

# Gene–Environment Interactions and the Etiology of Common Complex Disease

Richard S. Cooper, MD

Genetic epidemiology has greatly expanded its scope as a result of major technological innovations in the past decade. Laboratory capacity to determine DNA variation and archival information on the human genome sequence are now readily available. A wide range of research projects have been launched on chronic disease and health problems of aging, on the assumption that a better understanding of mechanisms will improve treatment and prevention. In many instances, the actions of genes are known to be modified by environmental conditions, and considerable emphasis has now been placed on finding specific interactions between genes and the environment. Studies in agriculture and animals provide clear empirical evidence on the importance of this con-

cept. Describing gene–environment interactions in studies of humans is still very challenging, however, given the difficulties in study design and measurement. Despite the theoretical value of characterizing both intrinsic and extrinsic components of the causal process in the development of disease, the argument can also be made that main effects of each component separately are much more important. For these reasons, gene–environment interactions are likely to remain a conceptual framework for health research rather than a practical goal for the foreseeable future.

*Ann Intern Med.* 2003;139:437–440.

[www.annals.org](http://www.annals.org)

For author affiliation, see end of text.

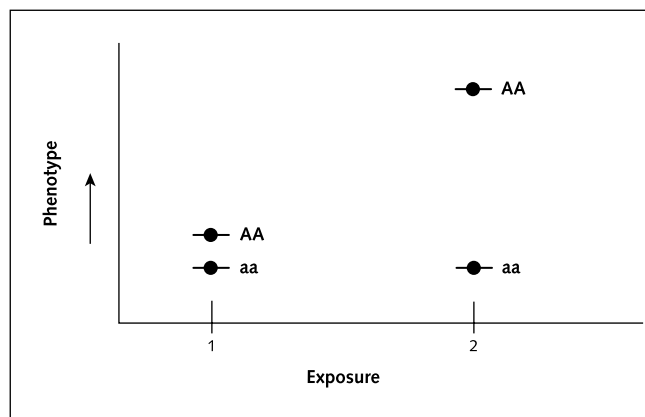
The technology that makes it possible to define DNA variants associated with human disease has opened a new universe of research opportunities. All organisms represent the expression of an encrypted genetic message in a particular environment. If we discount the minimal contribution of random processes that occur during development of an organism, the DNA sequence and the local environmental conditions theoretically include all the information necessary to explain life forms. These abstract generalities stand at a great distance from practical biology, however, given the complexity of the intervening steps. Nonetheless, the ability to genotype efficiently on a large scale has generated intense interest in conducting epidemiologic studies that simultaneously measure genetic and environmental effects and account for their interactions. If this approach is successful, it is hoped that the scope of research opportunities on basic public health questions such as healthy aging will be substantially increased.

Perhaps because of human mental predisposition, or simply lack of experience in practical applications, the study of interactions is plagued by confusion on many levels. The formal study of gene–environment interactions has its roots in the quantitative genetics of agriculture, in which control over breeding and growing conditions makes it possible to conduct the necessary experiments (1). For example, the relative yield of a strain of wheat under conditions where rainfall is limited has obvious importance. The attempt to apply the same research questions to studies of human health and disease faces serious difficulties, however. Neither breeding patterns nor environmental conditions can be controlled for most conditions that affect chronic disease. Epidemiologists have instead focused on identifying so-called main effects—the risk associated with a particular environmental exposure or genotype considered on its own. In many cases, simply demonstrating an elevated risk is a sufficiently useful outcome, since the question is whether to, in the case of an environmental

factor, limit exposure or alter risk prediction on the basis of a genetic test. As a rule, this success has come by studying sufficiently large groups so that measurement error does not undermine our ability to discern main effects. Interactions require that the analysis be stratified by subgroups, and the sample size, on average, needs to be increased fourfold to achieve the same statistical power available for studying main effects. Studies focused on how genes and environment interact to influence the development of a phenotype can also impose more difficult measurement requirements. If the goal is to assess the impact of lifestyle factors, then it becomes necessary to characterize these effects over the long term among individuals. For many reasons, therefore, few examples exist of successful studies of gene–environment interaction effects for health-related traits.

What in fact are gene–environment interactions? In common usage, gene–environment interactions imply a change in the direction or magnitude of the effect of a genetic variant when the environment changes. (Of course, biological complexity also includes the possibility of gene–gene and environment–environment interactions, but those are not addressed here.) Biologists use the concept of the *norm of reaction* to represent the genome as a curved reflecting surface, mediating the effect of the environment on the phenotype (2). Alternatively, as summarized in a basic textbook, the norm of reaction is “the function relating mean phenotypic response of a genotype to a change in the environment” (1). The effect of the environment cannot therefore be predicted as a simple linear function. For an evolutionary biologist, for example, reaction norms are of interest when it is important to know whether selection acts directly to expand plasticity and make organisms capable of nonstereotyped responses or does so in the face of challenges from altered environments. Concerns in epidemiology are generally more limited. Perhaps a more effective term to convey the notion of gene–environment interaction effects in studies of human health is *context dependency*,

Figure 1. Gene–environment interaction.



The design of this experiment asks whether organisms with different genotypes (that is, AA vs aa) undergo a shift in the relative relationship between their expressed phenotypes when the environment changes.

which describes the conditional nature of the interrelationship between genes and environments (3). In this research, we are interested in the effect of a causal factor “conditional on”—or in the context of—another factor. Thus, development in a setting of deficient nutrition might alter the expression of a specific set of genes. As suggested earlier, the challenge is to begin to give the study of gene–environment interaction effects an empirical basis. It is hoped that the accumulation of evidence will clarify how widespread and important these effects are and how they might improve our understanding of disease mechanisms. Collecting this empirical evidence requires that we conduct appropriately designed studies and use rigorous statistical tests to interpret the results.

The importance of combining into a single analytic process the joint effect of genes and environment thus has a solid theoretical foundation. At present, however, only very limited tests can be conducted for most conditions related to the degenerative disorders of aging. *Genotype* is defined in practice as the presence of A, T, C, or G at a polymorphic site thought to be of functional significance or at a site physically adjacent to such a mutation. Except for the rare mendelian disorders, it is also unlikely that these single nucleotide polymorphisms code for large phenotypic effects, so it must be recognized that genotype as measured by this approach is a very weak approximation of the “norm of reaction.” The genetic component of all complex traits, such as hypertension, arthritis, and cognitive function, will instead be influenced by a broad range of effects spread across many single nucleotide polymorphisms in many genes. As the sophistication of genetic epidemiology continues to develop, there is the expectation that the information that can be quantified under the rubric of “genotype” will markedly improve, at least making it possible to account for multiple variants at the same time.

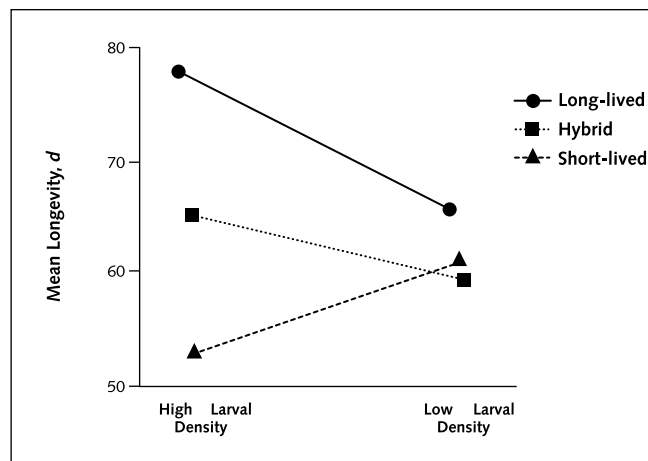
The description of the formal design for a study of gene–environment interactions can sound more like a riddle than a medical research study, given the layers of com-

plexity. The question is whether different genotypes have a different relationship to the phenotype in different environments. The basic design thus requires that the contrasting genotypes (for example, AA vs. TT at a given DNA site) be studied under at least two sets of conditions. The analysis then compares the difference in the genotype–phenotype relationships in the different environments for persons with contrasting genotypes (Figure 1). Research on longevity in strains of *Drosophila melanogaster* has produced one of the classic examples of a gene–environment interaction effect (Figure 2) (4). In these experiments, crowding during the developmental period markedly affected the longevity of a strain that had been bred for longevity. In contrast, for a short-lived strain, larval density had little effect. The hybrid form responded to the change of conditions in a manner similar to the long-lived strain.

To provide the basis for hypothesis testing for experiments such as these, it is necessary to define the expected null genotype–phenotype relationship across environments and test whether the observed data deviate from that pattern. For statistical purposes, a multiplicative interaction is usually assumed, as in Figure 2. Deviation from that pattern is interpreted as evidence for a gene–environment interaction. When the elements of this test are considered, it becomes apparent why the research questions that can be asked are relatively limited; clearly the underlying biological processes could potentially be much more complicated than those described in Figures 1 and 2.

Within this design, what are the challenges and opportunities for medical research? The first requirement is the discovery of an informative variant in a gene of interest. Polymorphisms at the apolipoprotein E locus, for example, have been shown to modestly influence serum lipids and cognitive function in several populations, and the phenotype is jointly modified by environmental factors (5, 6). The second requirement is then to stratify the sample by

Figure 2. Response of longevity to density treatment for strains of *Drosophila melanogaster* selected for short and long life spans and for F<sub>1</sub> hybrids.



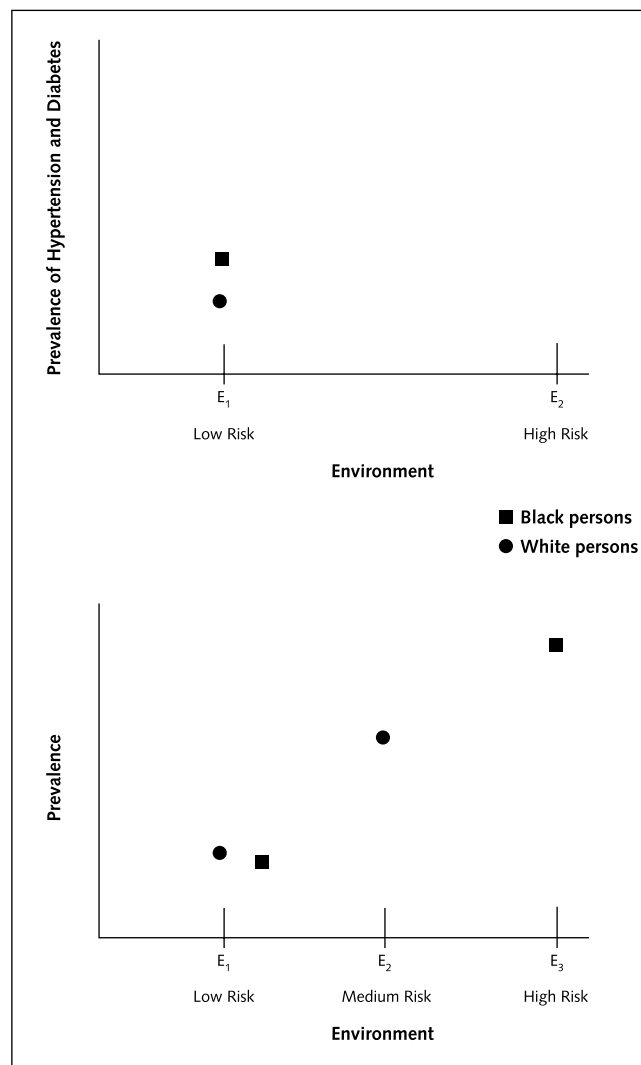
environmental factors. Standard epidemiologic survey methods are used for this purpose to assess smoking, obesity, and diet. As suggested earlier, however, the precision of these methods in characterizing lifelong exposures can be very limited, as in the case of dietary patterns. Intervention trials are another alternative, although by definition the change in environment will be transient and recent, and thereby testing will involve only a subset of genetically moderated responses (7). Studies of related populations living in different social settings can also provide the necessary contrasts (8). Since many exposures associated with chronic degenerative illnesses tend to be correlated, it will often be difficult to distinguish the relevant factors from confounders with this design. In all these instances, it is obviously important to apply standard, objective definitions in the classification of case-patients and non–case-patients.

The magnitude of the phenotypic response in contrasting genotypic classes (Figure 1) is not the only way in which context dependency can be manifested. It is also of interest to know the contribution of a genotype to the overall variation in a trait (1). It might be assumed, for example, that where the environmental exposures are weak, a genetic effect could be relatively more important. This process mirrors the relationship between increasing age and the declining proportion of cases of dementia that are familial or genetic. Genetic influences are relatively more important in determining premature dementia; cases in which genetics has a weak role and environment is more important may not be manifested until much later.

Heritability is often used as a nonspecific measure of the familial aggregation of a trait. Since first-degree relatives on average share 50% of their genetic material, the correlations among phenotypes should represent half of the genetic effect, after discounting environmental factors shared from living together. In accord with this proposition, the heritability of blood pressure and related traits appears to be quite a bit higher in low-risk social environments (9), which could in part reflect the greater impact of individual-level risk factors in westernized cultural settings. Obesity, on the other hand, has been shown to have more similar degrees of heritability across environments (10). Analyses using heritability alone cannot be unambiguously interpreted, however, since a variety of factors could alter the result. Direct measurement of DNA variants is needed to give precision and meaning to the genetic effects.

An important related application of the general concept of gene–environment interactions in public health involves the explanation offered for racial and ethnic variation in common disease (11). Studies of migrants, populations undergoing transition, or related groups who live in societies at different levels of industrial development demonstrate complex differential patterns of disease prevalence. In the United States, for example, African-American persons have more vascular disease and dementia than white persons, but both conditions are relatively infrequent in West Africa (12–14). As an extension of the debate over

Figure 3. Explanations of racial and ethnic differentials.



Top. Gene–environment interactions. Bottom. Alternative environmental explanation.

whether the racial differentials in the United States represent primarily genes or environment, the explanation of low disease rates in African persons versus African-American persons is often framed as an example of gene–environment interactions (Figure 3, top). To sustain the argument that the African population somehow has more genetic variants that lead to hypertension, the change in environment must be invoked as the cause of the large contrasts in hypertension risk in the social contexts in which persons of African heritage live today. Thus, West African persons are seen as harboring a latent susceptibility to the diseases that are fully expressed in the U.S. environment. This broad hypothesis cannot currently be subjected to any practical test, given our ignorance of the relevant genes and mutations, and serves primarily to structure the discussion. It must be remembered, however, that an equally if not more plausible framework might be offered whereby the genetic component of susceptibility across the large population groups we call races is equal, yet the

groups themselves occupy different environmental niches (Figure 3, *bottom*). The result could then be explained as a result of additive main effects. In fact, a bias exists toward giving excessive weight to loosely formulated genetic hypotheses on racial and ethnic differentials, and the use of the construct of gene–environment interactions is no exception (15, 16). In our current state of ignorance, when only environmental effects have been documented as causes of large population differences in prevalence, explanations restricted to these known factors are much more secure.

So why do we want to describe gene–environment interaction effects? Enormous public resources are being invested in genomics at the present time, and we must ask, What use can we make of the new knowledge that genomic research might produce? As suggested earlier, because we understand living organisms to be what genes use the environment to make, a long-term goal of biology must include a description of the mechanics of that process. For public health and medicine, however, these fundamental questions have little immediate relevance. Given the early stage of genetic research on conditions such as aging, can we predict the value this research program will ultimately have? Widely varying opinions have been voiced about the potential impact of genomics on health and medicine, and that debate cannot be summarized here. One consequence of the advent of genomics has in fact been to blur the distinction between basic research in human biology and the search for tools to improve health. We have accumulated a historical record that confirms that basic research and an improved understanding of physiologic mechanisms have value that cannot be judged by their application in the predictable future. However, the principles of prevention science that have led to such enormous gains in the healthy life span still apply in the genomic era: Organizing social life to avoid noxious environmental effects will ensure healthy aging for the greatest number of people. Gene–environment interactions, once identified, are likely to be useful primarily as aids to improving risk prediction for subgroups. Thus, persons with a given genotype might be predicted to have an exaggerated response to a common exposure. While identification of “high-risk” genotypes has already made a useful contribution to the practice of medicine in some areas, the benefits are likely to be limited to fairly small sets of high-risk families. Most of the interest will remain focused on main effects, that is, those that are simply the result of adding an environmental exposure to a particular genotype (17). In the broader public health context, common diseases result from common exposures to which we are all susceptible, albeit in varying degrees (18). For the time being at least, the study of gene–environment interaction effects will remain a tool to aid genetic epidemiologists in their efforts to more precisely define the role of specific genetic variants and the relative impact of genes and environment on the distribution of phenotypes.

From Loyola Stritch School of Medicine, Maywood, Illinois.

**Grant Support:** By grants from the National Heart, Lung, and Blood Institute (HL 45508 and HL 47910) and the Donald W. Reynolds Cardiovascular Clinical Research Center at the Dallas/University of Texas Southwestern Medical Center.

**Potential Financial Conflicts of Interest:** None disclosed.

**Requests for Single Reprints:** Richard S. Cooper, MD, Department of Preventive Medicine and Epidemiology, Loyola Stritch School of Medicine, 2160 South First Avenue, Maywood, IL 60153.

## References

- Lynch M, Walsh B. Gene  $\times$  Environment Interactions. Genetics and Analysis of Quantitative Traits. Sunderland, MA: Sinauer Associates; 1998:567-685.
- Lewontin R. Human Diversity. New York: Scientific American; 1997.
- Kardia SL. Context-dependent genetic effects in hypertension. *Curr Hypertens Rep.* 2000;2:32-8. [PMID: 10981124]
- Clare MJ, Luckinbill LS. The effects of gene-environment interaction on the expression of longevity. *Heredity.* 1985;55(Pt 1):19-26. [PMID: 3930428]
- Wilson PW, Schaefer EJ, Larson MG, Ordovas JM. Apolipoprotein E alleles and risk of coronary disease. A meta-analysis. *Arterioscler Thromb Vasc Biol.* 1996;16:1250-5. [PMID: 8857921]
- Evans RM, Emsley CL, Gao S, Sahota A, Hall KS, Farlow MR, et al. Serum cholesterol, APOE genotype, and the risk of Alzheimer's disease: a population-based study of African Americans. *Neurology.* 2000;54:240-2. [PMID: 10636159]
- Turner ST, Schwartz GL, Chapman AB, Boerwinkle E. C825T polymorphism of the G protein beta(3)-subunit and antihypertensive response to a thiazide diuretic. *Hypertension.* 2001;37:739-43. [PMID: 11230366]
- Rotimi CN, Cooper RS, Marcovina SM, McGee D, Owoaje E, Ladipo M. Serum distribution of lipoprotein(a) in African Americans and Nigerians: potential evidence for a genotype-environmental effect. *Genet Epidemiol.* 1997;14:157-68. [PMID: 9129961]
- Cooper RS, Guo X, Rotimi CN, Luke A, Ward R, Adeyemo A, et al. Heritability of angiotensin-converting enzyme and angiotensinogen: a comparison of US blacks and Nigerians. *Hypertension.* 2000;35:1141-7. [PMID: 10818078]
- Luke A, Guo X, Adeyemo AA, Wilks R, Forrester T, Lowe W Jr, et al. Heritability of obesity-related traits among Nigerians, Jamaicans and US black people. *Int J Obes Relat Metab Disord.* 2001;25:1034-41. [PMID: 11443503]
- Cooper RS. The role of genetics in ethnic disparities in health. In: *Disparities in Aging Health.* Washington, DC: National Academy of Sciences/National Research Council Pr; [In press].
- Cooper R, Rotimi C, Ataman S, McGee D, Osotimehin B, Kadiri S, et al. The prevalence of hypertension in seven populations of west African origin. *Am J Public Health.* 1997;87:160-8. [PMID: 9103091]
- Hendrie HC, Ogunniyi A, Hall KS, Baiyewu O, Unverzagt FW, Gureje O, et al. Incidence of dementia and Alzheimer disease in 2 communities: Yoruba residing in Ibadan, Nigeria, and African Americans residing in Indianapolis, Indiana. *JAMA.* 2001;285:739-47. [PMID: 11176911]
- Cooper RS, Rotimi CN, Kaufman JS, Owoaje EE, Fraser H, Forrester T, et al. Prevalence of NIDDM among populations of the African diaspora. *Diabetes Care.* 1997;20:343-8. [PMID: 9051385]
- Cooper RS, Kaufman JS, Ward R. Race and genomics. *N Engl J Med.* 2003;348:1166-70. [PMID: 12646675]
- Cooper RS, Kaufman JS. Race and hypertension: science and nescience. *Hypertension.* 1998;32:813-6. [PMID: 9822436]
- Clayton D, McKeigue PM. Epidemiological methods for studying genes and environmental factors in complex diseases. *Lancet.* 2001;358:1356-60. [PMID: 11684236]
- Vasan RS, Beiser A, Seshadri S, Larson MG, Kannel WB, D'Agostino RB, et al. Residual lifetime risk for developing hypertension in middle-aged women and men: The Framingham Heart Study. *JAMA.* 2002;287:1003-10. [PMID: 11866648]