Bray, Molly S. Genomics, genes, and environmental interaction: the role of exercise. J. Appl. Physiol. 88: 788–792, 2000.—Regular exercise has been shown to improve control of lipid abnormalities, diabetes mellitus, hypertension, and obesity, with the greatest benefits realized by sedentary individuals who begin to exercise. Responses to exercise interventions are often highly variable among individuals, however, and research has indicated that response to exercise may be mediated in large part by variation in genes. As we strive to unravel the complex etiology of diseases like obesity, diabetes, and cardiovascular disease through the use of molecular and genetic tools now available, understanding the interaction and influence of environmental factors, such as exercise, on gene expression and function has taken on increasing importance. This review briefly summarizes strategies presently being used to elucidate genes and genetic effects that may be mediated or influenced by exercise and serves to illustrate the importance of considering the effect of exercise when investigating genes related to health or other physiological outcomes.

Research focused on exercise at the genetic or genomic level has typically involved investigations of genes that affect quantitative measures known to be directly influenced by exercise (e.g., muscle mass, bone density, and so forth) or investigations of disease outcomes that are influenced by both genetic effects and exercise (e.g., hypertension in exercising and nonexercising individuals). A limited number of researchers have investigated the genetic basis of exercise or activity level as a phenotype itself. Gottschaldt (13) studied the concordance, or similarity, of various mental and physical activity traits in monozygotic twins over the span of 30 yr and found that, although cognitive skills remained highly concordant throughout a lifetime, the physical activity level did not, suggesting that environmental factors rather than genetic factors are more important in determining an individual’s propensity to be active in adulthood. Although the genetic basis of activity level has not been as well studied as personality traits and thinking skills, other researchers have reported similar findings (38). Family studies have provided heritability estimates for physical activity, measured by self-report or by observation, ranging from 0.29 to 0.62, with the wide range in estimates likely due to differences in the age and type of the activity.
To understand how genes and exercise can interact to modify a phenotypic trait or health outcome, it is necessary to consider multiple levels of interaction. Figure 1 illustrates the many complex ways in which genes and exercise, both together and separately, can influence the health status of an individual. In this model, exercise can produce direct and immediate effects on health status, without necessarily altering gene expression or function. Such is the case in exercise-induced asthma, in which the increased air flow and temperature resulting from rapid breathing during exercise can very quickly alter the short-term health status in genetically susceptible individuals (27). Another example of an acute exercise effect is that of exercise-related sudden cardiac death occurring in individuals with genetic defects leading to hypertrophic cardiomyopathy or coronary artery anomalies (8). These examples are illustrative of gene-environment interactions in which the same environment (i.e., exercise) produces differential effects in genetically different individuals.

Exercise can also affect health status indirectly by altering the expression or action of one or more genes that influence intermediate phenotypes (e.g., cholesterol level) that ultimately produce disease outcomes. This is an example of biological interaction, in which two or more factors influence a phenotype interdependently (41). Biological interaction, in which multiple genetic and environmental factors are interconnected in complicated ways, is inherent in complex chronic diseases, such as cardiovascular disease. Because of the complex nature of many diseases and intermediate phenotypes, and of health status in general, the model in Fig. 1 necessarily includes other genes and other environments, in addition to exercise, that may or may not interact with one another in determining overall health.

Several strategies have been used to detect genes whose effects may be altered by exercise. Many studies have investigated the effect of exercise on intermediate traits related to chronic diseases such as hypertension, obesity, and cardiovascular disease. Although the positive effects of exercise on reducing hyperlipidemia, high blood pressure, obesity, and related factors are well documented (4, 14, 17, 21), there is typically a great amount of heterogeneity in the responsiveness to exercise intervention. A limited number of studies have investigated the role of genetic variation in the control of traits such as blood pressure and lipid processing within the context of exercise, and several gene-exercise interactions have been identified. In a recent study (16), sedentary men underwent 9 mo of endurance training and were genotyped for the apolipoprotein E (apoE) gene. The apoE gene has three common alleles or variant forms (ε2, ε3, and ε4), which result from combinations of single nucleotide changes at amino acids 112 and 158. Although all subjects experienced increases in total high-density lipoprotein cholesterol subsequent to exercise training, subjects with an ε2 allele had significantly greater increases in high-density lipoprotein cholesterol than those with either an ε3 or ε4 allele (16). Conversely, another study reported that hypertensive individuals with an ε2 allele exhibited a lesser response in both systolic and diastolic blood pressure after exercise training compared with hypertensive individuals with either an ε3 or ε4 allele. The results of these studies suggest that the same apoE genotype can interact both positively and negatively with exercise to differentially influence physiological responses to exercise training (15).

The Gln27Glu polymorphism within the β2-adrenoceptor gene has been associated with obesity and body size measures in several studies, with somewhat conflicting results among the studies (6, 20, 33). Recently, Meyerhaeghe et al. (30) reported that this association was evident only in subjects who did not exercise, suggesting that exercise may attenuate the deleterious effects of the Glu27 genotype. In another study, obese women were genotyped for the Arg16Gly variant in the β2-adrenoceptor gene and randomly placed on a weight-loss program that included both diet and exercise (42). After 3 mo of the diet and exercise regimen, subjects with Arg/Gly or Gly/Gly genotype showed significantly greater weight loss than did Arg/Arg homozygotes. These studies serve as examples of the progress being made in elucidating gene-exercise interactions that modify disease outcomes. The identification of genetic effects that are modified by exercise and the application of this genetic information to exercise prescription may be one way to improve the efficacy of exercise as a preventive measure for chronic disease.

Another approach to identify genes that are likely to interact with exercise is through genome-wide searches for genes that influence exercise-related phenotypes, such as muscle mass, maximum O2 uptake (V̇O2max), energy expenditure, and so forth. To analyze an entire genome, family or random samples are genotyped for DNA marker sequences with known locations along the chromosomes, and a disease or trait gene is “mapped” to a genomic region through the use of statistical linkage or association analyses. DNA markers commonly con-
exist of highly polymorphic repeated sequences of two, three, four, or more nucleotides (microsatellite DNA markers) or single nucleotide changes in DNA sequence (single nucleotide polymorphism markers) distributed randomly throughout the genome. Mapping strategies are based on the assumption that both the marker sequence and the DNA sequence immediately surrounding it are transmitted intact from one generation to the next. Therefore, genotype at a marker locus incorporates genotype at the surrounding sequence, which may contain a functional variant.

Genetic linkage analyses typically employ the use of genotypes at microsatellite DNA markers to infer allele sharing or allele transmission in samples of related individuals. Significant evidence for linkage to a trait or disease is then determined based on the type of family sample being tested (e.g., consistent transmission of a disease or trait-influencing allele from parent to offspring, proportion of alleles shared between affected vs. unaffected siblings, or the relationship between allele sharing and means or differences in quantitative trait values for pairs of relatives). Alternatively, linkage disequilibrium mapping and association analyses can be used to localize genes in samples of unrelated individuals who are genotyped for single nucleotide polymorphisms that span the genome at smaller intervals than do microsatellite markers. Linkage disequilibrium is a measure of the probability that two variants in DNA sequence will be transmitted together from one generation to the next. In either type of genetic mapping analysis, significant evidence for linkage or association is an indicator of a genomic region that may contain trait- or disease-influencing genes.

Genome or chromosomal scans have been conducted for metabolic rate (35), bone density (34), body fat measures (36), insulin (19), blood pressure (24, 25, 45), VO$_{2\text{max}}$ (9), and other metabolic intermediates. However, a functional gene/variant has yet to be elucidated within candidate gene regions identified by these studies. One explanation for the lack of findings may be that detection of the effects of any one polymorphism on a complex trait, such as metabolic rate, for example, may be dependent on the context in which the variant is being analyzed (e.g., exercisers vs. nonexercisers). That more systematic attention has not been focused on the study of gene-environment interaction probably reflects the difficulties involved in conducting such studies in humans: large samples are needed, multiple polymorphisms must be typed in all subjects, demographic and environmental variables must be measured, and complex analytic methods are required. The primary limiting factor in genetic analyses stratified by activity level or other environmental factors is sample size, as even moderately large samples (e.g., 1,000 individuals) may have strata that contain only a handful of subjects. Nevertheless, studies of gene-environment interaction are critical if we are to make the most efficacious use of genetic information to improve overall health.

An alternative strategy to the use of genome-wide scans to identify trait- or disease genes that may interact with exercise is through candidate gene studies, which focus on genes involved in the metabolic pathways and physiological systems known to influence the exercise-related trait of interest. With the use of this strategy, association studies of single or multiple gene variants have identified a limited number of genes that appear to influence exercise-related phenotypes. For example, the renin-angiotensin system is a key regulator of plasma volume and blood pressure and may play a role in cardiac and skeletal muscle growth. The angiotensin I-converting enzyme (ACE) gene, a key component of the renin-angiotensin system, has been one of the most well studied of these candidate gene approaches.

Montgomery et al. (32) analyzed genotype at the insertion/deletion locus within the ACE gene for association to exercise response and reported that individuals with one or two insertion alleles showed significantly greater duration in repetitive arm flexion time after exercise training than did deletion/deletion homozygotes. The frequency of the insertion allele has also been shown to be elevated among elite Australian rowers and British mountaineers compared with the population at large (10, 32). Other studies have indicated that the ACE deletion allele is associated with exercise-induced left ventricular hypertrophy in young and middle-aged men (31, 43). Bouchard and colleagues (39, 40) investigated several polymorphisms in the genes for enzymes that regulate energy metabolism and cellular ion exchange and reported a significant association between a variant in the muscle-specific creatine kinase gene and VO$_{2\text{max}}$ and significant associations between polymorphisms present in the Na-K-ATPase $\alpha_2$ and $\beta_1$ genes and resting heart rate, resting and exercise blood pressure, and rate pressure product. These examples illustrate that selection of candidate genes based on physiological pathways has proved to be a useful strategy to detect genes that are modified by exercise.

In addition to statistical analyses of gene-exercise interactions, laboratory studies of cellular function have also revealed genes whose expression is directly altered by exercise. Genes encoding insulin signaling intermediates, intermediates of energy metabolism, procollagens and collagen-processing enzymes, enzymes involved in glucose and lipid metabolism, and others have demonstrated alterations in expression levels subsequent to exercise (5, 18, 23, 26). As new technologies presently being developed for differential gene expression research, such as chip-based microarrays, become more widely used, an increasing number of genes whose expression is modified by exercise will be identified. Nevertheless, the mechanisms by which these alterations in gene expression subsequent to exercise occur have not been well studied. In an elegant series of experiments, Goldspink and colleagues (11, 29, 47) have demonstrated that stretch is an important stimulator of gene expression in muscle tissue and that a newly identified variant of insulin growth factor 1, mechano growth factor, is responsible for local tissue repair, maintenance, and remodeling in stretched
muscles. These researchers have shown that the type and duration of mechanical movement experienced by exercising muscles regulate the expression of various isoforms of the myosin heavy chain gene that ultimately determine muscle-fiber phenotype and are one way in which muscle tissue can adapt to a specific type of physical activity (12, 46). Exercise may also influence gene expression through changes that occur subsequent to exercise that moderate nuclear protein binding or relocation of transcription factors to the nucleolus, such as fluctuations in circulating catecholamine levels, the release of neurotransmitter signals, or the conversion of energy substrate utilization (7, 44). Elucidating the means by which exercise can influence gene expression is an important area for further study.

In this brief review, several strategies for detecting gene-exercise interactions have been summarized. These include candidate gene studies of intermediate disease-related traits, genome scans for exercise-related phenotypes, selection of candidate genes based on exercise physiology, and identification of genes through differential expression. In many instances, these strategies have proven successful, but they have provided limited knowledge of the specific mechanisms underlying the differential effects of genes, either subsequent to exercise or in physically active vs. inactive individuals. The factors involved in human physiology and health are extremely complex. Traits such as energy metabolism, VO₂max, blood pressure, blood lipid levels, body fat, and many others, as well as disease states such as hypertension, coronary heart disease, and stroke, are determined by complicated and life-long factors. This fact necessitates the inclusion of multiple genetic and environmental factors. This fact necessitates the inclusion of multiple genetic and environmental factors. This fact necessitates the inclusion of multiple genetic and environmental factors. This fact necessitates the inclusion of multiple genetic and environmental factors. This fact necessitates the inclusion of multiple genetic and environmental factors. This fact necessitates the inclusion of multiple genetic and environmental factors. This fact necessitates the inclusion of multiple genetic and environmental factors. This fact necessitates the inclusion of multiple genetic and environmental factors. This fact necessitates the inclusion of multiple genetic and environmental factors. This fact necessitates the inclusion of multiple genetic and environmental factors. This fact necessitates the inclusion of multiple genetic and environmental factors. This fact necessitates the inclusion of multiple genetic and environmental factors. This fact necessitates the inclusion of multiple genetic and environmental factors. This fact necessitates the inclusion of multiple genetic and environmental factors. This fact necessitates the inclusion of multiple genetic and environmental factors.