

Advancing sports and exercise genomics: moving from hypothesis-driven single study approaches to large multi-omics collaborative science

Authors: Masashi Tanaka¹, Guan Wang², Yannis P. Pitsiladis^{2*}

Affiliations: ¹Department of Longevity and Health, Tokyo Metropolitan Institute of Gerontology, Tokyo, Japan; ²FIMS Reference Collaborating Centre of Sports Medicine for Anti-Doping Research, University of Brighton, Eastbourne, United Kingdom.

***Corresponding Author:** Professor Yannis P. Pitsiladis

Professor of Sport and Exercise Science

FIMS Reference Collaborating Centre of Sports Medicine

for Anti-Doping Research

University of Brighton

Eastbourne, BN20 7SN, United Kingdom

Email: y.pitsiladis@brighton.ac.uk

Key Words: Genetics, Performance, Athlome, Omics

Abstract

The primary use of the traditional candidate gene approach over the past decades in sports genetics has had limited success in identifying genes associated with elite athletic performance. Advances in high-throughput technologies now permit the application of “omics” (e.g. genomics, transcriptomics, metabolomics, proteomics and epigenomics) approaches to examine the global features of a cell, tissue or organism. “Omics” approaches are being applied with some success to a wide range of pertinent biomedical problems such as cancer diagnosis but also in sports science and sports medicine such as for the identification of biomarkers of trainability or blood doping. There is good evidence to suggest that a combined “omics” solution will greatly facilitate discovery of the genetic influences on sporting performance, training response, injury predisposition and other potential determinants of successful human performance. In this regard, large-scale, collaborative efforts involving well-phenotyped cohorts will be essential for major progress to be made. A recent consensus emerged among 15 research groups active in the field of sports genetics to unite their efforts under one new collaborative initiative named the *Athlome Project Consortium*. The primary aim of the

Athlome Project is to combine resources from individual studies and consortia worldwide to collectively study the genotype and phenotype data available on elite athletes, the adaptation to exercise training (in both human and animal models) and the determinants of exercise-related musculoskeletal injuries. This editorial summarizes the challenges and opportunities facing the *Athlome Project Consortium* and the field of sports and exercise genomics in general.

Rare versus common variant hypothesis and the dilemma of rare phenotypes

There is much interest in the fact that athletes dominating certain sports originate from specific geographical areas for instance, endurance runners from Ethiopia or Kenya (East Africa), sprinters from Jamaica or USA (West Africa), and gymnasts and figure skaters from Japan, Korea, or China (Far East Asia). Although these occurrences are more likely a reflection of historic socioeconomic and cultural characteristics of each location, it is presumed that the area-specific segregation of sports performance is based on uneven distribution of the genetic characteristics that relate to the physical and metabolic properties of individuals in each area (see (8)).

In recent decades, genetic approaches in sports and exercise science have primarily focused on the candidate gene approach, with some attempts of applying the genome-wide association study (GWAS). Despite these efforts, there has been little progress in identifying genes that are locally or universally linked with athletic performance, with the likely exceptions the variants in the *ACTN3* and *ACE* genes (1,11). Each of the common single nucleotide polymorphisms (SNPs) usually has only a

small functional effect, thus initially resulted in the development of the genomic score approach that sums up the numbers of alleles of potential influential SNPs to predict performance (17), training response (2), or risk of obesity, hyperlipidemia, or hyperglycemia with reference to the effect of physical fitness (14-16). The genetic score approach usually exhibits a Gaussian distribution, and although the very best athletes score higher, their extreme performances cannot be adequately explained by the summation of common SNPs with small effects (3). Based on the current literature, it is reasonable to hypothesize that the rare and extreme phenotypes of elite athletes can be explained by rare functional variations or mutations that emerged and were selected rather recently in human history, so that these SNPs or mutations are detectable only in some global regions. This notion is similar to the idea that is used for the discovery of inherited diseases caused by mutations in single or multiple genes (18). Shifting the preferred approach from the candidate gene approach to the genome-wide association study resulted in the emergence of the multi-comparison problem where the Bonferroni correction or the conventional genome-wide significance threshold of 5×10^{-8} in European populations or 1×10^{-8} in African populations (13) needed to be overcome. If

the target region is expanded to include the whole genome of 3×10^9 base pairs, and assuming the frequency of sequence variations in our genome is about 1 out of 1000 bases, each individual is likely to have 3×10^6 variations in their genome. Consequently, identifying a SNP or structural variation linked with an extreme phenotype would be exceedingly difficult even if hundreds or thousands of DNA samples were collected from elite athletes. This is the major problem facing sports and exercise genomics. It is likely that the rare SNPs are more region-specific and emerged more recently in the history of human evolution than the common SNPs. Other views include the notion that the capacity for endurance running is a newly acquired common phenotype of all modern humans (4,12). It is unknown at the moment whether the extreme phenotypes of elite athletes are ascribable to a limited number of rare variants or to high combinations of common variants. To discover the secret of the genomes of elite athletes will require a dedicated international collaborative efforts such as the newly formed *Athlome Project Consortium* that was founded as a result of a symposium held in Athens and on the Greek island of Santorini from 14-17th May 2015. The purpose of the Santorini meeting was to review the main findings in exercise genetics and genomics and to explore

promising trends and possibilities.

New era of whole genome sequencing and future directions

A detailed analysis of the genome reveals unexpected findings in the genome data of each individual such as a few hundred disease-causing mutations. It is recommended that researchers involved in whole genome sequencing report mutations in the list of genes that are considered linked with inherited diseases. Not all individuals are symptomatic of disease and many will never experience any symptoms during their entire life (The MedSeq Project, (9)). Whole genome analysis of elite athletes is going to reshape our current understanding of the genetic basis of human performance and by association health-related fitness. This knowledge will help generate new solutions to modern-day health problems such as inactivity, obesity and sarcopenia; all of which are at the opposite spectrum of the elite athlete phenotype.

In order to facilitate maximum progress recent developments in information technology that guarantee both the safety of unique genetic information of individual athletes as well as the ease of access for researchers exploring links between genotypes and

phenotypes must be applied. Vital lessons are already being learned from the field of cancer genomics. For example, in the COSMIC catalogue storing the somatic mutation information (5,7), the somatic mutations in the cancer tissues and the germ line variations are analyzed but only the somatic mutations present in the cancer tissue are open access. Data generated from the *Athlome Project Consortium* will be made publically available for sharing of resources with the wider scientific community (<http://staging.athlomeconsortium.org/>), particularly when whole genome sequencing of hundreds of elite athletes takes place, as is currently planned (e.g. 1000 Athlomes project (10) in this series). Whole genome sequencing studies of elite athletes alone will not, however, be sufficient to meaningfully understand the mechanisms underlying the extreme athlete phenotype. This will require genomic and epigenomic information that can be potentially linked with transcriptomic and proteomic information (e.g. (6) paper by Durussel et al. in this series). Establishing lymphocyte or myocyte clones isolated from elite athletes will also be important in this regard as the transcriptome or proteome data from such resources are less affected by environmental conditions including exercise training or unknown doping.

This issue of *Physiological Genomics* is a pivotal landmark in the field of sports and exercise genomics, contributed by the participants of the Santorini Conference 2015 that abridges the current literature and the future trends in sports and exercise science and medicine. This series of manuscripts will be a useful resource for scientists, researchers, students and others who have an interest in genetics, genomics and exercise biology. It is hoped that this shift in current approach that has led to the creation of the *Athlome Project Consortium* and ideas brought forth in this series will encourage a new standard of excellence and motivate more international collaboration in sport and exercise genomics.

References

1. **Berman, Y, North, KN.** A gene for speed: the emerging role of alpha-actinin-3 in muscle metabolism. *Physiology (Bethesda)* 25: 250-259, 2010.
2. **Bouchard, C.** Genomic predictors of trainability. *Exp Physiol* 97: 347-352, 2012.
3. **Bouchard, C.** Exercise genomics-a paradigm shift is needed: a commentary. *Br J Sports Med* 49: 1492-1496, 2015.
4. **Bramble, DM, Lieberman, DE.** Endurance running and the evolution of Homo. *Nature* 432: 345-352, 2004.
5. Catalogue of somatic mutations in cancer. Retrieved 13/01/2016, from <http://cancer.sanger.ac.uk>
6. **Durussel, J, Haile, DW, Mooses, K, Daskalaki, E, Beattie, W, Mooses, M, Mekonen, W, Ongaro, N, Anjila, E, Patel, RK, Padmanabhan, N, McBride, MW,**

McClure, JD, Pitsiladis, YP. The blood transcriptional signature of recombinant human erythropoietin administration and implications for anti-doping strategies. *Physiol Genomics* (January 12, 2016), doi: 10.1152/physiolgenomics.00108.2015

7. **Forbes, SA, Beare, D, Gunasekaran, P, Leung, K, Bindal, N, Boutselakis, H, Ding, M, Bamford, S, Cole, C, Ward, S, Kok, CY, Jia, M, De, T, Teague, JW, Stratton, MR, McDermott, U, Campbell, PJ.** COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res* 43: D805-811, 2015.

8. **MacArthur, DG, Seto, JT, Raftery, JM, Quinlan, KG, Huttley, GA, Hook, JW, Lemckert, FA, Kee, AJ, Edwards, MR, Berman, Y, Hardeman, EC, Gunning, PW, Easteal, S, Yang, N, North, KN.** Loss of ACTN3 gene function alters mouse muscle metabolism and shows evidence of positive selection in humans. *Nat Genet* 39: 1261-1265, 2007.

9. National Human Genome Research Institute (NHGRI) Clinical Sequencing Exploratory Research (CSER) The MedSeq Project: Integration of Whole Genome Sequencing into Clinical Medicine (HG006500). Retrieved 13/01/2016, from http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000958.v1.p1

10. **Pitsiladis, YP, Tanaka, M, Eynon, N, Bouchard, C, North, KN, Williams, AG, Collins, M, Moran, CN, Britton, SL, Fuku, N, Ashley, EA, Klissouras, V, Lucia, A, Ahmetov, II, de Geus, EJ, Alsayrafi, M.** The Athlome Project Consortium: A Concerted Effort to Discover Genomic and other "OMIC" Markers of Athletic Performance. *Physiol Genomics* (December 29, 2015), doi: 10.1152/physiolgenomics.00105.2015

11. **Puthuchery, Z, Skipworth, JR, Rawal, J, Loosemore, M, Van Someren, K, Montgomery, HE.** The ACE gene and human performance: 12 years on. *Sports Med* 41: 433-448, 2011.

12. **Ruxton, GD, Wilkinson, DM.** Endurance running and its relevance to scavenging by early hominins. *Evolution* 67: 861-867, 2013.

13. **Sham, PC, Purcell, SM.** Statistical power and significance testing in large-scale genetic studies. *Nat Rev Genet* 15: 335-346, 2014.

14. **Tanisawa, K, Ito, T, Sun, X, Cao, ZB, Sakamoto, S, Tanaka, M, Higuchi, M.** Polygenic risk for hypertriglyceridemia is attenuated in Japanese men with high fitness levels. *Physiol Genomics* 46: 207-215, 2014.

15. **Tanisawa, K, Ito, T, Sun, X, Ise, R, Oshima, S, Cao, ZB, Sakamoto, S, Tanaka, M, Higuchi, M.** Strong influence of dietary intake and physical activity on body fatness in elderly Japanese men: age-associated loss of polygenic resistance against obesity. *Genes Nutr* 9: 416, 2014.
16. **Tanisawa, K, Ito, T, Sun, X, Ise, R, Oshima, S, Cao, ZB, Sakamoto, S, Tanaka, M, Higuchi, M.** High cardiorespiratory fitness can reduce glycated hemoglobin levels regardless of polygenic risk for Type 2 diabetes mellitus in nondiabetic Japanese men. *Physiol Genomics* 46: 497-504, 2014.
17. **Williams, AG, Folland, JP.** Similarity of polygenic profiles for elite human physical performance. *J Physiol* 586: 113-121, 2008.
18. **Zimran, A, Gelbart, T, Westwood, B, Grabowski, GA, Beutler, E.** High frequency of the Gaucher disease mutation at nucleotide 1226 among Ashkenazi Jews. *Am J Hum Genet* 49: 855-859, 1991.