OPEN ACCESS

Sport and exercise genomics: the FIMS 2019 consensus statement update

Kumpei Tanisawa,[1] Guan Wang,[2] Jane Seto,[3,4] Ioanna Verdouka,[2] Richard Twycross-Lewis,[5] Antonia Karanikolou,[2] Masashi Tanaka,[6] Mats Borjesson,[7,8] Luigi Di Luigi,[9] Michiko Dohi ,[10] Bernd Wolfarth,[11] Jeroen Swart,[12] James Lee John Bilzon ,[13] Victoriya Badtieva,[14,15] Theodora Papadopoulou,[16,17] Maurizio Casasco,[18] Michael Geistlinger,[19] Norbert Bachl,[20,21] Fabio Pigozzi,[22] Yannis Pitsiladis [2]

For numbered affiliations see end of article.

Correspondence to
Professor Yannis Pitsiladis, Collaborating Centre of Sports Medicine, University of Brighton, Eastbourne BN20 7UR, UK; Y.Pitsiladis@Brighton.ac.uk

Accepted 17 December 2019
Published Online First 22 March 2020

ABSTRACT

Rapid advances in technologies in the field of genomics such as high throughput DNA sequencing, big data processing by machine learning algorithms and gene-editing techniques are expected to make precision medicine and gene-therapy a greater reality. However, this development will raise many important new issues, including ethical, moral, social and privacy issues. The field of exercise genomics has also advanced by incorporating these innovative technologies. There is therefore an urgent need for guiding references for sport and exercise genomics to allow the necessary advancements in this field of sport and exercise medicine, while protecting athletes from any invasion of privacy and misuse of their genomic information. Here, we update a previous consensus and develop a guiding reference for sport and exercise genomics based on a SWOT (Strengths, Weaknesses, Opportunities and Threats) analysis. This SWOT analysis and the developed guiding reference highlight the need for scientists/clinicians to be well-versed in ethics and data protection policy to advance sport and exercise genomics without compromising the privacy of athletes and the efforts of international sports federations. Conducting research based on the present guiding reference will mitigate to a great extent the risks brought about by inappropriate use of genomic information and allow further development of sport and exercise genomics in accordance with best ethical standards and international data protection principles and policies. This guiding reference should regularly be updated on the basis of new information emerging from the area of sport and exercise medicine as well as from the developments and challenges in genomics of health and disease in general in order to best protect the athletes, patients and all other relevant stakeholders.

To cite: Tanisawa K, Wang G, Seto J, *et al*. *Br J Sports Med* 2020;**54**:969–975.

INTRODUCTION

Recent advances in DNA sequencing technology enable the analysis of a large number of genomes[1]—a genome is a complete set of DNA, including all the genes of an organism. The generation of large data combined with new and improved methods of analysis, which includes machine learning and artificial intelligence, is collectively predicted to advance precision medicine considerably to facilitate optimal tailored medical therapies for the individual based on the individuals complete clinical and risk profiles which includes their l genomic information. This new reality has the potential to revolutionise healthcare by substantially enhancing the efficacy of treatment with a promise to significantly reduce the costs associated with healthcare provision.[2] To achieve the necessary progress in precision medicine, many countries have established large scale biobanks and are performing analyses on large datasets (table 1).[3] For example, the UK Biobank (2006–2010) has already genotyped approximately 500 000 participants using the UK BiLEVE Axiom Array and the UK Biobank Axiom Array and performed genome-wide association studies (GWASs) in the largest to date single population-based cohort involving more than 20 000 traits.[4 5] The 100 000 Genomes Project, launched in the UK in 2012, is the first national whole-genome sequencing project targeting National Health Service (NHS) patients to complete the sequencing of 100 000 whole genomes.[6]

These rapid advances in DNA sequencing technology have also introduced many new ethical and confidentiality issues such as reidentification of anonymised genotype data,[7] data ownership,[8] newborn screening[9] and incidental findings.[10] These advances and the anticipation of a true revolution in precision medicine have created a lively market for direct to consumer (DTC) genetic testing companies.[11] At present, the vast majority of company claims are more in line with future aspiration and promise rather than current evidence-based reality. Most DTC companies are too small to have any significant research and development (R&D) and therefore are solely dependent on the scientific community to generate new and clinically relevant data. There are also no guarantees that these companies will be allowed to freely exploit the data that will emerge due to patient confidentiality and data protection issues. It is likely that elaborate algorithms will be developed using big data processing methods and controlled by the larger companies that have the R&D resources to invest in the necessary analytical technology such as supercomputers, programmers and specialist bioinformaticians. At present, DNA sequencing technologies are able to generate data at a much faster rate than our ability to interpret and therefore appropriately exploit these data.

In addition to DNA sequencing technologies, gene-editing technology has also made significant

Table 1 Current large biobanks with populations over 100 000 individuals

Biobank	Cohort size	Phenotyping data	Genotyping platform	References
Million Veteran Program	550 000 (400,000 genotyped) with goal for 1 million individuals	Baseline survey data, EHR structured data	Affymetrix Axiom Biobank	Gaziano, 2016[122]
All of Us	Goal for 1 million individuals	Baseline physical examination, baseline survey data, sensor based observations (wearable devices), EHR structured data, social media	Not yet determined, whole genome sequencing likely to factor prominently from outset	Precision Medicine Working Group, 2015[123]
UK BioBank	502 632 genotyped	Web-based questionnaires, sensor based observations (wearable devices), EHR structured data	UK BiLEVE, UK Biobank Axiom	http://www.ukbiobank.ac.uk/
Kaiser: Research Program on Genes, Environment, and Health	257 686 (176 200 genotyped), with goal for 500 000 individuals	Baseline survey data, EHR structured data	Affymetrix Axiom Genome- Wide EUR Array	https://researchbank.kaiserpermanente.org/
Geisinger Health System MyCode Community Health Initiative	145 165 (92 455 genotyped)	EHR structured data	Illumina HumanExome array V1.1	Carey, 2016[124]
Vanderbilt: BioVU	225 000 genotyped	EHR structured data	Illumina Exome BeadChip, Illumina MEGA BeadChip	https://victr.vanderbilt.edu/pub/biovu/
China: Kadoorie Biobank	>500 000 (32 000 genotyped)	Baseline survey data, baseline physical exam, health insurance information	Affymetrix Axiom Biobank	http://www.ckbiobank.org/
Japan Biobank	200 000 genotyped	Baseline survey data, Annual review of incident disease	Multiple	https://biobankjp.org/
National Biobank of Korea	525 416 (genotyped number unclear)	Repeated surveys and exams	Not described	Cho, 2012[125]
deCode	160 000 genotyped	Geneologies, EHR structured data	Illumina Omni-1 Quad BeadChips	https://www.decode.com/
Finngen	Goal for 500 000 individuals	EHR structured data	Not described	https://www.finngen.fi/

This table was cited from Small *et al*.[3]
EHR, electronic health record.

advancements in recent years. In particular, clustered regularly interspaced short palindromic repeats (CRISPR) together with CRISPR-associated proteins (eg, Cas9, CasX, Cas12a and Cas13), known as CRISPR/Cas systems are poised to make gene-editing truly revolutionary by enabling easy, rapid and cost effective editing of DNA sequences.[12–15] Since inception in 1993[16] CRISPR technology has now advanced to the point of smart technology gels for drug delivery.[17] This pioneering approach, particularly CRISPR/Cas9, is already being used to develop lifesaving/altering gene therapy in monogenic diseases such as sickle cell disease,[18] Huntington's disease,[19] cystic fibrosis[20] and Duchenne muscular dystrophy[21] and poised to make big advances in the near future also in cancer treatment[22] This promising and effective tool also allows the editing of DNA sequences of human germlines.[23 24] CRISPR, however, is not without limitations. For example, insertion-deletions (INDELS) delivered through CRISPR/Cas9 mechanism have been shown to induce foreign mRNA or proteins in approximately 50% of cell lines through ribosomal entry, thereby causing mutations and reduced production of viable genes.[25] While gene-editing tools in human germline have been restricted to research and prohibited in human reproduction by all countries that have established gene editing regulations, CRISPR/Cas9 gene-editing was recently used for reproductive purposes with claims of generating the first gene-edited babies.[26] This was a direct breach of Chinese law and led to a joint statement of 122 Chinese scientists calling for urgent legislation against further 'direct human experimentation'.[27] There is therefore an urgent need to create the necessary regulatory framework to safeguard against the real threat of 'genetic pollution' of the human gene pool, a controversial term to describe the process of intentional, uncontrolled and potentially unlawful, introduction of genetic material for the purpose of increasing the 'fitness' of a population or subsample of a population.[28]

In the field of sport and exercise sciences and medicine, dissecting the relationship between genetic factors and health-related fitness, athletic performance, trainability and susceptibility for exercise-related health risks (eg, musculoskeletal injury) were previously attempted. Identification of specific sport and exercise-related genes are expected to be used for precision sports medicine to provide tailor-made training as well as to select optimal sports and/or other exercise activities for each individual. However, from previous candidate gene approaches and GWAS, there are very limited outcomes with clinical utility, and therefore a paradigm shift in sports genomics is urgently needed.[29–33] However, with the exception of the Genetic-Biological Physical Activity Consortium (GenBioPAC),[30] which is aimed at understanding genetic and other biological factors in the regulation of physical activity, there are no significant funded international consortia to meet this aim. Progress towards such a significant development in the field of sport and exercise genomics will require a paradigm shift in line with recent recommendations for international collaborations such as the Athlome Project Consortium (see www.athlomeconsortium.org) which was launched in 2015 for the advancement of 'omics' in exercise sciences and medicine.[31] The Athlome Consortium aims to collectively study the genotype and phenotype data currently available on elite athletes, in adaptation to exercise training and on exercise-related musculoskeletal injuries both from individual studies and from consortia worldwide. One of the consortium projects, the 1000 Athlomes Project, aims to sequence 1000 genomes of sprinters and distance runners of West and East African descent to clarify the genetic architecture of extreme athletic performance. To date, 72 world class Ethiopian and Kenyan distance runners have been sequenced and their genotype distribution has been compared with that of region-matched general population from the 1000 Genomes Project.[31 34]

As in other biomedical fields, large-scale genomic research is helping develop sport and exercise science and medicine in realising goals towards precision sports medicine and exercise prescription. Terms such as individualised response, personalised

medicine, stratified medicine, personalised prescription are increasingly being used as primarily inspirational concepts rather than current realities in terms of genomic technologies. Ross *et al* recently reported large interindividual differences in cardiorespiratory fitness (CRF) in response to standardised exercise and introduced measures to identify determinants and modifiers of CRF response.[32] However, these necessary advances also introduce many ethical problems that would be an obstacle to achieve precision sports medicine, particularly in terms of ethical and data protection issues. Given the rapid advances in gene-editing technology such as CRISPR/Cas9, and the fact that gene-edited babies are now a reality,[23 24] gene doping or creating talented sports children by using gene-editing technique combined with the knowledge of sport and exercise genomics would be realistic in the very near future, however currently immature at this stage. Therefore, advances in human genomics if left completely unregulated would inevitably become a 'weapon' that would threaten the health and well-being of athletes and the general worldwide population.

Considering the rapid changes in circumstances surrounding genomic research and the threats described above, ethics and data protection policies (eg, GDPR/Data Protection Act 2018 in the UK and EU) should continuously be updated. As the first sequencing studies of elite athletes are being conducted,[31 34] there is an urgent need for a guiding reference for sport and exercise genomics to allow the necessary advancement in sports genomics while also protecting athletes from any invasion of privacy and misuse of their genomic data. This reference guide attempted here is presented as a SWOT (Strengths, Weaknesses, Opportunities and Threats) analysis. This guiding reference will be regularly updated, based on new and emerging evidence, from the area of sport and exercise medicine and from valuable lessons being learnt from the developments and challenges in genomics of health and disease in general.[35 36]

SWOT ANALYSIS

Strengths

Previous twin and familial studies suggest that there is moderate heritability of 'sport and exercise-related traits' (eg, athletic performance, response to exercise training and fitness level).[37–40] Identification of genetic variants determining variabilities in sport and exercise-related traits may offer significant benefits to athletes and the general population. For example, the Heritage Family Study demonstrated a considerable heterogeneity in the change in maximal oxygen uptake (VO_2max) in response to a 20 week standardised exercise training programme (the range in training response: −114 to 1097 mL/min).[32 37] Similarly, large individual variabilities in response to resistance training,[41] and high-intensity interval training were also reported.[42] Collectively, evidence from human twin and family studies suggest that there are considerable interindividual differences in the response of CRF and other cardiometabolic traits to a given dose of exercise, and are partly dependent on genetic factors.[32] If the genetic variants that predict which type of exercise training is the most effective for each person are identified, then individualised therapeutic exercise programme can be used in early intervention and chronic disease prevention.

From the perspective of athletes, musculoskeletal injuries such as soft tissue disruption (eg, Achilles tendon injury, anterior cruciate ligament ruptures and shoulder dislocations), muscle strain, and stress fracture are serious medical conditions that inhibit regular training and may shorten an athlete's career. As genetic factors have been suggested to contribute to the susceptibility of musculoskeletal injuries,[43] identification of musculoskeletal injury-related genetic loci may provide information required to optimise training load that is tailored for training volume and intensity. Previous candidate gene approaches have demonstrated that several single nucleotide polymorphisms (SNPs) were associated with soft tissue ruptures,[44–46] muscle strain[47 48] and stress fracture.[49 50] However, these SNPs have no current clinical utility because they have been replicated in limited independent populations. Functional analysis of these SNPs is, however, needed to achieve a greater understanding of the mechanisms of susceptibility to musculoskeletal injuries.

Genetic variants associated with elite athlete status in various sporting disciplines may contribute to talent identification or selection of optimal sports to maximise the talent of specific athletes in the future, notwithstanding the serious ethical concerns that athletes should have the right to freely select sports they want to play regardless of their genes as well as the fact that athletic performance is polygenic (no single gene should be held accountable for athletic success). For example, Mikami *et al* reported that Japanese sprinters with the RR+RX genotype of alpha-actinin-3 (*ACTN3*) gene had significantly faster personal best times for the 100 m than those with XX genotype; however, no such association was found in the 400 m sprinters.[51] Nevertheless, given the polygenic nature of athletic performance and sports skill, talent identification or selection of optimal sports by using only limited genetic variants is unlikely to ever be a possibility.

Although cardiomyopathies and channelopathies (eg, hypertrophic cardiomyopathy, congenital long-QT syndrome)[52] are usually non-fatal diseases, they are the major causes of sudden cardiac death (SCD) in young athletes.[53] Most cases of SCD in young athletes are probably caused by the combination of inherited cardiomyopathy or channelopathy with intensive exercise training. Therefore, cardiovascular screening is essential to prevent SCD in young athletes. Currently, preparticipation screening of young athletes for prevention of SCD includes screening for family history and cardiovascular symptoms, physical examination, and often a 12-lead ECG, as recommended by the European Society of Cardiology.[54] Cardiac screening, which includes the ECG, has been shown to have a high sensitivity for detecting conditions at elevated risk of SCD such as cardiomyopathy or channelopathy and has been associated with reduced mortality in competitive athletes.[55] Nevertheless, accurate diagnosis of cardiomyopathy and channelopathy in athletes remains a challenge because of the difficulty of distinguishing between a so-called athlete's heart based on physiological adaptation to intense training and cardiac diseases.[56] However, most cases of inherited cardiomyopathy and channelopathy are monogenic and numerous causative mutations have been identified for each disorder.[57–59] Thus, genetic testing may be an important tool in the evaluation of athletes with abnormal cardiovascular screening, inconclusive cardiac imaging and in athletes with a family history of an inheritable cardiac disorder. In the future, genetic testing may also have a potential role in cardiac screening.

Identification of genetic variants associated with sport and exercise-related traits is of great importance in terms of understanding the molecular basis of trainability. For example, deficiency of *ACTN3*, the most replicated and studied sports performance-related gene,[60 61] turned out to influence metabolic enzyme activity in skeletal muscle and a shift in the properties of fast fibres towards those characteristics of slow twitch fibres.[62] These findings are consistent with that the null allele of the *ACTN3* p.R577X polymorphism (ie, a point mutation that usually results in a non-functional protein product) are

over-represented in endurance athletes than in power athletes in ancestrally diverse populations. Thus, identification of genetic variants associated with individual variabilities in sport and exercise-related traits could provide novel insights into molecular adaptations in skeletal muscle. Furthermore, integrating other 'omics' responses to exercise such as transcriptomics and proteomics will undoubtedly enhance our understanding of the mechanisms of adaptative response to exercise and its individual variability.

Weaknesses

Although a large number of studies have been conducted to identify sport and exercise-related genes, the findings are mostly inconclusive because of a lack of replication. Some studies reported completely opposite associations across different populations (eg, *ACE* gene).[63–65] Sport and exercise-related genes and genetic loci confirmed to influence the health status of the athlete have not been identified to date mainly due to the lack of replication in independent populations. Furthermore, GWAS of world class endurance athletes identified no genetic variants associated with extreme endurance performance at the genome-wide level of significance although several SNP associations might be missed in that study because low-density arrays (Illumina Cardio-MetaboChip) were used for genotyping without imputation.[66] Previous studies teach us that two major problems underlie the lack of discovery of novel sport and exercise-related genes.

One of the major problems is small sample size. Common SNPs associated with polygenic traits (including sport and exercise-related traits) generally show modest OR of 1.1–1.5[67]. For example, approximately 5500 cases and equal number of controls are needed to detect an OR of 1.2 at alpha error of 5.0×10^{-8} and power of 80% in a case-control GWAS if minor allele frequency is assumed to be 0.3.[68] The difficulty of recruiting a large number of elite athletes for sufficient power explains one of the bottlenecks in the discovery of novel variants of small effects associated with athletic performance. Another major problem is the deficiency in how the phenotype is being assessed. The factors shaping athletic performance are diverse. For example, endurance performance that is one of the more simple traits in sports, is shaped by VO_2max, VO_2 at the lactate threshold, economy of movement and other parameters.[69] However, each physiological marker of performance is also a complex trait, regulated by a network of genes and pathways. In most case-control studies of elite athletes, the physiological, anthropometric and biomechanical characteristics of the athletes are not well phenotyped. Consequently, the definition of 'elite athlete' in such studies is often ambiguous. The experience of participating in a world championship or national competition is usually used for defining elite athlete status.[70] However, there is no biological explanation for clearly distinguishing world class athletes from national level athletes. The opportunity and level of athletic achievement needed to participate in the Olympics or World Championships varies considerably depending on the country of origin. This raises the importance of 'phenomics', which is defined as the acquisition of high-dimensional phenotypic data on an organism-wide scale,[71] in the field of sport and exercise genomics.

Both large sample size and precise phenotyping are necessary to reduce the SE and increase statistical power to detect a significant SNP-trait association. However, it is difficult to perform comprehensive and precise phenotyping while keeping adequate sample size. For example, The Heritage Family Study, which is the only large-scale standardised exercise intervention study consisting of well phenotyped participants (n=720) to explore genetic variants associated with response to exercise training by using Illumina HumanCNV370-Quad v3.0 BeadChips (containing approximately 370 000 markers),[72] did not yield any SNPs associated with VO_2max response to exercise training at the genome-wide level of significance.[73] This study suggests that a sample size of less than 1000 is still insufficient despite a well standardised intervention protocol and precise phenotyping, while the use of a low-density array without imputation may also have contributed to the finding of no significant SNP associations. The development of technology such as multifaceted wearable devices[74] is needed for comprehensive and precise high-throughput phenotyping of sport and exercise-related traits while keeping adequate sample size for genetic association analyses. Furthermore, multiple large-scale cohorts with well phenotyped participants are needed to replicate and validate the genetic variants detected in a discovery cohort, which might require a substantial budget. In the current environment of a general 'research grant famine', it is often difficult to obtain funds in the field of sport and exercise science and medicine to perform such large-scale studies compared with medical science which directly contribute to health and disease prevention for the general population. Financial constraints may also prevent the introduction of large-scale genetic approaches into screening programmes.[75–78] Whether to screen and what precisely the preparticipation screening should comprise will be hotly debated for years to come, but in the meantime, progress in this field would be greatly advanced if in SCD cases in in sport, molecular considerations were part of the standardised routine autopsy[79] as well as the genetic screening of first-degree relatives.[80]

Understanding the genetic architecture of human athletic performance, the widely ranging sport phenotype and other sport and exercise-related traits is challenging because of its high complexity. Human athletic performance has long been assumed to be polygenic except for some rare cases where single mutation(s) confers extreme phenotype as in the example of a gain-of-function mutation erythropoietin receptor in the Olympic cross-country skiing gold medalist.[81] An accumulation of common variants with small effect size has been suggested to explain a large proportion of phenotypic variance of complex traits.[82] However, to date, only a very small number of common variants have been reported to be associated with sport and health-related fitness phenotypes.[83] Furthermore, recent advances in genomics have revealed that an individual carries approximately 40 000–200 000 of rare variants (minor allele frequency of <0.5%) per individual genome,[84] which can help explain the phenotypic variance of complex trait,[85] justifying the necessity to adopt whole genome sequencing technology in the field of sport and exercise genomics. In addition to single nucleotide variants in the gene regions, other type of genomic variation such as structural variation (eg, copy number variation, large insertion and deletion),[86] variants in non-coding RNA (eg, micro RNA, long non-coding RNA)[87] may also contribute to the complexity of the athletic phenotype. Integration of various types of genomic variation by multiomics approach will be needed to fully elucidate the complexity of athletic performance.

Opportunities

There is considerable commercial opportunity associated with the use of genomics in the sport and exercise sciences. A means to attract research funding for large-scale sport and exercise genomic studies is to collaborate with industry given the recent

increase in public interest and use of commercial genetic testing. For example, one of the largest DTC genetic testing companies 23andMe, has collected genetic data from approximately 5 million consumers. Although the primary service of 23andMe is to provide genetic health risks and ancestry information to consumers, they also conduct research by using genotype and phenotype data obtained from consumers that are collected through internet surveys.[88] 23andMe analyse these data[89] and also share the data with academic institutions to enhance large-scale genetic association studies. Several research groups have managed to identify additional loci for various diseases or traits using these large datasets.[90–92] Furthermore, 23andMe sell their data to the pharmaceutical industry for drug development research[93] as well as support scientists through collaboration agreements whereby scientists collect data in the field for analysis by 23andMe with the fundamental aim of elucidating a greater understanding of the diversity of genomics data globally. In this way, scientists and the wider community benefit from this additional access to grant funding.[94] In addition, new research investments by industry into genomics of sport and exercise has real potential to impact the field of genetics of disease with particular emphasis on lifestyle-related disorders by helping, for example, identify risk factors associated with sedentary lifestyles for wider society and public health gain. Such collaborations and commercial partnerships with industry should be pursued, although with care. To date, 23andMe is the only personal genomics and biotechnology company to offer these opportunities but others are expected soon to follow this example given their success.

Threats

Collaboration with industry for genomic research in the sport and exercise sciences is not without serious threats. Industries including DTC genetic testing companies may support researchers with the expectation of handing over some of the intellectual property that may be generated during the study life cycle. Commercial pressure in most cases results in the premature exploitation of data that have limited or no scientific bases given no or limited replication and validation. For example, DNAFit, a DTC genetic testing company active in the UK, has recently performed a GWAS of sprint performance in collaboration with Russian and Polish scientists, aimed at identifying novel genetic markers for their genetic testing product.[95] These authors reported several SNPs to be associated with sprint performance; however, the clinical significance of these results is unknown given the small sample size of the discovery cohort and the inappropriate replication study (ie, small sample size, different outcomes and heterogeneous characteristics of participants among the cohorts). Many DTC genetic testing companies have already offered genetic testing products for predicting athletic performance and talent identification although sport and exercise genomics has provided very limited evidence and predicting athletic performance and talent identification by using genetic information is almost impossible to this date.[96] Some athletes, coaches and parents of young individuals may believe the results of genetic testing regardless of the accuracy and quality of genetic testing products commercially available at present. Use of such unproven technology can lead to incorrect decisions such as inappropriate early specialisation for sports, inappropriate training, genetic discrimination and even increased health risks. Genetic testing should therefore be provided with appropriate genetic counselling as described in the statement of The European Society of Human Genetics[97] and The American Society of Human Genetics.[98] The vast majority of DTC companies sell genetic testing to consumers without providing adequate genetic counselling.[99] Thus, collaboration with industry has the potential danger to misuse the data despite the intention of scientists, consequently misleading athletes, their coaches and families.

In terms of the privacy and data protection, reidentification of anonymised genotype data has become a real concern. For example, US law enforcement authorities have begun exploiting genetic databases and publicly available family trees to identify suspects via distant familial relatives and have succeeded in arresting numerous suspects.[7] This practical use of publicly available genetic information for criminal investigation raises awareness that reidentification of anonymised genotype data are already technically possible. Individual genome information of elite athletes must be of special interest to many people. There is therefore a possibility that someone may attempt reidentifying anonymised genotype data of elite athletes to abuse this data. Given the recent rapid development of artificial intelligence, reidentification of anonymised genotype data would be much easier in the near future. A flow of genetic information from a DTC genetic testing company to a third party is another potential problem in privacy and data protection. DTC genetic testing companies provide genetic data to third-party scientists or pharmaceutical industry for research purposes without consumer's explicit consent. Although several DTC genetic testing companies obtain additional consent for the secondary use of data, the majority of them do not consistently meet international transparency guidelines related to privacy and secondary use of data.[100] Furthermore, the use of third-party interpretation services also increases the risk of privacy invasion and misuse of data. Consumers can download their raw genetic data from DTC genetic testing company website and freely upload it to third-party interpretation services for further explanation of genetic data. Because the data usage and privacy policy is less prominent in the third-party interpretation services than in DTC genetic testing companies,[101] the risk of data exploitation by someone could be high, as US law enforcement authorities have already exploited it for criminal investigation.[7] Thus, considering the feasibility of reidentification of genotype data and unexpected use of this, a lack of transparency in the provision of information to consumers is a serious problem because it inhibits them from recognising the threat of privacy invasion.

Given the development of gene-editing technology such as CRISPR/Cas9,[102] genetic variants determining athletic performance and elite athlete status may be used for gene-doping or creating talented sports children in the future. In fact, gene-editing of Myostatin in zygotes successfully enhanced muscle hypertrophy in several adult mammals.[103–105] Even in humans, researchers have used gene-editing techniques targeting human adult cell for disease therapy[18–21] and human germlines.[23 24] A pertinent recent example is the aforementioned revelations of the use of CRISPR/Cas9 gene editing technology to delete both copies of the *CCR5* gene in embryos to give twin babies resistance to HIV infection in violation of the laws and regulations in respective countries. This should not have been performed because there is no consensus on how to counsel 'gene-edited individuals' as well as a limitation in understanding of the long-term effects of gene-editing on mature body. In fact, although basic research involving gene-editing in human germlines has been admitted in several countries (eg, UK, USA, Sweden, China and Japan),[106–110] human gene-editing for reproduction is prohibited by law or regulation in many countries.[111] The fact that there are gene-edited babies alive today confirms that gene-edited human babies are already technically possible. Designing athletes with extraordinary athletic performance by

using gene-editing technique would be a real threat in terms of keeping sport fair, clean and protecting athlete health.

Guiding reference for sport and exercise genomics

The present SWOT analysis suggests that sport and exercise genomics has the potential to contribute to the health and well-being of athletes as well as real and necessary advances in the field of sport and exercise sciences and medicine. Large-scale studies in collaboration with industry may help to provide sufficient scientific evidence to adequately and ethically use genetic information of athletes, mainly to protect their health status. On the other hand, there are many potential dangers associated with the use of genomics in sport and exercise medicine such as reidentification of anonymised genotype data, inaccurate genetic testing based on insufficient evidence, discrimination and gene doping by using novel gene-editing technique. These threats must be addressed to protect privacy and health of athletes and to keep sports clean.

We therefore propose the following rules of conduct:

All research on sport and exercise genomics should be conducted in strict accordance with the local university and any associated medical trust ethical guidelines, relevant Data Protection Acts and EU General Data Protection Regulations (GDPR) or similar instruments in other regions of the world. Given the transnational nature of genomic work, policies of sponsors and partners must also be GDPR compliant.

Scientists must establish strict rules about acquisition of data, data flow management, anonymisation, security and data release policy before starting the project and collaboration with industries and other research partners.

Scientists must not receive funds from industries to develop the project unless industries completely agree with the scientist's host institute rules based on independent ethical committees and are prepared to sign an comprehensive agreement that include ethics, data protection, legal safe guards, intellectual property.

Scientists must respect rules regarding the acquisition of biological material to prevent exploitation of vulnerable individuals and societies.

Scientists must not receive funds from industries if they aim to exploit the data in return for giving funds.

Scientists must not release any data to industries and other research groups unless there are strict rules about data flow management, anonymisation, security and data release policy. Scientists should handle the data to protect individuals from privacy invasion and abuse of their personal data.

All experiments and analyses should be performed in house or the analysis process and data management are clearly protected by rules set out by the service providers and agreed by the researchers before commencing the analysis to minimise the risk of data leakage, privacy invasion and misuse of personal data.

The IP landscape on CRISPR/Cas9 is complex and constantly evolving. Issues surrounding intellectual property rights (IPRs) have broad international legal implications; however, for the purpose of these guidelines, consideration is given to domestic legislation under the Patent Act 1977 (as amended 2005) and the European Patent Convention 2000. Legally, IPRs relating to the research project belong to the institution of the principal investigator although the inventor may be entitled to compensation. However, a patent holder may infringe their own IPRs where rights are transferred to the licensee; therefore, explicit contract terms will need to be negotiated ensuring IPRs, where possible, remain with the academic institution.

Feedback of genetic data to individuals is not recommended unless the accuracy and precision of prediction by genetic information is assured by replication and validation studies. Scientists must minimise the risk of misinterpretation of genetic information by proper genetic counselling if feedback of genetic data is required or beneficial for the individuals (eg, incidental findings of mutations causing genetic disorders).

Scientists in sport and exercise genomics should perform replication and validation studies as much as possible to verify the results to improve our understanding of the scientific merit of the findings.

Scientists should keep enhancing their knowledge of ethics and data protection policies pertaining to existing big genome projects.

Scientists should try to implement best practice and develop a secure encrypted domain to reduce the risk of data leakage.

There should be regular interactions between scientists and practicing sports medicine doctors or practitioners to facilitate the transfer knowledge of any advancement in the arena for example on tools such as gene editing and gene therapy—or any other tools to reduce risk and promote the health of the athlete.

Scientists must not use gene editing techniques in somatic human cell aimed at enhancing athletic performance. In addition, scientists must not use gene editing techniques to modify DNA in human germlines for creating talented sports children. Scientists should keep learning from current guidelines for gene editing and gene therapy to establish the regulation to protect sports and athletes from threat of gene-doping and creating talented sports children.

Sport and exercise genomics is in transition from focused research performed by single research groups to large-scale discovery research involving many research groups and industry partners.[29 33] The advancements in sport and exercise genomics parallel the increase in the risk of data leakage, privacy invasion and abuse of personal data. At this moment, without strict rules about data flow management, anonymisation, security and data release policy that is standard practice in the large biobank studies,[112 113] releasing any genotype and phenotype data to industry, other research groups and public databases is not recommended. All scientists of sport and exercise genomics need to be well-versed in ethics and data protection policy to protect individuals from threats of privacy invasion and abuse of personal data in preparation for the era of large-scale collaborative science.

The application of CRISPR/Cas9 gene editing techniques to skeletal muscle as well as hematopoietic stem cells for treatment of monogenic diseases is becoming more commonplace.[18 21] Given these rapid advances in gene editing, it is expected and totally desirable that these techniques are harnessed by the sports medicine physician to treat sports-related injuries. This broader application of gene editing techniques to sports medicine will inevitably result in gene-doping being a real prospect.[114] Knowledge gained from gene-editing research for disease therapy could be misused for enhancement of athletic performance. To the best of our knowledge, no gene editing techniques have been applied in healthy individuals to enhance athletic performance. This threat poses new ethical dilemmas and hence the urgent need for gene-editing guidelines and regulation constantly updated to deal with all eventualities. It is the responsibility of those involved in the field of sport and exercise sciences and medicine to keep abreast of the gene-editing guidelines.[115] It is necessary to prioritise research in antidoping with particular reference to gene-doping. Although gene doping is already prohibited on the list of banned doping agents developed by the World Anti-Doping

Agency,[116] robust and effective antidoping measures to detect the use of gene doping have not been developed. Several PCR-based strategies to detect vector-mediated gene transfer of several candidate genes (eg, vascular endothelial growth factor, erythropoietin, insulin-like growth factors 1, growth hormone) for gene doping have been developed.[117–119] However, gene transfer of unexpected target gene cannot be detected by using these candidate gene approaches. Furthermore, the strategy to detect gene doping by CRISPR/Cas9 based on DNA modification has not been developed so far. Omics technologies have been shown to enhance the detection of blood doping,[120 121] and these cutting-edge technologies could be further harnessed to develop effective methods for the detection of gene doping in sport. Funding institutions should be encouraged to offer grants for further developments in the rapidly emerging field of antidoping strategies providing the antidoping laboratories with the tools to significantly improve their abilities.

CONCLUDING REMARKS

The present guidelines in sport and exercise genomics developed following an extensive SWOT analysis, advocates the need for clear and universal standards as they relate to the collection, management and storage of DNA/data with the overriding objective to protect individuals from privacy invasion and misuse of genomic information. Given the increased availability of high-throughput genomic information, there is an urgent need for such a guiding reference in sport and exercise genomics with a clear and consistent data handling and release policy for all individuals who potentially handle any genetic information. It is essential that sports physicians, scientists and all those involved in supporting the athlete keep abreast with new developments in genomics including new technologies and methods such as CRISPR/Cas9 and are well informed of the laws and regulations pertaining to the collection, storage and use of genetic data. Given the rapidly advancing field of sports genomics, regular updates to this guiding reference will be needed in order to best protect the athletes and all the relevant stakeholders. Conducting research in accordance with the present guiding reference will reduce the threat brought about by inappropriate use of genomic information and allow further development of sport and exercise genomics in accordance with ethical principles.

Author affiliations
[1]Faculty of Sport Sciences, Waseda University, Tokorozawa, Japan
[2]Collaborating Centre of Sports Medicine, University of Brighton, Eastbourne, UK
[3]Murdoch Children's Research Institute, Parkville, Victoria, Australia
[4]Department of Paediatrics, University of Melbourne, Parkville, Victoria, Australia
[5]School of Engineering and Materials Science, Queen Mary University of London, London, UK
[6]Department for Health and Longevity Research, National Institutes of Biomedical Innovation, Health and Nutrition, Tokyo, Japan
[7]Department of Neuroscience and Physiology, Center for Health and Performance, Goteborg University, Göteborg, Sweden
[8]Sahlgrenska University Hospital/Ostra, Göteborg, Sweden
[9]Unit of Endocrinology, Department of Movement, Human and Health Sciences, University of Rome "Foro Italico", Rome, Italy
[10]Sport Medical Center, Japan Institute of Sports Sciences, Tokyo, Japan
[11]Department of Sport Medicine, Humboldt University and Charité University School of Medicine, Berlin, Germany
[12]UCT Research Unit for Exercise Science and Sports Medicine, Cape Town, South Africa
[13]Department for Health, University of Bath, Bath, UK
[14]I.M. Sechenov First Moscow State Medical University (Sechenov University), Ministry of Health of Russia, Moscow, Russian Federation
[15]Moscow Research and Practical Center for Medical Rehabilitation, Restorative and Sports Medicine, Moscow Healthcare Department, Moscow, Russian Federation
[16]Defence Medical Rehabilitation Centre, Stanford Hall, Loughborough, UK
[17]British Association of Sport and Exercise Medicine, Doncaster, UK
[18]Italian Federation of Sports Medicine (FMSI), Rome, Italy
[19]Unit of International Law, Department of Constitutional, International and European Law, University of Salzburg, Salzburg, Salzburg, Austria
[20]Institute of Sports Science, University of Vienna, Vienna, Austria
[21]Austrian Institute of Sports Medicine, Vienna, Austria
[22]Sport Medicine Unit, Department of Movement, Human and Health Sciences, University of Rome "Foro Italico", Rome, Italy

Twitter Michiko Dohi @rinsho, James Lee John Bilzon @JBilzon and Theodora Papadopoulou @Dora_Sportmed

Contributors All authors contributed significantly to this manuscript and all fulfilled the requirements for authorship as declared in the BJSM instructions to authors.

Funding This study was supported in part by the Grant-in-Aid for JSPS Fellows from the Ministry of Education, Culture, Sports, Science, and Technology (16J09593 to KT) and the JSPS KAKENHI Grant Number (16H01872 to MT) (Tokyo, Japan). Open access was paid by MEXT-Supported Program for the Strategic Research Foundation at Private Universities, 2015-2019 from the Ministry of Education, Culture, Sports, Science and Technology (S1511017).

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iDs
Michiko Dohi http://orcid.org/0000-0002-1126-7849
James Lee John Bilzon http://orcid.org/0000-0002-6701-7603
Yannis Pitsiladis http://orcid.org/0000-0001-6210-2449

REFERENCES

1 Shendure J, Balasubramanian S, Church GM, *et al*. Dna sequencing at 40: past, present and future. *Nature* 2017;550:345–53.
2 Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med Overseas Ed* 2015;372:793–5.
3 Small AM, O'Donnell CJ, Damrauer SM. Large-Scale genomic biobanks and cardiovascular disease. *Curr Cardiol Rep* 2018;20:22.
4 Bycroft C, Freeman C, Petkova D, *et al*. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018;562:203–9.
5 Elsworth B. Collaboratively analysing thousands of phenotypic traits using UK Biobank data, Google Apis and HPC, 2018. Available: https://ieup4.blogs.bristol.ac.uk/2018/10/01/ukb_gwas/ [Accessed 1 Aug 2019].
6 GOV.UK. 100,000 whole genomes sequenced in the NHS, 2018. Available: https://www.gov.uk/government/news/100000-whole-genomes-sequenced-in-the-nhs [Accessed 1 Aug 2019].
7 Erlich Y, Shor T, Pe'er I, *et al*. Identity inference of genomic data using long-range familial searches. *Science* 2018;362:690–4.
8 Gurwitz D, Bregman-Eschet Y. Personal genomics services: whose genomes? *Eur J Hum Genet* 2009;17:883–9.
9 Goldenberg AJ, Sharp RR. The ethical hazards and programmatic challenges of genomic newborn screening. *JAMA* 2012;307:461–2.
10 Green RC, Berg JS, Grody WW, *et al*. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med* 2013;15:565–74.
11 Phillips AM. 'Only a click away - DTC genetics for ancestry, health, love…and more: A view of the business and regulatory landscape'. *Appl Transl Genom* 2016;8:16–22.
12 Mali P, Yang L, Esvelt KM, *et al*. Rna-Guided human genome engineering via Cas9. *Science* 2013;339:823–6.
13 Cox DBT, Gootenberg JS, Abudayyeh OO, *et al*. Rna editing with CRISPR-Cas13. *Science* 2017;358:1019–27.
14 Chen JS, Ma E, Harrington LB, *et al*. CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* 2018;360:436–9.
15 Liu J-J, Orlova N, Oakes BL, *et al*. CasX enzymes comprise a distinct family of RNA-guided genome editors. *Nature* 2019;566:218–23.
16 Mojica FJM, Juez G, Rodriguez-Valera F. Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites. *Mol Microbiol* 1993;9:613–21.
17 Callaway E. Crispr cuts turn gels into biological watchdogs. *Nature* 2019;572:574.

18 Dever DP, Bak RO, Reinisch A, *et al*. Crispr/Cas9 β-globin gene targeting in human haematopoietic stem cells. *Nature* 2016;539:384–9.
19 Shin JW, Kim K-H, Chao MJ, *et al*. Permanent inactivation of Huntington's disease mutation by personalized allele-specific CRISPR/Cas9. *Hum Mol Genet* 2016;25:4566–76.
20 Firth AL, Menon T, Parker GS, *et al*. Functional gene correction for cystic fibrosis in lung epithelial cells generated from patient iPSCs. *Cell Rep* 2015;12:1385–90.
21 Ousterout DG, Kabadi AM, Thakore PI, *et al*. Multiplex CRISPR/Cas9-based genome editing for correction of dystrophin mutations that cause Duchenne muscular dystrophy. *Nat Commun* 2015;6:6244.
22 Zhan T, Rindtorff N, Betge J, *et al*. Crispr/Cas9 for cancer research and therapy. *Semin Cancer Biol* 2018.
23 Liang P, Xu Y, Zhang X, *et al*. Crispr/Cas9-Mediated gene editing in human tripronuclear zygotes. *Protein Cell* 2015;6:363–72.
24 Ma H, Marti-Gutierrez N, Park S-W, *et al*. Correction of a pathogenic gene mutation in human embryos. *Nature* 2017;548:413–9.
25 Tuladhar R, Yeu Y, Tyler Piazza J, *et al*. CRISPR-Cas9-based mutagenesis frequently provokes on-target mRNA misregulation. *Nat Commun* 2019;10:4056.
26 Cyranoski D, Ledford H. Genome-Edited baby claim provokes international outcry. *Nature* 2018;563:607–8.
27 China Daily. Investigation of human gene editing claim justified: China daily editorial, 2018. Available: http://www.chinadaily.com.cn/a/201811/27/WS5bfd3203a310eff30328b5a3.html [Accessed 10 Sep 2019].
28 Waller DM. Genetic rescue: a safe or risky BET? *Mol Ecol* 2015;24:2595–7.
29 Eynon N, Voisin S, Lucia A, *et al*. Preface: genomics and biology of exercise is undergoing a paradigm shift. *BMC Genomics* 2017;18:825.
30 Lightfoot JT, DEG EJC, Booth FW, *et al*. Biological/Genetic regulation of physical activity level: consensus from GenBioPAC. *Med Sci Sports Exerc* 2018;50:863–73.
31 Pitsiladis YP, Tanaka M, Eynon N, *et al*. Athlome Project Consortium: a concerted effort to discover genomic and other "omic" markers of athletic performance. *Physiol Genomics* 2016;48:183–90.
32 Ross R, Goodpaster BH, Koch LG, *et al*. Precision exercise medicine: understanding exercise response variability. *Br J Sports Med* 2019.
33 Tanaka M, Wang G, Pitsiladis YP. Advancing sports and exercise genomics: moving from hypothesis-driven single study approaches to large multi-omics collaborative science. *Physiol Genomics* 2016;48:173–4.
34 Tanana M, Tanisawa K, Wang G, *et al*. *The 1000 Athlomes project for 2020 summer Olympics in Tokyo. 35th FIMS world Congress of sports medicine*. Rio de Janeiro, 2018.
35 Patch C, Middleton A. Genetic counselling in the era of genomic medicine. *Br Med Bull* 2018;126:27–36.
36 Price WN. 2Nd, Cohen Ig. privacy in the age of medical big data. *Nat Med* 2019;25:37–43.
37 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.
38 De Moor MHM, Spector TD, Cherkas LF, *et al*. Genome-Wide linkage scan for athlete status in 700 British female DZ twin pairs. *Twin Res Hum Genet* 2007;10:812–20.
39 Miyamoto-Mikami E, Zempo H, Fuku N, *et al*. Heritability estimates of endurance-related phenotypes: A systematic review and meta-analysis. *Scand J Med Sci Sports* 2018;28:834–45.
40 Zempo H, Miyamoto-Mikami E, Kikuchi N, *et al*. Heritability estimates of muscle strength-related phenotypes: a systematic review and meta-analysis. *Scand J Med Sci Sports* 2017;27:1537–46.
41 Ahtiainen JP, Walker S, Peltonen H, *et al*. Heterogeneity in resistance training-induced muscle strength and mass responses in men and women of different ages. *Age* 2016;38:10.
42 Yan X, Eynon N, Papadimitriou ID, *et al*. The gene smart study: method, study design, and preliminary findings. *BMC Genomics* 2017;18:821.
43 Collins M, September AV, Posthumus M. Biological variation in musculoskeletal injuries: current knowledge, future research and practical implications. *Br J Sports Med* 2015;49:1497–503.
44 Collins M, Posthumus M, Schwellnus MP. The COL1A1 gene and acute soft tissue ruptures. *Br J Sports Med* 2010;44:1063–4.
45 Posthumus M, September AV, O'Cuinneagain D, *et al*. The *COL5A1* Gene Is Associated With Increased Risk of Anterior Cruciate Ligament Ruptures in Female Participants. *Am J Sports Med* 2009;37:2234–40.
46 Posthumus M, September AV, O'Cuinneagain D, *et al*. The association between the COL12A1 gene and anterior cruciate ligament ruptures. *Br J Sports Med* 2010;44:1160–5.
47 Miyamoto N, Miyamoto-Mikami E, Hirata K, *et al*. Association analysis of the *ACTN3* R577X polymorphism with passive muscle stiffness and muscle strain injury. *Scand J Med Sci Sports* 2018;28:1209–14.
48 Kumagai H, Miyamoto-Mikami E, Hirata K, *et al*. Esr1 rs2234693 polymorphism is associated with muscle injury and muscle stiffness. *Med Sci Sports Exerc* 2019;51:19–26.
49 Varley I, Hughes DC, Greeves JP, *et al*. RANK/RANKL/OPG pathway: genetic associations with stress fracture period prevalence in elite athletes. *Bone* 2015;71:131–6.
50 Varley I, Hughes DC, Greeves JP, *et al*. The association of novel polymorphisms with stress fracture injury in elite athletes: further insights from the SFEA cohort. *J Sci Med Sport* 2018;21:564–8.
51 Mikami E, Fuku N, Murakami H, *et al*. Actn3 R577X genotype is associated with sprinting in elite Japanese athletes. *Int J Sports Med* 2014;35:172–7.
52 Basavarajaiah S, Shah A, Sharma S. Sudden cardiac death in young athletes. *Heart* 2007;93:287–9.
53 Semsarian C, Sweeting J, Ackerman MJ. Sudden cardiac death in athletes. *BMJ* 2015;350:h1218.
54 Corrado D, Pelliccia A, Bjørnstad HH, *et al*. Cardiovascular pre-participation screening of young competitive athletes for prevention of sudden death: proposal for a common European protocol. consensus statement of the study group of sport cardiology of the Working group of cardiac rehabilitation and exercise physiology and the Working group of myocardial and pericardial diseases of the European Society of cardiology. *Eur Heart J* 2005;26:516–24.
55 Corrado D, Basso C, Pavei A, *et al*. Trends in sudden cardiovascular death in young competitive athletes after implementation of a preparticipation screening program. *JAMA* 2006;296:1593–601.
56 Chandra N, Bastiaenen R, Papadakis M, *et al*. Sudden cardiac death in young athletes: practical challenges and diagnostic dilemmas. *J Am Coll Cardiol* 2013;61:1027–40.
57 Charron P, Arad M, Arbustini E, *et al*. Genetic counselling and testing in cardiomyopathies: a position statement of the European Society of cardiology Working group on myocardial and pericardial diseases. *Eur Heart J* 2010;31:2715–26.
58 Tester DJ, Ackerman MJ. Genetic testing for potentially lethal, highly treatable inherited cardiomyopathies/channelopathies in clinical practice. *Circulation* 2011;123:1021–37.
59 Mayosi BM, Fish M, Shaboodien G, *et al*. Identification of Cadherin 2 (*CDH2*) Mutations in Arrhythmogenic Right Ventricular Cardiomyopathy. *Circulation* 2017;10.
60 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.
61 Del Coso J, Hiam D, Houweling P, *et al*. More than a 'speed gene': ACTN3 R577X genotype, trainability, muscle damage, and the risk for injuries. *Eur J Appl Physiol* 2019;119:49–60.
62 MacArthur DG, Seto JT, Raftery JM, *et al*. Loss of ACTN3 gene function alters mouse muscle metabolism and shows evidence of positive selection in humans. *Nat Genet* 2007;39:1261–5.
63 Amir O, Amir R, Yamin C, *et al*. The *ACE* deletion allele is associated with Israeli elite endurance athletes. *Exp Physiol* 2007;92:881–6.
64 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.
65 Wang G, Mikami E, Chiu L-L, *et al*. Association analysis of ACE and ACTN3 in elite Caucasian and East Asian swimmers. *Med Sci Sports Exerc* 2013;45:892–900.
66 Rankinen T, Fuku N, Wolfarth B, *et al*. No evidence of a common DNA variant profile specific to world class endurance athletes. *PLoS One* 2016;11:e0147330.
67 Manolio TA, Collins FS, Cox NJ, *et al*. Finding the missing heritability of complex diseases. *Nature* 2009;461:747–53.
68 Chapman K, Ferreira T, Morris A, *et al*. Defining the power limits of genome-wide association scan meta-analyses. *Genet Epidemiol* 2011;35:781–9.
69 Bassett DR, Jr., Howley ET. Limiting factors for maximum oxygen uptake and determinants of endurance performance. *Med Sci Sports Exerc* 2000;32:70–84.
70 MacArthur DG, North KN. Genes and human elite athletic performance. *Hum Genet* 2005;116:331–9.
71 Houle D, Govindaraju DR, Omholt S. Phenomics: the next challenge. *Nat Rev Genet* 2010;11:855–66.
72 Bouchard C, Leon AS, Rao DC, *et al*. The heritage family study. aims, design, and measurement protocol. *Med Sci Sports Exerc* 1995;27:721–9.
73 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.
74 Muniz-Pardos B, Sutehall S, Gellaerts J, *et al*. Integration of wearable sensors into the evaluation of running economy and foot mechanics in elite runners. *Curr Sports Med Rep* 2018;17:480–8.
75 Dhutia H, Malhotra A, Gabus V, *et al*. Cost Implications of Using Different ECG Criteria for Screening Young Athletes in the United Kingdom. *J Am Coll Cardiol* 2016;68:702–11.
76 McKinney J, Lithwick DJ, Morrison BN, *et al*. Detecting underlying cardiovascular disease in young competitive athletes. *Can J Cardiol* 2017;33:155–61.
77 Riding NR, Sharma S, Salah O, *et al*. Systematic echocardiography is not efficacious when screening an ethnically diverse cohort of athletes in West Asia. *Eur J Prev Cardiol* 2015;22:263–70.
78 Vessella T, Zorzi A, Merlo L, *et al*. The Italian preparticipation evaluation programme: diagnostic yield, rate of disqualification and cost analysis. *Br J Sports Med* 2019:bjsports-2018-100293.

79 Harmon KG, Drezner JA, Wilson MG, *et al*. Incidence of sudden cardiac death in athletes: a state-of-the-art review. *Br J Sports Med* 2014;48:1185–92.
80 Thiene G. Sudden cardiac death in the young: a genetic destiny? *Clin Med* 2018;18:s17–23.
81 de la Chapelle A, Traskelin AL, Juvonen E. Truncated erythropoietin receptor causes dominantly inherited benign human erythrocytosis. *Proc Natl Acad Sci U S A* 1993;90:4495–9.
82 Golan D, Lander ES, Rosset S. Measuring missing heritability: inferring the contribution of common variants. *Proc Natl Acad Sci U S A* 2014;111:E5272–81.
83 Sarzynski MA, Loos RJF, Lucia A, *et al*. Advances in exercise, fitness, and performance genomics in 2015. *Med Sci Sports Exerc* 2016;48:1906–16.
84 Auton A, Brooks LD, Durbin RM, *et al*. A global reference for human genetic variation. *Nature* 2015;526:68–74.
85 Wainschtein P, Jain DP, Yengo L, *et al*. Recovery of trait heritability from whole genome sequence data. *ESPE* 2019.
86 Conrad DF, Pinto D, Redon R, *et al*. Origins and functional impact of copy number variation in the human genome. *Nature* 2010;464:704–12.
87 Taft RJ, Pang KC, Mercer TR, *et al*. Non-Coding RNAs: regulators of disease. *J Pathol* 2010;220:126–39.
88 23andMe. Research. Available: https://research.23andme.com/research/ [Accessed 1 Aug 2019].
89 Hu Y, Shmygelska A, Tran D, *et al*. Gwas of 89,283 individuals identifies genetic variants associated with self-reporting of being a morning person. *Nat Commun* 2016;7:10448.
90 Chang D, Nalls MA, Hallgrímsdóttir IB, *et al*. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat Genet* 2017;49:1511–6.
91 Ferreira MA, Vonk JM, Baurecht H, *et al*. Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. *Nat Genet* 2017;49:1752–7.
92 Tsoi LC, Stuart PE, Tian C, *et al*. Large scale meta-analysis characterizes genetic architecture for common psoriasis associated variants. *Nat Commun* 2017;8:15382.
93 GlaxoSmithKline. Gsk and 23andMe sign agreement to leverage genetic insights for the development of novel medicines, 2018. Available: https://www.gsk.com/en-gb/media/press-releases/gsk-and-23andme-sign-agreement-to-leverage-genetic-insights-for-the-development-of-novel-medicines/ [Accessed 1 Aug 2019].
94 23andMe. 23andMe populations collaborations program, 2018. Available: https://research.23andme.com/populations-collaborations/ [Accessed 1 Aug 2019].
95 Pickering C, Suraci B, Semenova EA, *et al*. A genome-wide association study of sprint performance in elite youth football players. *J Strength Cond Res* 2019;33:2344–51.
96 Webborn N, Williams A, McNamee M, *et al*. Direct-To-Consumer genetic testing for predicting sports performance and talent identification: consensus statement. *Br J Sports Med* 2015;49:1486–91.
97 European Society of Human Genetics. Statement of the ESHG on direct-to-consumer genetic testing for health-related purposes. *Eur J Hum Genet* 2010;18:1271–3.
98 Hudson K, Javitt G, Burke W, *et al*. ASHG Statement* on direct-to-consumer genetic testing in the United States. *Obstet Gynecol* 2007;110:1392–5.
99 Harris A, Kelly SE, Wyatt S. Counseling customers: emerging roles for genetic counselors in the direct-to-consumer genetic testing market. *J Genet Couns* 2013;22:277–88.
100 Laestadius LI, Rich JR, Auer PL. All your data (effectively) belong to us: data practices among direct-to-consumer genetic testing firms. *Genet Med* 2017;19:513–20.
101 Badalato L, Kalokairinou L, Borry P. Third Party interpretation of raw genetic data: an ethical exploration. *Eur J Hum Genet* 2017;25:1189–94.
102 Cong L, Ran FA, Cox D, *et al*. Multiplex genome engineering using CRISPR/Cas systems. *Science* 2013;339:819–23.
103 Crispo M, Mulet AP, Tesson L, *et al*. Efficient generation of myostatin knock-out sheep using CRISPR/Cas9 technology and microinjection into zygotes. *PLoS One* 2015;10:e0136690.
104 Lv Q, Yuan L, Deng J, *et al*. Efficient generation of myostatin gene mutated rabbit by CRISPR/Cas9. *Sci Rep* 2016;6:25029.
105 Wang K, Ouyang H, Xie Z, *et al*. Efficient generation of myostatin mutations in pigs using the CRISPR/Cas9 system. *Sci Rep* 2015;5:16623.
106 Callaway E. Gene-editing research in human embryos gains momentum. *Nature* 2016;532:289–90.
107 Cyranoski D. Japan set to allow gene editing in human embryos. *Nature* 2018 https://www.nature.com/articles/d41586-018-06847-7 (accessed 1 Aug 2019).
108 Francis Crick Institute. HFEA approval for new "gene editing" techniques, 2016. Available: https://www.crick.ac.uk/news/2016-02-01-hfea-decision [Accessed 1 Aug 2019].
109 Servick K. First U.S. team to gene-edit human embryos revealed, 2017. Science. Available: https://www.sciencemag.org/news/2017/07/first-us-team-gene-edit-human-embryos-revealed [Accessed 1 Aug 2019].
110 Zhai X, Ng V, Lie R. No ethical divide between China and the West in human embryo research. *Dev World Bioeth* 2016;16:116–20.
111 Isasi R, Kleiderman E, Knoppers BM. Genetic technology regulation. editing policy to fit the genome? *Science* 2016;351:337–9.
112 UK Biobank. Access to the UK Biobank resource, 2019. Available: https://www.ukbiobank.ac.uk/key-documents/ [Accessed 1 Aug 2019].
113 All of Us. All of US research program protocol, 2019. Available: https://allofus.nih.gov/about/all-us-research-program-protocol [Accessed 1 Aug 2019].
114 Neuberger EWI, Gene SP. and Cell Doping: The New Frontier - Beyond Myth or Reality. *Med Sport Sci* 2017;62:91–106.
115 National Institutes of Health. NIH Guidelines for Research Involving Recombinant DNA Molecules (NIH Guidelines), 2011. Available: https://web.archive.org/web/20120915075423/http://oba.od.nih.gov/oba/rac/Guidelines/NIH_Guidelines.pdf [Accessed 1 Aug 2019].
116 World Anti-Doping Agency. 2019 list of Prohibited substances and methods, 2019. Available: www.wada-ama.org/sites/default/files/wada_2019_english_prohibited_list.pdf [Accessed 1 Aug 2019].
117 Beiter T, Zimmermann M, Fragasso A, *et al*. Establishing a novel single-copy primer-internal intron-spanning PCR (spiPCR) procedure for the direct detection of gene doping. *Exerc Immunol Rev* 2008;14:73–85.
118 Beiter T, Zimmermann M, Fragasso A, *et al*. Direct and long-term detection of gene doping in conventional blood samples. *Gene Ther* 2011;18:225–31.
119 Moser DA, Neuberger EW, Simon P. A quick one-tube nested PCR-protocol for EPO transgene detection. *Drug Test Anal* 2012;4:870–5.
120 Durussel J, Haile DW, Mooses K, *et al*. Blood transcriptional signature of recombinant human erythropoietin administration and implications for antidoping strategies. *Physiol Genomics* 2016;48:202–9.
121 Wang G, Durussel J, Shurlock J, *et al*. Validation of whole-blood transcriptome signature during microdose recombinant human erythropoietin (rHuEpo) administration. *BMC Genomics* 2017;18:817.
122 Gaziano JM, Concato J, Brophy M, *et al*. Million veteran program: a mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol* 2016;70:214–23.
123 Precision Medicine Initiative (PMI) Working Group. The Precision Medicine Initiative Cohort Program - Building a Research Foundation for 21st Century Medicine. Precision Medicine Initiative (PMI) Working Group Report to the Advisory Committee to the Director, NIH, 2015. Available: https://acd.od.nih.gov/documents/reports/DRAFT-PMI-WG-Report-9-11-2015-508.pdf [Accessed 1 Aug 2019].
124 Carey DJ, Fetterolf SN, Davis FD, *et al*. The Geisinger MyCode community health Initiative: an electronic health record–linked Biobank for precision medicine research. *Genet Med* 2016;18:906–13.
125 Cho SY, Hong EJ, Nam JM, *et al*. Opening of the National Biobank of Korea as the infrastructure of future biomedical science in Korea. *Osong Public Health Res Perspect* 2012;3:177–84.