempted to use novel alternative approaches to focus their search for genetic variants. However, genome-wide examinations for DNA and RNA expression profiling for $\dot{V}O_2$ max response to endurance training in the HERITAGE Family Study and other cohorts yielded exciting but inconsistent molecular findings (29, 77). Again this is attributable to some inherent differences between the studies such as small samples, population heterogeneity, different training programmes, and potential false positive findings among others (29, 73, 77). Despite the observed inconsistencies, combining gene expression profiling with targeted genotyping showed improved explanatory power in identifying genes associated with $\dot{V}O_2$ max training response (77). Indeed, a similar approach combining the genomics and transcriptomics data has produced a strong gene signature of triglyceride response to exercise training in HERITAGE white participants (72). Performing GWAS on elite performance is not straightforward. Other large-scale sequencing efforts at the DNA, RNA, and protein level and beyond (e.g. studies of epigenetic marks and DNA folding) involving recruiting participants of the highest performance calibre (such as the ELITE (9) and the Athlome (1) projects) should shed light on our understanding of the underlying genetic mechanisms of elite human performance.

How much do we understand?

This discovery journey on our understanding elite human performance has initially been recorded in a series of annual reviews and reports — The Human Gene Map for Performance and Health-related Fitness Phenotypes (30, 59, 62, 64-66, 83) and the Advances in Exercise, Fitness and Performance Genomics (39, 44, 58, 67, 71, 74, 84) were annually published in 2001–2007 and in 2008–2015, respectively. These reviews aimed to access and summarise genetic/genomic findings associated with human performance and health, identifying existing caveats in the literature, and explore trends for understanding the genetic basis of human performance/fitness. The authors reviewed a number of traits,

including physical activity behaviour, muscle strength and power, cardiorespiratory fitness and endurance performance, body weight and adiposity, insulin and glucose metabolism, lipid and lipoprotein metabolism, and hemodynamic traits. Despite this there are seemingly hundreds of genes associated with the various traits (but often a lack of replication), the exact casual genetic variants underlying elite athletic performance remain undetected. Small sample size, the primarily used candidate gene approach, and the failure to take into account multiple testing correction in the early studies are the primary causes for the inconclusive findings. Further, human genetic variation is now recognised to be individual, and within-population variation larger than between-population variation (21, 69, 70). The multifactorial nature of elite athleticism demands collaborative research studies with a versatile approach integrating the different layers of the molecular and cellular data in multiple and relevant tissues to discover and validate the results within and across populations; underpinned by concerted efforts from the research communities, funding bodies and other stakeholders (e.g. local governments) and across the globe. In summary, across the whole of sports genetics, regardless of approach, only a few more than 200 genes have been associated with performance, and only 20 of those with elite performance (30). Even fewer have been robustly reproduced in multiple studies or cohorts (19).

Gene Doping

So far, we have concentrated on how natural genetic variation contributes to sporting performance. However, the knowledge gained from studying the genetics of sporting performance could ultimately be used to enhance an individual's ability to perform. Whilst, Gene Doping is not currently believed to be possible, in 2003 Gene Doping was added to the World Antidoping Agency (WADA) Prohibited List (80). Gene Doping is defined as "the non-therapeutic use of genes, genetic elements and/or cells that have the capacity to enhance athletic performance." More simply put, Gene Doping is Gene Therapy in people who have no medical need for it. Gene Therapy is a medical technique that involves either transferring modified DNA into an individual, or modifying the DNA of an individual, to treat a medical condition. In its simplest form, this would be to provide a functional version of a missing or damaged protein.

The first clinical trial of Gene Therapy was conducted in 1990 (55). In the thirty years since then there have been >3000 clinical trials of gene therapy; although very few of them have led to successful mainstream clinical applications (categorised by the four clinical trial phases during the drug-development process) (11). Despite the complexities, great hope is held for the potential for Gene Therapy and many trials are currently ongoing. It seems very likely that Gene Therapy will have more widespread success in the future. Gene Doping will likely follow close behind. At that point, sport will have to consider its response. In fact, it is possible, perhaps likely, that some individuals have already tried gene doping.

Notably, Gene Doping has the potential to go beyond providing individual athletes with natural variants that they were not born with to providing them with versions of genes or

amounts of protein not found naturally in the global human population. Either modified proteins known to manipulate physiology in ways beneficial to performance (e.g. an always on *EPOR* similar to the Mäntyranta family variant above) or expression of proteins at supraphysiological levels to provide performance enhancements (e.g. *IGF-1* leading to enhanced muscle growth (43)).

A significant recent advancement in the field of Gene Therapy is the discovery and development of the CRISPR-Cas9 system for modifying DNA. The Cas9 protein assembles with a guide RNA enabling DNA binding and cutting much more precisely than has previously been possible in humans. This opens up possibilities for gene therapy, but also for Gene Doping. The CRISPR-Cas9 system can create either permanent or temporary changes to the genome causing insertion, deletion, replacement of gene(s), single-base changes, or gene suppression/activation (36). New potentially even better genome editing tools are also emerging, such as prime editing, CRISPR-Cas3 and EvolvR (36). Whilst Gene Therapy is not currently main stream, we move ever closer to that scenario.

Despite the profound advantages of genome editing in treating and preventing genetic disease, caveats such as challenges of delivering the system to the affected tissue or target site and off-target genetic changes face the applications of genome editing tools in many research fields. Therefore, an important consideration of Gene Doping is the potential risk to the health of the athletes. Gene Doping is highly likely to be untested. If it works at all, it may include unwanted and potentially lethal side effects. Although, many of the risks are not different from more conventional doping, changes may be permanent and side effects more severe. Despite this, people are trying it. Josiah Zayner publicly attempted to modify his muscles using a CRISPR DNA kit; although he did later regret his actions (88). It is worth also considering that genetic modification with CRISPR or other tools will likely eventually succeed and become commonplace. The desire to provide one off treatments for individuals with chronic genetic conditions drives the field forward inevitably. Sport will have to deal with it. Whilst currently morally and ethically highly questionable, the ethical viewpoint of society will likely change as treatments become possible and commonplace.

Limits of performance

It is notable that the genetic variants most strongly linked to performance appear to have a trade-off within them. Having the right *ACTN3* R577X genotype for speed means an individual doesn't have the right *ACTN3* genotype for endurance and *vice versa*. The same is true for the *ACE* I/D polymorphism and strength versus endurance associated genotypes. Similarly, although Eero Mäntyranta gained a performance advantage through his rare *EPOR* variant, the variant comes at a cost to health or at least health risk. In this way there may be a genetic limit to performance. It is also important to remember that there is no such thing as a perfect genetic profile for anything. The 'best' genetic profile is only best in the environmental context in which it is measured and the environmental context, even for one individual, is constantly changing. Elite athletes, who train multiple times a day whilst at their peak, do not continue with such intensive training into their retirement and may be accelerating undesirable conditions in later life if they alter their genetics.

In summary, elite athletic performance is within the constraints of both genetics and the environment; but currently only parts of the environment are within the athlete's control. Athletes should continue to train hard, frequently, with the best coaches and take care of their nutrition in an effort to achieve their potential. Meanwhile genetic research will uncover the knowledge required to best train and guide athletes to achieve their peak performance; although, we have only begun to scratch the surface. This may allow athletes to bypass some unwanted limitations and stimulate achievement of their full potentials. Consequently, genetic research should gain support from athletes, coaches and other stakeholders who wish to drive performance forward. Genetics may limit an individual's performance potential, but with so many genes likely to be involved and so few of them currently identified, athletes should not consider any genetic information they have to be a hindrance to their performance; but they should want to know more as it may help them or others perform better in the future.

References:

1. The Athlome Project Consortium. [cited 2020 6th April]; Available from: <http://www.athlomeconsortium.org/about/>.
2. Auria Biobank. [cited 2019 26 August]; Available from: <https://www.auria.fi/en/index/php?lang=en>.
3. Cristiano Ronaldo (The commons). [cited 2020 8th April]; Available from: [https://commons.wikimedia.org/wiki/File:Cristiano_Ronaldo_\(163461623\).jpeg](https://commons.wikimedia.org/wiki/File:Cristiano_Ronaldo_(163461623).jpeg).
4. dbSNP - rs12722. [cited 2020 8th April]; Available from: <https://www.ncbi.nlm.nih.gov/snp/rs12722>.
5. dbSNP - rs13946. [cited 2020 8th April]; Available from: <https://www.ncbi.nlm.nih.gov/snp/rs13946>.
6. deCODE genetics. [cited 2019 23rd August 2019]; Available from: www.decode.com.
7. Eero Mäntyranta (Wikipedia). [cited 2020 8th April]; Available from: https://en.wikipedia.org/wiki/Eero_Mäntyranta.
8. Eero Mäntyranta earned his first gold medal in 1960 in Squaw Valley as a member of Finland's 4 x 10 km relay team. [cited 2020 8th April]; Available from: www.olympic.org/eero-mantyranta.
9. ELITE - Exercise at the Limit - Inherited Traits of Endurance. [cited 2020 6th April]; Available from: <http://med.stanford.edu/elite.html>.
10. The ENCODE Consortium. [cited 2019 23rd August]; Available from: <https://www.encodeproject.org/>.
11. Gene Therapy Clinical Trials Worldwide. 2019 [cited 2020 8th April]; Available from: <http://www.abedia.com/wiley/phases.php>.
12. Genome Reference Consortium. [cited 2019 23rd August 2019]; Available from: <https://www.ncbi.nlm.nih.gov/grc/human>.
13. The Geuvadis Consortium. [cited 2019 23rd August 2019]; Available from: <https://www.internationalgenome.org/data-portal/data-collection/geuvadis>.
14. HERITAGE - Genetics, Response to Exercise, Risk Factors. [cited 2019 27 August]; Available from: <https://www.pbrc.edu/heritage/index.html>.
15. The Human Genome Project. [cited 2019 22nd August]; Available from: <http://www.genome.gov/human-genome-project>.
16. Lionel Messi (The commons). [cited 2020 8th April]; Available from: https://commons.wikimedia.org/wiki/File:Messi_Barcelona_-_Valladolid.jpg.
17. Matt Hancock announces ambition to map 5 million genomes. 2018 [cited 2020 8th April]; Available from: <https://www.gov.uk/government/news/matt-hancock-announces-ambition-to-map-5-million-genomes>.
18. Polycythaemia (NHS). [cited 2020 8th April]; Available from: <https://www.nhs.uk/conditions/polycythaemia/>.
19. Ahmetov, II, Egorova ES, Gabdrakhmanova LJ, and Fedotovskaya ON. Genes and Athletic Performance: An Update. *Med Sport Sci* 61: 41-54, 2016.
20. Ahmetov I, Kulemin N, Popov D, Naumov V, Akimov E, Bravy Y, et al. Genome-wide association study identifies three novel genetic markers associated with elite endurance performance. *Biol Sport* 32: 3-9, 2015.
21. Ahn S-M, Kim T-H, Lee S, Kim D, Ghang H, Kim D-S, et al. The first Korean genome sequence and analysis: Full genome sequencing for a socio-ethnic group. *Genome Res* 19: 1622-1629, 2009.

22. Al-Khelaifi F, Diboun I, Donati F, Botrè F, Abraham D, Hingorani A, et al. Metabolic GWAS of elite athletes reveals novel genetically-influenced metabolites associated with athletic performance. *Sci Rep* 9: 19889, 2019.
23. Alfred T, Ben-Shlomo Y, Cooper R, Hardy R, Cooper C, Deary IJ, et al. ACTN3 genotype, athletic status, and life course physical capability: meta-analysis of the published literature and findings from nine studies. *Hum Mutat* 32: 1008-1018, 2011.
24. Altinisik J, Meric G, Erduran M, Ates O, Ulusal AE, and Akseki D. The BstUI and DpnII Variants of the COL5A1 Gene Are Associated With Tennis Elbow. *Am J Sports Med* 43: 1784-1789, 2015.
25. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, et al. A global reference for human genetic variation. *Nature* 526: 68-74, 2015.
26. Birk DE. Type V collagen: heterotypic type I/V collagen interactions in the regulation of fibril assembly. *Micron* 32: 223-237, 2001.
27. Bouchard C, An P, Rice T, Skinner JS, Wilmore JH, Gagnon J, et al. Familial aggregation of VO(2max) response to exercise training: results from the HERITAGE Family Study. *J Appl Physiol (1985)* 87: 1003-1008, 1999.
28. Bouchard C, Daw EW, Rice T, Pérusse L, Gagnon J, Province MA, et al. Familial resemblance for VO2max in the sedentary state: the HERITAGE family study. *Med Sci Sports Exerc* 30: 252-258, 1998.
29. Bouchard C, Sarzynski MA, Rice TK, Kraus WE, Church TS, Sung YJ, et al. Genomic predictors of the maximal O₂ uptake response to standardized exercise training programs. *J Appl Physiol (1985)* 110: 1160-1170, 2011.
30. Bray MS, Hagberg JM, Pérusse L, Rankinen T, Roth SM, Wolfarth B, et al. The human gene map for performance and health-related fitness phenotypes: the 2006-2007 update. *Med Sci Sports Exerc* 41: 35-73, 2009.
31. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562: 203-209, 2018.
32. Chan TF, Poon A, Basu A, Addleman NR, Chen J, Phong A, et al. Natural variation in four human collagen genes across an ethnically diverse population. *Genomics* 91: 307-314, 2008.
33. Danser AH, Schalekamp MA, Bax WA, van den Brink AM, Saxena PR, Riegger GA, et al. Angiotensin-converting enzyme in the human heart. Effect of the deletion/insertion polymorphism. *Circulation* 92: 1387-1388, 1995.
34. de la Chapelle A, Träskelin AL, and Juvonen E. Truncated erythropoietin receptor causes dominantly inherited benign human erythrocytosis. *Proc Natl Acad Sci U S A* 90: 4495-4499, 1993.
35. De Moor MH, Spector TD, Cherkas LF, Falchi M, Hottenga JJ, Boomsma DI, et al. Genome-wide linkage scan for athlete status in 700 British female DZ twin pairs. *Twin Res Hum Genet* 10: 812-820, 2007.
36. Doudna JA. The promise and challenge of therapeutic genome editing. *Nature* 578: 229-236, 2020.
37. Epstein D, *The Sports Gene: Talent, Practice and the Truth about Success*. Great Britain: Yellow Jersey Press, 2013, p. 338.
38. Evangelou E, Warren HR, Mosen-Ansorena D, Mifsud B, Pazoki R, Gao H, et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat Genet* 50: 1412-1425, 2018.

39. Hagberg JM, Rankinen T, Loos RJ, Pérusse L, Roth SM, Wolfarth B, et al. Advances in exercise, fitness, and performance genomics in 2010. *Med Sci Sports Exerc* 43: 743-752, 2011.
40. Hussein K, Percy M, and McMullin MF. Clinical utility gene card for: familial erythrocytosis. *Eur J Hum Genet* 20, 2012.
41. Kikuchi N, Miyamoto-Mikami E, Murakami H, Nakamura T, Min SK, Mizuno M, et al. ACTN3 R577X genotype and athletic performance in a large cohort of Japanese athletes. *Eur J Sport Sci* 16: 694-701, 2016.
42. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, et al. Complement factor H polymorphism in age-related macular degeneration. *Science* 308: 385-389, 2005.
43. Lee S, Barton ER, Sweeney HL, and Farrar RP. Viral expression of insulin-like growth factor-I enhances muscle hypertrophy in resistance-trained rats. *J Appl Physiol* (1985) 96: 1097-1104, 2004.
44. Loos RJ, Hagberg JM, Pérusse L, Roth SM, Sarzynski MA, Wolfarth B, et al. Advances in exercise, fitness, and performance genomics in 2014. *Med Sci Sports Exerc* 47: 1105-1112, 2015.
45. Lv ZT, Gao ST, Cheng P, Liang S, Yu SY, Yang Q, et al. Association between polymorphism rs12722 in COL5A1 and musculoskeletal soft tissue injuries: a systematic review and meta-analysis. *Oncotarget* 9: 15365-15374, 2018.
46. Ma F, Yang Y, Li X, Zhou F, Gao C, Li M, et al. The association of sport performance with ACE and ACTN3 genetic polymorphisms: a systematic review and meta-analysis. *PLoS One* 8: e54685, 2013.
47. Malfait F and De Paepe A. Molecular genetics in classic Ehlers-Danlos syndrome. *Am J Med Genet C Semin Med Genet* 139c: 17-23, 2005.
48. Malfait F, Wenstrup RJ, and De Paepe A. Clinical and genetic aspects of Ehlers-Danlos syndrome, classic type. *Genet Med* 12: 597-605, 2010.
49. Marouli E, Graff M, Medina-Gomez C, Lo KS, Wood AR, Kjaer TR, et al. Rare and low-frequency coding variants alter human adult height. *Nature* 542: 186-190, 2017.
50. Mills M, Yang N, Weinberger R, Vander Woude DL, Beggs AH, Eastell S, et al. Differential expression of the actin-binding proteins, alpha-actinin-2 and -3, in different species: implications for the evolution of functional redundancy. *Hum Mol Genet* 10: 1335-1346, 2001.
51. Montero D and Lundby C. Refuting the myth of non-response to exercise training: 'non-responders' do respond to higher dose of training. *J Physiol* 595: 3377-3387, 2017.
52. Montgomery HE, Marshall R, Hemingway H, Myerson S, Clarkson P, Dollery C, et al. Human gene for physical performance. *Nature* 393: 221-222, 1998.
53. Moran CN, Yang N, Bailey ME, Tsiokanos A, Jamurtas A, MacArthur DG, et al. Association analysis of the ACTN3 R577X polymorphism and complex quantitative body composition and performance phenotypes in adolescent Greeks. *Eur J Hum Genet* 15: 88-93, 2007.
54. NHGRI-EBI. GWAS catalog - studies with available summary statistics. [cited 2020 8th April]; Available from: <https://www.ebi.ac.uk/gwas/downloads/summary-statistics>.
55. NIH. Results From First Human Gene Therapy Clinical Trial. 1995 [cited 2020 8th April]; Available from: <https://www.genome.gov/10000521/1995-release-first-human-gene-therapy-results>.

56. North KN, Yang N, Wattanasirichaigoon D, Mills M, Easteal S, and Beggs AH. A common nonsense mutation results in alpha-actinin-3 deficiency in the general population. *Nat Genet* 21: 353-354, 1999.
57. Pabalan N, Tharabenjasin P, Phababpha S, and Jarjanazi H. Association of COL5A1 gene polymorphisms and risk of tendon-ligament injuries among Caucasians: a meta-analysis. *Sports Med Open* 4: 46, 2018.
58. Pérusse L, Rankinen T, Hagberg JM, Loos RJ, Roth SM, Sarzynski MA, et al. Advances in exercise, fitness, and performance genomics in 2012. *Med Sci Sports Exerc* 45: 824-831, 2013.
59. Pérusse L, Rankinen T, Rauramaa R, Rivera MA, Wolfarth B, and Bouchard C. The human gene map for performance and health-related fitness phenotypes: the 2002 update. *Med Sci Sports Exerc* 35: 1248-1264, 2003.
60. Phillips BE, Kelly BM, Lilja M, Ponce-González JG, Brogan RJ, Morris DL, et al. A Practical and Time-Efficient High-Intensity Interval Training Program Modifies Cardio-Metabolic Risk Factors in Adults with Risk Factors for Type II Diabetes. *Front Endocrinol (Lausanne)* 8: 229, 2017.
61. Puthucherry Z, Skipworth JR, Rawal J, Loosemore M, Van Someren K, and Montgomery HE. Genetic influences in sport and physical performance. *Sports Med* 41: 845-859, 2011.
62. Rankinen T, Bray MS, Hagberg JM, Pérusse L, Roth SM, Wolfarth B, et al. The human gene map for performance and health-related fitness phenotypes: the 2005 update. *Med Sci Sports Exerc* 38: 1863-1888, 2006.
63. Rankinen T, Fuku N, Wolfarth B, Wang G, Sarzynski MA, Alexeev DG, et al. No Evidence of a Common DNA Variant Profile Specific to World Class Endurance Athletes. *PLoS One* 11: e0147330, 2016.
64. Rankinen T, Pérusse L, Rauramaa R, Rivera MA, Wolfarth B, and Bouchard C. The human gene map for performance and health-related fitness phenotypes. *Med Sci Sports Exerc* 33: 855-867, 2001.
65. Rankinen T, Pérusse L, Rauramaa R, Rivera MA, Wolfarth B, and Bouchard C. The human gene map for performance and health-related fitness phenotypes: the 2001 update. *Med Sci Sports Exerc* 34: 1219-1233, 2002.
66. Rankinen T, Pérusse L, Rauramaa R, Rivera MA, Wolfarth B, and Bouchard C. The human gene map for performance and health-related fitness phenotypes: the 2003 update. *Med Sci Sports Exerc* 36: 1451-1469, 2004.
67. Rankinen T, Roth SM, Bray MS, Loos R, Pérusse L, Wolfarth B, et al. Advances in exercise, fitness, and performance genomics. *Med Sci Sports Exerc* 42: 835-846, 2010.
68. Rigat B, Hubert C, Alhenc-Gelas F, Cambien F, Corvol P, and Soubrier F. An insertion/deletion polymorphism in the angiotensin I-converting enzyme gene accounting for half the variance of serum enzyme levels. *J Clin Invest* 86: 1343-1346, 1990.
69. Romualdi C, Balding D, Nasidze IS, Risch G, Robichaux M, Sherry ST, et al. Patterns of Human Diversity, within and among Continents, Inferred from Biallelic DNA Polymorphisms. *Genome Res* 12: 602-612, 2002.
70. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, et al. Genetic Structure of Human Populations. *Science* 298: 2381-2385, 2002.
71. Roth SM, Rankinen T, Hagberg JM, Loos RJ, Pérusse L, Sarzynski MA, et al. Advances in exercise, fitness, and performance genomics in 2011. *Med Sci Sports Exerc* 44: 809-817, 2012.

72. Sarzynski MA, Davidsen PK, Sung YJ, Hesselink MK, Schrauwen P, Rice TK, et al. Genomic and transcriptomic predictors of triglyceride response to regular exercise. *Br J Sports Med* 49: 1524-1531, 2015.
73. Sarzynski MA, Ghosh S, and Bouchardcorresponding C. Genomic and transcriptomic predictors of response levels to endurance exercise training. *J Physiol* 595: 2931-2939, 2017.
74. Sarzynski MA, Loos RJ, Lucia A, Pérusse L, Roth SM, Wolfarth B, et al. Advances in Exercise, Fitness, and Performance Genomics in 2015. *Med Sci Sports Exerc* 48: 1906-1916, 2016.
75. September AV, Cook J, Handley CJ, van der Merwe L, Schwellnus MP, and Collins M. Variants within the COL5A1 gene are associated with Achilles tendinopathy in two populations. *Br J Sports Med* 43: 357-365, 2009.
76. Seto JT, Quinlan KG, Lek M, Zheng XF, Garton F, MacArthur DG, et al. ACTN3 genotype influences muscle performance through the regulation of calcineurin signaling. *J Clin Invest* 123: 4255-4263, 2013.
77. Timmons JA, Knudsen S, Rankinen T, Koch LG, Sarzynski M, Jensen T, et al. Using molecular classification to predict gains in maximal aerobic capacity following endurance exercise training in humans. *J Appl Physiol (1985)* 108: 1487-1496, 2010.
78. Tucker R and Collins M. What makes champions? A review of the relative contribution of genes and training to sporting success. *Br J Sports Med* 46: 555-561, 2012.
79. Visscher PM, Brown MA, McCarthy MI, and Yang J. Five years of GWAS discovery. *Am J Hum Genet* 90: 7-24, 2012.
80. WADA. Gene Doping. 2005 [cited 2020 8th April]; Available from: https://www.wada-ama.org/sites/default/files/resources/files/PlayTrue_2005_1_Gene_Doping_EN.pdf.
81. Wenstrup RJ, Florer JB, Brunskill EW, Bell SM, Chervoneva I, and Birk DE. Type V collagen controls the initiation of collagen fibril assembly. *J Biol Chem* 279: 53331-53337, 2004.
82. White D and Rabago-Smith M. Genotype-phenotype associations and human eye color. *J Hum Genet* 56: 5-7, 2011.
83. Wolfarth B, Bray MS, Hagberg JM, Pérusse L, Rauramaa R, Rivera MA, et al. The human gene map for performance and health-related fitness phenotypes: the 2004 update. *Med Sci Sports Exerc* 37: 881-903, 2005.
84. Wolfarth B, Rankinen T, Hagberg JM, Loos RJ, Pérusse L, Roth SM, et al. Advances in exercise, fitness, and performance genomics in 2013. *Med Sci Sports Exerc* 46: 851-859, 2014.
85. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* 46: 1173-1186, 2014.
86. Yang N, MacArthur DG, Gulbin JP, Hahn AG, Beggs AH, Eastaugh S, et al. ACTN3 genotype is associated with human elite athletic performance. *Am J Hum Genet* 73: 627-631, 2003.
87. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum Mol Genet* 27: 3641-3649, 2018.
88. Zhang S. A Biohacker Regrets Publicly Injecting Himself With CRISPR. 2018 [cited 2020 8th April]; Available from:

<https://www.theatlantic.com/science/archive/2018/02/biohacking-stunts-crispr/553511/>.