COURSE NOTES: Understanding genetics for improving health outcomes

Course Code: CEUGH

Instructor Name: Dr. Rahul Kushwah

Phone Number: (647) 889 6916

Instructor Email: rahulkushwah@anantlife.com

Course Description

The course is intended to provide the tools that holistic nutritionists can use to deliver personalized healthcare to their clients, catered to the genetic makeup of the client. Genetics plays a very robust role in nutrition, detoxification, weightloss and overall health and wellness. Furthermore, genetics can be used not only to improve the health of the clients but also to develop DNA based nutritional plans that can potentially prevent development of chronic diseases. Course participants will gain knowledge on how genetic information can be used to deliver nutritional plans, weightloss strategies, detoxification plans, hormonal balance plans along with nutritional plans to prevent development of chronic disease such as diabetes and cardiovascular diseases.

IHN has partnered with Anantlife Canada Inc., a leader in clinical grade genetic testing for healthcare providers all over the world, to offer a Certified Genetic Testing Provider Certificate upon successful completion of the course. Successful completion of the course implies that the candidates have received the education and training to not only understand genetic concepts pertaining to diet, nutrition, detoxification, fitness, hormonal health and metabolic disorders but have also been trained on interpretation of the genetic testing reports along with development of a DNA based health plan for better health outcomes.

SESSION 3:

FITNESS GENETICS AND GENETICS OF CHRONIC DISEASES: UNDERSTANDING HOW OUR GENES REGULATE OUR FITNESS AND HOW OUR GENES REGULATE THE RISK OF CHRONIC DISEASES – APPLICATIONS IN DEVELOPING PERSONALIZED HEALTH PLANS TO IMPROVE HEALTH

Human metabolism as well as response to different exercises including potential of performing various fitness exercises is regulated by genes and as such genetic differences among individuals impacts how their body reacts to different fitness regimens. Furthermore, chronic metabolic diseases such as diabetes and cardiovascular issues are controlled by genes involved in several metabolic processes and mutations in these genes play a role in determining one's predisposition to these chronic diseases. The readings herein are to provide an extensive understanding exercise and fitness genetics along with how genetics plays a role in predisposition to chronic metabolic disorders.

BMJ Open Sport & Exercise Medicine

Exercise genetics: seeking clarity from noise

Craig Pickering,[1,2] John Kiely[1]

To cite: Pickering C, Kiely J. Exercise genetics: seeking clarity from noise. *BMJ Open Sport & Exercise Medicine* 2017;**3**:e000309. doi:10.1136/bmjsem-2017-000309

Accepted 3 November 2017

[1]School of Sport and Wellbeing, Institute of Coaching and Performance, University of Central Lancashire, Preston, UK
[2]Exercise and Nutritional Genomics Research Centre, DNAFit Ltd, London, UK

Correspondence to
Mr Craig Pickering; craig@dnafit.com

The potential of recent advances in genetics research to supplement elite sport decision-making has potentially extensive implications, but remains highly controversial. One potential application is the use of genetic information to enhance exercise prescription, thereby positively influencing athletic performance and public health domains. Recent research suggests that this is both feasible and potentially beneficial.[1 2] However, such an effective use of genetic information requires a clear understanding of the mechanism by which each reported single nucleotide polymorphism (SNP) mediates physical performance. In the absence of such a clear, mechanistic explanation, we are left with vague associations without causative roots. While uncovering gene associations is necessary, it is not sufficient to presume causation. Given the complex entangled routes through which genes and environment interact to express phenotype, superficial association-based logical inferences are likely to be misleading.

ASSOCIATION OR CAUSATION?

As an example, within the HEalth, RIsk factors, exercise Training And GEnetics (HERITAGE) Family Study, variation in *CREB1* (rs2253206) predicted heart rate (HR) response to exercise.[3] Specifically, the A allele associated with a smaller reduction in HR during a submaximal exercise test following training, with the proposed mechanism relating to long-term cardiac memory. However, research in a separate cohort associated the A allele with a greater exercise-induced temperature increase— contributing to a less pleasant subjective experience of exercise, potentially reducing motivation to train or carry out an aerobic test.[4] Accordingly, it is unclear whether HR responsiveness was modified via biologically mediated adaptations or an increased perception of effort.

Similarly, a SNP within *COL5A1*—rs12722—has been linked to exercise-associated muscle cramps (EAMC), with the CC genotype associated with protection from EAMC during an ultra-marathon.[5] However, CC genotypes also recorded significantly slower ultra-marathon times compared with TT genotypes.[5] Does this genetic variation directly protect against EAMC or does it result in slower race times, which, given that EAMC is associated with increased neuromuscular fatigue, is what acts in a protective manner? Again, the biological impact of this SNP on EAMC is not clear, requiring more evidence before advice can be given.

ARE THESE RELATIONSHIPS CONSISTENT?

In addition to resolving the biological mechanisms underpinning the impact of genetic variation on exercise, we must also consider whether these genetic associations are consistent over time and across different cohorts. Much is made of non-responders to exercise, and yet is unclear whether this non-response is consistent, or whether it is a one-time response to an intervention. In addition, it is unclear whether SNPs associated with exercise response in sedentary individuals have similar effects in trained subjects. An SNP in *ACSL1*, rs6552828, had the strongest association with training-induced VO_{2max} improvements in HERITAGE,[6] a sedentary cohort. However, in an elite athlete cohort, there was no association between this SNP and elite endurance status (a proxy of high VO_{2max}) in Caucasians.[7] No further *ACSL1* replications exist. Does variation in *ACSL1* impact exercise adaptation in all humans or only the subset of humans who took part in HERITAGE? If HERITAGE were to be repeated with the same subjects, would the *ACSL1* and aerobic fitness association remain constant? Does this variation affect trained and untrained subjects to the same extent? Answers to these questions are needed before these SNPs should be used to modify the training process.

EFFECTIVE UTILISATION

Despite these issues, there are a number of SNPs in which the biological mechanisms are well understood. A common SNP in *ACTN3*, the gene that encodes for α-actinin-3, a protein found exclusively in fast-twitch muscle fibres, results in a premature stop

codon. Individuals homozygous for this polymorphism are unable to produce the protein, and as a result tend to have fewer fast-twitch fibres.[8] This in turn affects the response to strength training.[9] The utilisation of this information holds promise; a recent paper used this SNP in conjunction with 14 others to enhance resistance training response,[2] and evidence-based guidelines have been proposed.[10] This underscores both the effectiveness and the utility of genetic information in informing training methodologies *when* the biological mechanism is well understood.

SUMMARY

Research into the genetics of exercise adaptation is both exciting and promising. As each SNP exerts its influence potentially through a multitude of pathways, some identified gene–trait associations may be spurious. Conceptual clarity therefore requires that the causative mechanisms directly linking genotype to phenotype are more clearly deciphered; simply revealing associations are insufficient when the aim is to better inform practice.

Perspectives on the promise of exercise genetics vary widely, with polarised extremes of staunch advocates and deniers. For the majority, the complex relationship between genotype and phenotype promotes a healthy scepticism; nevertheless, a total rejection of the potential utility of gene panels to categorise adaptive subtypes, given promising preliminary findings,[1 2 9 10] is premature. Beyond a formulaic statement of the obvious—that correlation is not causation—it seems wise to proceed cautiously, sceptically, but with an open mind as more evidence unfolds.

Contributors CP conceived the idea for this manuscript and wrote the first draft. JK provided feedback on the initial idea, first draft and rewrote large parts of final draft. Both authors gave final approval for the version to be published.

Competing interests CP is an employee of DNAFit Ltd. He received no financial incentive for the production of this manuscript which was produced as part of his Professional Doctorate studies.

Provenance and peer review Not commissioned; internally peer reviewed.

REFERENCES

1 Timmons JA, Knudsen S, Rankinen T, *et al*. Using molecular classification to predict gains in maximal aerobic capacity following endurance exercise training in humans. *J Appl Physiol* 2010;108:1487–96.
2 Jones N, Kiely J, Suraci B, *et al*. A genetic-based algorithm for personalized resistance training. *Biol Sport* 2016;33:117–26.
3 Rankinen T, Argyropoulos G, Rice T, *et al*. CREB1 is a strong genetic predictor of the variation in exercise heart rate response to regular exercise: the HERITAGE Family Study. *Circ Cardiovasc Genet* 2010;3:294–9.
4 Karoly HC, Stevens CJ, Magnan RE, *et al*. Genetic influences on physiological and subjective responses to an aerobic exercise session among sedentary adults. *J Cancer Epidemiol* 2012;2012:1–12.
5. O'Connell K, Posthumus M, Schwellnus MP, *et al*. Collagen genes and exercise-associated muscle cramping. *Clin J Sport Med* 2013;23:64–9.
6. Bouchard C, Sarzynski MA, Rice TK, *et al*. Genomic predictors of the maximal O□ uptake response to standardized exercise training programs. *J Appl Physiol* 2011;110:1160–70.
7 Yvert T, He ZH, Santiago C, *et al*. Acyl coenzyme A synthetase long-chain 1 (ACSL1) gene polymorphism (rs6552828) and elite endurance athletic status: a replication study. *PLoS One* 2012;7:e41268.
8 Vincent B, De Bock K, Ramaekers M, *et al*. ACTN3 (R577X) genotype is associated with fiber type distribution. *Physiol Genomics* 2007;32:58–63.
9 Delmonico MJ, Kostek MC, Doldo NA, *et al*. Alpha-actinin-3 (ACTN3) R577X polymorphism influences knee extensor peak power response to strength training in older men and women. *J Gerontol A Biol Sci Med Sci* 2007;62:206–12.
10 Kikuchi N, Nakazato K. Effective utilization of genetic information for athletes and coaches: focus on ACTN3 R577X polymorphism. *J Exerc Nutrition Biochem* 2015;19:157–64.

Genetics and sport performance: current challenges and directions to the future

CDD. 20.ed. 575.1
796.092
796.5

João Paulo Limongi França GUILHERME*
Aline Cristina Capparelli TRITTO*
Kathryn Nance NORTH**
Antonio Herbert LANCHA JUNIOR*
Guilherme Giannini ARTIOLI*

*Escola de Educação Física e Esporte, Universidade de São Paulo.
**Institute for Neuroscience and Muscle Research, University of Sidney - Austrália.

Abstract

In recent years there has been a great progress in molecular biology techniques, which has facilitated the researches on influence of genetics on human performance. There are specific regions of DNA that can vary between individuals. Such variations (i.e., polymorphisms) may, in part, explain why some individuals have differentiated responses to certain stimuli, including the responses to sports training. In a particular sport, the presence of specific polymorphisms may contribute to high levels of performance. Since 1998, several polymorphisms have been associated with athletic phenotypes; however the accumulation of information generated over these 15 years shows that the influence of genetics to sport is extremely complex. In this review, we will summarise the current status of the field, discussing the implications of available knowledge for the practice of professionals involved with the sport and suggesting future directions for research. We also discuss topics related to the importance of polygenic profile characterization of athletes, methods for the identification of new polymorphisms associated with physical performance, the use of genetic testing for predicting competitive success, and how crucial is the genetic profile for the success athletes in competition.

Key words: Genetic; Polymorphism; Athletes; Sports; Performance.

Introduction

The determinants of human athletic performance have long been a challenging field of study in sport sciences. Sports performance is an enormously complex multifactorial phenomenon, and is determined by numerous intrinsic (e.g., genetics, motor behavior, physiological and psychological profile) and extrinsic factors (e.g., training, nutrition, development opportunities and overall health conditions) as well as by the interaction between them[1]. Although it is impossible to set a unique formula to make anyone becoming a successful athlete, it is widely accepted that any individual who is highly committed and dedicated to training is able to improve athletic performance. Likewise, to be a top-level athlete, several years of dedication to an organized and rigid training system is indeed a prerequisite, although not a guarantee of success. However, a few athletes seem to be exceptionally gifted and demonstrate extraordinarily high performance levels even before taking part in training programs; some athletes demonstrate better responses to training than others, or may be able to consistently sustain high levels of performance over their competitive career[2-3].

Despite the awareness of the genetic influences on competitive success, genetics of sports performance is a quite recent area of investigation. As a consequence, the currently available knowledge is largely incipient and some authors consider we are at infant stages of the area[4]. Hence, every effort aiming at improving our understanding on this phenomenon is of great importance.

The earlier studies on genetics of human performance were focused on estimating the heritability of different complex traits. Using approaches such as twin-studies

and familial aggregation studies, investigators were able to estimate the percentage contribution of genetic factors to muscle fibre type distribution and enzyme activities[5-6], bone density and muscle strength[7], aerobic[8] and anaerobic capacities[9], among other performance-relevant variables. Although unquestionably relevant, these studies did not provide specific information on which particular genes and genetic variants would be involved in such genetic influences. The first polymorphism related to sport performance (i.e., angiotensin-conversion enzyme, ACE) was not identified until 1998 [10-11].

Over the past 15 years, the advances in biotechnology and molecular biology tools have facilitated a rapid increase in the identification of structural genetic variations capable of exerting some influence on the phenotypes related to athletic performance[12]. Association studies made possible the establishment of a human gene map for exercise[13] and constant updates reflect the rapid increase in the number of polymorphisms that entered in this map[14-15]. To date, more than 200 polymorphisms have been associated with some feature related to physical exercise[15], and it is expected that this number will increase in the following years[15]. However, only about 20 of these >200 genetic variants were specifically observed in athletes[4]. Furthermore, most of these genes and variants have failed to confirm association in replication studies[16], so that less than 10 genetic variants have been consistently associated with sports performance[17]. Considering that the human genome has over 20,000 genes and that each gene may present an enormous diversity of common variants that could theoretically influence some performance-related phenotype, it is extremely likely that our currently knowledge represents only a small fraction of the genetic factors that influence sports performance. Hence, numerous new genetic variants are yet to be discovered, and we still barely understand how genes interact with each other and with environmental factors.

High-level sports performance is an extremely complex phenotype and genetic background is only one of its multiple contributory factors. It is likely that the contribution of heritability to a particular phenotype will largely depend on the specific sport discipline, among other factors. Even if only the genetic factors are considered, sports success remains an extremely complex phenomenon because it is a multigenic trait[18].

Understanding the influence of genetics in sport is a critical step to unravel the determinants of sports excellence. However, the challenges are enormous and the experimental approaches to address this highly complex and multifaceted phenotype are somewhat limited. In this critical review, we discuss the potential and limitations of research methods on genetics of sports performance, the implications of this knowledge in "real-world" sport settings, as well as directions for future researchs.

Basic concepts

For the benefit of reader, the terms and concepts most important for understanding this discussion are briefly revised below.

Genome refers to the entire collection of genetic material hereditarily transmitted to the next generation that a given species possesses[19]. In the case of most eukaryotes, including humans, the genome is encoded in DNA sequences. These sequences are composed by four different nitrogen bases (Adenine, Thymine, Cytosine and Guanidine - A, T, C and G, respectively) that, by binding in sequence, make a single DNA strand (e.g., AACGGT is a sequence of nitrogen bases forming a single-stranded DNA). Each nitrogen base also binds to its complimentary base (i.e., A binds to T and C binds to G) so that a single DNA strand is attached to a complimentary single DNA strand, forming a double-stranded DNA. The DNA molecules are predominantly found inside the nucleus, although a small portion is also found inside mitochondria. Inside the nucleus, the genome is organized in 23 different pairs of chromosomes.

The human genome has over 3 billion nitrogen base pairs. Of this total, only ~5% are encoding regions. A gene is considered a specific region of the genome whose DNA sequence encodes a biologically active product, which is, in most cases, a RNA molecule that ultimately results in a protein. Every individual has two copies of each gene, which are called alleles. In the coding sequences, a sequence of 3 nucleotides encodes for 1 specific amino acid, which will take place in the peptide chain. These 3-nucleotide sequences encoding for 1 amino acid are called codons.

Although all human beings share all the same genes, they display some slight structural variations in

their nitrogen base sequences. It is estimated that only 0.1% of the genome varies between individuals[20]. However, this minor portion of the genome explains the enormous phenotypic diversity that exists among humans. These variations may occur as: 1) changes in a single base pair (for example, in the sequence AAC***G***GT the nucleotide G is swapped by a nucleotide A, so the variant sequence is AAC***A***GT); 2) deletions of a single base pair (for example, in the sequence AAC***G***GT, the nucleotide G is deleted, so the variant sequence is AACGT); 3) insertions of single base pairs (for example, in the sequence AACGGT, the nucleotide C is inserted, so the variant sequence is AAC***C***GT); or 4) changes, deletions or insertions in two or more base pairs. The most common type of variation in the human genome consists in a change in a single base pair[21]. This type of variant can be a single nucleotide polymorphism (SNP) or a point mutation, depending on how prevalent it is in the population and on the impact it has on phenotype (see further discussion).

Despite being non-coding sequences, 95% of the genome do play fundamental physiological roles, especially by regulating the rate of genes expression (i.e., whether or not a gene will be transcribed into RNA to produce a protein, and how much RNA will be produced from its corresponding gene). Therefore, genetic variants may encompass both coding and non-coding regions of the genome[22]. Genetic variants that occur in coding regions may affect the sequence of amino acids in a protein and, depending on the type of variant, the protein structure can be slightly or severely affected. Obviously, the more a protein structure is affected, the greater its physiological impact tends to be. The variant protein may be less functional or even not functional at all. On the other hand, when genetic variants occur in non-coding regions of the genome, the protein structure is normally unaffected and its physiological impact tends to be less pronounced. In these cases, it is more likely that the rate of the gene expression is affected.

Some genetic variations are rare and some are common. When a variant appears in less than 1% of the population, it is considered a mutation; when its frequency in the population is greater than 1%, it is considered a polymorphism[19]. Normally, a mutation has a greater impact on physiological functions than a polymorphism[23]. As a consequence, mutations tend to have some health impact whereas polymorphisms tend to account for normal phenotypic variation. However, there are cases of rare mutations that do not lead to a disease, as well as there are common polymorphisms that increase the likelihood of an individual to develop a disease. Whether these variants should be considered mutations or polymorphisms is still a matter under debate and are not under the scope of this review.

As discussed earlier, there are two copies of each gene in the genome[a]. One allele is found in a specific region of a specific chromosome whilst the other allele (which is exactly the same gene but not formed by exactly the same sequence) is found in the same region of the homologous chromosome. Considering a given variation in a given gene, an individual may have one or two copies of the most frequent variant (which is often referred as the "normal" copy) and/or one or two copies of the least frequent variant (which is often referred as the polymorphic or mutated copy). Therefore, for this specific variation, the genotype of an individual can be: 1) homozygous (two alleles of a "normal" copy of the gene); 2) heterozygous (one allele of a "normal" copy and the other of a polymorphic copy of the gene); or 3) homozygous (two alleles of a polymorphic or mutated copy of the gene).

Phenotypic traits are observable characteristics controlled by genes. Thus, a given genotype affects a given phenotype to some extent. Some traits are controlled by one single gene, and they are referred to as monogenic traits. Normally, it is relatively easy to establish a link between a genotype and a phenotype in monogenic traits since they obey the Mendelian logic of inheritance. On the other hand, polygenic traits are far more complex because they are influenced by several genes, as well as by multiple non-genetic environmental factors. Due to its multifactorial nature, it is normally difficult to establish a strong association between one single genetic variant and a complex phenotype, which often imposes a hurdle to the studies attempting to identify specific genes that influence a complex phenotype. This explains, at least in part, why the associations between genetic polymorphisms and athletic performance are normally weak and frequently not confirmed in replication studies.

Genetic influences on quantitative traits and sports performance

Sports performance is an extremely complex phenotypic trait, which is in turn influenced, although not determined, by many other traits, such as muscle fibre type distribution, aerobic power and capacity, anaerobic power and capacity, and trainability of physical capacities[24].

Most traits that are relevant to sports performance are quantitative, meaning that they are possible to be measured and quantified. Some examples of quantitative traits that are relevant to physical performance are: body composition, aerobic power and muscle strength. In some cases, the final outcome of sport performance can also be a quantitative trait. For examples, swimming distance times, running races, jumps, throws and all other sports in which final performance is quantifiable can be considered quantitative traits. In other cases, however, sports performance "per se" is not a quantitative trait. This is the case of unpredictable sports, such as team sports, individual sports that depend on natures' conditions (e.g., surfing and sailing) and individual sports that depend on opponents' actions (e.g., combat sports). Theoretically, some performance-relevant quantitative traits are strongly influence by genetic factors, which is also the case of some "predictable sports" (FIGURE 1). On the other hand, other traits as well as "unpredictable sports" are less influenced by genetic factors (FIGURE 1) and, therefore, genotype-phenotype relationships are less likely to be established. This must be kept in mind when performing association studies, as latter discussed in this review.

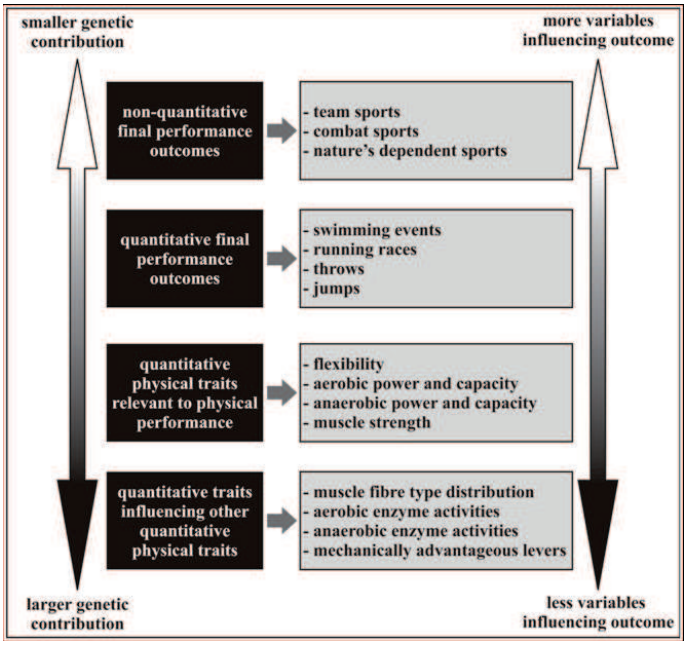


FIGURE 1 - Contribution of genetic factors to performance-relevant quantitative traits.

Experimental approaches for studying genetics of sport performance

Over the past 15 years, there has been a great effort to identify, at the molecular level, variations in DNA sequence that may contribute to sports performance or to any trait that is relevant to sport performance, so-called genotype-phenotype correlations. However, in complex polygenic multifactorial traits, genotype-phenotype correlations are often elusive and difficult to be clearly identified.

Candidate genes association studies

One of the most frequent experimental approaches for assessing genotype-phenotype correlations is the genetic association. In genetic association studies, a candidate polymorphism is correlated with a performance-relevant trait. For example, the frequency of a candidate polymorphism is compared between two highly distinct populations: elite athletes and non-athletes. If the polymorphism is significantly more frequent in the athletic group, it is assumed that this polymorphism is associated with athletic status and contributes to elite performance. In general, a polymorphism is considered a candidate based on the physiological role of the gene and on how the difference in nucleotide sequence affects gene function and/or expression.

Association studies can be divided into three main categories: 1) case-control studies, which compare the frequency of genotypes in a cohort of controls (non-athletes) and a cohort of elite athletes; 2) cross-sectional studies, which compare selected physiological and/or performance data between different genotypes[25] and 3) longitudinal studies, in which responses to a given intervention (e.g., exercise training or diet) are compared between genotype groups. All approaches are important to demonstrate the relevance of a genetic variant to performance[26]. However, polymorphisms emerging from association studies remain as "candidate" genetic variants until the association is replicated in other independent cohorts and a plausible biological explanation for the impact of the polymorphism is formulated[25].

Case-control association studies are relatively cost-effective and easy to be performed, especially when a large number of DNA samples from top-level athletes and controls is readily available, which makes this approach interesting to initially screen potential candidate polymorphisms. In contrast, association studies do not provide cause-effect relationships, meaning that establishing an association between a candidate polymorphism and elite athletic status is not sufficient to accept a candidate polymorphism as valid. Thus, providing further evidence on the influence of that polymorphism on physical performance it is extremely important. This evidence can be produced using the cross-sectional and longitudinal prospective approaches detailed in the previous paragraph, as well as by determining a physiological role of the polymorphism on sports-related phenotypes. The critical steps for producing compelling evidence on the role of a genetic variant in enhancing sports performance are schematised in FIGURE 2.

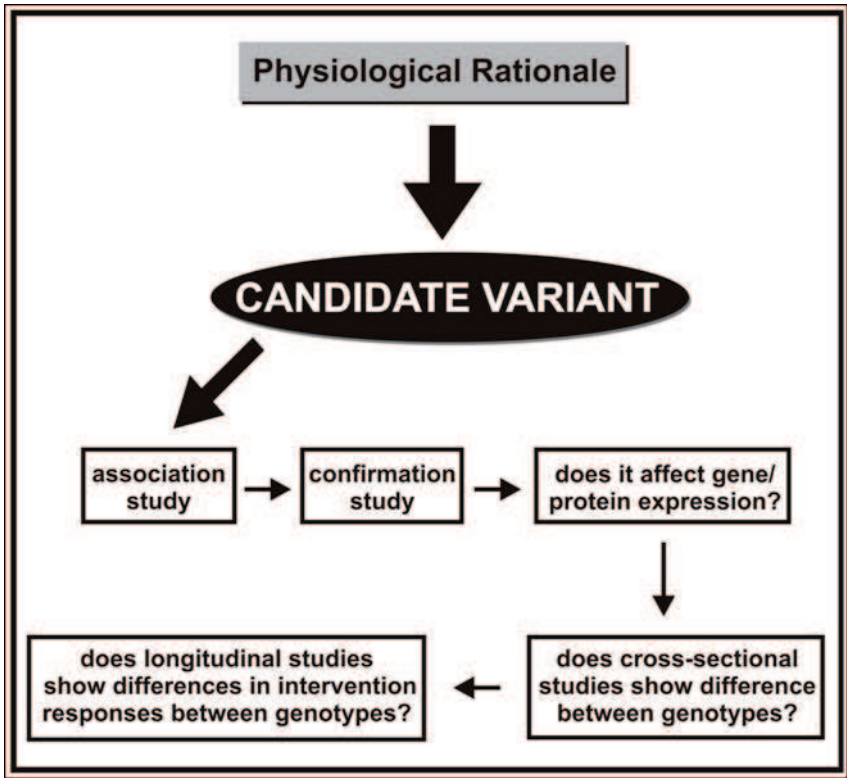


FIGURE 2 - Critical steps and experimental approaches necessary to validate a candidate gene as relevant to sports performance.

Although relatively cost-effective and straightforward, association studies often face some obstacles that may hamper researchers to securely draw conclusions from the data obtained. To circumvent these problems, some measures are recommended when performing association studies.

In some cases, a particular polymorphism correlates with athletic phenotypes in a group of individuals from a specific region and, later, the same association is observed in a different set of individuals from a distinct region[27-28]. This positive replication (i.e., a consistent association) strengthens the evidence for the influence of the polymorphism on phenotype. However, even if the replication occurs in more than one population, it does not mean that the same association will be found in every population of the world[29]. In fact, associations found in a study are frequently not replicated in subsequent studies[30-33]. Depending on the characteristics of the studies that showed the association (e.g., sample size, ethnic background and homogeneity of athletic cohort in terms of competitive level and sports disciplines), such inconsistencies may be interpreted as: 1) the polymorphism is not relevant to physical performance; 2) the polymorphism is relevant in a population with a specific ethnic background; 3) the polymorphism is relevant for some specific sports disciplines.

According to the ethnic background of the studies cohort, the frequency of each genotype can vary dramatically[33-34]. Therefore, all polymorphisms reported in the literature should be replicated in different populations[4,16]. It is possible that some polymorphisms are relevant to performance only in some specific regions or under some specific conditions, whereas other polymorphisms may have a more "universal" effect.

Genome-wide studies

One major disadvantage of candidate-gene studies is that only one or a few (in the case of polygenic profile studies) genes can be assessed at a time. In view of this, new methods allowing the screening of the whole genome were developed.

Genome-wide linkage studies (GWLS) were the first approach to analyse genetic markers across the entire genome. This method identifies chromosomal regions that harbour genes affecting quantitative traits over generations (i.e., it identifies quantitative trait loci)[24]. GWLS have been used to discover QTL associated with a variety of diseases and other phenotypes[25].

More recently, technological advances originated another technique, the so-called genome-wide association study (GWAS), capable of identifying genes, rather than genomic loci, associated with a phenotype[35]. Unlike GWLS, which requires familial data (the basic unit of observation is a pair of parents, usually brothers), GWAS studies analyse individual data[24]. This new approach is becoming increasingly popular in the search for variations that contribute to complex traits[24].

While candidate gene studies are driven by the theoretical impact that a variant would have on physical performance, GWAS do not make any prior assumption regarding genes and variants involved with physical performance[2]. Due to our limited ability to select candidate polymorphisms for association with performance based only on available theory, the design of candidate-genes association studies will be always restricted to a certain degree[24]. In this sense, the fact that GWAS studies are "theory-free" and that polymorphism selection is based on observational data makes this approach more robust. This increases the chances of finding new and perhaps unexpected genes and variants affecting physical performance, which might open new venues for investigation.

Despite being considered more robust than candidate genes studies, GWAS are very expensive and, therefore, not widely used in sports sciences or not used in truly large athletic cohorts[35]. In 2008, the cost of a run containing two human genomes (30x coverage - Illumina sequencing machine - HiSeq 2000) was around US $60.000 [36]. Even though this value has been falling and in 2011, the price of a similar run was around US $10.000 [36], meaning that a very large study with adequate statistical power would reach extremely high costs.

Because of the multivariate nature of GWAS, the p value must be times lower than usual to accept the correlation as significant (normally $5x\ 10^{-8}$ rather than 0.05) [2,37], which obviously require a very large sample size to achieve a desirable statistical power. By their very nature, cohorts of fairly homogenous top-elite athletes are small. Thus, reaching desirable statistical power in GWAS studies can be an enormous challenge in sports sciences. The study by Pitsiladis and Wang[36] describes a good example on how sample size can affect GWAS outcomes, reporting two studies designed to quantify the heritability of stature, a very stable and easy to be measured variable, using the GWAS approach. In one study, a set of ~30,000 individuals was analysed,

and it was found that less than 5% of the variation in the phenotype is explained by genetics. In contrast, the other study using a much larger set of individuals (n ≅ 180.000) showed that genetics explained 10% of the variance in height.

Likewise, the number of SNPs included in the analysis affects the GWAS statistical power. The greater the number of SNPs, the greater the number of comparisons carried out by statistical analysis, so the value accepted to consider statistical significance is reduced[16]. Another limitation of GWAS is that, despite screening for genetic variants along the entire genome, only SNPs are detected in this analysis, meaning that not all types of genetic variations will be captured[4,24], such as copy number variants.

A massive amount of information is generated by one GWAS. However, as with any other association study, the results of GWAS can be vague or uninformative[38]. Within the same study, the following situations can be possible: 1) a significant association is found between a SNP and a phenotype, and the result is replicated in other cohorts (true-positive); 2) a significant association between the SNP and the phenotype is found, but replication studies fail to confirm it in other cohorts (false-positive) ; 3) GWAS was not sensitive enough to detect associations which were already confirmed in previous studies (lost-results)[38]. It has been shown that non-replicating results and inconsistency in the magnitude of results (heterogeneity) are commons flaws in GWAS[38].

In the last 5 years, several investigations were carried out using GWAS for complex phenotypes, mainly pathologies[35]. In 2011, Bouchard et al.[39] conducted the first study applying GWAS in a group of individuals (non-athletes) before and after a training period. They examined the association of SNPs with the VO_{2max} responses to physical training. The authors managed to perform the study in a quite large sample (> 1000 subjects) and to compare the results in different cohorts. The study successfully identified genetic variants associated with training responses, and then it was possible to construct a panel of 21 SNPs, which accounted for 49% of the variance in VO_{2max} responses ($p < 0.05$). To date, there are no studies using the GWAS approach in a cohort of athletes. Although it is acceptable that access to large cohorts of truly elite athletes represents a tremendous barrier to high-quality genome-wide studies in this population, a collective effort from research groups around world is indeed necessary to undertake this extremely relevant type of study.

The role of sample size

The number of participants is probably the most important limitation of genetic association studies[26], especially when referring to elite athletes[40]. Because performance-relevant polymorphisms are normally found in low frequencies in a given population[12], reduced sample sizes will probably return very small numbers of individuals presenting the rare genotype. Each polymorphism exhibits distinct frequencies in different populations, so the sample size necessary to detect an association can vary according to the polymorphism and to the population analysed[17,41]. Sample size is directly related to the statistical power of the study, and a reduced number of participants can hamper the drawing of firm conclusions[4]. This limitation becomes even more evident when multiple comparisons enter into statistical model (e.g., in GWAS and polygenic studies), which makes the analysis more rigid and lowers the set level of significance[16].

Some authors claim that this limitation is justifiable since there is a very limited number of high-level athletes in most regions and countries. This makes the collection of a sufficiently large cohort almost impossible and it explains why studies with elite athletes usually assess a small number of individuals (i.e., $n > 100$)[42]. Indeed, this is a very strong and truthful argument and, regrettably, no much can be done to enlarge athletic cohorts, especially in countries where competitive sport is less prominent. However, researchers in this area should endeavour to maximize the number of athletes in their cohorts. In fact, some authors advocate that the research groups around the world should create an international consortium with DNA samples from worldwide elite athletes[4,16]. This is probably the best way to circumvent this relevant limitation in sample size. However, researchers should be conscious to the fact that this procedure could result in heterogeneous groups regarding both athletic and ethnic backgrounds, which would add extra confounding variables to the analysis.

Alternatively, increasing the number of participants in the non-athletic (control) group can be an appealing manner to minimize problems with low sample sizes and underpowered studies. For example, by running a statistical simulation, we observe that a group of athletes (n = 100; frequency of the rare allele = 30%) when compared to a control group (n = 100; frequency of the rare allele = 20%), no significant differences are observed for the allele frequency between groups ($x^2 = 2.16$, $p = 0.14$). However, as shown in FIGURE 3, even if the number of athletes

remains unchanged (n = 100; frequency of the rare allele = 30%), it is possible to detect a significant difference between groups by increasing the number of controls to 325 (frequency of the rare allele remains 20%) ($x^2 = 3.85$, $p = 0.0497$). As displayed in FIGURE 3, successive increases in the number of controls are paralleled by increases in statistical power. However, at a certain point, further large increases in sample size result in merely slight increases in power. In view of this, researchers are advised to increase the number of controls to a maximum in their analysis. Possibly, a control sample size between 1000 and 1500 will yield a desirably good statistical power for the majority of the situations.

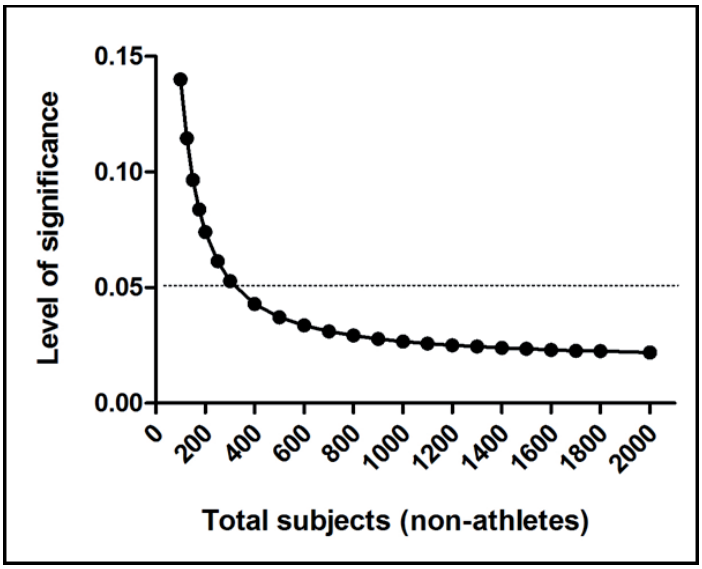


FIGURE 3 - Level of significance found considering a fixed sample of athletes and varying the amount of non-athletic subjects (controls).

The importance of making comparisons between homogeneous groups

It is widely accepted that each sport discipline has its particular physiological, psychological, anthropometrical and biomechanical demands, which directly influence the characteristics that would most contribute to competitive success[43-45]. In fact, it has been considered that the polymorphisms influencing endurance sports are different from the polymorphisms influencing sprint sports, which display a greater demand for muscle strength and power[46]. Considering that some (if not all) of these characteristics are in part genetically determined, studies comparing genotypes between athletes should carefully select which sports disciplines will be included in a given group when categorizing athletes.

Most studies have been categorized athletes into two opposing groups, namely sprint/power-orientated sports vs. endurance-orientated sports[31,47]. Although this is an interesting approach, since it would group together sports under the influence of the same genes and polymorphisms, the inclusion of sports in the same group that are not really analogous can represent a major flaw in the analysis. For example, football, 200-m swimming and powerlifting could be classified as power/sprint-orientated sports, even though the determinants of these disciplines are clearly different. Even nearly identical sprint/power-orientated sports could have quite different demands, such as 50-m and 400-m swimming. In view of this, it becomes evident that researchers should wisely include in the same group only athletes who compete in truly comparable disciplines.

The study by Muniesa et al.[48] compared the distribution of 6 polymorphisms between endurance runners (> 5000 meters) and endurance cyclists, sports that are usually clustered on the same group due to their similar metabolic demand. Interestingly, the frequency of the I and D alleles (indel polymorphism in the ACE gene) was different between these groups. Despite the similarities regarding the metabolic demands, factors such as mechanical efficiency and movement economy may have influenced the differential role of the ACE polymorphism in these sports[48].

Another matter of concern is how to categorize athletes according to competitive levels. Ideally, a polymorphism that exerts some influence on sports performance should be capable of differentiating not

only athletes from non-athletes, but also elite athletes from non-elite athletes. Thus, the criteria utilised to categorise athletes into competitive groups can be also confounding factors in association studies. Usually, "high-elite" athletes are those who participate in Olympic Games or World Championships; "elite" are those who participate in international-level competitions (e.g., continental championships), "sub-elite" participate at the national-level; and "non-elite" in state- and regional-level competitions. Although slightly different classifications have been used in literature[40], this is an easy and straightforward manner to classify athletes according to competitive levels. However, researchers from different countries should be cautious in relation to how a particular sport is developed in that country. For example, judo is a very popular and developed sport in Brazil, whilst rugby is not. On the other hand, judo is not a popular and developed sport in New Zealand, whereas rugby is probably the most prominent sport. In this context, reaching national judo level in Brazil is probably more difficult than reaching international high-elite levels in judo in New Zealand. Likewise, a world-cup rugby player in Brazil would probably have less of the necessary characteristics to succeed in rugby than a national-level New Zealand athlete. Because of these national disparities, authors should ponder the most appropriate way to classify the athletes from theirs cohorts taking into consideration how the specific sport is developed.

Single gene approach vs. polygenic profile

Most studies on genetics of sports performance have assessed the impact of only one variant on athletic performance[46]. Due to the multifactorial nature of sports performance, the effect of a single variant is most likely to be small[49]. According to Flueck et al.[50], one important drawback of the single gene approach in sports sciences is that the selected gene may only marginally contribute to performance, or it may not be the "bottleneck" of performance, since many other genes may compensate the altered function of the polymorphic gene. The regulation of every physiological system involved in exercise performance and training responses is dependent on a complex network of interconnected genes[51]. Thus, it is assumed that the contribution of genetics to physical performance relies on the combined action of different genes and, consequently, a number of different polymorphisms accounts to the variation in sports excellence and training responses.

Despite their limitations, single gene studies are important to first identify polymorphisms associated with performance, especially when it is not possible to perform GWAS. However, because one single polymorphism would only account for a minor part of the total variation, a combination of polymorphisms (i.e., a polygenic profile) would provide a better model to explain how changes at the molecular level in DNA would affect the phenotype.

The concept of polygenic profile is now becoming more solid. It was recently shown that healthy individuals who possess more alleles associated with aerobic metabolism present better responses to aerobic training (according to VO_{2max} responses to training)[39]. Likewise, endurance athletes with a greater amount of alleles already associated with endurance performance have a greater chance of being successful in endurance-orientated sports[52]. Therefore, it is becoming increasingly clear that high-level athletes are individuals genetically distinguished, meaning that they probably present a combination of numerous polymorphic alleles associated with physical performance[12].

In 2008, Williams and Folland[17] proposed the calculation of the total score genotyping (TGS). To calculate TGS, the polymorphisms that have been previously associated with a target phenotype must be selected. Each allele receives a score, which ranges from 0 to 2, according to the existing genotype. Homozygous genotypes associated with the phenotype receive the score "2", while heterozygotes receive the score "1". Homozygous for the allele not associated with performance returns the score "zero". After determining the scores for each polymorphism, these are summed and the result is expressed in a scale ranging from 0 to 100, so that the TGS is obtained. Using this score, it is possible to assess the balance between several selected different polymorphisms that are knowingly relevant to performance[41].

From 2009 to the present day, the TGS has been used by some research groups and has proved to be a sensitive tool to differentiate endurance from strength/power athletes[31,47,53-54]. Additionally, the use of TGS has revealed that high-elite athletes have a polygenic profile significantly different from the general population, which probably makes athletes more favourable to sports success[18].

The TGS assumes a dose-response effect, i.e., the more associated alleles an athlete has, the better his/her genetic profile for sports success. Hence, it is assumed that there is an additive effect

of polymorphisms[4]. Moreover, the sensibility of the TGS to differentiate individuals with different genetic predisposition to excel in a given type of sport seems to be dependent on the number of polymorphisms included in the calculation[18].

The choice of which polymorphisms are included in the TGS calculation should be made in light of the characteristics of the studied population. For example, if a study aims to determine the polygenic profile of sprinters, polymorphisms associated with muscle size and strength, as well as anaerobic energy metabolism and other sprint-orientated phenotypes should be inserted in the formula[41]. Besides, the choice of the polymorphisms should include only those truly consistently associated with the phenotype. Otherwise, the TGS will lack power in differentiating athletes between non-athletes or sprinters endurance athletes[17,41].

A potential limitation of the TGS is the fact that it considers all polymorphisms as influencing athletic performance to the same extent[4]. With the currently available knowledge, it is impossible to determine the exact weight that each polymorphism should have in TGS calculation, simply because it is not yet determined the weight of each polymorphism in the regulation of the different sports-related phenotypes[55].

The number of genes that are modulated in response to physical training is relatively large. In contrast, the scientific literature shows a relatively limited number of polymorphisms associated with physical performance in athletes, i.e., about 20 genes[17,41]. The scores so far published have included 6-10 polymorphisms in the TGS calculation, which clearly represents only a small fraction of the genes that putatively affect sports-related phenotypes[52,56]. As a consequence, the optimal combination of polymorphisms that would result in an ideal TGS model is yet to be determined[49]. However, this does not preclude the TGS to be successfully used to distinguish endurance athletes from strength/power athletes and from non-athletes. Furthermore, the TGS will probably be updated and optimized in the upcoming years as the knowledge on single genes associated with performance will improve[42]. One important caveat, however, is that as the number of polymorphisms is included in the calculation increases, the likelihood of finding an individual with optimum genotypes decreases exponentially[17,41], which could cause TGS to lose sensitivity. Thus, an ideal model of TGS will include only the genes that most contribute to that specific trait. The exact number and the most important genes for each phenotype are fundamental questions that researchers should address in the future.

Identification of new candidate polymorphisms

To choose a new polymorphism that will be analysed in an association study, it is first necessary to choose a candidate gene to be explored. A convenient way to choose a candidate gene is to observe which genes are modulated during or after exercise [57]. If a particular gene is consistently modulated, it would be worthy to search for structural variations in this gene that could influence phenotypes related to physical performance[58].

Recently, Timmons et al.[57] screened, through the use of microarray, the mRNA expression in "vastus lateralis" muscle of healthy subjects who underwent 20 weeks of aerobic training. Afterwards, the authors searched for polymorphism in the genes whose expression accounted for the variation in VO_{2max} responses. Six SNPs that may account for the variability in the cardiorespiratory response to aerobic training were identified.

Although useful, microarrays are somewhat expensive and time-consuming. The enormous amount of data generated might also represent another obstacle to the use of this technique. In order to identify new candidate polymorphisms influencing athletic performance, alternative methods other than microarrays can be employed. A considerable amount of information on mRNA expression and exercise is available in literature and potential genes are waiting to be explored[59].

After choosing one candidate gene to be explored, an investigator should examine all genetic variations described for the gene, which can easily be done with online public databases (e.g., NCBI and UCSC)[4]. For the vast majority of the genes, several structural variations will be found, including single nucleotide, *indel* and copy number polymorphisms. The selection of the best candidates should consider previously available data. Nonetheless, multiple polymorphisms for each gene should be tested in order to verify which one presents the best correlation with the phenotype[25].

There is a natural tendency in sports research to study genes involved in motor activities. However, more recent studies indicate that genes involved

in psychological characteristics may also influence athletic performance[60-61]. Some prime examples are the genes *5HTT, BDNF* and *UCP2*[62] and future studies should start focusing on the psychological aspects implicated in sports performance when selecting new candidate genes.

Strategies to identify physiological roles of a polymorphism

As previously discussed, unravelling the underlying physiological mechanisms by which a genetic variant affect performance is crucial to definitively associate that variant to performance. However, this is often a very laborious and elusive task. Nevertheless, a few experimental approaches might be of great use for this purpose.

Knowing whether the structural DNA variant affects or not gene and protein expression at the tissue level is probably the most fundamental question. If the polymorphism severely affects protein structure (e.g., a non-sense or a frameshift polymorphism), it is usually easier to observe its impact on physiology and on phenotype. In fact, the best-characterized polymorphism impacting physical performance is the R577X in the *ACTN3* gene, which is a non-sense polymorphism leading to the synthesis of a non-functional protein[63]. This is because the absence of the alpha-actinin-3 protein in humans carrying both polymorphic alleles can be mimicked in knockout animals, providing a very interesting model to evaluate how the lack of the protein impacts muscle structure, muscle metabolism and performance[64]. However, most polymorphisms are SNPs, indel or copy number variants, and they may occur at non-coding genomic regions. Consequently, their impact on protein structure is less evident and the creation of a good animal model is often unfeasible. In these cases, human studies are indeed necessary.

The knowledge of the tissues and cells where the polymorphic gene is expressed should dictate the best approach to examine how the polymorphism affects gene and protein expression. For example, considering a variant in a gene that expresses exclusively in skeletal muscle, then muscle biopsies should be collected from a group of individuals with all genotypes. These samples could be submitted to a variety of analysis, including gene expression (e.g., qPCR), protein expression (e.g., western blotting), and morphological inspection (e.g., light or electron microscopy). Further analysis should also be carried out depending on the role of the gene. Exemplifying, if the gene encodes for an enzyme, then comparing enzyme activity between genotypes is probably a very reasonable approach. Obviously, in some cases human tissue collection is not an option, especially if the gene is expressed in heart, bone, kidney, liver brain or any other organ where a biopsy is too invasive and unjustifiable. This would imply the necessity for alternative and more indirect approaches. These scenarios illustrate quite well how difficult establishing a physiological link between a polymorphism and a phenotype can be.

In cases that an animal model can be generated so the genetic variation in humans is properly mimicked, acceptable sample sizes for these animal studies are substantially low, since all other sources of variation are controlled. In contrast, when human studies are performed and groups of individual with different genotypes are compared, sample size becomes critical. Because studies with humans, especially the cross-sectional ones, normally have poor control of major intervenient variables (e.g., diet, exercise background, genetic background, use of medications and development conditions), it is imperative that all genotypes have a large enough number of participants in order to minimize the chances of groups being different due to confounding factors rather than to genotype itself. In this case, the importance of large samples is not only about statistical power, but to conveniently decrease the likelihood of assuming that genotype groups are different when the cause of the difference is not the genotype "per se".

Genetic interactions

A still unexplored, yet interesting and promising area of research in sports genetics is how genes interact with other genes to modulate exercise responses or to modulate a phenotype that is strongly related to performance[4]. It is well recognised that genes do not act in isolation. Rather, there are

complex interactions among many genes modulating phenotypes[42]. Some polymorphisms might not have a meaningful impact on athletic performance alone, but the presence of other polymorphisms, through gene-gene interactions, may enlarge its impact upon phenotype, so that not a presence of a single polymorphism but, instead, the combination of some specific polymorphisms may be potentially beneficial to some performance- related phenotypes[4].

In addition to the interactions between genes, gene-environment interactions must be taken into consideration in future research, especially in light of the multifactorial nature of sports performance. After certain environmental stimuli, some polymorphisms may have greater or lesser influence on the responses to those stimuli[26]. The replication of the results in different cohorts, under different environmental stimuli, is an indirect way to assess this interaction. The researchers' primary efforts have been focused on the discovery of novel polymorphisms with the capability to influence physical fitness components. Subsequently establishing a panorama for one or more polymorphisms, possible gene-gene and gene-environment interactions should be investigated.

Rare variants

It is well accepted that the influence of common genetic variants on performance is slight, but the combination of several polymorphisms probably is more influential[16]. On the other hand, it is highly plausible that rare genetic variants or mutations may eventually confer a very significant advantage to performance. Rare variants, therefore, are probably a very important aspect of the genetic variability that underlies sports excellence. However, due to the extreme difficulty to identify and characterize these variants, this is still poorly understood[15].

In most instances, mutations lead to loss of function, disease or disabilities, which adversely affects sports-related phenotypes[24]. A good example is the mutation found in the *PYGM* gene that results in a deficit in carbohydrate metabolism (i.e., McArdle's disease) and, as a consequence, the patient presents intolerance to physical exercise. Many other examples of mutations that would cause incompatibility with athletic phenotypes could be provided. However, not all mutations are detrimental to physical performance. The notorious case of a mutation resulting in the absence of the myostatin protein illustrates how a rare variant can represent an extreme advantage to some performance-related phenotypes without causing any harm to health[65]. It must be noted, however, that it is currently unknown whether this specific mutation in the myostatin gene is in fact favourable to physical performance.

Because each rare variant will be carried by no more than a few individuals, identifying these rare variants and correlate them to a specific sports-related phenotype is not an easy task. Nonetheless, efforts should be made in order to identify new rare variants and broaden our knowledge about the genetics of the sports excellence. Maybe performing a genome-wide scan of individuals presenting abnormally high phenotypes such as muscle mass, strength, flexibility and aerobic capacity and extraordinarily excellent performance in sports is a reasonable approach to search for new rare variants.

The use of genetic markers to detect sports talent

The molecular mechanisms influencing athletic performance take place and are regulated by genes; these are gradually being revealed. Some authors argue that testing for the presence of key genetic variants in youth can be used as a way to select potential athletes[61]. The early identification of potential elite athletes could, in theory, optimize training plans and competitive stimuli during growth and development, therefore increasing the chances of reaching the peak of physical performance[61]. This process of identifying talent by means of polymorphisms (genetic testing) could, in principle, be revolutionary to the field of sport[25].

The publication of Yang et al.[66] describing the influence of the polymorphism R577X in *ACTN3* gene on athletic performance had a major impact on the scientific community and in general population. Some authors[67] have stated that the association of *ACTN3* with the physical performance is sufficient for using this polymorphism as a tool to select potential athletes. In 2004, just one year after

the publication of Yang and colleague's paper, an Australian company and after, in 2008, an American company began to develop and sell, at a cost of US $170, claiming that the genetic test would determine to which sport the person would have a higher "vocation"[68]. Currently, there are at least 7 companies that sell genetic tests, now with a broader range of polymorphisms being offered[61].

Despite the strong appeal of these genetic tests, the information provided by them is useless to anyone who seeks the competitive career in any sport. As discussed throughout this paper, sports performance is multifactorial. Genetics comprises only one of many contributing factors. Moreover, the genes and variants that may have a positive effect on performance are numerous, all of them under the regulation of an extremely complex network of other genes and variants, among other factors. Therefore, the presence or the absence of a few polymorphisms will never have such a predictive power. The literature presents many good examples of very successful athletes who, nevertheless, presented a genetic profile that could be considered "unfavourable". Regarding the *ACTN3*, which is one of the best-characterized polymorphisms[69], there are athletes achieving great results in strength activities that do not display the RR genotype[70].

The Council of Europe Bioethics Convention and the Genetic Information Non-Discrimination Act in the US consider the use of the genetic testing to predict performance an unethical action and they are currently evaluating whether they should be banned[71]. Besides the absence of scientific validity and the lack of useful information, three other important ethical issues must be highlighted in relation to information misuse: 1) the autonomy of the individual to not to know their genetic information, since some polymorphisms associated physical performance may also be associated with diseases; 2) invasion of privacy, as a personal information can be passed to other people and used in a discriminative way; 3) professional misconduct and diminished opportunities, as coaches and trainers may not be willing to invest their time and efforts in an individual without "genetic predisposition" to sports[61].

With the intention to find gifted individuals at early stages, the major target of genetic testing would be children and youths. Some authors point that fanatic relatives could take their children to train very hard, if they believe that the children have more changes based on results of genetic testing[72]. Knowingly, excessive training loads at young ages could represent a health harm, and result in overtraining, burnout and sport abandonment[73]. It is important to note that, as highlighted by Hawley[74], the athlete who knows your genetic information may feel discouraged to keep practicing, especially if the information is negative for his/her expectation.

A potential use of genetic testing is related to the individualized prescription of exercise[26]. Several researchers argue that genetics should be used as an aid for improving exercise prescription and optimizing training responses in non-athletes[75-80]. Although some authors have already published training recommendations according to athletes' genotype[67], it must be noted that there is no currently available information to support any genotype-directed training for athletes. In fact, there is no study addressing how genotypes influence training responses in athletes. Although in principle the proposal of genetic tests is revolutionary, at the moment, we still have no evidence supporting the creation of any model with acceptable predictive value. At present, the use of genetic tests has no advantage over the traditional methods for talent identification already used by trainers[25].

Conclusion

The future of genetic studies involving athletes is promising. In recent years, many polymorphisms have been associated with athletic phenotypes, but definitive confirmation of association and the underlying physiological mechanisms are proven difficult tasks. The challenges to progress in this novel area are enormous, but a variety of experimental approaches can be used to unravel part of the mystery. Researchers and the general population should be conscious about the implications of the misuse of the genetic information. While some people may claim that genetic information could be used to detect talent and to drive athletic development, it must be noted that there is no scientific evidences for the predictive value of genetic in sports. The most appropriate statement at the moment is that genetics is only one out of many contributing factors to the athletic performance, and sometimes it may play only

secondary roles. It will be a long way until we know exactly what is the role of genetics for each sport and which are, at the molecular level, the variants accounting for this and how they work.

Note

a. In fact, some genes are present in more than one copy per haploid genome (considering only the 23 non-redundant chromosomes), which means that there are more than two copies of these genes per diploid genome (considering the entire 46 chromosomes).

References

1. Kiss MAPDM, Böhme MTS, Mansoldo AC, Degaki E, Regazzini M. Performance and sports talent. Rev Paul Educ Fís. 2004;19(n.Esp):89-100.
2. Gibson WT. Key concepts in human genetics: understanding the complex phenotype. Med Sport Sci. 2009;54:1-10.
3. Tucker R, Collins M. What makes champions? A review of the relative contribution of genes and training to sporting success. Br J Sports Med. 2012;46:555-61.
4. Eynon N, Ruiz JR, Oliveira J, Duarte JA, Birk R, Lucia A. Genes and elite athletes: a roadmap for future research. J Physiol. 2011;589:3063-70.
5. Komi PV, Viitasalo JH, Havu M, Thorstensson A, Sjodin B, Karlsson J. Skeletal muscle fibres and muscle enzyme activities in monozygous and dizygous twins of both sexes. Acta Physiol Scand. 1977;100:385-92.
6. Bouchard C, Simoneau JA, Lortie G, Boulay MR, Marcotte M, Thibault MC. Genetic effects in human skeletal muscle fiber type distribution and enzyme activities. Can J Physiol Pharmacol. 1986;64:1245-51.
7. Arden NK, Spector TD. Genetic influences on muscle strength, lean body mass, and bone mineral density: a twin study. J Bone Miner Res. 1997;12:2076-81.
8. Bouchard C, An P, Rice T, et al. Familial aggregation of VO(2max) response to exercise training: results from the HERITAGE Family Study. J Appl Physiol. 1999;87:1003-8.
9. Calvo M, Rodas G, Vallejo M, et al. Heritability of explosive power and anaerobic capacity in humans. Eur J Appl Physiol. 2002;86:218-25.
10. Gayagay G, Yu B, Hambly B, et al. Elite endurance athletes and the ACE I allele-the role of genes in athletic performance. Hum Genet. 1998;103:48-50.
11. Montgomery HE, Marshall R, Hemingway H, et al. Human gene for physical performance. Nature. 1998;393:221-2.
12. Ostrander EA, Huson HJ, Ostrander GK. Genetics of athletic performance. Annu Rev Genomics Hum Genet. 2009;10:407-29.
13. Bray MS, Hagberg JM, Perusse L, et al. The human gene map for performance and health-related fitness phenotypes: the 2006-2007 update. Med Sci Sports Exerc. 2009;41:35-73.
14. Hagberg JM, Rankinen T, Loos RJ, et al. Advances in exercise, fitness, and performance genomics in 2010. Med Sci Sports Exerc. 2011;43:743-52.
15. Roth SM, Rankinen T, Hagberg JM, et al. Advances in exercise, fitness, and performance genomics in 2011. Med Sci Sports Exerc. 2012;44:809-17.
16. Bouchard C. Overcoming barriers to progress in exercise genomics. Exerc Sport Sci Rev. 2011;39:212-7.
17. Williams AG, Folland JP. Similarity of polygenic profiles limits the potential for elite human physical performance. J Physiol. 2008;586:113-21.
18. Lucia A, Moran M, Zihong H, Ruiz JR. Elite athletes: are the genes the champions? Int J Sports Physiol Perform. 2010;5:98-102.
19. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Molecular biology of the cell. New York: Garland Science; 2008.
20. IHGSC. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature. 2004;431:931-45.
21. Salisbury BA, Pungliya M, Choi JY, Jiang R, Sun XJ, Stephens JC. SNP and haplotype variation in the human genome. Mutat Res. 2003;526:53-61.

22. Mills RE, Pittard WS, Mullaney JM, et al. Natural genetic variation caused by small insertions and deletions in the human genome. Genome Res. 2011;21:830-9.
23. Gonzaga-Jauregui C, Lupski JR, Gibbs RA. Human genome sequencing in health and disease. Annu Rev Med. 2012;63:35-61.
24. Bouchard C, Rankinen T, Timmons JA. Genomics and genetics in the biology of adaptation to exercise. Compr Physiol. 2011;1:1603-48.
25. MacArthur DG, North KN. Genes and human elite athletic performance. Hum Genet. 2005;116:331-9.
26. Rankinen T, Bouchard C. Gene-exercise interactions. Prog Mol Biol Transl Sci. 2012;108:447-60.
27. Lucia A, Gomez-Gallego F, Barroso I, et al. PPARGC1A genotype (Gly482Ser) predicts exceptional endurance capacity in European men. J Appl Physiol. 2005;99:344-8.
28. Maciejewska A, Sawczuk M, Cieszczyk P, Mozhayskaya IA, Ahmetov II. The PPARGC1A gene Gly482Ser in Polish and Russian athletes. J Sports Sci. 2012;30:101-13.
29. He Z, Hu Y, Feng L, et al. Is there an association between PPARGC1A genotypes and endurance capacity in Chinese men? Scand J Med Sci Sports. 2008;18:195-204.
30. Eynon N, Ruiz JR, Meckel Y, et al. Is the -174 C/G polymorphism of the IL6 gene associated with elite power performance? A replication study with two different Caucasian cohorts. Exp Psychol. 2010;96:156-62.
31. Ruiz JR, Eynon N, Meckel Y, et al. GNB3 C825T Polymorphism and elite athletic status: a replication study with two ethnic groups. Int J Sports Med. 2010;32:151-3.
32. He ZH, Hu Y, Li YC, et al. Are calcineurin genes associated with athletic status? A function, replication study. Med Sci Sports Exerc. 2011;43:1433-40.
33. Puthucheary Z, Skipworth JR, Rawal J, Loosemore M, Van Someren K, Montgomery HE. The ACE gene and human performance: 12 years on. Sports Med. 2011;41:433-48.
34. Ash GI, Scott RA, Deason M, et al. No association between ACE gene variation and endurance athlete status in Ethiopians. Med Sci Sports Exerc. 2011;43:590-7.
35. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. Am J Hum Genet. 2012;90:7-24.
36. Pitsiladis Y, Wang G. Necessary advances in exercise genomics and likely pitfalls. J Appl Physiol. 2011;110:1150-1.
37. McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet. 2008;9:356-69.
38. Ioannidis JP. Non-replication and inconsistency in the genome-wide association setting. Hum Hered. 2007;64:203-13.
39. Bouchard C, Sarzynski MA, Rice TK, et al. Genomic predictors of the maximal O(2) uptake response to standardized exercise training programs. J Appl Physiol. 2011;110:1160-70.
40. Puthucheary Z, Skipworth JR, Rawal J, Loosemore M, Van Someren K, Montgomery HE. Genetic influences in sport and physical performance. Sports Med. 2011;41:845-59.
41. Hughes DC, Day SH, Ahmetov II, Williams AG. Genetics of muscle strength and power: polygenic profile similarity limits skeletal muscle performance. J Sports Sci. 2011;29:1425-34.
42. Ruiz JR, Arteta D, Buxens A, et al. Can we identify a power-oriented polygenic profile? J Appl Physiol. 2010;108:561-6.
43. Artioli GG, Gualano B, Franchini E, Batista RN, Polacow VO, Lancha Junior AH. Physiological, performance, and nutritional profile of the Brazilian Olympic Wushu (kung-fu) team. J Strength Cond Res. 2009;23:20-5.
44. Franchini E, Del Vecchio FB, Matsushigue KA, Artioli GG. Physiological profiles of elite judo athletes. Sports Med. 2011;41:147-66.
45. Alves CR, Pasqua L, Artioli GG, et al. Anthropometric, physiological, performance, and nutritional profile of the Brazil National Canoe Polo Team. J Sports Sci. 2012;30:305-11.
46. Buxens A, Ruiz JR, Arteta D, et al. Can we predict top-level sports performance in power vs endurance events? A genetic approach. Scand J Med Sci Sports. 2011;21:570-9.
47. Ruiz JR, Gomez-Gallego F, Santiago C, et al. Is there an optimum endurance polygenic profile? J Physiol. 2009;587:1527-34.
48. Muniesa CA, Gonzalez-Freire M, Santiago C, et al. World-class performance in lightweight rowing: is it genetically influenced? A comparison with cyclists, runners and non-athletes. Br J Sports Med. 2010;44:898-901.
49. Gonzalez-Freire M, Santiago C, Verde Z, et al. Unique among unique. Is it genetically determined? Br J Sports Med. 2009;43:307-9.
50. Flueck M, Vaughan D, Westerblad H. Linking genes with exercise: where is the cut-off? Eur J Appl Physiol. 2010;110:1095-8.
51. Hoppeler H, Baum O, Lurman G, Mueller M. Molecular mechanisms of muscle plasticity with exercise. Compr Physiol. 2011;1:1383-412.

52. Aahmetov II, Williams AG, Popov DV, et al. The combined impact of metabolic gene polymorphisms on elite endurance athlete status and related phenotypes. Hum Genet. 2009;126:751-61.
53. Santiago C, Ruiz JR, Muniesa CA, Gonzalez-Freire M, Gomez-Gallego F, Lucia A. Does the polygenic profile determine the potential for becoming a world-class athlete? Insights from the sport of rowing. Scand J Med Sci Sports. 2010;20:e188-94.
54. Eynon N, Ruiz JR, Meckel Y, Moran M, Lucia A. Mitochondrial biogenesis related endurance genotype score and sports performance in athletes. Mitochondrion. 2011;11:64-9.
55. Dias RG. Genetics, human physical performance and gene doping: common sense versus the scientific reality. Rev Bras Med Esporte. 2011;17:62-70.
56. Gomez-Gallego F, Santiago C, Gonzalez-Freire M, et al. Endurance performance: genes or gene combinations? Int J Sports Med. 2009;30:66-72.
57. Timmons JA, Knudsen S, Rankinen T, et al. Using molecular classification to predict gains in maximal aerobic capacity following endurance exercise training in humans. J Appl Physiol. 2010;108:1487-96.
58. Keller P, Vollaard NB, Gustafsson T, et al. A transcriptional map of the impact of endurance exercise training on skeletal muscle phenotype. J Appl Physioly. 2010;110:46-59.
59. Coffey VG, Hawley JA. The molecular bases of training adaptation. Sports Med. 2007;37:737-63.
60. Pearson, H. Physiology: freaks of nature? Nature. 2006;444:1000-1.
61. Roth SM. Critical overview of applications of genetic testing in sport talent identification. Recent Pat DNA Gene Seq. 2012:6:247-55.
62. Lippi G, Longo UG, Maffulli N. Genetics and sports. Br Med Bull. 2009;93:27-47.
63. Yang N, Garton F, North K. alpha-actinin-3 and performance. Med Sport Sci. 2009;54:88-101.
64. MacArthur DG, Seto JT, Chan S, et al. An Actn3 knockout mouse provides mechanistic insights into the association between alpha-actinin-3 deficiency and human athletic performance. Hum Mol Genet. 2008;17:1076-86.
65. Schuelke M, Wagner KR, Stolz LE, et al. Myostatin mutation associated with gross muscle hypertrophy in a child. N Engl J Med. 2004;350:2682-8.
66. Yang N, MacArthur DG, Gulbin JP, et al. ACTN3 genotype is associated with human elite athletic performance. Am J Hum Genet. 2003;73:627-31.
67. Kambouris M, Ntalouka F, Ziogas G, Maffulli N. Predictive genomics DNA profiling for athletic performance. Recent Pat DNA Gene Seq. 2012;6:229-39.
68. Brooks MA, Tarini BA. Genetic testing and youth sports. JAMA. 2011;305:1033-4.
69. Pasqua LA, Artioli GG, Pires FO, Bertuzzi R. ACTN3 gene and sports performance: a candidate gene to success in short and long duration events. Rev Bras Cineantropom Desempenho Hum. 2011;13:477-83.
70. Lucia A, Olivan J, Gomez-Gallego F, Santiago C, Montil M, Foster C. Citius and longius (faster and longer) with no alpha-actinin-3 in skeletal muscles? Br J Sports Med. 2007;41:616-7.
71. McNamee MJ, Muller A, Van Hilvoorde I, Holm S. Genetic testing and sports medicine ethics. Sports Med. 2009;39:339-44.
72. Collier R. Testing the ethics of genetic testing in sports. CMAJ. 2012;184:E45-6.
73. Brenner JS. Overuse injuries, overtraining, and burnout in child and adolescent athletes. Pediatrics. 2007;119:1242-5.
74. Hawley JA. Commentary on viewpoint: perspective on the future use of genomics in exercise prescription. J Appl Physiol. 2008;104:1253.
75. Lightfoot JT. Commentary on viewpoint: perspective on the future use of genomics in exercise prescription. J Appl Physiol. 2008;104:1249.
76. Pescatello LS. Commentary on viewpoint: perspective on the future use of genomics in exercise prescription. J Appl Physiol. 2008;104:1247.
77. Roth SM. Perspective on the future use of genomics in exercise prescription. J Appl Physiol. 2008:104:1243-5.
78. Thomis M. Commentary on viewpoint: perspective on the future use of genomics in exercise prescription. J Appl Physiol. 2008;104:1251.
79. Timmons JA. Commentary on viewpoint: perspective on the future use of genomics in exercise prescription. J Appl Physiol. 2008;104:1250.
80. Wagner PD. Commentary on viewpoint: perspective on the future use of genomics in exercise prescription. J Appl Physiol. 2008;104:1246.

ENDEREÇO
João Paulo Limongi França Guilherme
Escola de Educação Física e Esporte - USP
Av. Prof. Mello Moraes, 65
05508-030 - São Paulo - SP - BRASIL
e-mail: jplfguilherme@usp.br

Recebido para publicação: 15/01/2013
Aceito: 11/02/2013

Genetics and Diabetes

Background

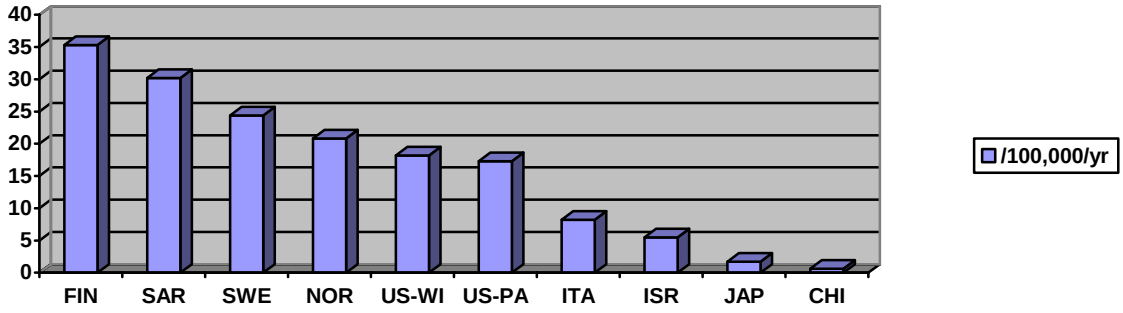
Diabetes mellitus is a heterogeneous group of disorders characterized by persistent hyperglycemia. The two most common forms of diabetes are type 1 diabetes (T1D, previously known as insulin-dependent diabetes or IDDM) and type 2 diabetes (T2D, previously known as non-insulin-dependent diabetes or NIDDM). Both are caused by a combination of genetic and environmental risk factors. However, there are other rare forms of diabetes that are directly inherited. These include maturity onset diabetes in the young (MODY), and diabetes due to mutations in mitochondrial DNA.

All forms of diabetes have very serious effects on health. In addition to the consequences of abnormal metabolism of glucose (e.g., hyperlipidemia, glycosylation of proteins, etc.), there are a number of long-term complications associated with the disease. These include cardiovascular, peripheral vascular, ocular, neurologic and renal abnormalities, which are responsible for morbidity, disability and premature death in young adults. Furthermore, the disease is associated with reproductive complications causing problems for both mothers and their children. Although improved glycemic control may decrease the risk of developing these complications, diabetes remains a very significant cause of social, psychological and financial burdens in populations worldwide.

Type 1 Diabetes

Epidemiology**.** T1D is caused by the autoimmune destruction of the beta cells of the pancreas, and represents approximately 10% of all cases with diabetes. At present, lifelong insulin therapy is the only treatment for the disease. Without exogenous insulin injections, individuals with T1D will not survive. Although the prevalence of T1D is <1% in most populations, the geographic variation in incidence is enormous, ranging from <1/100,000 per year in China to approximately 40/100,000 per year in Finland (Figure 1) (Karvonen et al., 1993). The only chronic childhood disorder more prevalent than T1D is asthma. It has been estimated that approximately 20 million people worldwide, mostly children and young adults, have T1D (Holt, 2004).

Figure 1. T1D Incidence Rates Worldwide



FIN = Finland, SAR = Sardinia, SWE = Sweden, NOR = Norway, US-WI = US-Wisconsin, US-PA = US-Pennsylvania, ITA = Italy, ISR = Israel, JAP = Japan, CHI = China

The incidence of T1D is increasing worldwide at a rate of about 3% per year (Onkamo et al., 1999). This trend appears to be most dramatic in the youngest age groups, and is completely unrelated to the current increase in T2D in children. More children with beta cell autoantibodies, a hallmark of T1D,

are being diagnosed with the T1D around the world each year. Although the peak age at onset is at puberty, T1D can also develop in adults. Epidemiologic studies have revealed no significant gender differences in incidence among individuals diagnosed before age 15 (Kyvik et al., 2004). However, after age 25, the male to female incidence ratio is approximately 1.5. There is also a notable seasonal variation in the incidence of T1D in many countries, with lower rates in the warm summer months, and higher rates during the cold winter (Dorman et al., 2003).

Environmental Risk Factors**.** The epidemiological patterns described above suggest that environmental factors contribute to the etiology of the T1D. In particular, the recent temporal increase in T1D incidence points to a changing global environment rather than variation in the gene pool, which require the passage of multiple generations. Twin studies also provide evidence for the importance of environmental risk factors for T1D. T1D concordance rates for monozygous twins are higher than those for dizygous twins (approximately 30% vs. 10%, respectively) (Hirschhorn, 2003). However, most monozygous twin pairs remain discordant. Thus, T1D cannot be completely genetically determined.

Environmental risk factors are thought to act as either 'initiators' or 'accelerators' of beta cell autoimmunity, or 'precipitators' of overt symptoms in individuals who already have evidence of beta cell destruction. They also may function by mechanisms that are directly harmful to the pancreas, or by indirect methods that produce an abnormal immune response to proteins normally present in cells. The T1D environmental risk factors that have received most attention are viruses and infant nutrition.

Enteroviruses, especially Coxsackie virus B (CVB), have been the focus of numerous ecologic and case-control studies (Dahlquist et al., 1998). CVB infections are frequent during childhood and are known to have systemic effects on the pancreas. Recent prospective studies are helping to elucidate the role of viruses to the etiology of T1D. For example, enteroviral infections occurring as early as *in utero* appear to increase a child's subsequent risk of developing the disease (Dahlquist et al., 1995, Hyoty et al., 1995). Other viruses, including mumps (Hyoty et al., 1993), cytomegalovirus (Pak et al., 1988), rotavirus (Honeyman et al., 2000) and rubella, (McIntosh and Menser, 1992) have also been associated with the disease.

Another hypothesis that has been the subject of considerable interest relates to early exposure to cow's milk protein and the subsequent development of T1D. The first epidemiologic observation of such a relationship was by Borch-Johnsen et al., who found that T1D children were breast-fed for shorter periods of time than their non-diabetic siblings or children from the general population (Borsh-Johnsen et al., 1984). The authors postulated that the lack of immunologic protection from insufficient breast-feeding may increase risk for T1D later during childhood. It was also postulated that shorter duration of breast feeding may indirectly reflect early exposure to dietary proteins that stimulate an abnormal immune response in newborns. Most recently it has been hypothesized that the protective effect of breast-feeding may be due, in part, to its role in gut maturation (Kolb and Pozzilli, 1999; Harrison and Honeyman, 1999; Vaarala, 1999). Breast milk contains growth factors, cytokines, and other substances necessary for the maturation of the intestinal mucosa. Breast-feeding also protects against enteric infections during infancy, and promotes proper colonization of the gut. Interestingly, enteroviral infections can also interfere with gut immunoregulation, which may explain the epidemiologic associations between viral infections and T1D.

The role of hygiene in the etiology of T1D is also currently being explored (McKinney et al., 1997; Marshall et al., 2004). It has been hypothesized that delayed exposure to microorganisms due to improvements in standard of living hinders the development of the immune system, such that it is more

likely to respond inappropriately when introduced to such agents at older (compared to younger) ages. This explanation is consistent with recent reports indicating that factors such as day care attendance (McKinney et al. 2000), sharing a bedroom with a sibling, and contact with pets are protective against T1D (Marshall et al., 2004). Further studies are needed to determine if improved hygiene can explain the temporal increase in the incidence of T1D worldwide.

Type 2 Diabetes

Epidemiology. T2D is the most common form of the disease, accounting for approximately 90% of all affected individuals. A diagnosis of T2D is made if a fasting plasma glucose concentration is $\geq$ 7.0 mmol/L ($\geq$ 126 mg/dl) or plasma glucose 2 hours after a standard glucose challenge is $\geq$ 11.1 mmol/L ($\geq$ 200 mg/dl) (WHO, 1999) T2D is caused by relative impaired insulin secretion and peripheral insulin resistance. Typically, T2D is managed with diet, exercise, oral hypoglycemic agents and sometimes exogenous insulin. However, it is associated with the same long-term complications as T1D.

The highest rates of T2D are found among Native Americans, particularly the Pima Indians who reside in Arizona in the US, and in natives of the South Pacific islands, such as Nauru (Wild et al., 2004). T2D is also known to be more predominant in Hispanic and African American populations than in Caucasians. In 2000, it is estimated that 171 million people (2.8% of the worlds population) had diabetes and that by 2030 this number will be 366 million (4.4% of the world's population). The vast majority of this increase will occur in men and women aged 45 to 64 years living in developing countries. According to Wild et al.(2004), the 'top' three countries in terms of the number of T2D individuals with diabetes are India (31.7 million in 2000; 79.4 million in 2030), China (20.8 million in 2000; 42.3 million in 2030) and the US (17.7 million in 2000; 30.3 million in 2030). Clearly, T2D has become an epidemic in the 21st century.

In addition to the burden of T2D there is an even larger number of people with raised levels of blood glucose but below the level for diabetes. The World Health Organization defines impaired fasting glucose as a fasting plasma glucose level of $\geq$ 6.1 mmoll^{-1} and less than 7 mmoll^{-1}, and impaired glucose tolerance as 2 hour plasma glucose, post glucose challenge, of 7.8 to less than 11.1 mmoll^{-1} (WHO, 1999).

The prevalence of T2D increases with age of population (Wild et al., 2004). In developing countries, the largest number of people with diabetes are in the age group 45 to 64 years, while in developed the largest number is found in those aged 65 years and over. These differences largely reflected differences in population age structure between developed and developing countries. Worldwide rates are similar in men and women, although they are slightly higher in men < 60 years of age and in women > age 65 years.

Of great concern is the recent increase in T2D in children (Bloomgarden, 2004). A report based on the Pima Indians in Arizona noted that between 1967-76 and 1987-96, the prevalence of T2D increased 6-fold in adolescents (Fagot-Campagna et al., 2000). In the US, the incidence of T2D increased from 0.3-1.2/100,000/yr before 1992 to 2.4/100,000/yr in 1994 (Weill et al., 2004). Most T2D children diagnosed during this period were females from minority populations, with a mean age of onset at around puberty. They were also likely to have a positive family history of the disease, particularly maternal diabetes.

Environmental Risk Factors. As early as 1962, Neel hypothesized that T2D represented a 'thrifty genotype', which had a selective advantage (Neel, 1962). He postulated that in primitive times, individuals who were 'metabolically thrifty' and able to store a high proportion of energy as fat when food was plentiful were more likely to survive times of famine. However, in recent years, most populations experience a continuous supply of calorie-dense processed foods, as well as a decrease in physical activity. This likely explains the rise in T2D prevalence worldwide.

The major environmental risk factors for T2D are obesity ($\geq$ 120% ideal body weight or a body mass index $\geq$ 30 k/m^2) and a sedentary lifestyle (van Dam, 2003; Shaw and Chisholm, 2003). Thus, the tremendous increase in the rates of T2D in recent years has been attributed, primarily, to the dramatic rise in obesity worldwide (Zimmet et al., 2001). It has been estimated that approximately 80% of all new T2D cases are due to obesity (Lean, 2000). This is true for adults and children. In the Pima Indians, 85% of the T2D children were either overweight or obese (Fagot-Campagna et al., 2000). Another study in the US reported that IGT was detected in 25% of obese children age 4-10 years, and in 21% of obese adolescents (Sinha et al., 2002). Undiagnosed T2D was detected in 4% of the adolescents.

In addition to general obesity, the distribution of body fat, estimated by the ratio of waist-to-hip circumference (WHR), also has an impact on T2D risk. WHR is a reflection of abdominal (central) obesity, which is more strongly associated with T2D than the standard measures of obesity, such as those based on body mass index.

The other major T2D risk factor is physical inactivity. In addition to controlling weight, exercise improves glucose and lipid metabolism, which decreases T2D risk. Physical activity, such as daily walking or cycling for more than 30 minutes, has been shown to significantly reduce the risk of T2D (Hu et al., 2003). Physical activity has also been inversely related to body mass index and IGT. Recently, intervention studies in China (Pan et al., 1997), Finland (Tuomilehto J et al., 2001) and the US (Diabetes Prevention Program Study Group, 2002) have shown that lifestyle interventions targeting diet and exercise decreased the risk of progression from IGT to T2D by approximately 60% . In contrast, oral hypoglycemic medication only reduced the risk of progression by about 30%.

There is also considerable evidence suggesting that the intrauterine environment is an important predictor of T2D risk (Hales and Barker, 2001; Sobngwi et al., 2003), Numerous studies have shown that low birth weight, which is an indicator of fetal malnutrition, is associated with IGT and T2D later in life. However, it is unclear whether low birth weight is causal or related to potential confounding factors that contribute to both poor fetal growth and T2D (Frayling and Hattersley, 2001).

Role of Genetics in the Development of Diabetes

Type 1 Diabetes

First degree relatives have a higher risk of developing T1D than unrelated individuals from the general population (approximately 6% vs. <1%, respectively) (Dorman and Bunker, 2000). These data suggest that genetic factors are involved with the development of the disease. At present, there is evidence that more than 20 regions of the genome may be involved in genetic susceptibility to T1D. However, none of the candidates identified have a greater influence on T1D risk than that conferred by genes in the HLA region of chromosome 6. This region contains several hundred genes known to be involved in

immune response. Those most strongly associated with the disease are the HLA class II genes (i.e., HLA-DR, DQ, DP).

IDDM1. The HLA class II genes, also referred to as *IDDM1*, contribute approximately 40-50% of the heritable risk for T1D (Hirschhorn et al., 2003). When evaluated as haplotypes, DQA1*0501-DQB1*0201 and DQA1*0301-DQB1*0302 are most strongly associated T1D in Caucasian populations. They are in linkage disequilibrium with DRB1*03 and DRB1*04, respectively. Specific DRB1*04 alleles also modify the risk associated with the DQA1*0301-DQB1*0302 haplotype. Other reported high risk haplotypes for T1D include DRB1*07-DQA1*0301-DQB1*0201 among African Americans, DRB1*09-DQA1*0301-DQB1*0303 among Japanese, and DRB1*04-DQA1*0401-DQB1*0302 among Chinese. DRB1*15-DQA1*0602-DQB1*0102 is protective and associated with a reduced risk of T1D in most populations. Recent reports suggest that other genes in the central, class I and extended class I regions may also increase T1D risk independent of HLA class II genes (Nejentsev et al., 1997; Lie et al., 1999).

Individuals with two high risk DRB1-DQA1-DQB1 haplotypes have a significantly higher T1D risk than individuals with no high risk haplotype. The T1D risk among those with only one susceptibility haplotype is also increased, but effect is more modest. Relative risk estimates range from 10 – 45 and 3-7, respectively, for these groups, depending on race (Dorman and Bunker, 2000). In terms of absolute risk, Caucasian individuals with two susceptibility haplotypes have an approximately 6% chance of developing T1D through age 35 years. However, this figure is substantially lower in populations where T1D is rare (i.e., < 1% among Asians). In addition to *IDDM1*, two other genes are now known to influence T1D risk (Anjos and Polychronakos, 2004). These include *INS* and *CTLA-4*.

Table 1. Several T1D Susceptibility Genes

Gene	Locus	Variant	Estimated RR[†]
HLA-DQB1	6p21.3	*0201 & *0302	3 – 45
INS	11p15. 5	Class I	1 – 2
CTLA4	2q31-35	Thr17Ala	1 – 2

[†]RR = relative risk

INS (insulin). The *INS* gene, located on chromosome 11p15.5, has been designated as *IDDM2*. Positive associations have been observed with a non-transcribed variable number of tandem repeat (VNTR) in the 5' flanking region (Bennett et al., 1997; Pugliese et al., 1997) . There are two common variants. The shorter class I variant predisposes to T1D (relative increase: 1 – 2), whereas the longer class III variant appears to be dominantly protective. The biological plausibility of these associations may relate to the expression of insulin mRNA in the thymus. Class III variants appear to generate higher levels of insulin mRNA than class I variants. Such differences could contribute to a better immune tolerance for class III positive individuals by increasing the likelihood of negative selection for autoreactive T-cell clones. The effect of *INS* appears to vary by ethnicity, with lesser effects in non-Caucasian populations (Undlien et al. 1994).

CTLA-4 (cytotoxic T lymphocyte-associated 4). The *CTLA-4* gene is located on chromosome 2q31-35 (Anjos and Polychronakos, 2004), where multiple T1D genes may be located. CTLA-4 variants have been associated with T1D, as well as other autoimmune disease. *CTLA-4* negatively regulates T-cell

function. However, impaired activity, which has been associated with the Thr17Ala variant, may increase T1D risk. Overall, the relative increase in risk for the CTLA-4Ala17 variant has been estimated as ~ 1.5.

Type 2 Diabetes

It has long been known that T2D is, in part, inherited. Family studies have revealed that first degree relatives of individuals with T2D are about 3 times more likely to develop the disease than individuals without a positive family history of the disease (Flores et al., 2003; Hansen 2003; Gloyn 2003). It has also been shown that concordance rates for monozygotic twins, which have ranged from 60-90%, are significantly higher than those for dizygotic twins. Thus, it is clear that T2D has a strong genetic component.

One approach that is used to identify disease susceptibility genes is based on the identification of candidate genes (Barroso et al., 2003; Stumvoll, 2004). Candidate genes are selected because they are thought to be involved in pancreatic β cell function, insulin action / glucose metabolism, or other metabolic conditions that increase T2D risk (e.g., energy intake / expenditure, lipid metabolism). To date, more than 50 candidate genes for T2D have been studied in various populations worldwide. However, results for essentially all candidate genes have been conflicting. Possible explanations for the divergent findings include small sample sizes, differences in T2D susceptibility across ethnic groups, variation in environmental exposures, and gene-environmental interactions. Because of current controversy, this review will focus only on a few of the most promising candidate genes. These include *PPARγ*, *ABCC8*, *KCNJ11*, and *CALPN10*.

Table 2. Several T2D Susceptibility Genes

Gene	Locus	Variant	Estimated RR†
PPARγ	3p25	Pro12Ala	1 – 3
ABCC8	11p15.1	Ser1369Ala	2 – 4
KCNJ11	11p15.1	Glu23Lys	1 – 2
CALPN10	2q37.3	A43G	1 - 4

†RR = relative risk

PPARγ (peroxisome proliferator-activated receptor-γ). This gene has been widely studied because it is important in adipocyte and lipid metabolism. In addition, it is a target for the hypoglycemic drugs known as thiazolidinediones. One form of the *PPARγ* gene (Pro) decreases insulin sensitivity and increases T2D risk by several fold. Perhaps more importantly is that this variant is very common in most populations. Approximately 98% of Europeans carry at least one copy of the Pro allele. Thus, it likely contributes to a considerable proportion (~25%) of T2D that occurs, particularly among Caucasians.

ABCC8 (ATP binding cassette, subfamily C, member 8). This gene encodes the high-affinity sulfonylurea receptor (SUR1) subunit that is coupled to the Kir6.2 subunit (encoded by *KCNJ11,* also known as the potassium channel, inwardly rectifying subfamily J, member 11). Both genes are part of the ATP-sensitive potassium channel, which plays a key role in regulating the release of hormones, such as insulin and glucagon, in the beta cell. Mutations in either gene can affect the potassium

channel's activity and insulin secretion, ultimately leading to the development of T2D. Interestingly, *ABCC8* and *KCNJ11* are only 4.5 kb apart, and not far from the *INS* gene. Variant forms of *KCNJ11* (Lys) and *ABCC8* (Ala) genes have been associated with T2D, as well as other diabetes-related traits. Because of the close proximity of these genes, current studies are evaluating whether they work in concert with each other, or rather have an independent effect on T2D susceptibility.

Since *PPARγ, ABCC8* and *KCNJ11* are the targets of drugs used routinely in the treatment of T2D, there are pharmacogenetic implications for maintaining good glycemic control. Response to hypoglycemic therapy may actually be related one's genotype. Thus, genetic testing may not only help determine who is at high risk for developing T2D, but may also be useful in guiding treatment regimens for T2D.

CAPN10 (calpain 10). *CAPN10* encodes an intracellular calcium-dependent cysteine protease that is ubiquitously expressed (Cox et al., 2004). A haplotype that was initially linked to T2D included an intronic A to G mutation at position 43, which appears to be involved in *CAPN10* transcription. Two amino acid polymorphisms (Thr504Ala and Phe200Thr) have also been associated with T2D risk. However, it has been suggested that the coding and noncoding polymorphisms do not independently influence T2D risk, but instead contribute to an earlier age at diagnosis. Physiological studies suggest that variations in calpain 10 activity effects insulin secretion, and therefore, susceptibility to T2D. Studies from different ethnic groups indicate that the contribution of this locus to increased T2D risk may be much larger in Mexican-American than Caucasian populations.

Maturity-Onset Diabetes of the Young

An uncommon form of T2D (accounting for <5% of all T2D cases) that generally occurs before age 25 years is MODY. MODY is characterized by a slow onset of symptoms, the absence of obesity, no ketosis, and no evidence of beta cell autoimmunity. It is most often managed without the need for exogenous insulin. MODY displays an autosomal dominant pattern inheritance, generally spanning three generations (Stride and Hattersley, 2002). Because of advances in molecular genetics, it is now known that there are at least six forms of MODY, each of which caused by a mutation in a different gene that is directly involved with beta cell function (Winter, 2003). Table 3 lists the MODY genes that have been identified to date. Because ~15% of MODY patients do not carry mutations in one of these genes, it is anticipated that other genes that cause MODY will be discovered in the near future (Demenais et al., 2003; Frayling et al., 2003; Kim et al., 2004).

Table 3. MODY Genes

Type	Gene	Locus	# Mutations	% MODY
MODY1	*HNF4A*	20q12-q13.1	12	~5%
MODY2	*GCK*	7p15-p13	~200	~15%
MODY3	*HNF1A*	12q24.2	>100	~65%
MODY4	*IPF1*	13q12.1	Few	
MODY5	*HNF1B*	17cen-q21.3	Few	<3%
MODY6	*NEUROD1*	2q32	Few	

GCK (glucokinase). The *GCK* gene is currently the only MODY gene that does not regulate the expression of other genes. Rather, the *GCK* gene plays a key role in glucose metabolism and insulin secretion. Thus, the clinical course of MODY2 patients differs from the prognosis associated with other types of MODY. MODY2 patients have a mild fasting hyperglycemia that is present from birth, and generally stable throughout life. There may be a mild deterioration of normoglycemia with age, but patients with

MODY2 mutations are usually asymptomatic. Most are detected during routine medical screening. Women with MODY2 mutations are often diagnosed during pregnancy. However, the outcome of the pregnancy can be influenced by whether the mother and / or fetus carry the mutation. When both mother and fetus are MODY2 positive, there is generally no effect on birth weight. However, MODY2 negative fetuses are carried by MODY2 positive mothers are typically large for gestational age due to maternal hyperglycemia. In contrast, if the fetus, but not the mother, carries the MODY2 mutation, their birth weight will be reduced by approximately 500g due to reduced fetal insulin secretion, which inhibits growth.

HNF4A (hepatocyte nuclear factor 4-α). Mutations in promoter and coding regions of the *HNF4A* gene cause MODY1. *HNF4A* is expressed in many tissues, including the liver and pancreas. It regulates hepatic gene expression, and influences the expression of other MODY genes such as *HNF1A,* which causes MODY3. In the beta cell of the pancreas, it directly activates insulin gene expression. Mutations in the *HNF4A* gene also have been associated with T2D (Silander et al., 2004).

HNF1A (hepatocyte nucleara factor 1-α). MODY3, the most frequent cause of the disease, results from mutations in the *HNF1A* gene. *HNF1A* is expressed in the liver and pancreas. It can also influence *HNF4A* expression, indicating a connection between MODY1 and MODY3. This suggests that the MODY transcription factors form a regulatory network that maintains glucose homeostasis. In addition to causing MODY3, *HNF1A* mutations have been associated with T1D (Moller et al., 1998; Lehto et al., 1999) and T2D (Pearson et al., 2004).

IPF1 (insulin promoter factor-1). MODY4, which is a rare form of the disease, is due to mutations in the *IPF1* gene. Homozygosity for such mutations has been associated with newborn pancreatic agenesis and neonatal diabetes. Therefore, infants who carry MODY4 mutations tend to be small for gestational age. Individuals with MODY4 may also develop T2D (Cockburn et al., 2004). *IPF1* regulates expression of glucokinase, insulin and other genes involved in glucose metabolism.

HNF1B (hepatocyte nucleara factor 1-β). MODY5, another rare form of MODY, has also been linked with MODY1 because *HNF1β* regulates *HNF4α*. However, unlike MODY1, MODY5 is also associated with renal cysts, proteinuria and renal failure.

NEUROD1 (neurogenic differentiation factor 1). Mutations in *NEUROD1* are responsible for MODY6. MODY6 is also rare. Together, MODY4, MODY5 and MODY6 comprise less then 3% of all MODY cases. *NERUOD1* is expressed in the beta cells of the pancreas, the intestine and the brain. In the pancreas, it contributes to the regulation of the expression of insulin.

To summarize, all MODY genes are expressed in the islet cells of the pancreas, and play a role in the metabolism of glucose, the regulation of insulin or other genes involved in glucose transport, and/or the development of the fetal pancreas. Because MODY phenotypes vary depending which gene is involved (Table 4), genetic testing may also assist in the treatment of the disease.

Table 4. MODY Phenotypes

Type	Disease Onset	Complications	Treatment
MODY1	Severe	Frequent	Diet, oral agents, insulin
MODY2	Mild	Rare	Diet

MODY3	Severe	Frequent	Diet, oral agents, insulin
MODY4	Moderate	Little data	Oral agents, insulin
MODY5	Severe	Renal cysts	Oral agents, insulin
MODY6	Severe	Little data	Diet, oral agents, insulin

Role of Genetics in the Treatment and Prevention of Diabetes

Type 1 Diabetes

At the present time, there is no way to prevent T1D. Lifelong insulin injections are the only available treatment for the disease. Thus, genetics does not currently play a role in the management or prevention of T1D.

Although a cure for T1D is currently unavailable, several large multi-national investigations have been designed to evaluate a variety of primary and secondary disease interventions (Devendra et al., 2004). The tested interventions have included prophylactic nasal insulin (Diabetes Prediction and Prevention Project (DIPP) in Finland), oral and injected insulin (Diabetes Prevention Trial – 1 (DPT-1) in the US), as well as high doses of nicotinamide (European Nicotinamide Diabetes Intervention Trial - ENDIT), and the avoidance of cow's milk exposure during the first six months of life (Trial to Reduce in Genetically At-Risk (TRIGR) in Finland, US and other countries). These investigations focus on 'prediabetic' individuals identified from families with at least one child with type 1 diabetes. DIPP and TRIGR use HLA-DQB1 screening and recruit only individuals at increased genetic risk. The remaining trials recruit relatives with evidence of beta cell autoimmunity as a pre-clinical marker for disease. To date, none of these interventions have prevented or delayed the onset of T1D (Diabetes Prevention Trial-Type 1 Study Group, 2002; NIDDK, 2003; The ENDIT Group, 2003; Paronen, et al., 2000). However, with the formation of *Type 1 Diabetes TrialNet (www.trialnet.com)*, a collaborative network of clinical centers and experts in diabetes and immunology, new intervention strategies are currently being planned. It is ultimately hoped that through genetic testing, individuals at high risk for T1D could be identified prior to the onset of the disease – at a time when primary prevention strategies could be safely administered. It is most likely that such predictive genetic testing would be offered to families with an affected individual before it was made available to the general population.

Type 2 Diabetes

Unlike T1D, T2D can generally be prevented by maintaining an age-appropriate body weight and engaging in physical activity. Although public health messages that emphasize a nutritious diet and regular physical activity are now commonplace, they have not been effective in terms of disease prevention. Given the recent obesity epidemic, it is obvious that current intervention strategies are being ignored by a majority of individuals in the general population.

Leaders of the Human Genome Project have predicted that genetic tests will become available for many common disorders during the first decade of the 21st century, permitting persons "to learn their individual susceptibilities and to take steps to reduce those risks" by applying interventions based on "medical surveillance, lifestyle modifications, diet or drug therapy" (Collins and McKusick, 2001). In

fact, several companies are now offering genetic susceptibility testing, which can be ordered online by any individual, for conditions such as cardiovascular disease and obesity (Khoury et al., 2004).

Although many scientists and health professionals share this optimistic perspective regarding genetics and disease prevention, others are more pessimistic for a variety of reasons. First, the predictive value of most genetic tests is low (Haga et al., 2003); and risk estimates do not account for well-known environmental determinants of disease. Secondly, it is unclear whether knowledge of one's genetic risk increases motivation to engage in disease interventions. Thirdly, genetic testing presents educational and information-dissemination challenges that were outlined in detail by the Secretary's Advisory Committee on Genetics, Health and Society (Holtzman and Watson, 1998). These include being able to communicate the validity and utility of proposed genetic tests, as well as the potential risks and benefits of being tested, to individuals who may have little knowledge of human genetics. Fourthly, most health care professionals are currently unqualified to interpret the results of genetic tests; and there are no standards for the use of molecular diagnostics in clinical practice. Fifthly, genetic testing may lead to significant distress, the magnitude of which is likely to vary as a function of actual test results, coping skills and resources, risk perception, optimism, health beliefs and pre-existing depression or anxiety.

These factors directly relate to other concerns such as insurance and employment discrimination, confidentiality and stigmatization based on knowing that one is at high genetic risk. In the near future, genetic testing for T2D and other chronic diseases will most certainly become available. Although it is unclear whether this will actually contribute to the prevention of T2D, it may be beneficial in terms of disease management. Many of the current T2D susceptibility genes of interest are drug targets. Evidence for the role of pharmacogenetics in diabetes is already apparent in treatment approaches for MODY.

Maturity-Onset Diabetes of the Young

The most common causes of MODY are related to mutations in MODY3, MODY2 and MODY1 genes. Although individuals who carry MODY2 mutations have a very mild form of the disease, those who carry MODY1 and MODY3 variants have a much more severe expression that is associated with long-term complications. In addition, there has been a link between MODY3 and MODY5 because of their interaction in terms of gene expression. However, it is now becoming clear that the metabolic phenotype of individuals with these two forms of MODY is actually quite different (Pearson et al., 2004). To date, little has been known about MODY5 other than its association with renal cysts. However, it now appears that MODY5 is more strongly associated with hyperinsulinemia and dyslipidemia (and more closely related to insulin resistance and T2D) than MODY3. Thus, knowledge about the underlying MODY defect is likely to lead to better management and an improved prognosis for individuals with the disease.

Given the autosomal dominant inheritance of all forms of MODY, individuals with a diabetic parent may also wish to have genetic testing. Early diagnosis MODY may also help reduce the likelihood of long-term complications. In addition, psychological and family adjustments to diabetes may also be improved when the specific form of the disease is known.

Approximately one-third of individuals with MODY3 and MODY1 are each treated by diet, oral agents and insulin. Some individuals with MODY3 have been previously classified as having T1D because of the severity of the disease (Moller et al., 1998; Lehto et al., 1999). It is now known that individuals with MODY3 mutations are extremely sensitive to the hypoglycemic effects of sulfonylureas. Thus, these oral agents are likely to be the treatment of choice of individuals with MODY3. Recently, there

have been a number of reports of MODY3 individuals being able to change treatment regimens from insulin injections to oral sulphonylurea agents, with considerable improvement in glycemic control (Shepherd, 2003a; Shepherd and Hattersley, 2004)). This is frequently associated with a positive impact on lifestyle and self image, as well as fear and anxiety about the possibility of stopping insulin. Some individuals, particularly those who have long-term complications, have become angry because they were previously misdiagnosed and/or treated inappropriately. These reactions have implications for health professionals who need to be knowledgeable about the potential psychological consequences of changing treatment regimens (Shepherd, 2003b).

Future Role of Genetics in Diabetes

Within the next decade, the genes that increase risk of developing all forms of diabetes will likely be known. It is, therefore, important that scientists, health professionals, and members of population at large consider how to maximize the advantages, and minimize the disadvantages of predictive genetic testing for diabetes.

In September 2004, the Office of Genomics and Disease Prevention at the Centers for Disease Control in the US held a meeting entitled "Public Health Assessment of Genetic Tests for Screening and Prevention". One of the objectives of this session was to discuss issues related to the evaluation and utilization of genetic tests. Emphasis was placed on three major barriers: 1) the lack of available population data regarding the contribution of genetic variants to disease susceptibility, 2) the lack of an evidence-based process for the integration of genomics into practice and, 3) the lack of readiness of the health care and public health systems to utilize genetic testing for disease prevention. At the end of the meeting it was apparent that we, as a society, are a long way from the practice of 'genomic medicine'.

With regards to diabetes, addressing the first barrier is most critical at the present time. This barrier pertains to the lack of consistent results across populations with regards to the genetic determinants of the disease. Failure to replicate study results may be due to a variety of factors, the most important of which may be that different gene-environment interactions operate different populations to increase risk of developing diabetes. Thus, considerably more epidemiologic research will be needed before we know the actual risk associated with particular genetic variants. This also likely means that we will not be able to apply a 'one size fits all' model when it comes to the genetic testing for any of the forms of diabetes.

To fulfill the promise of the Human Genome Project, several issues that warrant careful consideration. First, multidisciplinary teams will be required to translate genetic discoveries from the laboratory to the community. This is, perhaps, best exemplified by the development of new initiatives such as the NIH Roadmap in the US. Scientists will no longer be able to work in isolation, without input of individuals from other professions, if they are to maximize the impact of their research in terms of improving health. In particular, issues such as quality assurance, health risks and benefits, and economics need to be addressed. This will require expertise from persons who have typically worked outside the profession of science. Finally, the ethical, legal and social issues associated with widespread availability and use of predictive genetic tests must been addressed. These include confidentiality, discrimination, diversity, informed consent, keeping up with genetic discoveries and uncertainty. Ideally, consideration of such issues will lead to the development of practice guidelines for diabetes, which will hopefully serve as a model for genetic testing for other complex diseases.

Acknowledgement: We acknowledge the contribution of Dr Janice S. Dorman, Ph.D., University of Pittsburgh, School of Nursing, 350 Victoria Building, Pittsburgh, PA, USA, to this paper.

References

1. Alberti, K.G.M.M., Zimmett, P. *Definition, diagnosis and classification of diabetes mellitus and its complications part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation.* Diabet Med, 1998. **15**: 539-553.
2. Anjos, S., Polychronakos, C. *Mechanisms of genetic susceptibility to type 1 diabetes: beyond HLA.* Mol Genet Metab, 2004. **81**: 187-195.
3. Barroso, I., Luan, J., Middelberg, R.P.S., et al. *Candidate gene association study in type 2 diabetes indicates a role for genes involved in B-Cell function as well as insulin action.* PLoS Biol, 2003. **1**: 41-55.
4. Bennett, S.T., Wilson, A.J., Esposito, L. *Insulin VNTR allele-specific effect in type 1 diabetes depends on identity of untransmitted paternal allele.* Nat Genet, 1997. **17**: 350-352.
5. Bloomgarden, Z.T. *Type 2 diabetes in the young: the evolving epidemic.* Diabetes Care, 2004. **27**: 998-1010.
6. Borch-Johnsen, K., Joner, G., Mandrup-Poulsen, T., et al. *Relation between breast-feeding and incidence rates of insulin-dependent diabetes mellitus. A hypothesis.* Lancet, 1984. **2**: 1083-1086.
7. Cockburn, B.N., Bermano, G., Boodram, L.-L.G., et al. *Insulin promoter factor-1 mutation and diabetes in Trinidad: identification of a novel diabetes-associated mutation (E224K) in an Indo-Trinidadian family.* J Clin Endocrinol Metab, 2004. **89**: 971-978.
8. Collins, F.S., McKusick, V.A. *Implications of the Human Genome Project for medical science.* JAMA, 2001. **285**: 540-544.
9. Cox, N.J., Hayes, M.G., Roe, C.A., et al. *Linkage of calpain 10 to type 2 diabetes: the biological rationale.* Diabetes, 2004. **53**: S19-S25.
10. Dahlquist, G. *The aetiology of type 1 diabetes: an epidemiological perspective.* Acta Paediatr Suppl, 1998. **425**: 5-10.
11. Dahlquist, G., Frisk, G., Ivarsson, S.A., et al. *Indications that maternal coxsackie B Virus infection during pregnancy is a risk factor for childhood-onset IDDM.* Diabetologia, 1995. **38**: 1371-1373.
12. Demanais, F., Kanninen, T., Lindgren, C.M., et al. *A meta-analysis of four European genome screens (GIFT Consortium) shows evidence for a novel region on chromosome 17p11.2-q22 linked to type 2 diabetes.* Hum Mol Genet, 2003. **12**: 1865-1873.
13. Devendra, D., Liu, E., Eisenbarth, G.S. *Type 1 diabetes: recent developments.* BMJ, 2004. **328**: 750-754.
14. Dorman, J.S., Bunker, C.H. *HLA-DQ locus of the human leukocyte antigen complex and type 1 diabetes mellitus: a HuGE review.* Epidemiol Rev, 2000. **22**: 218-227.
15. Dorman, J.S., LaPorte, R.E., Songer, T.J. *Epidemiology of Type 1 Diabetes*, in *Type 1 Diabetes: Etiology and Treatment*, Mark A. Sperling, Editor. 2003, Humana Press: Totowa, NJ. p. 3-22.
16. Fagot-Campagna, A., Pettitt, D.J., Engelgau, M.M., et al. *Type 2 diabetes among North American children and adolescents: an epidemiologic review and a public health perspective.* J Pediatr, 2000. **136**: 664-672.
17. Flores, J.C., Hirschhorn, J., Altshuler, D. *The inherited basis of diabetes mellitus: implications for the genetic analysis of complex traits.* Annu Rev Genomics Hum Genet, 2003. **4**: 257-291.
18. Frayling, T.M., Hattersley, A.T. *The role of genetic susceptibility in the association of low birth weight with type 2 diabetes.* Br Med Bull, 2001. **60**: 89-101.

19. Frayling, T.M., Lindgren, C.M., Chevre, J.C., et al. *A genome-wide scan in families with maturity-onset diabetes of the young: evidence for further genetic heterogeneity.* Diabetes, 2003. **42**: 872-881.
20. Gloyn, A.L. *The search for type 2 diabetes genes.* Ageing Res Rev, 2003. **2**: 111-127.
21. Group, D.C.a.C.T.R. *The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus.* N Engl J Med, 1993. **329**: 977-986.
22. Group, D.P.T.-T.S. *Effects of insulin in relatives of patients with type 2 diabetes mellitus.* N Engl J Med, 2002. **346**: 1685-1691.
23. Group, T.E.N.D.I.T.E. *Intervening before the onset of type 1 diabetes: baseline data from the European Nicotinamide Diabetes Group.* Diabetologia, 2003. **46**: 339-346.
24. Haga, S.B., Khoury, M.J., Burke, W. *Genomic profiling to promote a healthy lifestyle: not ready for prime time.* Nat Genet, 2003. **34**: 347-350.
25. Hales, C.N., Barker, D.J. *The thrifty phenotype hypothesis.* Br Med Bull, 2001. **60**: 5-20.
26. Hansen, L. *Candidate genes and late-onset type 2 diabetes mellitus. Susceptibility genes or common polymorphisms?* Dan Med Bull, 2003. **50**: 320-346.
27. Harrison, L.C., Honeyman, M.C. *Cow's milk and type 1 diabetes.* Diabetes, 1999. **48**: 1501-1507.
28. Hirschhorn, J.N. *Genetic epidemiology of type 1 diabetes.* Pediatr Diabetes, 2003. **4**: 87-100.
29. Holt, R.I.G. *Diagnosis, epidemiology and pathogenesis of diabetes mellitus: an update for psychiatrists.* Br J Psychiatry, 2004. **184**: s55-s63.
30. Holztman, N.A., Watson, M.S. eds. *Promoting Safe and Effective Genetic Testing in the United States*. Johns Hopkins University Press, 1998.
31. Honeyman, M.C., Coulson, B.S., Stone, N.L. *Association between rotavirrus infection and pancreatic islet autoimmunity i nchildren at risk of developing type 1 diabetes.* Diabetes, 2000. **49**.
32. Hu, F.B. *Sedentary lifestyle and risk of obesity and type 2 diabetes.* Lipids, 2003. **38**: 103-108.
33. Hyoty, H., Hiltunen, M., Knip, M., et al. *A prospective study of the role of coxsackie B and other enterovirus infections in the pathogenesis of IDDM. Childhood Diabetes in Finland (DiMe) Study Group.* Diabetes, 1995. **44**: 652-657.
34. Hyoty, H., Hiltunen, M., Reuranen, A. *Decline of mumps antibodies in type 1 (insulin-dependent) diabetic children with a plateau in the rising incidence of type 1 diabetes after introduction of the mumps-measles-rubella vaccine in Finland.* Diabetologia, 1993. **41**: 40-46.
35. Karovonen, M., Tuomilehto, J., Libman, I., et al. *A review of the recent epidemiological data on the worldwide incidence of type 1 (insulin-dependent) diabetes mellitus. World Health Organization DiaMond Group.* Diabetologia, 1993. **36**: 883-892.
36. Khoury, M.J., Yang, Q., Gwinn, M., et al. *An epidemiologic assessment of genomic profilling for measuring susceptibility to common diseases and targeting interventions.* Genet Med, 2004. **6**: 38-47.
37. Kim, S.-H., Ma, X., Weremowicz, S., et al. *Identification of a locus for maturity-onset diabetes of the young on chromosome 8p23.* Diabetes, 2004. **53**: 1375-1384.
38. Kolb, H., Pozzilli, P. *Cow's milk and type 1 diabetes: the gut immune system deserves attention.* Immunol Today, 1999. **20**: 108-110.
39. Kyvik, K.O., Nystrom, L., Gorus, F., et al. *The epidemiology of type 1 diabetes mellitus is not the same in young adults as in children.* Diabetologia, 2004. **47**: 377-384.
40. Lean, M.E. *Obesity: burdens of illness and strategies for prevention or management.* Drugs Today (Barc), 2000. **36**: 773-784.

41. Lehto, M., Wipemo, C., Ivarsson, S.A., et al. *High frequency of mutations in MODY and mitochondrial genes in Scandinavian patients with familial early-onset diabetes.* Diabetologia, 1999. **42**: 1131-1137.
42. Lie, B.A., Todd, J.A., Pociot, F., et al. *The predisposition to type 1 diabetes linked to the human leukocyte antigen complex includes at least one non-class II gene.* Am J Hum Genet, 1999. **64**: 793-800.
43. Marshall, A.L., Chetwynd, A., Morris, J., et al. *Type 1 diabetes mellitus in childhood: a matched case control study in Lancashire and Cumbria, UK.* Diabet Med, 2004. **21**: 1035-1040.
44. McIntosh, E.D.G., Menser, M. *A fifty-year follow-up of congenital rubella.* Lancet, 1992. **340**: 414-415.
45. McKinney, P.A., Okasha, M., Parslow, R., et al. *Ante-natal risk factors for childhood diabetes mellitus, a case-control study of medical record data in Yorkshire, UK.* Diabetologia, 1997. **40**: 933-939.
46. Moller, A.M., Dalgaard, L.T., Pociot, F., et al. *Mutations in the hepatocyte nuclear factor-1a gene in Caucasian families originally classified as having type 1 diabetes.* Diabetologia, 1998. **41**: 1528-1531.
47. Neel, J. *Diabetes mellitus: a thrifty genotype rendered detrimental by "progress"?* Am J Hum Genet, 1962. **14**: 353-362.
48. Nejentsev, S., Reijonen, H., Adojaan, B., et al. *The effect of HLA-B allele on the IDDM risk defined by DRB1*04 subtypes and DQB1*0302.* Diabetes, 1997. **46**: 1888-1892.
49. (NIDDK), N.I.o.D.D.K.D. *Oral Insulin Does Not Prevent Type 1 Diabetes.* News Briefs (June 15, 2003): Available at: www.niddk.nih.gov/welcome/releases/6-15-03.htm March 8, 2004.
50. Onkamo, P., Vaananen, S., Karvonen, M., et al. *Worldwide increase in incidence of type 1 diabetes--the analysis of the data on published incidence trends.* Diabetologia, 1999. **42**: 1395-1403.
51. Pak, C.Y., McArthur, R.G., Eun, H.M. *Association of cytomegalovirus infection with autoimmune type 1 diabetes.* Lancet, 1988. **2**: 1-4.
52. Pan, X.R., Li, G.W., Hu, Y.H., et al. *Effects of diet and exercise in preventing NIDDM in people with impaired glucose tolerance. The Da Qing IGT and Diabetes Study.* Diabetes Care, 1997. **20**: 537-544.
53. Paronen, J., Knip, M., Savilahti, E., et al. *Effect of cow's milk exposure and maternal type 1 diabetes on cellular and humoral immunization to dietary insulin in infants at genetic risk for type 1 diabetes.* Diabetes, 2000. **49**: 1657-1665.
54. Pearson, E.R., Badman, M.K., Lockwood, C., et al. *Contrasting diabetes phenotypes associated with hepatocyte nuclear factor-1a and -1B mutations.* Diabetes Care, 2004. **27**: 1102-1107.
55. Pugliese, A., Zeller, M., Ferndandez, J.A. *The insulin gene is transcribed in human thymus and transcription levels correlated with allelic variation at the INS VNTR-IDDM 2 susceptibility locus for type 1 diabetes.* Nat Genet, 1997. **15**: 293-297.
56. Shaw, J., Chisholm, D. *Epidemiology and prevention of type 2 diabetes and the metabolic syndrome.* MJA, 2003. **179**: 379-383.
57. Shepherd, M. *'I'm amazed I've been able to come off injections': Patients' perceptions of genetic testing in diabetes.* Pract Diab Int, 2003. **20**: 338-342.
58. Shepherd, M. *Genetic testing in maturity onset diabetes of the young (MODY)-practical guidelines for professionals.* Pract Diab Int, 2003. **20**: 108-110.
59. Shepherd, M., Hattersley, A.T. *'I don't feel like a diabetic any more': the impact of stopping insulin in patients with maturity onset diabetes of the young following genetic testing.* Clin Med, 2004. **4**: 144-147.
60. Silander, K., Mohlke, K.L., Scott, L.J., et al. *Genetic variation near the hepatocyte nuclear factor-4a gene predicts susceptibility to type 2 diabetes.* Diabetes, 2004. **53**: 1141-1149.

61. Sinha, R., Fisch, G., Teague, B., et al. *Prevalence of impaired glucose tolerance among children and adolescents with marked obesity.* N Engl J Med, 2002. **346**: 802-810.
62. Sobngwi, E., Boudou, P., Mauvais-Jarvis, F., et al. *Effect of a diabetic environment in utero on predisposition to type 2 diabetes.* Lancet, 2003. **361**: 1861-1865.
63. Stride, A., Hattersley, A.T. *Different genes, different diabetes: lessons from maturity-onset diabetes of the young.* Ann Med, 2002. **34**: 207-216.
64. Stumvoll, M. *Control of glycaemia: from molecules to men. Minkowski Lecture 2003.* Diabetologia, 2004. **47**: 770-781.
65. Tuomilehto, J., Lindstrom, J., Eriksson, J.G., et al. *Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance.* N Engl J Med, 2001. **344**: 1343-1350.
66. Undlien, D.E., Hamaguchi, K., Kimura, A. *Type 1 diabetes susceptibility associated with polymorphism in the insulin gene region: a study of blacks, Caucasians, and orientals.* Diabetologia, 1994. **37**: 745-749.
67. Vaarala, O. *Gut and the induction of immune tolerance in type 1 diabetes.* Diabetes Metab Res Rev, 1999. **15**: 353-361.
68. van Dam, R.M. *The epidemiology of lifestyle and risk for type 2 diabetes.* Eur J Epidemiol, 2003. **18**: 1115-1125.
69. Weill, J., Vanderbecken, S., Froguel, P. *Understanding the rising incidence of type 2 diabetes in adolescence.* Arch Dis Child, 2004. **89**: 502-504.
70. WHO, 1999. Definition, Diagnosis and Classification of Diabetes Mellitus and its Complications. WHO/NCD/NCS/99.2.
71. Wild, S., Roglic, G., Green, A., et al. *Global prevalence of diabetes.* Diabetes Care, 2004. **27**: 1047-1053.
72. Winter, W.E. *Newly defined genetic diabetes syndromes: maturity onset diabetes of the young.* Rev Endocr Metab Disord, 2003. **4**: 43-51.
73. Zimmett, P., Alberti, K.G.M.M., Shaw, J. *Global and societal implication of the diabetes epidemic.* Nature, 2001. **414**: 782-787.