Sport and Exercise Genomics: the FIMS 2019 Consensus Statement Update

Kumpei Tanisawa[1, 2], Guan Wang[1], Jane Seto[3-4] , Ioanna Verdouka[1], Antonia Karanikolou[1], Richard Twycross-lewis[5], Masashi Tanaka[6], Mats Börjesson[7,17], Luigi Di Luigi[8,17], Michiko Dohi[9,17], Bernd Wolfarth[10,17], Jeroen Swart[11,17], James L. J. Bilzon[12,17], Victoriya Badtieva[13,17], Theodora Papadopoulou[14,17], Michael Geistlinger[15,17], Norbert Bachl[16,17], Fabio Pigozzi[8,17], Yannis P Pitsiladis[1,8,17]

Kumpei Tanisawa <tanisawa@waseda.jp>

Guan Wang <G.Wang2@brighton.ac.uk>

Jane Seto <jane.seto@mcri.edu.au>

Ioanna Verdouka <i.verdouka@gmail.com>

Richard Twycross-lewis <r.twycross-lewis@qmul.ac.uk>

Antonia Karanikolou <A.Karanikolou@brighton.ac.uk>

Masashi Tanaka <masashi_tanaka@me.com>

Mats Börjesson <mats.brjesson@telia.com>

Di Luigi Luigi <luigi.diluigi@uniroma4.it>

Michiko Dohi <michiko.dohi@jpnsport.go.jp>

Bernd Wolfarth <bernd.wolfarth@charite.de>

Jeroen Swart <jeroen.swart@uct.ac.za>

James L. J. Bilzon <j.bilzon@bath.ac.uk>

Victoriya Badtieva <Victoriya.Badtieva@rusada.ru>

Theodora Papadopoulou < Theodora.Papadopoulou100@mod.gov.uk>

Michael Geistlinger <Michael.Geistlinger@sbg.ac.at>

Norbert Bachl <norbert.bachl@univie.ac.at>

Yannis Pitsiladis <y.pitsiladis@brighton.ac.uk>

[1]Collaborating Centre of Sports Medicine, University of Brighton, Eastbourne, UK

[2]Faculty of Sport Sciences, Waseda University, Tokorozawa, Japan

[3]Murdoch Children's Research Institute, Melbourne, Victoria, Australia

[4]Department of Paediatrics, University of Melbourne, The Royal Children's Hospital, Melbourne, Victoria, Australia

[5]School of Engineering & 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]Institute of Neuroscience and Physiology, Department of Food and Nutrition, and Sport Science, University of Gothenburg, Sahlgrenska University Hospital/Östra, Gothenburg, Sweden.

[8]Department of Movement, Human and Health Sciences, University of Rome "Foro Italico", Rome, Italy

[9]Japan Institute of Sports Sciences, Tokyo, Japan

[10]Sportmedizin, Charité - Universitätsmedizin Berlin, Berlin, Germany

[11]Department of Human Biology, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa

[12]Department for Health, University of Bath, Bath, UK

[13]The First Moscow State Sechenov Medical University, Moscow, Russia

[14]Defence Medical Rehabilitation Centre, Stanford Hall, Loughborough, UK

[15]Department for Public Law/Public International Law/European Law. University of Salzburg, Austria

[16]Department of Sports and Exercise Physiology, Centre for Sports Science and University Sports of the University of Vienna, Vienna, Austria.

[17]International Federation of Sports Medicine (FIMS), Lausanne, Switzerland

Corresponding Author:

Professor Yannis Pitsiladis

Collaborating Centre of Sports Medicine

University of Brighton

Welkin House

30 Carlisle Road

Eastbourne

UK

Email: y.pitsiladis@brighton.ac.uk

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.

Keywords: Genetics, athletic performance, precision sports medicine, genetic testing, gene-editing technology, privacy, ethics

Introduction genome is an organism's complete set of DNA, including all of its genes

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 are 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) have 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[8 9]. 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[10].

These rapid advances in DNA sequencing technology have also introduced many new ethical and confidentiality issues such as re-identification of anonymised genotype data[11], data ownership[12], newborn screening[13], and incidental findings[14]. 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[15]. 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 advancements in recent years. In particular, clustered regularly interspaced short palindromic repeats (CRISPR) together with CRISPR-associated proteins (e.g. 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[16-19]. Since inception in 1993[20] CRISPR technology has now advanced to the point of smart technology gels for drug delivery[21]. This pioneering approach, particularly CRISPR/Cas9, is already being used to develop lifesaving/altering gene therapy in monogenic diseases such as sickle cell disease[22], Huntington's disease[23], Cystic Fibrosis[24], and Duchenne muscular dystrophy[25], and poised to make big advances in the near future also in cancer treatment[26] This promising and effective tool also allows the editing of DNA sequences of human germlines[27 28]. 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[29]. Whilst gene-editing tools in human germline have been

138 restricted to research and prohibited in human reproduction by all countries that have
139 established gene editing regulations, CRISPR/Cas9 gene-editing was recently used for
140 reproductive purposes with claims of generating the first gene-edited babies[30]. Not only was
141 this a direct breach of Chinese law, but led to a joint statement of 122 Chinese scientists calling
142 for urgent legislation against further "direct human experimentation"[31]. There is therefore an
143 urgent need to create the necessary regulatory framework to safeguard against the real threat of
144 "genetic pollution" of the human gene pool; a controversial term to describe the process of
145 intentional, uncontrolled and potentially unlawful, introduction of genetic material for the
146 purpose of increasing the "fitness" of a population or sub-sample of a population[32].
147
148 In the field of sport and exercise sciences and medicine, dissecting the relationship between
149 genetic factors and health-related fitness, athletic performance, trainability and susceptibility
150 for exercise-related health risks (e.g. musculoskeletal injury) were previously attempted.
151 Identification of specific sport and exercise-related genes are expected to be utilised for
152 precision sports medicine to provide tailor-made training as well as to select optimal sports
153 and/or other exercise activities for each individual. However, from previous candidate gene
154 approaches and GWAS, there are very limited outcomes with clinical utility, and therefore a
155 paradigm shift in sports genomics is urgently needed[33-37]. However, with the exception of the
156 Genetic-Biological Physical Activity Consortium (GenBioPAC)[34], which is aimed at
157 understanding genetic and other biological factors in the regulation of physical activity, there
158 are no significant funded international consortia to meet this aim. Progress towards such a
159 significant development in the field of sport and exercise genomics will require a paradigm
160 shift in line with recent recommendations for international collaborations such as the Athlome
161 Project Consortium (see www.athlomeconsortium.org) which was launched in 2015 for the
162 advancement of "omics" in exercise sciences and medicine[35]. 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 1,000 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[35 38].

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 inter-individual differences in cardiorespiratory fitness (CRF) in response to standardised exercise and introduced measures to identify determinants and modifiers of CRF response[36]. 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[27 28], 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 wellbeing of athletes and the general worldwide population.

188

189 Considering the rapid changes in circumstances surrounding genomic research and the threats
190 described above, ethics and data protection policies (e.g., GDPR/Data Protection Act 2018 in
191 the UK and EU) should continuously be updated. As the first sequencing studies of elite
192 athletes are being conducted[35 38], there is an urgent need for a guiding reference for sport and
193 exercise genomics to allow the necessary advancement in sports genomics while also
194 protecting athletes from any invasion of privacy and misuse of their genomic data. This
195 reference guide attempted here is presented as a SWOT Analysis. This guiding reference will
196 be regularly updated, based on new and emerging evidence, not only from the area of sport and
197 exercise medicine but also from valuable lessons being learned from the developments and
198 challenges in genomics of health and disease in general[39 40].

199

200 ***SWOT analysis***

201

202 ***1. Strengths***

203 Previous twin and familial studies suggest that there is moderate heritability of "sport and
204 exercise-related traits" (e.g. athletic performance, response to exercise training, and fitness
205 level)[41-44]. Identification of genetic variants determining variabilities in sport and
206 exercise-related traits may offer significant benefits, not only to athletes but also the general
207 population. For example, the Heritage Family Study demonstrated a considerable
208 heterogeneity in the change in maximal oxygen uptake ($\dot{V}O_2max$) in response to a 20-week
209 standardised exercise training programme (the range in training response: –114 to 1097
210 mL/min)[36 41]. Similarly, large individual variabilities in response to resistance training[45], and
211 high-intensity interval training were also reported[46]. Collectively, evidence from human twin
212 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[36]. 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 (e.g. 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[47], 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[48-50], muscle strain[51 52], and stress fracture[53 54]. 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[55]. 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 (e.g. hypertrophic cardiomyopathy, congenital long-QT syndrome)[56] are usually non-fatal diseases, they are the major causes of sudden cardiac death (SCD) in young athletes[57]. 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, pre-participation screening of young athletes for prevention of SCD includes screening for family history and cardiovascular symptoms, physical examination, and often a 12-lead electrocardiogram (ECG), as recommended by the European Society of Cardiology[58]. 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[59]. 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[60]. However, most cases of inherited cardiomyopathy and channelopathy are monogenic and numerous causative mutations have been identified for each disorder[61-63]. 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.

263 Identification of genetic variants associated with sport and exercise-related traits is of great
264 importance in terms of understanding the molecular basis of trainability. For example,
265 deficiency of *ACTN3*, the most replicated and studied sports performance-related gene[64 65],
266 turned out to influence metabolic enzyme activity in skeletal muscle and a shift in the
267 properties of fast fibres towards those characteristics of slow twitch fibres[66]. These findings are
268 consistent with that the null allele of the *ACTN3* p.R577X polymorphism (i.e. a point mutation
269 that usually results in a non-functional protein product) are overrepresented in endurance
270 athletes than in power athletes in ancestrally diverse populations. Thus, identification of
271 genetic variants associated with individual variabilities in sport and exercise-related traits
272 could provide novel insights into molecular adaptations in skeletal muscle. Furthermore,
273 integrating other "omics" responses to exercise such as transcriptomics and proteomics will
274 undoubtedly enhance our understanding of the mechanisms of adaptative response to exercise
275 and its individual variability.
276
277 ***2. Weaknesses***
278 Although a large number of studies have been conducted to identify sport and exercise-related
279 genes, the findings are mostly inconclusive because of a lack of replication. Some studies
280 reported completely opposite associations across different populations [e.g. angiotensin
281 converting enzyme (*ACE*) gene][67-69]. Sport and exercise-related genes and genetic loci
282 confirmed to influence the health status of the athlete have not been identified to date mainly
283 due to the lack of replication in independent populations. Furthermore, GWAS of world class
284 endurance athletes identified no genetic variants associated with extreme endurance
285 performance at the genome-wide level of significance although several SNP associations might
286 be missed in that study because low-density arrays (Ilumina Cardio-MetaboChip) were used

for genotyping without imputation[70]. 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 odds ratio of 1.1-1.5[71]. For example, approximately 5500 cases and equal number of controls are needed to detect an odds ratio 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.[72]. 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 $\dot{V}O_2$max, $\dot{V}O_2$ at the lactate threshold, economy of movement, and other parameters[73]. 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[74]. 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[75], in the field of sport and exercise genomics.

312 Both large sample size and precise phenotyping are necessary to reduce the standard error and
313 increase statistical power to detect a significant SNP-trait association. However, it is difficult
314 to perform comprehensive and precise phenotyping while keeping adequate sample size. For
315 example, The Heritage Family Study, which is the only large-scale standardised exercise
316 intervention study consisting of well phenotyped participants (n = 720) to explore genetic
317 variants associated with response to exercise training by using Illumina HumanCNV370-Quad
318 v3.0 BeadChips (containing approximately 370,000 markers)[76], did not yield any SNPs
319 associated with $\dot{V}O_2$max response to exercise training at the genome-wide level of
320 significance[77]. This study suggests that a sample size of less than 1,000 is still insufficient
321 despite a well standardised intervention protocol and precise phenotyping, while the use of a
322 low-density array without imputation may also have contributed to the finding of no significant
323 SNP associations. The development of technology such as multifaceted wearable devices[78] is
324 needed for comprehensive and precise high-throughput phenotyping of sport and
325 exercise-related traits while keeping adequate sample size for genetic association analyses.
326 Furthermore, multiple large-scale cohorts with well phenotyped participants are needed to
327 replicate and validate the genetic variants detected in a discovery cohort, which might require a
328 substantial budget. In the current environment of a general "research grant famine", it is often
329 difficult to obtain funds in the field of sport and exercise science and medicine to perform such
330 large-scale studies compared with medical science which directly contribute to health and
331 disease prevention for the general population. Financial constraints may also prevent the
332 introduction of large scale genetic approaches into screening programmes[79-82]. Whether to
333 screen and what precisely the preparticipation screening should comprise will be hotly debated
334 for years to come, but in the meantime, progress in this field would be greatly advanced if in
335 SCD cases in in sport, molecular considerations were part of the standardized routine autopsy[83],
336 as well as the genetic screening of first degree relatives[84].

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 (EPOR) in the Olympic cross-country skiing gold medalist[85]. An accumulation of common variants with small effect size has been suggested to explain a large proportion of phenotypic variance of complex traits[86]. However, to date, only a very small number of common variants have been reported to be associated with sport and health-related fitness phenotypes[87]. 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[88], which can help explain the phenotypic variance of complex trait[89], 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 (e.g. copy number variation, large insertion and deletion)[90], variants in non-coding RNA (e.g. micro RNA, long non-coding RNA)[91] may also contribute to the complexity of the athletic phenotype. Integration of various types of genomic variation by multi-omics approach will be needed to fully elucidate the complexity of athletic performance.

3. 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[92]. 23andMe analyse these data[93] but 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[94-96]. Furthermore, 23andMe sell their data to the pharmaceutical industry for drug development research[97] 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[98]. 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, albeit 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.

4. 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 has limited or no scientific bases given no or limited

387 replication and validation. For example, DNAFit, a DTC genetic testing company active in the
388 UK, has recently performed a GWAS of sprint performance in collaboration with Russian and
389 Polish scientists, aimed at identifying novel genetic markers for their genetic testing product[99].
390 These authors reported several SNPs to be associated with sprint performance; however, the
391 clinical significance of these results are unknown given the small sample size of the discovery
392 cohort and the inappropriate replication study (i.e., small sample size, different outcomes and
393 heterogeneous characteristics of participants amongst the cohorts). Many DTC genetic testing
394 companies have already offered genetic testing products for predicting athletic performance
395 and talent identification although sport and exercise genomics has provided very limited
396 evidence and predicting athletic performance and talent identification by using genetic
397 information is almost impossible to this date[100]. Some athletes, coaches, and parents of young
398 individuals may believe the results of genetic testing regardless of the accuracy and quality of
399 genetic testing products commercially available at present. Use of such unproven technology
400 can lead to incorrect decisions such as inappropriate early specialisation for sports,
401 inappropriate training, genetic discrimination and even increased health risks. Genetic testing
402 should therefore be provided with appropriate genetic counselling as described in the statement
403 of The European Society of Human Genetics[101] and The American Society of Human
404 Genetics[102]. The vast majority of DTC companies sell genetic testing to consumers without
405 providing adequate genetic counselling[103]. Thus, collaboration with industry has the potential
406 danger to misuse the data despite the intention of scientists, consequently misleading athletes,
407 their coaches, and families.
408
409 In terms of the privacy and data protection, re-identification of anonymised genotype data has
410 become a real concern. For example, US law enforcement authorities have begun exploiting
411 genetic databases and publicly available family trees to identify suspects via distant familial

412 relatives and have succeeded in arresting numerous suspects[11]. This practical use of publicly
413 available genetic information for criminal investigation raise awareness that re-identification
414 of anonymised genotype data is already technically possible. Individual genome information of
415 elite athletes must be of special interest to many people. There is therefore a possibility that
416 someone may attempt re-identifying anonymised genotype data of elite athletes to abuse this
417 data. Given the recent rapid development of artificial intelligence, re-identification of
418 anonymised genotype data would be much easier in the near future. A flow of genetic
419 information from a DTC genetic testing company to a third party is another potential problem
420 in privacy and data protection. DTC genetic testing companies provide genetic data to
421 third-party scientists or pharmaceutical industry for research purposes without consumer's
422 explicit consent. Although several DTC genetic testing companies obtain additional consent
423 for the secondary use of data, the majority of them do not consistently meet international
424 transparency guidelines related to privacy, and secondary use of data[104]. Furthermore, the use
425 of third-party interpretation services also increases the risk of privacy invasion and misuse of
426 data. Consumers can download their raw genetic data from DTC genetic testing company
427 website and freely upload it to third-party interpretation services for further explanation of
428 genetic data. Because the data usage and privacy policy is less prominent in the
429 third-party interpretation services than in DTC genetic testing companies[105], the risk of data
430 exploitation by someone could be high, as US law enforcement authorities have already
431 exploited it for criminal investigation[11]. Thus, considering the feasibility of re-identification of
432 genotype data and unexpected use of this, a lack of transparency in the provision of information
433 to consumers is a serious problem because it inhibits them from recognising the threat of
434 privacy invasion.
435

436 Given the development of gene-editing technology such as CRISPR/Cas9 [106], genetic variants
437 determining athletic performance and elite athlete status may be used for gene-doping or
438 creating talented sports children in the future. In fact, gene-editing of Myostatin in zygotes
439 successfully enhanced muscle hypertrophy in several adult mammals[107-109]. Even in humans,
440 researchers have used gene-editing techniques targeting not only human adult cell for disease
441 therapy[22-25], but also human germlines[27 28]. A pertinent recent example is the aforementioned
442 revelations of the use of CRISPR/Cas9 gene editing technology to delete both copies of the
443 CCR5 gene in embryos to give twin babies resistance to HIV infection in violation of the laws
444 and regulations in respective countries. This should not have been performed because there is
445 no consensus on how to counsel "gene-edited individuals" as well as a limitation in
446 understanding of the long term effects of gene-editing on mature body. In fact, although basic
447 research involving gene-editing in human germlines has been admitted in several countries (e.g.
448 UK, US, Sweden, China, and Japan)[110-114], human gene-editing for reproduction is prohibited
449 by law or regulation in many countries[115]. The fact that there are gene-edited babies alive today
450 confirms that gene-edited human babies are already technically possible. Designing athletes
451 with extraordinary athletic performance by using gene-editing technique would be a real threat
452 in terms of keeping sport fair, clean and protecting athlete health.
453
454 ***Guiding reference for sport and exercise genomics***
455 The present SWOT analysis suggests that sport and exercise genomics has the potential to
456 contribute to the health and wellbeing of athletes as well as real and necessary advances in the
457 field of sport and exercise sciences and medicine. Large-scale studies in collaboration with
458 industry may help to provide sufficient scientific evidence to adequately and ethically utilise
459 genetic information of athletes, mainly to protect their health status. On the other hand, there
460 are many potential dangers associated with the use of genomics in sport and exercise medicine

such as re-identification 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, anonymization, 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 (e.g. 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.

535 Sport and exercise genomics is in transition from focused research performed by single
536 research groups to large-scale discovery research involving many research groups and industry
537 partners[33 37]. The advancements in sport and exercise genomics parallel the increase in the risk
538 of data leakage, privacy invasion, and abuse of personal data. At this moment, without strict
539 rules about data flow management, anonymisation, security, and data release policy that is
540 standard practice in the large biobank studies[116 117], releasing any genotype and phenotype data
541 to industry, other research groups, and public databases is not recommended. All scientists of
542 sport and exercise genomics need to be well-versed in ethics and data protection policy to
543 protect individuals from threats of privacy invasion and abuse of personal data in preparation
544 for the era of large-scale collaborative science.
545
546 The application of CRISPR/Cas9 gene editing techniques to skeletal muscle as well as
547 hematopoietic stem cells for treatment of monogenic diseases is becoming more
548 commonplace[22 25]. Given these rapid advances in gene editing, it is expected and totally
549 desirable that these techniques are harnessed by the sports medicine physician to treat
550 sports-related injuries. This broader application of gene editing techniques to sports medicine,
551 will inevitably result in gene-doping being a real prospect[118]. Knowledge gained from
552 gene-editing research for disease therapy could be misused for enhancement of athletic
553 performance. To the best of our knowledge, no gene editing techniques have been applied in
554 healthy individuals to enhance athletic performance. This threat poses new ethical dilemmas
555 and hence the urgent need for gene-editing guidelines and regulation constantly updated to deal
556 with all eventualities. It is the responsibility of those involved in the field of sport and exercise
557 sciences and medicine to keep abreast of the gene-editing guidelines[119]. It is necessary to
558 prioritise research in anti-doping with particular reference to gene-doping. Although gene
559 doping is already prohibited on the list of banned doping agents developed by the World

560 Anti-Doping Agency (WADA)[120], robust and effective anti-doping measures to detect the use
561 of gene doping have not been developed. Several PCR-based strategies to detect
562 vector-mediated gene transfer of several candidate genes [e.g., vascular endothelial growth
563 factor (*VEGF*), erythropoietin (*EPO*), insulin-like growth factors 1 (*IGF1*), growth hormone
564 (*GH*)] for gene doping have been developed[121-123]. However, gene transfer of unexpected target
565 gene cannot be detected by using these candidate gene approaches. Furthermore, the strategy to
566 detect gene doping by CRISPR/Cas9 based on DNA modification has not been developed so
567 far. Omics technologies have been shown to enhance the detection of blood doping[124 125], and
568 these cutting-edge technologies could be further harnessed to develop effective methods for the
569 detection of gene doping in sport. Funding institutions should be encouraged to offer grants for
570 further developments in the rapidly emerging field of anti-doping strategies providing the
571 anti-doping laboratories with the tools to significantly improve their abilities.
572
573 **Concluding Remarks**
574 The present guidelines in sport and exercise genomics developed following an extensive
575 SWOT analysis, advocates the need for clear and universal standards as they relate to the
576 collection, management and storage of DNA/data with the overriding objective to protect
577 individuals from privacy invasion and misuse of genomic information. Given the increased
578 availability of high-throughput genomic information, there is an urgent need for such a guiding
579 reference in sport and exercise genomics with a clear and consistent data handling and release
580 policy for all individuals who potentially handle any genetic information. It is essential that
581 sports physicians, scientists and all those involved in supporting the athlete keep abreast with
582 new developments in genomics including new technologies and methods such as
583 CRISPR/Cas9 and are well informed of the laws and regulations pertaining to the collection,
584 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.

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 (14J03810 to KT) and the JSPS KAKENHI Grant Number (16H01872 to MT) (Tokyo, Japan).

Competing interests

One of the authors, YP, is the founding member of the Athlome Consortium (www.athlomeconsortium.org).

Patient consent for publication

Not required.

Provenance and peer review

Not commissioned; externally peer reviewed.

References

1. Shendure J, Balasubramanian S, Church GM, et al. DNA sequencing at 40: past, present and future. *Nature* 2017;550(7676):345-53. doi: 10.1038/nature24286
2. Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med* 2015;372(9):793-5. doi: 10.1056/NEJMp1500523

610 3. Small AM, O'Donnell CJ, Damrauer SM. Large-Scale Genomic Biobanks and
611 Cardiovascular Disease. *Curr Cardiol Rep* 2018;20(4):22. doi:
612 10.1007/s11886-018-0969-8
613 4. Carey DJ, Fetterolf SN, Davis FD, et al. The Geisinger MyCode community health
614 initiative: an electronic health record-linked biobank for precision medicine research.
615 *Genet Med* 2016;18(9):906-13. doi: 10.1038/gim.2015.187
616 5. Cho SY, Hong EJ, Nam JM, et al. Opening of the national biobank of Korea as the
617 infrastructure of future biomedical science in Korea. *Osong Public Health Res Perspect*
618 2012;3(3):177-84. doi: 10.1016/j.phrp.2012.07.004
619 6. Gaziano JM, Concato J, Brophy M, et al. Million Veteran Program: A mega-biobank to
620 study genetic influences on health and disease. *J Clin Epidemiol* 2016;70:214-23. doi:
621 10.1016/j.jclinepi.2015.09.016
622 7. Precision Medicine Initiative (PMI) Working Group. The Precision Medicine Initiative
623 Cohort Program - Building a Research Foundation for 21st Century Medicine.
624 Precision Medicine Initiative (PMI) Working Group Report to the Advisory Committee
625 to the Director, NIH; 2015.
626 https://acd.od.nih.gov/documents/reports/DRAFT-PMI-WG-Report-9-11-2015-508.p
627 df (accessed 1 Aug 2019)
628 8. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping
629 and genomic data. *Nature* 2018;562(7726):203-09. doi: 10.1038/s41586-018-0579-z
630 9. Elsworth B. Collaboratively analysing thousands of phenotypic traits using UK Biobank
631 data, Google APIs and HPC. 2018.
632 https://ieup4.blogs.bristol.ac.uk/2018/10/01/ukb_gwas/ (accessed 1 Aug 2019)

633 10. GOV.UK. 100,000 whole genomes sequenced in the NHS. 2018.
634 https://www.gov.uk/government/news/100000-whole-genomes-sequenced-in-the-nhs
635 (accessed 1 Aug 2019)
636 11. Erlich Y, Shor T, Pe'er I, et al. Identity inference of genomic data using long-range familial
637 searches. *Science* 2018;362(6415):690-94. doi: 10.1126/science.aau4832
638 12. Gurwitz D, Bregman-Eschet Y. Personal genomics services: whose genomes? *Eur J Hum*
639 *Genet* 2009;17(7):883-9. doi: 10.1038/ejhg.2008.254
640 13. Goldenberg AJ, Sharp RR. The ethical hazards and programmatic challenges of genomic
641 newborn screening. *JAMA* 2012;307(5):461-2. doi: 10.1001/jama.2012.68
642 14. Green RC, Berg JS, Grody WW, et al. ACMG recommendations for reporting of incidental
643 findings in clinical exome and genome sequencing. *Genet Med* 2013;15(7):565-74. doi:
644 10.1038/gim.2013.73
645 15. Phillips AM. 'Only a click away - DTC genetics for ancestry, health, love...and more: A
646 view of the business and regulatory landscape'. *Appl Transl Genom* 2016;8:16-22. doi:
647 10.1016/j.atg.2016.01.001
648 16. Mali P, Yang L, Esvelt KM, et al. RNA-guided human genome engineering via Cas9.
649 *Science* 2013;339(6121):823-6. doi: 10.1126/science.1232033
650 17. Cox DBT, Gootenberg JS, Abudayyeh OO, et al. RNA editing with CRISPR-Cas13.
651 *Science* 2017;358(6366):1019-27. doi: 10.1126/science.aaq0180
652 18. Chen JS, Ma E, Harrington LB, et al. CRISPR-Cas12a target binding unleashes
653 indiscriminate single-stranded DNase activity. *Science* 2018;360(6387):436-39. doi:
654 10.1126/science.aar6245
655 19. Liu JJ, Orlova N, Oakes BL, et al. CasX enzymes comprise a distinct family of
656 RNA-guided genome editors. *Nature* 2019;566(7743):218-23. doi:
657 10.1038/s41586-019-0908-x

20. Mojica FJ, Juez G, Rodriguez-Valera F. Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites. *Mol Microbiol* 1993;9(3):613-21. doi: 10.1111/j.1365-2958.1993.tb01721.x

21. Callaway E. CRISPR cuts turn gels into biological watchdogs. *Nature* 2019;572(7771):574. doi: 10.1038/d41586-019-02542-3

22. Dever DP, Bak RO, Reinisch A, et al. CRISPR/Cas9 beta-globin gene targeting in human haematopoietic stem cells. *Nature* 2016;539(7629):384-89. doi: 10.1038/nature20134

23. Shin JW, Kim KH, Chao MJ, et al. Permanent inactivation of Huntington's disease mutation by personalized allele-specific CRISPR/Cas9. *Hum Mol Genet* 2016;25(20):4566-76. doi: 10.1093/hmg/ddw286

24. 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(9):1385-90. doi: 10.1016/j.celrep.2015.07.062

25. 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. doi: 10.1038/ncomms7244

26. Zhan T, Rindtorff N, Betge J, et al. CRISPR/Cas9 for cancer research and therapy. *Semin Cancer Biol* 2018 doi: 10.1016/j.semcancer.2018.04.001

27. Liang P, Xu Y, Zhang X, et al. CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. *Protein Cell* 2015;6(5):363-72. doi: 10.1007/s13238-015-0153-5

28. Ma H, Marti-Gutierrez N, Park SW, et al. Correction of a pathogenic gene mutation in human embryos. *Nature* 2017;548(7668):413-19. doi: 10.1038/nature23305

681 29. Tuladhar R, Yeu Y, Tyler Piazza J, et al. CRISPR-Cas9-based mutagenesis frequently
682 provokes on-target mRNA misregulation. *Nat Commun* 2019;10(1):4056. doi:
683 10.1038/s41467-019-12028-5
684 30. Cyranoski D, Ledford H. Genome-edited baby claim provokes international outcry. *Nature*
685 2018;563(7733):607-08. doi: 10.1038/d41586-018-07545-0
686 31. China Daily. Investigation of human gene editing claim justified: China Daily editorial.
687 2018.
688 http://www.chinadaily.com.cn/a/201811/27/WS5bfd3203a310eff30328b5a3.html
689 (accessed 10 Sep 2019)
690 32. Waller DM. Genetic rescue: a safe or risky bet? *Mol Ecol* 2015;24(11):2595-7. doi:
691 10.1111/mec.13220
692 33. Eynon N, Voisin S, Lucia A, et al. Preface: genomics and biology of exercise is undergoing
693 a paradigm shift. *BMC Genomics* 2017;18(Suppl 8):825. doi:
694 10.1186/s12864-017-4184-6
695 34. Lightfoot JT, EJC DEG, Booth FW, et al. Biological/Genetic Regulation of Physical
696 Activity Level: Consensus from GenBioPAC. *Med Sci Sports Exerc*
697 2018;50(4):863-73. doi: 10.1249/MSS.0000000000001499
698 35. Pitsiladis YP, Tanaka M, Eynon N, et al. Athlome Project Consortium: a concerted effort to
699 discover genomic and other "omic" markers of athletic performance. *Physiol Genomics*
700 2016;48(3):183-90. doi: 10.1152/physiolgenomics.00105.2015
701 36. Ross R, Goodpaster BH, Koch LG, et al. Precision exercise medicine: understanding
702 exercise response variability. *Br J Sports Med* 2019 doi:
703 10.1136/bjsports-2018-100328

704 37. Tanaka M, Wang G, Pitsiladis YP. Advancing sports and exercise genomics: moving from
705 hypothesis-driven single study approaches to large multi-omics collaborative science.
706 *Physiol Genomics* 2016;48(3):173-4. doi: 10.1152/physiolgenomics.00009.2016
707 38. Tanana M, Tanisawa K, Wang G, et al. The 1000 Athlomes Project for 2020 Summer
708 Olympics in Tokyo. 35th FIMS World Congress of Sports Medicine. Rio de Janeiro,
709 2018.
710 39. Patch C, Middleton A. Genetic counselling in the era of genomic medicine. *Br Med Bull*
711 2018;126(1):27-36. doi: 10.1093/bmb/ldy008
712 40. Price WN, 2nd, Cohen IG. Privacy in the age of medical big data. *Nat Med*
713 2019;25(1):37-43. doi: 10.1038/s41591-018-0272-7
714 41. Bouchard C, An P, Rice T, et al. Familial aggregation of VO(2max) response to exercise
715 training: results from the HERITAGE Family Study. *J Appl Physiol (1985)*
716 1999;87(3):1003-8. doi: 10.1152/jappl.1999.87.3.1003
717 42. De Moor MH, Spector TD, Cherkas LF, et al. Genome-wide linkage scan for athlete status
718 in 700 British female DZ twin pairs. *Twin Res Hum Genet* 2007;10(6):812-20. doi:
719 10.1375/twin.10.6.812
720 43. Miyamoto-Mikami E, Zempo H, Fuku N, et al. Heritability estimates of endurance-related
721 phenotypes: A systematic review and meta-analysis. *Scand J Med Sci Sports*
722 2018;28(3):834-45. doi: 10.1111/sms.12958
723 44. Zempo H, Miyamoto-Mikami E, Kikuchi N, et al. Heritability estimates of muscle
724 strength-related phenotypes: A systematic review and meta-analysis. *Scand J Med Sci*
725 *Sports* 2017;27(12):1537-46. doi: 10.1111/sms.12804
726 45. Ahtiainen JP, Walker S, Peltonen H, et al. Heterogeneity in resistance training-induced
727 muscle strength and mass responses in men and women of different ages. *Age (Dordr)*
728 2016;38(1):10. doi: 10.1007/s11357-015-9870-1

729 46. Yan X, Eynon N, Papadimitriou ID, et al. The gene SMART study: method, study design,
730 and preliminary findings. *BMC Genomics* 2017;18(Suppl 8):821. doi:
731 10.1186/s12864-017-4186-4
732 47. Collins M, September AV, Posthumus M. Biological variation in musculoskeletal injuries:
733 current knowledge, future research and practical implications. *Br J Sports Med*
734 2015;49(23):1497-503. doi: 10.1136/bjsports-2015-095180
735 48. Collins M, Posthumus M, Schwellnus MP. The COL1A1 gene and acute soft tissue
736 ruptures. *Br J Sports Med* 2010;44(14):1063-4. doi: 10.1136/bjsm.2008.056184
737 49. Posthumus M, September AV, O'Cuinneagain D, et al. The COL5A1 gene is associated
738 with increased risk of anterior cruciate ligament ruptures in female participants. *Am J*
739 *Sports Med* 2009;37(11):2234-40. doi: 10.1177/0363546509338266
740 50. Posthumus M, September AV, O'Cuinneagain D, et al. The association between the
741 COL12A1 gene and anterior cruciate ligament ruptures. *Br J Sports Med*
742 2010;44(16):1160-5. doi: 10.1136/bjsm.2009.060756
743 51. Miyamoto N, Miyamoto-Mikami E, Hirata K, et al. Association analysis of the ACTN3
744 R577X polymorphism with passive muscle stiffness and muscle strain injury. *Scand J*
745 *Med Sci Sports* 2018;28(3):1209-14. doi: 10.1111/sms.12994
746 52. Kumagai H, Miyamoto-Mikami E, Hirata K, et al. ESR1 rs2234693 Polymorphism Is
747 Associated with Muscle Injury and Muscle Stiffness. *Med Sci Sports Exerc*
748 2019;51(1):19-26. doi: 10.1249/MSS.0000000000001750
749 53. Varley I, Hughes DC, Greeves JP, et al. RANK/RANKL/OPG pathway: genetic
750 associations with stress fracture period prevalence in elite athletes. *Bone* 2015;71:131-6.
751 doi: 10.1016/j.bone.2014.10.004

752 54. Varley I, Hughes DC, Greeves JP, et al. The association of novel polymorphisms with
753 stress fracture injury in Elite Athletes: Further insights from the SFEA cohort. *J Sci*
754 *Med Sport* 2018;21(6):564-68. doi: 10.1016/j.jsams.2017.10.038
755 55. Mikami E, Fuku N, Murakami H, et al. ACTN3 R577X genotype is associated with
756 sprinting in elite Japanese athletes. *Int J Sports Med* 2014;35(2):172-7. doi:
757 10.1055/s-0033-1347171
758 56. Basavarajaiah S, Shah A, Sharma S. Sudden cardiac death in young athletes. *Heart*
759 2007;93(3):287-9. doi: 10.1136/hrt.2006.100107
760 57. Semsarian C, Sweeting J, Ackerman MJ. Sudden cardiac death in athletes. *BMJ*
761 2015;350:h1218. doi: 10.1136/bmj.h1218
762 58. Corrado D, Pelliccia A, Bjornstad HH, et al. Cardiovascular pre-participation screening of
763 young competitive athletes for prevention of sudden death: proposal for a common
764 European protocol. Consensus Statement of the Study Group of Sport Cardiology of
765 the Working Group of Cardiac Rehabilitation and Exercise Physiology and the
766 Working Group of Myocardial and Pericardial Diseases of the European Society of
767 Cardiology. *Eur Heart J* 2005;26(5):516-24. doi: 10.1093/eurheartj/ehi108
768 59. Corrado D, Basso C, Pavei A, et al. Trends in sudden cardiovascular death in young
769 competitive athletes after implementation of a preparticipation screening program.
770 *JAMA* 2006;296(13):1593-601. doi: 10.1001/jama.296.13.1593
771 60. Chandra N, Bastiaenen R, Papadakis M, et al. Sudden cardiac death in young athletes:
772 practical challenges and diagnostic dilemmas. *J Am Coll Cardiol* 2013;61(10):1027-40.
773 doi: 10.1016/j.jacc.2012.08.1032
774 61. Charron P, Arad M, Arbustini E, et al. Genetic counselling and testing in
775 cardiomyopathies: a position statement of the European Society of Cardiology

776 Working Group on Myocardial and Pericardial Diseases. *Eur Heart J*
777 2010;31(22):2715-26. doi: 10.1093/eurheartj/ehq271
778 62. Tester DJ, Ackerman MJ. Genetic testing for potentially lethal, highly treatable inherited
779 cardiomyopathies/channelopathies in clinical practice. *Circulation*
780 2011;123(9):1021-37. doi: 10.1161/CIRCULATIONAHA.109.914838
781 63. Mayosi BM, Fish M, Shaboodien G, et al. Identification of Cadherin 2 (CDH2) Mutations
782 in Arrhythmogenic Right Ventricular Cardiomyopathy. *Circ Cardiovasc Genet*
783 2017;10(2) doi: 10.1161/CIRCGENETICS.116.001605
784 64. Yang N, MacArthur DG, Gulbin JP, et al. ACTN3 genotype is associated with human elite
785 athletic performance. *Am J Hum Genet* 2003;73(3):627-31. doi: 10.1086/377590
786 65. Del Coso J, Hiam D, Houweling P, et al. More than a 'speed gene': ACTN3 R577X
787 genotype, trainability, muscle damage, and the risk for injuries. *Eur J Appl Physiol*
788 2019;119(1):49-60. doi: 10.1007/s00421-018-4010-0
789 66. MacArthur DG, Seto JT, Raftery JM, et al. Loss of ACTN3 gene function alters mouse
790 muscle metabolism and shows evidence of positive selection in humans. *Nat Genet*
791 2007;39(10):1261-5. doi: 10.1038/ng2122
792 67. Amir O, Amir R, Yamin C, et al. The ACE deletion allele is associated with Israeli elite
793 endurance athletes. *Exp Physiol* 2007;92(5):881-6. doi:
794 10.1113/expphysiol.2007.038711
795 68. Gayagay G, Yu B, Hambly B, et al. Elite endurance athletes and the ACE I allele--the role
796 of genes in athletic performance. *Hum Genet* 1998;103(1):48-50.
797 69. Wang G, Mikami E, Chiu LL, et al. Association analysis of ACE and ACTN3 in elite
798 Caucasian and East Asian swimmers. *Med Sci Sports Exerc* 2013;45(5):892-900. doi:
799 10.1249/MSS.0b013e31827c501f

800 70. Rankinen T, Fuku N, Wolfarth B, et al. No Evidence of a Common DNA Variant Profile
801 Specific to World Class Endurance Athletes. *PLoS One* 2016;11(1):e0147330. doi:
802 10.1371/journal.pone.0147330
803 71. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex
804 diseases. *Nature* 2009;461(7265):747-53. doi: 10.1038/nature08494
805 72. Chapman K, Ferreira T, Morris A, et al. Defining the power limits of genome-wide
806 association scan meta-analyses. *Genet Epidemiol* 2011;35(8):781-9. doi:
807 10.1002/gepi.20627
808 73. Bassett DR, Jr., Howley ET. Limiting factors for maximum oxygen uptake and
809 determinants of endurance performance. *Med Sci Sports Exerc* 2000;32(1):70-84.
810 74. Macarthur DG, North KN. Genes and human elite athletic performance. *Hum Genet*
811 2005;116(5):331-9. doi: 10.1007/s00439-005-1261-8
812 75. Houle D, Govindaraju DR, Omholt S. Phenomics: the next challenge. *Nat Rev Genet*
813 2010;11(12):855-66. doi: 10.1038/nrg2897
814 76. Bouchard C, Leon AS, Rao DC, et al. The HERITAGE family study. Aims, design, and
815 measurement protocol. *Med Sci Sports Exerc* 1995;27(5):721-9.
816 77. Bouchard C, Sarzynski MA, Rice TK, et al. Genomic predictors of the maximal O(2)
817 uptake response to standardized exercise training programs. *J Appl Physiol (1985)*
818 2011;110(5):1160-70. doi: 10.1152/japplphysiol.00973.2010
819 78. Muniz-Pardos B, Sutehall S, Gellaerts J, et al. Integration of Wearable Sensors Into the
820 Evaluation of Running Economy and Foot Mechanics in Elite Runners. *Curr Sports*
821 *Med Rep* 2018;17(12):480-88. doi: 10.1249/JSR.0000000000000550
822 79. Dhutia H, Malhotra A, Gabus V, et al. Cost Implications of Using Different ECG Criteria
823 for Screening Young Athletes in the United Kingdom. *J Am Coll Cardiol*
824 2016;68(7):702-11. doi: 10.1016/j.jacc.2016.05.076

825 80. McKinney J, Lithwick DJ, Morrison BN, et al. Detecting Underlying Cardiovascular
826 Disease in Young Competitive Athletes. *Can J Cardiol* 2017;33(1):155-61. doi:
827 10.1016/j.cjca.2016.06.007
828 81. Riding NR, Sharma S, Salah O, et al. Systematic echocardiography is not efficacious when
829 screening an ethnically diverse cohort of athletes in West Asia. *Eur J Prev Cardiol*
830 2015;22(2):263-70. doi: 10.1177/2047487313506549
831 82. Vessella T, Zorzi A, Merlo L, et al. The Italian preparticipation evaluation programme:
832 diagnostic yield, rate of disqualification and cost analysis. *Br J Sports Med* 2019 doi:
833 10.1136/bjsports-2018-100293
834 83. Harmon KG, Drezner JA, Wilson MG, et al. Incidence of sudden cardiac death in athletes:
835 a state-of-the-art review. *Br J Sports Med* 2014;48(15):1185-92. doi:
836 10.1136/bjsports-2014-093872
837 84. Thiene G. Sudden cardiac death in the young: a genetic destiny? *Clin Med (Lond)*
838 2018;18(Suppl 2):s17-s23. doi: 10.7861/clinmedicine.18-2-s17
839 85. de la Chapelle A, Traskelin AL, Juvonen E. Truncated erythropoietin receptor causes
840 dominantly inherited benign human erythrocytosis. *Proc Natl Acad Sci U S A*
841 1993;90(10):4495-9. doi: 10.1073/pnas.90.10.4495
842 86. Golan D, Lander ES, Rosset S. Measuring missing heritability: inferring the contribution of
843 common variants. *Proc Natl Acad Sci U S A* 2014;111(49):E5272-81. doi:
844 10.1073/pnas.1419064111
845 87. Sarzynski MA, Loos RJ, Lucia A, et al. Advances in Exercise, Fitness, and Performance
846 Genomics in 2015. *Med Sci Sports Exerc* 2016;48(10):1906-16. doi:
847 10.1249/MSS.0000000000000982
848 88. 1000 Genomes Project Consortium, Auton A, Brooks LD, et al. A global reference for
849 human genetic variation. *Nature* 2015;526(7571):68-74. doi: 10.1038/nature15393

850 89. Wainschtein P, Jain D, Yengo L, et al. Recovery of trait heritability from whole genome
851 sequence data. *bioRxiv* 2019 doi: 10.1101/588020
852 90. Conrad DF, Pinto D, Redon R, et al. Origins and functional impact of copy number
853 variation in the human genome. *Nature* 2010;464(7289):704-12. doi:
854 10.1038/nature08516
855 91. Taft RJ, Pang KC, Mercer TR, et al. Non-coding RNAs: regulators of disease. *J Pathol*
856 2010;220(2):126-39. doi: 10.1002/path.2638
857 92. 23andMe. Research. https://research.23andme.com/research/ (accessed 1 Aug 2019)
858 93. Hu Y, Shmygelska A, Tran D, et al. GWAS of 89,283 individuals identifies genetic
859 variants associated with self-reporting of being a morning person. *Nat Commun*
860 2016;7:10448. doi: 10.1038/ncomms10448
861 94. Chang D, Nalls MA, Hallgrimsdottir IB, et al. A meta-analysis of genome-wide association
862 studies identifies 17 new Parkinson's disease risk loci. *Nat Genet* 2017;49(10):1511-16.
863 doi: 10.1038/ng.3955
864 95. Ferreira MA, Vonk JM, Baurecht H, et al. Shared genetic origin of asthma, hay fever and
865 eczema elucidates allergic disease biology. *Nat Genet* 2017;49(12):1752-57. doi:
866 10.1038/ng.3985
867 96. Tsoi LC, Stuart PE, Tian C, et al. Large scale meta-analysis characterizes genetic
868 architecture for common psoriasis associated variants. *Nat Commun* 2017;8:15382.
869 doi: 10.1038/ncomms15382
870 97. GlaxoSmithKline. GSK and 23andMe sign agreement to leverage genetic insights for the
871 development of novel medicines. 2018.
872 https://www.gsk.com/en-gb/media/press-releases/gsk-and-23andme-sign-agreement-t
873 o-leverage-genetic-insights-for-the-development-of-novel-medicines/ (accessed 1 Aug
874 2019)

875 98. 23andMe. 23andMe Populations Collaborations Program. 2018.
876 https://research.23andme.com/populations-collaborations/ (accessed 1 Aug 2019)
877 99. Pickering C, Suraci B, Semenova EA, et al. A Genome-Wide Association Study of Sprint
878 Performance in Elite Youth Football Players. *J Strength Cond Res* 2019 doi:
879 10.1519/JSC.0000000000003259
880 100. Webborn N, Williams A, McNamee M, et al. Direct-to-consumer genetic testing for
881 predicting sports performance and talent identification: Consensus statement. *Br J*
882 *Sports Med* 2015;49(23):1486-91. doi: 10.1136/bjsports-2015-095343
883 101. European Society of Human Genetics. Statement of the ESHG on
884 direct-to-consumer genetic testing for health-related purposes. *Eur J Hum Genet*
885 2010;18(12):1271-3. doi: 10.1038/ejhg.2010.129
886 102. Hudson K, Javitt G, Burke W, et al. ASHG Statement* on direct-to-consumer genetic
887 testing in the United States. *Obstet Gynecol* 2007;110(6):1392-5. doi:
888 10.1097/01.AOG.0000292086.98514.8b
889 103. Harris A, Kelly SE, Wyatt S. Counseling customers: emerging roles for genetic
890 counselors in the direct-to-consumer genetic testing market. *J Genet Couns*
891 2013;22(2):277-88. doi: 10.1007/s10897-012-9548-0
892 104. Laestadius LI, Rich JR, Auer PL. All your data (effectively) belong to us: data practices
893 among direct-to-consumer genetic testing firms. *Genet Med* 2017;19(5):513-20. doi:
894 10.1038/gim.2016.136
895 105. Badalato L, Kalokairinou L, Borry P. Third party interpretation of raw genetic data: an
896 ethical exploration. *Eur J Hum Genet* 2017;25(11):1189-94. doi:
897 10.1038/ejhg.2017.126
898 106. Cong L, Ran FA, Cox D, et al. Multiplex genome engineering using CRISPR/Cas systems.
899 *Science* 2013;339(6121):819-23. doi: 10.1126/science.1231143

900 107. Crispo M, Mulet AP, Tesson L, et al. Efficient Generation of Myostatin Knock-Out Sheep
901 Using CRISPR/Cas9 Technology and Microinjection into Zygotes. *PLoS One*
902 2015;10(8):e0136690. doi: 10.1371/journal.pone.0136690
903 108. Lv Q, Yuan L, Deng J, et al. Efficient Generation of Myostatin Gene Mutated Rabbit by
904 CRISPR/Cas9. *Sci Rep* 2016;6:25029. doi: 10.1038/srep25029
905 109. Wang K, Ouyang H, Xie Z, et al. Efficient Generation of Myostatin Mutations in Pigs
906 Using the CRISPR/Cas9 System. *Sci Rep* 2015;5:16623. doi: 10.1038/srep16623
907 110. Callaway E. Gene-editing research in human embryos gains momentum. *Nature*
908 2016;532(7599):289-90. doi: 10.1038/532289a
909 111. Cyranoski D. Japan set to allow gene editing in human embryos. Nature; 2018.
910 https://www.nature.com/articles/d41586-018-06847-7 (accessed 1 Aug 2019)
911 112. Francis Crick Institute. HFEA approval for new "gene editing" techniques. 2016.
912 https://www.crick.ac.uk/news/2016-02-01-hfea-decision (accessed 1 Aug 2019)
913 113. Servick K. First U.S. team to gene edit human embryos revealed. Science; 2017.
914 https://www.sciencemag.org/news/2017/07/first-us-team-gene-edit-human-embryos-r
915 evealed (accessed 1 Aug 2019)
916 114. Zhai X, Ng V, Lie R. No ethical divide between China and the West in human embryo
917 research. *Dev World Bioeth* 2016;16(2):116-20. doi: 10.1111/dewb.12108
918 115. Isasi R, Kleiderman E, Knoppers BM. GENETIC TECHNOLOGY REGULATION.
919 Editing policy to fit the genome? *Science* 2016;351(6271):337-9. doi:
920 10.1126/science.aad6778
921 116. UK biobank. Access to the UK Biobank Resource. 2019.
922 https://www.ukbiobank.ac.uk/key-documents/ (accessed 1 Aug 2019)
923 117. All of Us. All of Us Research Program Protocol. 2019.
924 https://allofus.nih.gov/about/all-us-research-program-protocol (accessed 1 Aug 2019)

925 118. Neuberger EWI, Simon P. Gene and Cell Doping: The New Frontier - Beyond Myth or
926 Reality. *Med Sport Sci* 2017;62:91-106. doi: 10.1159/000465456
927 119. National Institutes of Health. NIH Guidelines for Research Involving Recombinant
928 DNA Molecules(NIH Guidelines). 2011.
929 https://web.archive.org/web/20120915075423/http://oba.od.nih.gov/oba/rac/Guideline
930 s/NIH_Guidelines.pdf (accessed 1 Aug 2019)
931 120. World Anti-Doping Agency. 2019 List of Prohibited Substances and Methods. 2019.
932 www.wada-ama.org/sites/default/files/wada_2019_english_prohibited_list.pdf
933 (accessed 1 Aug 2019)
934 121. Beiter T, Zimmermann M, Fragasso A, et al. Establishing a novel single-copy
935 primer-internal intron-spanning PCR (spiPCR) procedure for the direct detection of
936 gene doping. *Exerc Immunol Rev* 2008;14:73-85.
937 122. Beiter T, Zimmermann M, Fragasso A, et al. Direct and long-term detection of gene
938 doping in conventional blood samples. *Gene Ther* 2011;18(3):225-31. doi:
939 10.1038/gt.2010.122
940 123. Moser DA, Neuberger EW, Simon P. A quick one-tube nested PCR-protocol for EPO
941 transgene detection. *Drug Test Anal* 2012;4(11):870-5. doi: 10.1002/dta.1348
942 124. Durussel J, Haile DW, Mooses K, et al. Blood transcriptional signature of recombinant
943 human erythropoietin administration and implications for antidoping strategies.
944 *Physiol Genomics* 2016;48(3):202-9. doi: 10.1152/physiolgenomics.00108.2015
945 125. Wang G, Durussel J, Shurlock J, et al. Validation of whole-blood transcriptome signature
946 during microdose recombinant human erythropoietin (rHuEpo) administration. *BMC*
947 *Genomics* 2017;18(Suppl 8):817. doi: 10.1186/s12864-017-4191-7

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 JM, 2016[6]
All of Us	Goal for 1 million individuals	Baseline physical exam, 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[7]
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 DJ, 2016[4]
Vanderbilt: BioVU	225,000 genotyped	EHR structured data	Ilumnia 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 SY, 2012[5]
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/

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