



Editor's choice
Scan to access more
free content

Correspondence to

Dr Claude Bouchard, Human Genomics Laboratory, Pennington Biomedical Research Center, 6400 Perkins Road, Baton Rouge, LA 70808-4124, USA; claudie.bouchard@pbrc.edu

Accepted 10 September 2015

Published Online First

1 October 2015

Exercise genomics—a paradigm shift is needed: a commentary

Claude Bouchard

ABSTRACT

The overarching goal of exercise genomics is to illuminate exercise biology and behaviour in order to better understand the preventive and therapeutic values of exercise. An ancillary aim is to understand the role of genomic variation in human physical attributes and sports performance. The aim of this report is to briefly comment on the current status of exercise genetics and genomics and to suggest potential improvements to the research agenda and translational activities. First, the genomic features of interest to the biology of exercise are defined. Then, the limit of the current focus on common variants and their implications for exercise genomics is highlighted. The need for a major paradigm shift in exercise genomics research is discussed with an emphasis on study designs and appropriately powered studies as well as on more mechanistic and functional research. Finally, a summary of current practices in translational activities compared with what best practice demands is introduced. One suggestion is that the research portfolio of exercise genomics be composed of a larger fraction of experimental and mechanistic investigations and a smaller fraction of observational studies. It is also recommended that research should shift to unbiased exploration of the genome using all the power of genomics, epigenomics and transcriptomics in combination with large observational but preferably experimental study designs, including Mendelian randomisation. In all cases, emphasis on replications is of paramount importance. This represents an extraordinary challenge that can only be met with large-scale collaborative and multicentre research programmes.

The aim of this report is to briefly comment on the current status of exercise genetics and genomics and to suggest potential improvements to the research agenda and translational activities. First, the nature of the genomic features of interest to the biology of exercise will be defined. Then, the potential role and limitation of the current focus on common variants and their implications for exercise genomics will be highlighted. The need for a major paradigm shift in exercise genomics research will be discussed with an emphasis on study designs and appropriately powered studies as well as on more mechanistic and functional research. Finally, a summary of current practices in translational activities compared with what best practice demands will be introduced.

HUMAN GENOMIC FEATURES OF INTEREST

With the availability of the almost complete human genome sequence in 2001,^{1 2} one could begin to look for deviations from the reference sequence and to begin exploring in-depth interindividual

differences at the DNA level. This was later greatly facilitated by the HAP-MAP³ and the 1000 Genome⁴ projects. Important advances towards this goal soon occurred when the whole genome sequence of literally thousands of individuals became available for detailed analyses. A high altitude view of the current understanding of the extent of sequence differences among people suggests that the DNA features listed in box 1 are of particular importance to exercise genomics. There are about 40 million polymorphic DNA sites in the genome of populations studied to date where the minor allele is observed at a frequency of at least 1%. Any given human being studied thus far has been shown to carry from 3 to 4 million of these common DNA variants. Moreover, estimates of the number of rare variants carried by any of us range from 200 000 to 500 000 variants. The latter variants are said to be of recent origin and are characterised as 'private' or limited to a pedigree or a 'clan'.⁵ Interestingly, rare variants tend on average to exhibit larger effect sizes than common variants. One last feature that is highlighted in this cursory account would be that there are extraordinary large numbers of short and large sequences repeat variants in the human genome, and this class of variants has been shown to have profound functional consequences at times.⁶ These copy number variable regions represent about 10% of the whole human genome.

In 2012, a whole series of reports published in leading journals (*Nature*, *Science*, *Journal of Biological Chemistry*, *Genome Research*, etc) described the progress achieved in the Noncoding Human Genome Sequence (ENCODE) project.⁷ The findings are of direct relevance to our expose. ENCODE has concluded that about 1% of the human genome sequence encodes an estimated 20 687 protein-coding genes. Another major observation is that about 80% of the genome is transcribed with encoded products participating in the regulation of gene expression and other cellular events. The human genome harbours almost three million protein-binding sites along its DNA. The 1800 or so transcription factors identified to date have been shown to bind at DNA sites representing about 8% of the genome. There are about 8800 small RNAs, including about 1000 miRNAs, and more than 9600 long RNAs being transcribed in at least one type of cells and most participate in the regulation of transcription and translation. One last set of numbers to illustrate the inherent structural complexity of the genome: human DNA encodes about 70 000 promoter regions and 400 000 enhancer regions, and those sequences are at times at substantial genomic distance from the genes they



CrossMark

To cite: Bouchard C. *Br J Sports Med* 2015;**49**:1492–1496.

Box 1 An overview of human genomic features of interest to physiologists

- ▶ About 40 million common polymorphic sites in the genome of *Homo sapiens*
- ▶ An individual carries 3–4 million common variant alleles
- ▶ An individual carries 200 000–500 000 rare variants
- ▶ Copy number variants of repeat sequences are common and cover about 10% of the human genome
- ▶ Less than 21 000 protein-coding genes
- ▶ About 2.9 million DNA protein-binding sites or an average of about 250 per protein-coding gene
- ▶ About 1000 miRNAs, 9000 other small RNAs and 10 000 long RNAs are transcribed
- ▶ About 1800 transcription factors binding at about 8% of the genome: an average of about 12 per protein-coding gene
- ▶ About 70 000 promoter sequences and 400 000 enhancer regions regulating gene expression have been identified or an average of about 22 of these sites per protein-coding gene

are known to regulate. Thus, a complex network of regulatory molecules and DNA binding sequences are involved in the widely distributed regulation of less than 21 000 protein-coding genes. To illustrate this extraordinary complexity, there are on average as many as 250 DNA protein-binding sites per protein-coding gene, each transcription factor could relate to an average of 15 protein-coding genes, and there are on average more than 23 promoter and enhancer regions per protein-coding gene.

All of the above strongly suggest that the relationship between the genome and a given phenotype is highly complex. In retrospect, exercise genomics research paradigms were naïve in the early days as it was assumed that there was a more or less direct and unequivocal relations between DNA sequence variants and traits of interest. This simple model did not recognise the highly complex distributed regulation of gene expression and of post-transcription and post-translation events. The effect of a common DNA variant on exercise-related traits is likely to pale in comparison to the non-genetic events impacting cells and tissues such as energy demands of exercise, energy status of cells, effects of intracrine, autocrine, paracrine and endocrine factors, including cytokines, inflammatory factors and others, or levels of activation of autophagy and apoptosis pathways to name some of the most impactful.

COMMON VARIANTS: LESSONS FROM A DECADE OF GENOME-WIDE ASSOCIATION STUDIES

Over the last decade, with the advent of technological advances allowing to genotype millions of common single nucleotide polymorphisms (SNPs) in each individual of a study, the investigation of the contribution of common variants to human characteristics and disease outcomes has grown exponentially. Microarray chips have been designed for the genotyping of large number of SNPs providing a dense covering of all chromosomes (at times only the autosomes), and these tools have made it possible to undertake unbiased and hypothesis-free genome-wide association explorations for continuous and discontinuous traits. Based on genome-wide association studies (GWAS) results and importantly on meta-analysis of GWAS for human traits, there are multiple lessons that can be carried forward for the benefit of exercise genomics.

At the outset, it should be emphasised that the above does not imply that rare genomic variants are not important for

exercise genomics. Rare variants are undoubtedly critical to exercise genomics research but they are extremely challenging to study. To uncover a rare variant associated with an exercise-related trait, one would typically have to sequence the genome of a large number of individuals on which the relevant trait has been measured. Since the rare variants tend to cluster in families, pedigrees or clans and often exhibit large effect size in comparison to common variants, alternate approaches have been proposed.⁵ The poster child for a rare variant impacting sports performance predates modern genomics era. It was triggered by observations made on Eero Antero Mäntyranta, a world class cross-country skier from Finland in the 1960s. Mäntyranta won a total of seven medals in the 1960, 1964 and 1968 Winter Olympic Games plus five more in World Championships. He was shown to have primary familial and congenital polycythaemia (familial erythrocytosis) resulting in a major increase in haematocrit (68%) and haemoglobin (231 g/L) caused by a hypersensitivity of erythroid progenitors to erythropoietin.⁸ In the extended pedigree of the proband, 29 of the 200 individuals were shown to be affected by familial erythrocytosis and participants were shown to be generally healthy. The causal mutation was localised by linkage studies and was identified as a truncation of the 70 C-terminal amino acids of the erythropoietin receptor gene (*EPOR*), where a G to A transition converted TGG (tryptophan) to TAG (stop codon).^{9–10} This is clearly the most vivid example of the impact of a rare variant on an exercise-related trait, although other cases have been described (eg, variants in *MSTN* (myostatin) or *EPAS1* (endothelial PAS domain protein 1)) but their true effect size and physiological significance remain to be established in humans. One can also argue that mutations in genes resulting in defective or diminished gene products causing intolerance to exercise fall into the category of rare variants affecting sport performance potential.¹¹

The textbook model of rare versus common DNA variants affecting biology or behaviour posits that most common variants are ancient and have been exposed to the forces of selection for hundreds and perhaps thousands of generations. A variant allele being present at a reasonable frequency in a given population is a reflection of the fact that it does not cause sufficient deviations from the norm, say homeostasis, to have been strongly selected against. On the other hand, it does not carry sufficient benefits to have been strongly selected for either. These variants are typically characterised by small effect sizes. In contrast, rare variants are typically of a recent origin and may have arisen only once in the gamete of a given individual and may have become more frequent in a given family line or a large pedigree. Evolutionary pressures against such variants can be extremely strong if they are lethal or diminish Darwinian fitness, but they can also be selected for if they carry substantial benefits. Not all rare variants have large effect sizes, but some do and they have not been totally eliminated from the gene pool or fixed in a population of interbreeding individuals (if they are very advantageous) because of their recent origin.

Common variants have received the most attention in exercise genomics studies and we will now focus on their properties. We have recently learned a great deal about the genetic architecture of common traits from large-scale GWAS reports and the lessons have serious implications for exercise genomics. For instance, in a meta-analysis of GWAS data from multiple cohorts encompassing 253 288 adults of European descent, 697 SNPs were associated with adult human height at the genome-wide significance level of 5×10^{-8} and these SNPs collectively accounted for about 20% of the variance in adult height.¹² It

was estimated that with a much larger sample size, it would take 9500 SNPs with statistical significance to explain about 30% of the variance in height and, with an almost infinite sample size, all common SNPs together would account for a maximum of 60% of the trait variance. Similar observations have been reported for body mass index (BMI). The most recent effort based on a meta-analysis of GWAS (total number of 339 224) identified 97 SNPs (at $p=5 \times 10^{-8}$ or better) accounting for only 2.7% of the BMI variance.¹³ The SNPs had comparable effects across genders and ancestries. Close examinations of the consistency in p levels and direction of the effect of the minor allele at all SNPs suggested that with a substantially larger sample size, more than 1000 SNPs would prove to be true BMI SNPs and collectively they would account for about 20% of the BMI variance. Although this would represent a substantial improvement, it would still be well short of the commonly reported heritability levels for BMI.¹⁴ Interestingly, a >3 BMI (about 12 kg) unit difference was observed between the adults who carried 78 risk alleles or less compared with those with 104 and more risk alleles at the 97 SNPs. In other words, despite the small effect size of these alleles considered individually, the susceptibility load arising from the presence of a large number of risk alleles can be biologically meaningful. Similar observations have been made for other traits including anthropometric indicators of upper body fat and abdominal fat¹⁵ and resting heart rate.¹⁶ Lessons to keep in mind for exercise genomics!

The examples described above are not extreme cases. Rather they seem to be truly representative of the contribution of common SNPs to the biology of complex traits. Table 1 summarises the data on a panel of traits that were recently reviewed.¹⁷ Shown are the typically reported heritability level ranges, the sample sizes for the various meta-analysis GWAS, the number of SNPs found with significance at the genome-wide p value, the percentage of the trait variance explained by these panels of SNPs and the sources for these data. Despite the fact that these meta-analyses were based on multiple cohorts and large sample sizes, the variance accounted for by significantly associated SNPs ranges from a low of 1.4% to a high of 14%. It is expected that both the number of statistically significant SNPs and the per cent variance they account for in a given trait will increase as the studies are performed with growing sample sizes. However, it is also recognised, as was clearly shown recently for BMI,¹³ that the effect size of statistically significant SNPs being accrued becomes progressively smaller, reaching at times only a small fraction of 1%.

The variance explained by common SNPs may increase to some extent once SNP-age, SNP-gender, SNP-exercise, SNP-diet,

SNP-smoking, SNP-alcohol intake and other potentially relevant SNP-environmental and behavioural factor interaction effects are incorporated in the models. The caveat here is that even larger sample sizes will be required to investigate the main effect plus potential interaction effects of environmental and behavioural factors with a SNP on a complex trait; and the sample size will have to be extraordinary large to incorporate multiple SNPs or a genome-wide panel of SNPs in the model.

One other important issue that deserves consideration is that of the potential pleiotropic effects of a SNP on two or more traits. This is a topic that is likely to be of relevance to exercise genomics as adaptation to exercise depends on a complex interplay of multiple organs and systems. In a recent investigation of the question for cardiovascular risk factors and coronary heart disease as an outcome, all the SNPs identified at the genome-wide significance level in the most comprehensive GWAS for coronary heart disease, BMI, C reactive protein, blood pressure, total cholesterol, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, triglycerides and type 2 diabetes were retrieved.¹⁷ A total of 181 SNPs mapping to 56 gene loci showed statistical evidence of pleiotropism as they were associated with two or more traits. Most of these pleiotropic SNPs related to coronary heart disease and lipids but some associated with BMI and other CVD risk factors. There was enrichment for liver X receptor, retinoid X receptor and nuclear signalling genes among the pleiotropic loci.

The lessons for exercise genomics from these recent studies and others are numerous and have to be taken very seriously. In brief, all traits are complex and multifactorial. Common variants have small effect sizes. Focusing on one or a few loci/SNPs in cohort and small-scale association studies is unlikely to generate useful and reproducible findings. However, focusing on one or a few loci/SNPs in intensive mechanistic studies is essential for progress to occur in exercise genomics. In large cohort studies, it is preferable to use unbiased genome-wide screen approaches than a priori and arbitrarily defined candidate loci/SNPs. The regulation of a given gene (and of all genes) is not dependent on a single sequence element. Quite contrary, the regulation is widely distributed as shown by the early findings of the ENCODE project. This is of course in line with the small effect size reported for significant SNPs for just about any trait investigated to date. One implication for exercise genomics is that it has to bring into the tent more computational biology and bioinformatics expertise. Finally, rare variants may be the most clinically relevant type of genomic variants, but they are also among the most challenging to uncover.

Table 1 Heritability levels, number of loci with genome-wide significance and percentage of trait variance explained by these loci

Phenotype	Heritability (%)	Number of participants	Number of loci at $p < 5 \times 10^{-8}$	Per cent of variance explained	Reference
BMI	25–60	339 224	97	2.7	13
WHR adjusted BMI	40–55	224 459	49	1.4	15
T2 diabetes	45–90	149 821	65	5.7	18
Total cholesterol	50–60	188 577	40	12–14	19
Triglycerides	25–60	188 577	32	10–13	19
hs-CRP	25–55	81 870	18		20
Systolic BP	20–70	203 056	25	<1.0	21
Hypertension	50	203 056	11	<1.0	21

The numbers are from the latest and largest meta-analysis of GWAS report for each trait.

Adapted and expanded from Rankinen *et al.*¹⁷

BMI, body mass index; BP, blood pressure; GWAS, genome-wide association studies; hs-CRP, high-sensitivity C reactive protein; WHR, waist to hip ratio.

The genomic architecture of complex exercise-related traits will undoubtedly in the end include informative rare and common polymorphisms. But one can also expect that other genomic variable features such as variability in the number of copies of short (dinucleotides and trinucleotides) and long sequences repeats will be playing a role. This is a line of investigation that has been largely ignored thus far in exercise genomics.

PARADIGM SHIFT IN EXERCISE GENOMICS

What are the directions to be embraced by exercise genomics research in order to increase understanding of genes, pathways and networks contributing to human heterogeneity in adaptation to physical activity programmes aimed at disease prevention or rehabilitation, and exercise training regimens designed to improve performance including elite athletic performance? By now it should be clear to all that doing more of the same will not lead to transformative results. This section offers a few suggestions on study designs and technologies aimed at improving the quality of the science in exercise genomics. It is by no means a complete expose of all measures that would augment the credibility of our science.

One obvious measure that would immediately improve the credibility of exercise genomics research would be to abandon the current practice of pursuing the current study and publication pipeline characterised by an almost total reliance on observational study designs and poorly justified candidate genes with small sample sizes. There was a time when it was justifiable to pursue these types of enquiries, but those days are long gone and their potential to generate credible answers to the many open questions in exercise genomics is almost nil. There are situations in which observational study designs can play a useful role. They can provide us with opportunities to explore in unbiased manners the whole genome, quantify the effect size of given genomic and epigenomic markers, test whether there are gene-behaviour interaction effects or gene-gene (or variant-variant) interactions impacting population variance in a given physical activity-related trait. However, a prerequisite for such studies is that they are based on very large sample sizes.²² There is a place for candidate gene studies, but careful considerations must be given to study design and statistical power issues. One approach with considerable merit, known as Mendelian randomisation, consists using genotype or haplotype information in observational or experimental study designs.^{23 24} However, testing the true contribution of a given candidate gene is best done in a setting in which an experimental (exercise) group is compared with control participants. Experiments based on animal models can often provide robust evidence for the potential involvement of a candidate gene and the potential mechanisms of a causal relationship for an exercise-related phenotype. Overall, it would be in the best interest of all if the research portfolio of exercise genomics was composed of a very large fraction of experimental and mechanistic investigations and a much smaller fraction of observational studies.

In the end, in all studies of complex traits, whether at a GWAS level or for a given candidate gene, even when the sample size is deemed adequate, it is critical that replication of the findings be provided. Replication studies can take multiple forms as was outlined in a recent report.²² In this regard, collaborative research and data sharing practices offer hope that exercise genomics can develop the kind of large resources needed not only for the discovery process but also for the replication phase of genomic research. If these basic conditions became the rule rather than the exception in exercise genomics research,

one would soon see the emergence of a solid body of knowledge that would gain acceptance and respect not only among exercise scientists and sports medicine physicians but also in the scientific community at large.

GWAS have provided us with a unique opportunity to explore the whole genome in an unbiased manner in the search for sequence variants associated with traits of interest. The ultimate unbiased screen of the genome is obviously the whole genome sequencing of all participants of a study. This is not yet common practice, but it is increasingly becoming the preferred approach and will likely be best practice in a not too distant future as the cost of sequencing the whole human genome continues to come down.²⁵ Exploration of genomic features can also be undertaken in an unbiased manner with whole transcriptomic profiling in relevant tissues, such as skeletal muscle, or whole epigenome screen using methylome arrays. Gene expression profiling can yield unbiased (when done objectively) transcript signatures and molecular targets that could constitute candidates for subsequent genomic and genetic studies.²⁶ An epigenome-wide snapshot of all methylated sites, particularly those located in gene regulatory sequences, has the potential to generate hypotheses and targets that could lend themselves to further genomic and mechanistic investigations.

In summary, we propose that exercise genomics abandon the current practice of focusing on candidate genes typically defined by authors' preference or from biases in the published scientific literature, and the reliance on small, statistically underpowered, observational studies. Instead, we recommend that exercise genomic science shifts to unbiased exploration of the genome using all the power of genomics (both GWAS and whole genome sequencing), epigenomics and transcriptomics in combination with large observational (preferably prospective) and experimental study designs, including Mendelian randomisation. In all cases, emphasis on replications is of paramount importance.

A final comment: The overarching aim of exercise genomics studies is to identify genomic variants impacting cellular functions in such a way that their discrete effects on physiology or behaviour can be observed if they truly exist. This represents an extraordinary challenge as most DNA variants do not seem to impact gene expression and do not seem to relate to epigenomic markers. Moreover, considering that the regulation of gene expression is complex, multifactorial and widely distributed, DNA variants with small effect sizes are not likely to impact a phenotype in an easily detectable, Mendelian way. Thus, as has been shown repeatedly even in simpler model organisms,²⁷ it is and will undoubtedly remain one of the great challenges of current-day biology to be able to link mechanistically genotype and phenotype.

TRANSLATIONAL OPPORTUNITIES

We will conclude with brief comments on exercise genomics and translational opportunities. Up-to-now, it is fair to summarise that exercise genomics has not generated evidence of a quality that can be actionable, particularly in the context of translating basic science information acquired at the 'bench' to patients (or athletes) or populations. It is not that the need for actionable exercise genomic information is not there, it is simply a reflection of the quality of the evidence accumulated thus far. Although the main goal of exercise genomics research should be to illuminate exercise biology and behaviour in order to better understand the preventive and therapeutic values of regular exercise, there is no doubt that a number of practical applications would arise if valid and replicated exercise genomic evidence could become the norm in the field.

One can speculate that the availability of strong and commonly accepted exercise genomics facts could lead to the development of genomic-based diagnostics that would over time grow into powerful and reliable classifiers. Such tools could be used in clinical settings to better match patients to therapies and in the broad context of secondary prevention. One situation where such diagnostic instruments are urgently needed is in area of the detection of potential adverse responses in those who are physically active or are beginning an exercise programme, so that appropriate preventive means can be deployed.²⁸ At the other end of the activity spectrum, there are lifelong endurance athletes who are more prone than others to develop cardiac arrhythmias²⁹ and other cardiac ailments.^{30–31} Being able to identify who is at risk of developing any of these conditions could lead eventually to individualised preventive measures or alternative therapies.

Needless to say, there is a strong interest in the world of sports for meaningful information that could improve the probability of identifying children and adolescents who are gifted for high level sports and athletic performance. This is of course an area where translational opportunities abound but also one where the current level of credibility of exercise genomics is at its lowest. The fact that industry has been marketing diagnostic tools that are way ahead of the science and have essentially no diagnostic value is not foreign to this state of affair. As is always the case with science, it does not pay to embrace the quick fix. Exercise genomics is no exception.

In summary, although translational opportunities are ubiquitous, exercise genomics has not progressed to the point that actionable findings are commonly recognised. There is no substitute for strong study designs, ample statistical power, replication of the most important observations in multiple settings, protection against publication bias (specially under-reporting of negative findings), and powerful diagnostic tools with excellent sensitivity and specificity. It is an extraordinary challenge to meet all these expectations. It is unlikely that a single laboratory can be persistently successful in present-day exercise genomics without close collaboration with other investigators. The future of exercise genomics lies in large-scale collaborative and multi-centre research programmes.

Summary

- Exercise genomics has the potential to make substantive contributions to our understanding of exercise biology.
- However, exercise genomics has not delivered thus far the high quality data required to meet expectations.
- It is not sufficient to launch larger studies using the same research designs: a paradigm shift is needed.
- Exercise genomics would benefit from a greater reliance on experimental studies and unbiased technologies to identify genomics, epigenomics and transcriptomics targets.
- Engaging in translational activities is a worthy pursuit but it is highly premature at this time to use genomic markers to advise or guide a decision making process related to fitness or sports performance goals.

Funding CB received book royalties from Elsevier, CRC Press and Human Kinetics. He received honoraria for lectures from Gatorade, Brazil DASA and Brazil Biogenetika. He served on advisory boards within the past 5 years for Pathway Genomics, NIKE and Weight Watchers.

Competing interests CB received research funding from the National Institutes of Health (HL-45670) and the Prince Faisal Foundation, Saudi Arabia. CB is partially funded by the John W Barton Sr Chair in Genetics and Nutrition.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

- 1 Lander ES, Linton LM, Birren B, *et al.* Initial sequencing and analysis of the human genome. *Nature* 2001;409:860–921.
- 2 Venter JC, Adams MD, Myers EW, *et al.* The sequence of the human genome. *Science* 2001;291:1304–51.
- 3 International HapMap Consortium. The international HapMap project. *Nature* 2003;426:789–96.
- 4 Abecasis GR, Auton A, Brooks LD, *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;491:56–65.
- 5 Lupski JR, Belmont JW, Boerwinkle E, *et al.* Clan genomics and the complex architecture of human disease. *Cell* 2011;147:32–43.
- 6 Henrichsen CN, Chaingat E, Reymond A. Copy number variants, diseases and gene expression. *Hum Mol Genet* 2009;18(R1):R1–8.
- 7 ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;489:57–74.
- 8 Juvonen E, Ikkala E, Fyhrquist F, *et al.* Autosomal dominant erythrocytosis caused by increased sensitivity to erythropoietin. *Blood* 1991;78:3066–9.
- 9 de la Chapelle A, Sistonen P, Lehväläho H, *et al.* Familial erythrocytosis genetically linked to erythropoietin receptor gene. *Lancet* 1993;341:82–4.
- 10 de la Chapelle A, Traskelin AL, Juvonen E. Truncated erythropoietin receptor causes dominantly inherited benign human erythrocytosis. *Proc Natl Acad Sci USA* 1993;90:4495–9.
- 11 Bouchard C, Rankinen T, Timmons JA. Genomics and genetics in the biology of adaptation to exercise. *Compr Physiol* 2011;1:1603–48.
- 12 Wood AR, Tuke MA, Nalls MA, *et al.* Another explanation for apparent epistasis. *Nature* 2014;514:E3–5.
- 13 Locke AE, Kahali B, Berndt SI, *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* 2015;518:197–206.
- 14 Perusse L, Rice TK, Bouchard C. Genetic component to obesity: evidence from genetic epidemiology. In: Bray G, Bouchard C, eds. *Handbook of obesity*. Boca Raton: Taylor and Francis Group, 2014:91–104.
- 15 Shungin D, Winkler TW, Croteau-Chonka DC, *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* 2015;518:187–96.
- 16 den Hoed M, Eijgelsheim M, Esko T, *et al.* Identification of heart rate-associated loci and their effects on cardiac conduction and rhythm disorders. *Nat Genet* 2013;45:621–31.
- 17 Rankinen T, Sarzynski MA, Ghosh S, *et al.* Are there genetic paths common to obesity, cardiovascular disease outcomes, and cardiovascular risk factors? *Circ Res* 2015;116:909–22.
- 18 Morris AP, Voight BF, Teslovich TM, *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* 2012;44:981–90.
- 19 Willer CJ, Schmidt EM, Sengupta S, *et al.* Discovery and refinement of loci associated with lipid levels. *Nat Genet* 2013;45:1274–83.
- 20 Dehghan A, Dupuis J, Barbalic M, *et al.* Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for c-reactive protein levels. *Circulation* 2011;123:731–8.
- 21 Ehret GB, Munroe PB, Rice KM, *et al.* Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* 2011;478:103–9.
- 22 Loos RJ, Hagberg JM, Pérusse L, *et al.* Advances in exercise, fitness, and performance genomics in 2014. *Med Sci Sports Exerc* 2015;47:1105–12.
- 23 Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;32:1–22.
- 24 Timpson NJ, Wade KH, Smith GD. Mendelian randomization: application to cardiovascular disease. *Curr Hypertens Rep* 2012;14:29–37.
- 25 Roach JC, Glusman G, Smit AF, *et al.* Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* 2010;328:636–9.
- 26 Bouchard C. Overcoming barriers to progress in exercise genomics. *Exerc Sport Sci Rev* 2011;39:212–17.
- 27 Dowell RD, Ryan O, Jansen A, *et al.* Genotype to phenotype: a complex problem. *Science* 2010;328:469.
- 28 Bouchard C, Blair SN, Church TS, *et al.* Adverse metabolic response to regular exercise: is it a rare or common occurrence? *PLoS ONE* 2012;7:e37887.
- 29 Andersen K, Farahmand B, Ahlborn A, *et al.* Risk of arrhythmias in 52 755 long-distance cross-country skiers: a cohort study. *Eur Heart J* 2013;34:3624–31.
- 30 O'Keefe JH, Patil HR, Lavie CJ, *et al.* Potential adverse cardiovascular effects from excessive endurance exercise. *Mayo Clin Proc* 2012;87:587–95.
- 31 Wilson M, O'Hanlon R, Prasad S, *et al.* Diverse patterns of myocardial fibrosis in lifelong, veteran endurance athletes. *J Appl Physiol* 2011;110:1622–6.