Published in final edited form as:
Exerc Sport Sci Rev. 2011 October ; 39(4): 212–217. doi:10.1097/JES.0b013e31822643f6.

Overcoming Barriers to Progress in Exercise Genomics

Claude Bouchard
Human Genomics Laboratory, Pennington Biomedical Research Center, Baton Rouge, LA 70808

Abstract

This commentary focuses on the issues of statistical power, the usefulness of hypothesis-free approaches such as in genome-wide association explorations, the necessity of expanding the research beyond common DNA variants, the advantage of combining transcriptomics with genomics, and the complexities inherent to the search for links between genotype and phenotype in exercise genomics research.

Keywords

exercise genomics; candidate genes; single nucleotide polymorphism; genome-wide association study; genomic predictors

INTRODUCTION

Human genomics is the science that investigates the physical features and properties of the human genome, while *human genetics* is the science of inheritance, i.e., the transmission of traits across generations. Both concepts are important for this report. When multiple genes contribute to variation in a trait, it becomes very challenging to define the mode of inheritance. Phenotypes of interest to exercise biologists are overwhelmingly of this class and are referred to as quantitative and polygenic traits.

The early 21st century is an interesting period to be involved in genetics and genomics research. The extraordinary progress made in the past decade on the sequence and information content of the human genome and the genomes of commonly used animal models, on high-throughput technologies for genotyping and sequencing whole genomes, and on transcriptomics has generated a lot of enthusiasm regarding our ability to elucidate the genetic, genomic, and molecular basis of human variation in health and disease. For instance, the availability of the human genomic sequence has played a key role in making it possible to identify more than 2850 Mendelian disease genes (16). Importantly, the cost of genomics technologies has decreased dramatically within the short span of about one decade. In fact, the cost of sequencing a first human genome was of the order of $3 billion, but this cost is soon predicted to reach only about $1,000 and perhaps even less. Even though it remains modest, the pace of exercise genomics research is also accelerating. There are a growing number of laboratories engaged in exercise genomics, and in the aggregate,

Corresponding author: Claude Bouchard, John W. Barton, Sr. Endowed Chair in Genetics and Nutrition, Human Genomics Laboratory, Pennington Biomedical Research Center, 6400 Perkins Road, Baton Rouge, LA 70806, Phone: 225-763-2663, fax: 225-763-2525, claude.bouchard@pbrc.edu .

they cover a wide range of experimental models and technologies. One can therefore predict that there will be significant advances in the near future. The relative contributions of genetic and environmental factors to phenotypic variation in a variety of human traits have long been a topic of interest. The first papers that dealt in some fashion with the magnitude of the genetic variance for phenotypes of interest to exercise science were published in the 1960s, and these types of papers dominated the field until the 1990s. Then began a second phase of research focused on the identification of genes and DNA sequence variants contributing to human variation in exercise-related traits. Exercise genetics research began with a flurry of underpowered candidate gene studies, as was the case in all other areas. But the field is now moving on to better and more powerful approaches.

The genetic epidemiology literature dealing with nature versus nurture issues for exercise-related traits up to the mid-1990s was reviewed in depth in a 1997 book (4). This was followed by a series on the gene map for performance and fitness traits published yearly from 2001 to 2009 in *Medicine and Science in Sports and Exercise (MSSE)* (e.g., 7, 21). More recently, the same group shifted its emphasis from a yearly update of the gene and marker compendium to a review of the most significant exercise genomics and genetics findings (12, 22). Finally, the latest and most comprehensive reviews cover either the whole field, as in a volume published in 2011 under the aegis of the Medical Commission of the International Olympic Committee (3), or emphasize the biology of adaptation to exercise (5).

This report offers comments on future progress in exercise genomics research, including issues of statistical power, the usefulness of hypothesis-free approaches such as in genome-wide association explorations, the necessity of expanding the research beyond common DNA variants, the advantage of combining transcriptomics with genomics, and the complexities inherent to the search for links between genotype and phenotype.

ACHIEVING ADEQUATE STATISTICAL POWER

One of the most critical attributes of a research study is its inherent statistical power to detect an effect or a difference when one exists in the real world. One of the consequences of an inadequate level of statistical power is that a study cannot reliably reject the null hypothesis or protect against declaring a difference or an effect that does not exist in reality. Moreover, a common weakness of exercise genomics reports is the failure to correct for the fact that multiple tests have been performed. A stronger level of evidence is required to guard against false-positive and false-negative results when multiple tests are performed, for instance in the comparison of athletes and sedentary people. These two issues are particularly critical in genomics studies, as large numbers of single nucleotide polymorphisms (SNPs) are typically tested in the same experiment. We have to recognize that failure to use an adequate sample size and to adjust the significance level for multiple testing, more than any other factors, are responsible for the lack of reproducibility of exercise genomics findings.

Exercise genomics research is generally based either on quantitative traits measured on continuous scales (e.g., skeletal muscle strength) or on dichotomous traits (e.g., active versus inactive participants). In discussing statistical power issues, it is useful to distinguish between these two situations. This report emphasizes some of the points made in a recent exercise genomics review published in *MSSE* and presents a key figure adapted from this publication (12). The relationship between the contribution of a SNP and the sample size required to find a significant association with 80% statistical power under an additive model is shown in figure 1 for continuous and dichotomized traits.

Panel A summarizes a simulation for a continuous trait with the SNP effect size expressed as R^2. In this case, the minor allele frequency (MAF) has no impact on the sample size. This figure illustrates only the scenario in which 500 SNPs were tested for associations with the trait of interest (assuming only one trait). The multiple testing-adjusted *P*-value threshold for significance is now set at 0.0001. Under these conditions, a minimum of 2,200 subjects are needed to detect a significant association with a SNP that accounts for 1% of the trait variance. A 1% effect size is quite realistic for observational studies based on the extensive experience accumulated in large-scale explorations of the genome for SNPs associated with human diseases and common traits. If the SNP had an effect size of 2%, it would necessitate approximately 1200 subjects for the same alpha level.

Panel B depicts the curve of the odds ratio (OR) versus the number of subjects per group in a case-control design assuming an approximately equal number of cases and controls. The figure is for an additive model and a MAF of 20%. Based on genome-wide association studies (GWAS)-derived ORs for common human diseases and traits from case-control studies, an OR of 1.2 is a realistic expectation. If 500 SNPs are tested (alpha = 1×10^{-4}), we need 4000 cases and 4000 controls to find an association with an OR of 1.2 with 80% power. If the MAF is less than 20%, we need even larger sample sizes, whereas the sample size requirements decrease by about one-third when MAF approaches 50% (not shown in the figure).

Other likely scenarios are discussed in Hagberg et al (12). Needless to say, if we raise the bar to a more conservative statistical power of 90%, the sample size requirements become markedly larger. One more issue: Figure 1 depicts situations in which the main effects of SNPs are being tested. The sample size requirement increases several folds when an interaction term is being investigated for its contribution to the variance of a phenotype. For instance, a test of a SNP–physical activity interaction effect on adiposity requires at least four times more subjects than a test of the main effect of the same SNP.

GENOME-WIDE ASSOCIATION STUDIES

It is clear by now that reliance on candidate genes and personal views of what constitutes a valid candidate is insufficient for a solid foundation upon which a complete genomic anatomy of complex traits can be built. New candidates need to be identified from unbiased exploration methods that are less sensitive to personal views and preferences. One such method is the genome-wide association exploration, which has been primarily used to date in observational studies. For a successful GWAS, a large sample size, a properly measured trait, and a sufficiently large and high-quality set of SNPs are needed.

The development of high-throughput SNP genotyping methods combined with the availability of the human genome sequence and the inventory of DNA sequence variants in the major ethnic groups (from the *HapMap* and the *1000 Genomes* projects) has made it possible to undertake genome-wide association screens for continuous and dichotomous traits at a reasonable cost. More than 25 million SNPs with a MAF of at least 1% have been identified in the human genome (1000genomes.org), with at least 10 million of them reported to have a MAF of 5% or more. Interestingly, based on the growing number of personal genomes that have been sequenced, a given individual carries about 4 million variants, most of which are SNPs.

Variants in close physical proximity are often transmitted together across generations as a unit or a block, i.e., recombination among them is infrequent. Because of this property, it is possible to tag a set of SNPs by genotyping only the most representative SNP, which is commonly referred to as the tagSNP. As a result, genotyping 500,000 to 1 million SNPs has been considered as sufficient to capture most of the common SNPs for a given ethnic group.

The first GWAS based on this technology was published in 2005 (15). Research has since advanced at a very rapid pace. There are currently (end of February 2011) 805 published GWASs in the peer-reviewed literature, covering more than 150 diseases and complex traits. It is common in observational studies, assuming 1 million SNPs, to use $P < 5 \times 10^{-8}$ as the threshold for genome-wide statistical significance adjusted for multiple testing, and almost 4000 SNPs have been associated with a trait at this genome-wide significance level (19, 20). The catalog of these studies and their findings can be accessed at the National Human Genome Research Institute Web site (20).

Exercise science has been slow in incorporating GWAS in its portfolio of research technologies. However, two such studies have been reported. The first GWAS of interest dealt with physical activity level in observational studies. It was based on two cohorts: 1644 unrelated individuals from the Netherlands Twin Register and 978 subjects living in Omaha, NE (10). Exercisers were defined as subjects who reported at least four MET-hours per week of leisure-time physical activity. An exerciser versus nonexerciser classification was used as the trait for the GWAS analyses. The final genotype data set included 1.6 million measured or imputed SNPs. None of the 1.6 million SNPs reached the threshold of genome-wide significance ($P = 5 \times 10^{-8}$). However, SNPs in three genomic regions showed *P*-values less than 1×10^{-5}. The strongest associations were observed on chromosome 10q23.2 at the 3′-phosphoadenosine 5′-phosphosulfate synthase 2 (*PAPSS2*) gene locus. The associations with previously reported physical activity candidate genes and linkage regions were explored, but they generated no conclusive evidence. As no SNP reached a genome-wide significance level, one key lesson from this first study is that much larger sample sizes are needed if a true genomic dissection of physical activity as a behavior is to be undertaken.

The second GWAS-based report focused on exercise training-induced changes in $\dot{V}O_2$max in the HERITAGE Family Study and their associations with a panel of more than 320,000 SNPs (6). Here the situation is quite different: we are dealing with a phenotype that was altered experimentally. The assumption was made that SNPs with larger effect sizes would be detected under such conditions. Again, none of the SNPs reached the stringent threshold *P*-level for genome-wide significance, even though the study was based on a much smaller set of SNPs than assumed when the alpha level is set at 5×10^{-8}. In single-SNP analysis, a total of 39 SNPs were . associated with $\dot{V}O_2$max training response at $P < 1.5 \times 10^{-4}$. The strongest evidence of association ($P = 1.3 \times 10^{-6}$) was observed with a SNP located in the first intron of the acyl-CoA synthetase long-chain family member 1 (*ACSL1)* gene, located on chromosome 4q35. When all 39 SNPs were analyzed simultaneously in multivariate regression models, 21 SNPs were retained in the final model. Of these, nine SNPs each explained at least 2% (range 2.2% to 7.0%) of the variance ($P < 0.0001$ for all), while seven markers contributed between 1% and 2% each. Collectively, these 16 SNPs explained 45% of the variance in $\dot{V}O_2$max training response, which is very close to the maximal heritability estimate of 47% reported previously in the HERITAGE Family Study (2). Attempts to replicate these associations in three smaller cohorts were only partially successful. A predictor score was constructed using the 21 SNPs from the final regression model derived from the HERITAGE cohort. Each SNP was recoded based on the number of high $\dot{V}O_2$max training response alleles, and the sum of the high-trainability alleles ranged from 7 to 31 in HERITAGE whites. The difference in $\dot{V}O_2$max training response between those with the lowest (9 or less, n = 36, mean = +221 mL/min) and the highest classes of the predictor score (19 or more, n = 52, mean = +604 mL/min) was almost 400 mL O_2/min (figure 2).

The genome-wide approach is not restricted by *a priori* hypotheses, as is the case in candidate gene studies. Moreover, a GWAS can cover almost the entire genome uniformly and at a dense level. A critical feature of any genetics study is replication, which implies that the findings of an individual study must be replicated in other large cohorts with a similar

phenotype and study design. If the associations are replicated again and again, the case for the contribution of a DNA sequence variant to the trait of interest becomes considerably stronger.

Exercise genomics would greatly benefit from the availability of large cohorts of individuals who have been phenotyped for appropriate exercise-related traits. Pooling the data from several cohorts would allow for even greater statistical power to detect loci with small effect sizes. Establishing such large cohorts in populations of whites, African-Americans, and other ancestries would also be useful for the definition of population heterogeneity. While this is an achievable goal for observational studies, it is a much greater challenge for genomics studies of the response to regular exercise, as maximum compliance with the exercise regimen is required for this type of research to be successful.

BEYOND COMMON VARIANTS

A hotly debated issue at the moment is how much of the genetic architecture of common complex traits is attributable to common and rare DNA sequence variants. In brief, the common variant hypothesis assumes that a trait is affected by several common DNA variants, each with minor effect size, while the rare variant scenario proposes that a large number of rare but relatively high-impact variants are contributing to the trait variance. Examples supporting both hypotheses can be found in the literature. However, interest in the rare variant hypothesis has been renewed by the consistent observation that common SNPs found to be associated with common diseases in genome-wide reports explain only a modest fraction of the estimated trait heritability (18).

One could argue that structural variants, such as small insertions and deletions, copy number variants (CNVs), and balanced rearrangements (e.g., inversions and translocations), need to be incorporated in our models if we are to fully account for the contribution of DNA sequence variation to the genetic architecture of common complex traits. CNVs can encompass duplications that range in size from a few to several million base pairs. Data from the full sequence of individual human genomes reported thus far and from the 1000 Genomes project show that insertions, deletions, and CNVs are quite common.

The next-generation sequencing techniques provide valuable tools to explore common variants, rare variants, CNVs, and other structural variants in relation to complex multifactorial traits of interest to human biologists. Exercise genomics needs to incorporate them all moving forward.

COMBINING TRANSCRIPTOMICS AND GENOMICS

Another strategy that can be used to identify gene targets in a relatively unbiased mode for subsequent genomics and genetics research relies on the study of transcript abundance in relevant tissues in order to derive molecular predictors of a trait. These new targets can then be probed for DNA sequence variants, and their associations with traits of interest can be investigated. In this regard, transcriptomics has been a common tool of exercise molecular biology studies for about a decade. For instance, one report dealt with cardiac expression levels of transcripts in middle-aged and old male mice derived from sedentary and spontaneously physically active breeding lines (8). In another study, the effects of unloading and low-intensity activity on the gene expression pattern of the soleus muscle were investigated in rats (1). Microarray technologies have also been used in human experiments. For instance, the effects of a strength training program on vastus lateralis gene expression profile were studied in sedentary subjects (23).

As an example of integration between transcript abundance, protein content and localization, and DNA differences, a series of studies focused on muscle damage mechanisms induced by eccentric exercise was performed (9, 13). Inflammatory factors were shown to be upregulated following eccentric maximal exercise in 157 sedentary men and women, with three genes (chemokine ligand 2 [*CCL2*], zinc finger protein 36, C3H type, homolog [*ZFP36*], and CCAAT/enhancer binding protein, delta [*CEBPD*]) exhibiting upregulation with the first exercise bout and further upregulation with the second exercise session. SNPs in *CCL2* and *CCR2* were associated with exercise-induced muscle damage.

Gene expression profiling studies were subsequently used to generate new candidate genes to be tested for the role of allelic variation in the HERITAGE Family Study cohort of whites. A first report focused on transcripts from the vastus lateralis muscle associated with the exercise training response of insulin sensitivity (25). Eight subjects who were high responders for insulin sensitivity and eight age-, sex-, and BMI-matched nonresponders, all with muscle biopsies, were compared for transcript abundance before and after the exercise program. Four genes (v-ski sarcoma viral oncogene homolog [*SKI*], four and a half LIM domains 1 [*FHL1*], titin [*TTN*], pyruvate dehydrogenase kinase, isozyme 4 [*PDK4*]) exhibited at least a 50% difference in expression between high responders and nonresponders. Three SNPs were genotyped in the *FHL1* gene encoded on Xq26 and tested for associations with exercise training-induced changes in insulin metabolism phenotypes (26). Two of these SNPs (rs9018 and rs2180062) were associated with insulin and glucose training responses. *FHL1* is the first gene (encoded on X) that differentiates between those who respond favorably to regular exercise and those who exhibit an adverse response pattern in terms of insulin and glucose homeostasis traits.

Recently, the strategy of combining transcriptomics and genomics was taken a step further. Timmons and colleagues used a combination of global skeletal muscle gene expression profiling and DNA sequence variants to identify genes associated with $\dot{V}O_2$max trainability (27). A panel of 29 transcripts was strongly associated with $\dot{V}O_2$max training response in a study of 24 sedentary males trained for 8 weeks and in a replication training study. Subsequently, tagSNPs of the 29 predictor transcripts were genotyped in the HERITAGE Family Study. A multivariable regression analysis using the predictor gene SNPs and a set of SNPs from positional cloning and candidate gene studies of the HERITAGE Family Study identified a set of 11 SNPs that explained 23% of the variance in $\dot{V}O_2$max training response. Seven of the SNPs were from the RNA predictor transcript set and four were from the original HERITAGE panel of candidates.

Timmons and colleagues have also recently reported on a “training responsive transcriptome” (TRT) and the key regulatory molecules governing this complex network of transcripts (14). Runt-related transcription factor 1 (RUNX1), SRY-box 9 (SOX9), and paired box 3 (PAX3) transcription factor abundance was overrepresented in the TRT. At least 100 of the 800 TRT transcripts were differentially regulated between low and high responders to aerobic training. Pro-angiogenic and tissue development networks were among the strongest candidate regulators of adaptation to endurance training. A panel of 3400 SNPs was genotyped to cover the genetic variance in and around the top 86 high-responder genes (as defined by the transcriptome) in the HERITAGE Family Study cohort of whites and those SNPs associated weakly with the $\dot{V}O_2$max response to exercise training.

Combined transcriptomics and genomics studies have the potential to identify new candidates for exercise-related traits in an unbiased manner. Not only can such an integrated strategy provide more powerful tools for the study of human variation, but it also has the potential to shed new light on the biology of exercise. Understanding the profile of acute exercise- or training-related changes in gene expression can provide new and exciting

candidates for genetics studies of questions central to exercise biology and perhaps even exercise as a behavior.

FROM GENOTYPE TO PHENOTYPE

A strong and replicated association with a SNP is not the end of the process. On the contrary, it is only the beginning of the scientific journey that leads to the resolution at the gene, DNA sequence, transcript, pathway, and biological system level of the signal identified in hypothesis-free and unbiased explorations of the genome.

One of the most challenging tasks of human genomics is to define the complex links between a genotype and a phenotype of interest. This is often referred to as functional genomics, but a full understanding of the direct and interactive pathways connecting given genomic characteristics to an endophenotype or a trait goes well beyond functional genomics. It will not be possible to address this topic in any detail in the present commentary, but its complexity and critical importance must be recognized if true progress in exercise genomics is to be achieved in the long term. A glimpse into the world of the genotype-phenotype complexities was provided recently in a short paper based on yeast (11). Deletion of thousands of genes in two strains revealed that the biological consequence of a deletion is conditional on other genes and that multiple mutations can lead to the same physiological state. If genotype-phenotype relationships turned out to be such an extraordinarily complex problem in an organism whose genome is only about 12 million base pairs, one can only imagine how intricate these relationships are in the human organism.

Exercise genomics will undoubtedly generate a large number of new gene targets whose further validation in human studies will be challenging. Finding a robust association between a genomic marker and an exercise-related trait is only a first step in the discovery process. Validating a new genomic target and defining the contributions of specific alleles are daunting tasks that have seldom been fulfilled in the global human genomics and genetics literature. In the exercise genomics research field, such advances have occurred for two genes thus far: the actinin, alpha 3 (*ACTN3)* (17) and the angiotensin I converting enzyme (*ACE)* (24) genes, although to date these are only partially validated as true causal genes.

Multiple technologies have to be used in order to validate a new gene target and define its biological functions. Embarking on this path requires *in silico* studies, cell-based investigations with variable levels of expression of the targeted gene, comparisons of alleles, exercise studies in informative strains of rodents, generation of transgenic, knockout, or knockdown mice for relevant exercise experiments, selective breeding for the level of expression of the targeted gene and its allelic variants, and other appropriate tools depending on the gene and pathways involved. This is an area where the experience of exercise physiologists and biochemists who have devoted decades of research to *in vitro* studies and animal experimentation constitutes a superb asset. These researchers could strongly support the efforts of exercise genomics and contribute in a significant way to the goal of clarifying the exact nature of the most robust genetic associations and understanding the mechanisms driving human heterogeneity.

FUTURE PERSPECTIVES

Future public health policies regarding physical activity as a behavior and physiological fitness as a state must recognize the importance of individual differences and the influence of DNA sequence variation. The next generation of exercise guidelines in preventive and therapeutic medicine will have to be grounded in the principles of personalized medicine.

Success will be measured to a large extent by our ability to identify the excellent, average, low, and even adverse responders for specific tissues, organs, and systems to given exercise regimens. Such advances will not occur without more and better exercise genomics and genetics research.

Historically, it all began with twin, family, and pedigree observational studies and expanded from genetic epidemiology studies to DNA sequence markers. There is evidence that DNA sequence heterogeneity contributes to human variation in exercise behavior, cardiorespiratory and muscular fitness, cardiovascular and metabolic adaptation to acute exercise, and responsiveness to regular exercise. It is now possible to undertake the molecular dissection of the genetic component of complex traits, such as those of interest to exercise biology, in terms of tissue-specific transcript expression profiles, genes, and allelic variants.

Although there are multiple issues that need attention, this short report proposes that progress needs to occur in four areas to accelerate the pace of discoveries in exercise genomics and genetics. First, investigators should give priority to the statistical power requirements before undertaking a study. Indeed, a common weakness of human exercise genomics studies is that they are based on small sample sizes. As most of the sequence variants have small effect sizes, it is particularly critical to have ample statistical power to identify them.

The second issue is that we should incentivize inter-laboratory collaborations so that it becomes possible to put together large cohorts of subjects with the proper set of traits, DNA samples, and other biological material for comprehensive, hypothesis-free genome-wide explorations. It would clearly be more productive for the small community of exercise genomics laboratories to work together in order to be able to undertake collaborative studies that would allow for large, adequately powered observational studies. This approach has proven to be highly successful for the conduct of GWAS focused on body mass index, diabetes, plasma lipids and lipoproteins, and many others (20). Working collaboratively presents many challenges, but it has to happen. For instance, study designs vary across laboratories, testing methodologies and assays will differ among the units involved, exercise programs may not be equivalent across studies, and recruitment strategies may not be comparable. The situation becomes even more complicated with intervention protocols, such as in exercise training studies. But there is no other credible solution, particularly in light of the present-day research funding restrictions.

Third, combining transcriptomics and genomics technologies has been shown to be more powerful in some recent studies. It has the potential to yield less biased and perhaps even totally unbiased panels, particularly in combination with a genome-wide exploration, of candidate genes and genomic markers. Fourth, following up on genomic marker discoveries, more resources should be devoted to functional studies in general so that the connections between genotype and endophenotypes and phenotypes become progressively established.

In conclusion, progress is being made, but more high-quality research designs and multiple replication studies with large sample sizes are urgently needed if exercise genomics is to deliver on its promises. Once these concerns have been successfully addressed, it will be of the utmost importance to incorporate epigenetic events in exercise genomics models. The addition of DNA methylation and histone modification data will undoubtedly increase the complexity of these models, but it will also enhance substantially their predictive values.

Acknowledgments

Thanks are expressed to Drs. Tuomo Rankinen and Mark Sarzynski for their critical reading of the manuscript. Gratitude is also expressed to Ms. Allison Templet for her careful editing of the manuscript. CB is partially funded by NIH (HL-45670) and the John W. Barton, Sr. Chair in Genetics and Nutrition.

Funding: CB is partially funded by NIH (HL-45670) and the John W. Barton, Sr. Chair in Genetics and Nutrition.

REFERENCES

1. Bey L, Akunuri N, Zhao P, Hoffman EP, Hamilton DG, Hamilton MT. Patterns of global gene expression in rat skeletal muscle during unloading and low-intensity ambulatory activity. Physiol Genomics. 2003; 13:157–67. [PubMed: 12582208]
2. Bouchard C, An P, Rice T, Skinner JS, Wilmore JH, Gagnon J, Pérusse L, Leon AS, Rao DC. Familial aggregation of VO(2max) response to exercise training: results from the HERITAGE Family Study. J Appl Physiol. 1999; 87(3):1003–8. [PubMed: 10484570]
3. Bouchard, C.; Hoffman, EP., editors. Genetic and Molecular Aspects of Sports Performance. International Olympic Committee/Blackwell Publishing Ltd.; West Sussex, UK: 2011.
4. Bouchard, C.; Malina, RM.; Perusse, L. Genetics of Fitness and Physical Performance. Human Kinetics; Champaign, IL: 1997.
5. Bouchard C, Rankinen T, Timmons JA. Genomics, genetics and gene-exercise interactions in health and disease. Compr Physiol. 2011 In press.
6. Bouchard C, Sarzynski MA, Rice TK, Kraus WE, Church TS, Sung YJ, Rao DC, Rankinen T. Genomic predictors of maximal oxygen uptake response to standardized exercise training programs. J Appl Physiol. 2011; 110:1160–70. [PubMed: 21183627]
7. Bray MS, Hagberg JM, Perusse L, Rankinen T, Roth SM, Wolfarth B, Bouchard C. The human gene map for performance and health-related fitness phenotypes: the 2006-2007 update. Med Sci Sports Exerc. 2009; 41:35–73. [PubMed: 19123262]
8. Bronikowski AM, Carter PA, Morgan TJ, Garland T Jr, Ung N, Pugh TD, Weindruch R, Prolla TA. Lifelong voluntary exercise in the mouse prevents age-related alterations in gene expression in the heart. Physiol Genomics. 2003; 12(2):129–38. [PubMed: 12429864]
9. Chen YW, Hubal MJ, Hoffman EP, Thompson PD, Clarkson PM. Molecular responses of human muscle to eccentric exercise. J Appl Physiol. 2003; 95:2485–94. [PubMed: 12937035]
10. De Moor MH, Liu YJ, Boomsma DI, Li J, Hamilton JJ, Hottenga JJ, Levy S, Liu XG, Pei YF, Posthuma D, Recker RR, Sullivan PF, Wang L, Willemsen G, Yan H, De Geus EJC, Deng HW. Genome-wide association study of exercise behavior in Dutch and American adults. Med Sci Sports Exerc. 2009; 41:1887–95. [PubMed: 19727025]
11. Dowell RD, Ryan O, Jansen A, Cheung D, Agarwala S, Danford T, Bernstein DA, Rolfe PA, Heisler LE, Chin B, Nislow C, Giaever G, Phillips PC, Fink GR, Gifford DK, Boone C. Genotype to phenotype: a complex problem. Science. 2010; 328(5977):469. [PubMed: 20413493]
12. Hagberg JM, Rankinen T, Loos RJF, Perusse L, Roth SM, Wolfarth B, Bouchard C. Advances in exercise, fitness, and performance genomics in 2010. Med Sci Sports Exer. 2011; 43(5):743–52.
13. Hubal MJ, Devaney JM, Hoffman EP, Zambraski EJ, Gordish-Dressman H, Kearns AK, Larkin JS, Adham K, Patel RR, Clarkson PM. CCL2 and CCR2 polymorphisms are associated with markers of exercise-induced skeletal muscle damage. J Appl Physiol. 2010; 108:1651–8. [PubMed: 20339010]
14. Keller P, Vollaard NB, Gustafsson T, Gallagher IJ, Sundberg CJ, Rankinen T, Britton SL, Bouchard C, Koch LG, Timmons JA. A transcriptional map of the impact of endurance exercise training on skeletal muscle phenotype. J Appl Physiol. 2011; 110:46–59. [PubMed: 20930125]
15. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J. Complement factor H polymorphism in age-related macular degeneration. Science. 2005; 308:385–9. [PubMed: 15761122]
16. Lander ES. Initial impact of the sequencing of the human genome. Nature. 2011; 470:187–97. [PubMed: 21307931]

17. MacArthur, DG.; North, KN. The ACTN3 gene and human performance. In: Bouchard, C.; Hoffman, EP., editors. Genetic and Molecular Aspects of Sports Performance. International Olympic Committee/Blackwell Publishing Ltd.; West Sussex, UK: 2011.

18. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. Nature. 2009; 461:747–53. [PubMed: 19812666]

19. Manolio TA. Genomewide association studies and assessment of the risk of disease. N Engl J Med. 2010; 363:166–76. [PubMed: 20647212]

20. National Human Genome Research Institute. [Accessed March 3, 2011] A catalog of published genome-wide association studies. http://www.genome.gov/gwastudies/

21. Rankinen T, Perusse L, Rauramaa R, Rivera MA, Wolfarth B, Bouchard C. The human gene map for performance and health-related fitness phenotypes. Med Sci Sports Exerc. 2001; 33:855–67. [PubMed: 11404647]

22. Rankinen T, Roth SM, Bray MS, Loos R, Perusse L, Wolfarth B, Hagberg JM, Bouchard C. Advances in exercise, fitness, and performance genomics. Med Sci Sports Exerc. 2010; 42:835–46. [PubMed: 20400881]

23. Roth SM, Ferrell RE, Peters DG, Metter EJ, Hurley BF, Rogers MA. Influence of age, sex, and strength training on human muscle gene expression determined by microarray. Physiol Genomics. 2002; 10:181–90. [PubMed: 12209020]

24. Skipworth, JRA.; Puthucheary, ZA.; Rawal, J.; Montgomery, HE. The ACE gene and performance. In: Bouchard, C.; Hoffman, EP., editors. Genetic and Molecular Aspects of Sports Performance. International Olympic Committee/Blackwell Publishing Ltd.; West Sussex, UK: 2011.

25. Teran-Garcia M, Rankinen T, Koza RA, Rao DC, Bouchard C. Endurance training-induced changes in insulin sensitivity and gene expression. Am J Physiol Endocrinol Metab. 2005; 288:E1168–78. [PubMed: 15687108]

26. Teran-Garcia M, Rankinen T, Rice T, Leon AS, Rao DC, Skinner JS, Bouchard C. Variations in the four and a half LIM domains 1 gene (FHL1) are associated with fasting insulin and insulin sensitivity responses to regular exercise. Diabetologia. 2007; 50:1858–66. [PubMed: 17589823]

27. Timmons JA, Knudsen S, Rankinen T, Koch LG, Sarzynski M, Jensen T, Keller P, Scheele C, Vollaard NB, Nielsen S, Akerstrom T, MacDougald OA, Jansson E, Greenhaff PL, Tarnopolsky MA, van Loon LJ, Pedersen BK, Sundberg CJ, Wahlestedt C, Britton SL, Bouchard C. Using molecular classification to predict gains in maximal aerobic capacity following endurance exercise training in humans. J Appl Physiol. 2010; 108:1487–96. [PubMed: 20133430]

Summary for Table of Contents Page

High-quality research designs and replication studies with large sample sizes are needed if exercise genomics is to deliver on its promises.

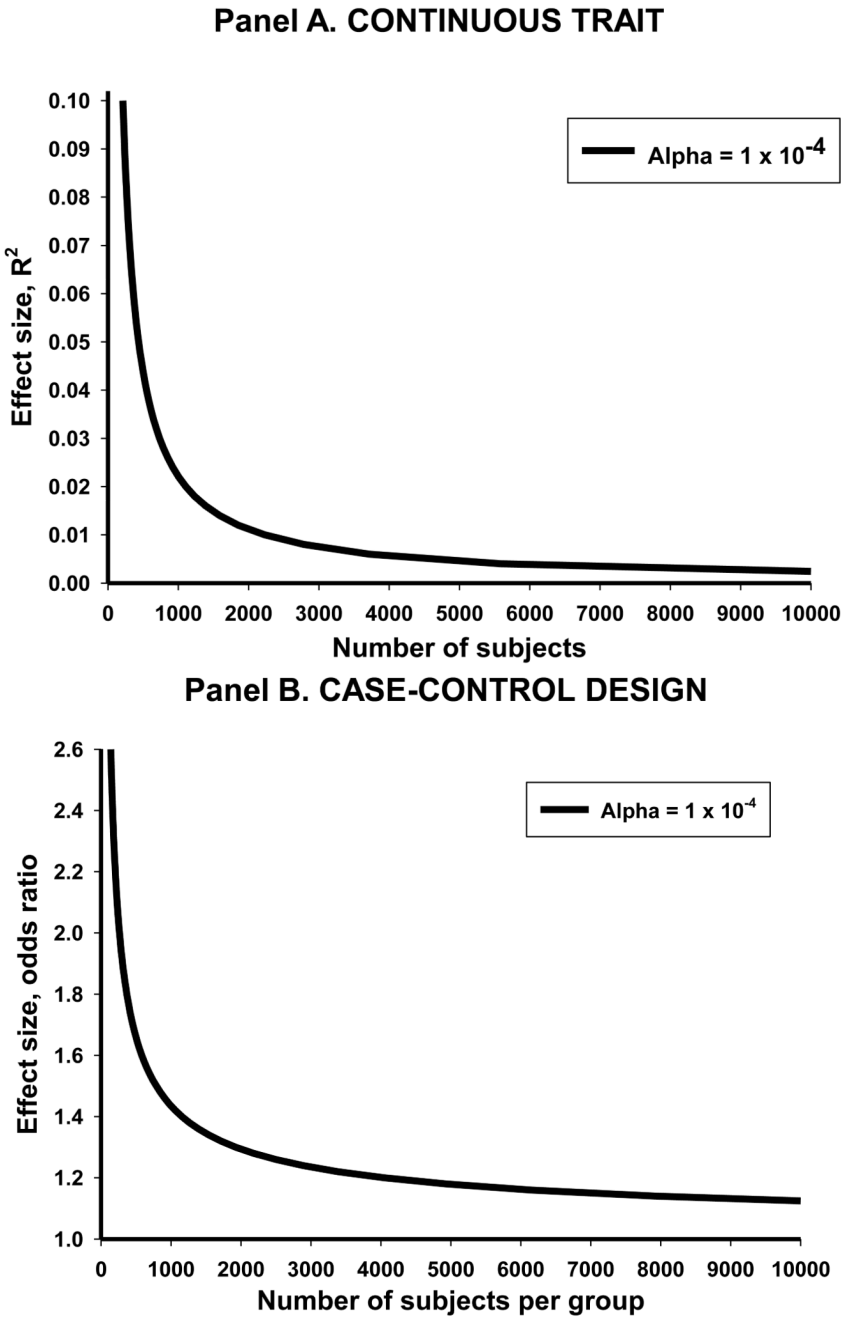

Figure 1.

Panel A. Number of subjects needed for a given effect size for a continuous trait assuming 80% statistical power, an additive model, and an alpha level of 0.0001 (500 single nucleotide polymorphisms (SNPs)).

Panel B. Number of subjects needed in each group for a given effect size (measured as an odds ratio [OR]) for a case-control design assuming 80% statistical power, a minor allele frequency of 20%, an additive model, and an alpha level of 0.0001 (500 SNPs). [Adapted from (12). Copyright © 2011 Wolters Kluwer Health/Lippincott Williams & Wilkins. Used with permission.]

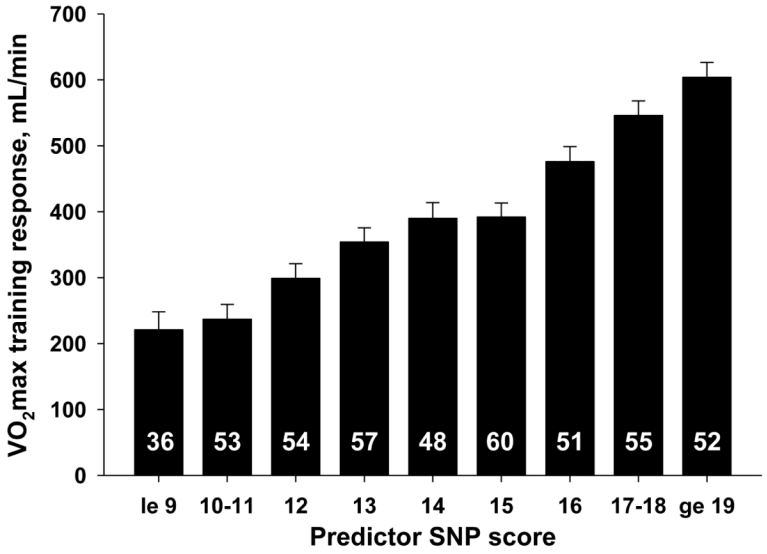

Figure 2.
Age, sex, and baseline V̇O2max-adjusted V̇O2max training responses across nine genome-wide association studies (GWAS) predictor single nucleotide polymorphisms (SNP) score categories in HERITAGE whites. The number of subjects within each SNP score category is indicated inside each histogram bar. Le stands for “less or equal to” and ge “greater or equal to.” [Adapted from (6). Copyright © 2011 The American Physiological Society. Used with permission.]