

Genes, Environment, and Cardiovascular Disease

Charles F. Sing, Jari H. Stengård, Sharon L.R. Kardia

Abstract—In this essay, we call to attention what every medical researcher knows about the etiology of cardiovascular disease but most deny, or choose to ignore, when designing, carrying out, and reporting genetic studies. Medical research is entering an era of synthesis that will take advantage of the successes of reductionism over the past decade in defining and describing human genome variations. Meaningful insights into the role of such variation requires a biological model of genome-phenotype relationships that incorporates interactions between subsets of possible genetic and environmental agents as causations in particular contexts indexed by time and space. We make recommendations for what needs to be done to cope with these complexities. (*Arterioscler Thromb Vasc Biol.* 2003;23:1190-1196.)

The triumph of the reductionism of the Greeks is a pyrrhic victory: We have succeeded in reducing all of ordinary physical behavior to a simple, correct Theory of Everything only to discover that it has revealed exactly nothing about many things of great importance.

R.B. Laughlin and D. Pines

The Problem

Common chronic multifactorial diseases are responsible for the greatest demand on medical services.¹⁻³ They also make the largest contribution to loss of human life and productivity in westernized societies (eg, see Murray and Lopez⁴ and the American Heart Association⁵⁻⁷). Deviations from health attributable to these diseases, which include cardiovascular disease (CVD), cancer, diabetes, and the psychiatric disorders, typically aggregate in families but they do not segregate as Mendelian single-gene disorders. The distribution of disease among individuals, families, and populations is a direct consequence of the distribution of interactions between the effects of many susceptibility genes and many environmental exposures, that, through dynamic, epigenetic, regulatory mechanisms, ultimately become integrated to generate the disease phenotype.⁸⁻¹⁸ The genetic analysis of a multifactorial disease presents the most difficult research challenge facing human geneticists today. We consider CVD in this presentation to illustrate the issues encountered in using genetic information in research to understand the etiology of most common chronic diseases as well as in the identification and treatment of individuals who are at increased risk. We call to attention what every cardiovascular researcher knows about the etiology of CVD but most deny, or choose to ignore, when designing, carrying out, and reporting genetic

studies of CVD. We close by suggesting steps that should be taken to cope with this inconsistency.

Disease Susceptibility

All cases of CVD have a complex multifactorial etiology. Neither genetic nor environmental agents acting independently cause disease. Full knowledge about an individual's genetic makeup or exposures to adverse environments cannot predict with certainty the onset, progression, or severity of disease. Disease develops as a consequence of interactions between the "initial" conditions, coded in the genotype, and exposures to environmental agents indexed by time and space¹⁹⁻²¹ that are integrated by dynamic, regulatory networks at levels above the genome.²² The interaction of an individual's environmental experiences with her/his genotype determines the history of her/his multidimensional phenotype, beginning at conception and continuing through adulthood. At a particular point in time, each genotype has a range of possible phenotypes determined by the range of possible environmental histories. To illustrate this relationship, by collapsing an individual's phenotype into a single dimension, two of the many possible phenotype histories for a genotype are given in Figure 1. The phenotype of an individual in a particular environmental niche, at a particular point in time, is influenced by the phenotype produced by previous genotype-environment interactions and the potential of the genotype-phenotype combination to react to contemporary environments. The potential to react is constantly changing throughout life from conception to death.^{18,23,24} The consequence of these interactions with exposures to environmental agents indexed by time and space is that many individuals who have a genotype that predicts an increased risk of disease

Received March 17, 2003; revision accepted April 22, 2003.

From the Departments of Human Genetics (C.F.S., J.H.S.) and Epidemiology (S.L.R.K.), University of Michigan, Ann Arbor, and the National Public Health Institute (J.H.S.), Helsinki, Finland.

Correspondence to Dr Charles F. Sing, Department of Human Genetics, University of Michigan, 1241 E Catherine St, 5928 Buhl Bldg, Ann Arbor, MI 48109-0618. E-mail csing@umich.edu

© 2003 American Heart Association, Inc.

Arterioscler Thromb Vasc Biol. is available at <http://www.atvbaha.org>

DOI: 10.1161/01.ATV.0000075081.51227.86

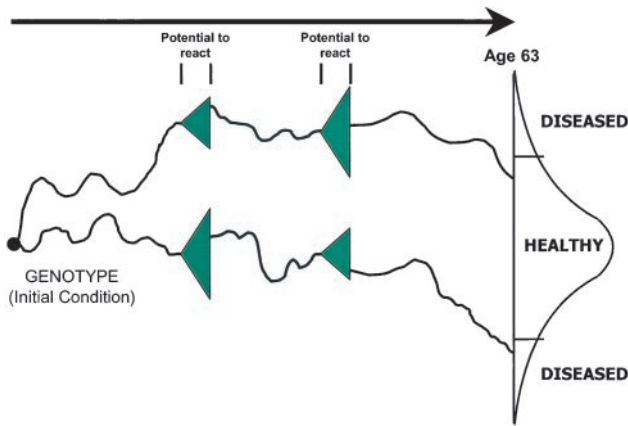


Figure 1. Two possible environmental histories in the time-space continuum encountered by a genotype.

will remain healthy because of exposures to compensatory environments. The converse will also be true. Individuals who have a genotype that has a low risk of disease might develop disease because of an adverse environmental history. The important role of the individualized history of exposures to environmental agents makes the *average* reaction of a group of individuals with the same genotype to the wide range of potential environmental agents over a lifetime a poor

predictor of the risk of disease for most individuals with that genotype. CVD research has revealed tens of high-risk environmental agents and hundreds of genes, each with many variations, that influence disease risk. As the number of interacting agents that are involved increases, a smaller number of cases of disease will have the same etiology and be associated with a particular multigene genotype.

The important role that biochemistry and physiology play in the connections between the genome and disease phenotypes brings into question the utility of the overused, simplistic view that the genome produces an independent, isolated, and fixed one-way flow of information from genome to phenotype. Figure 2 shows how a particular multigene genotype is connected to the domain of potential coronary artery phenotypes through the primary biochemical and physiological subsystems. The phenotypic measures of health are constantly being shaped, changed, and transposed as a consequence of epigenetic networks of cellular and organismal dimensions that evolve over the lifetime of the individual. At the level of the cell, these networks influence DNA methylation and repair; they also serve to organize coordinated responses to heat-shock, oxygen deprivation, and other environmental changes.²⁵ The relationships between subsystems influence the trajectory of an individual's phenotype across the potential reaction surface associated with a partic-

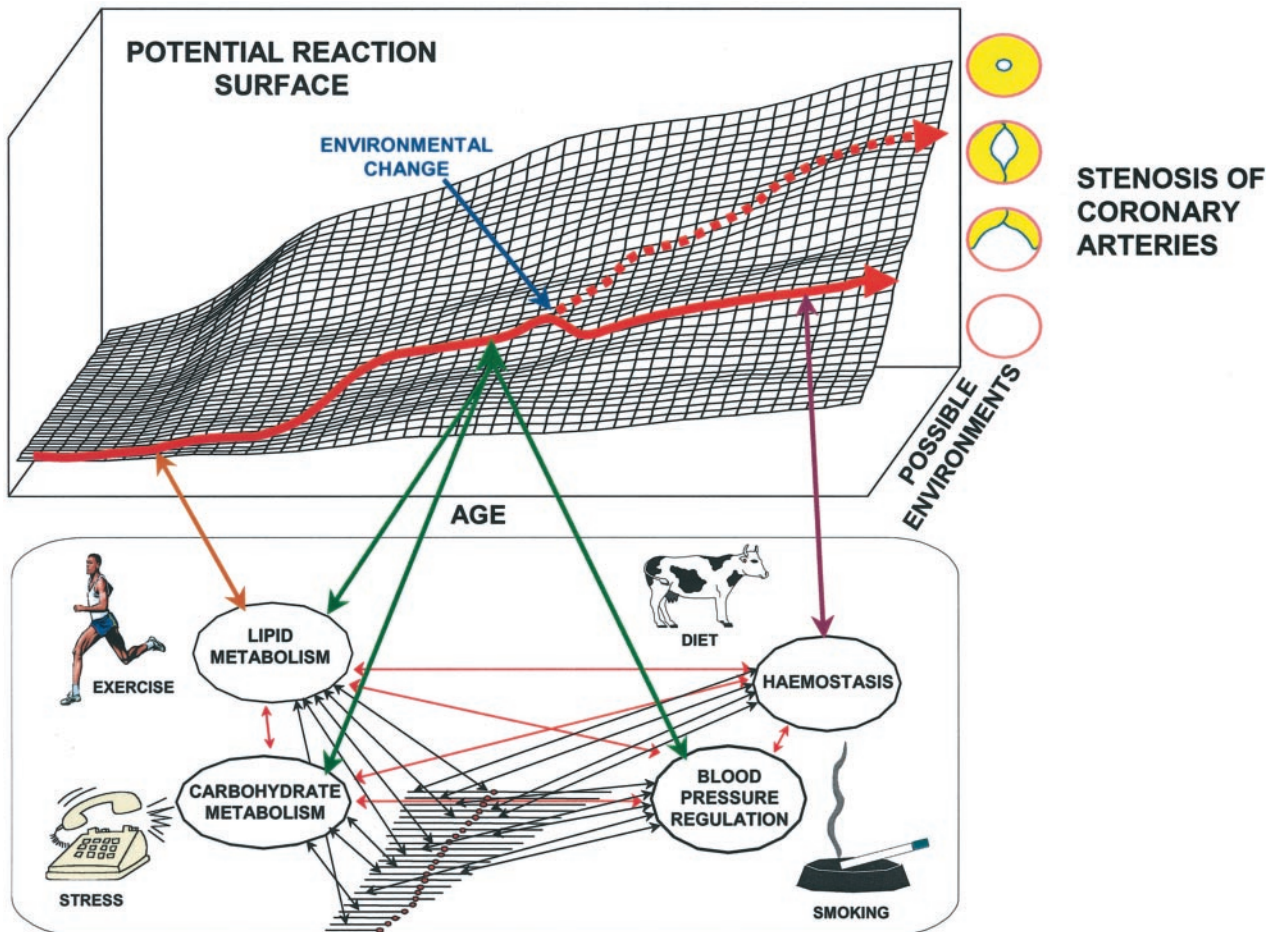


Figure 2. A model for an individual's propensity to develop coronary artery disease.

ular genotype. The phenotype produced by these subsystems continuously feeds back information to influence the expression of the participating genes and the relationships between the intermediate agents that make up the connecting subsystems. Predicting multifactorial disease outcomes without consideration of epigenetic networks is increasingly seen as naive.^{13,16–18} Few genetic studies of the CVDs recognize the realities of the dynamic relationships between an individual's genotype, her/his history of exposures to environmental agents, such as smoking, a high-fat diet, or a statin drug, and the contemporary phenotype in predicting phenotypic outcomes for a future point in time and a particular environmental niche.

Population Genetics

It is relevant to recognize that disease prevalence is a consequence of the intersection of the genetic variation that represents a population with the possible histories of environmental exposures that each member of that population might have experienced. Each population is expected to have a different distribution of relative genotype frequencies and a different constellation of possible environmental histories. The distribution of the relative frequencies of genotypes involved in determining the distribution of individual susceptibilities to disease in a particular population is defined by the number of segregating susceptibility genes, the number of alleles of each gene and their relative frequencies, and the correlation between alleles of each gene and between alleles of different genes. There are hundreds of genes known to have functional allelic variations that might contribute to determining an individual's susceptibility to CVD. All functional variations in a particular gene are not expected to be present in all populations.^{26–30} Because new DNA variations arise in isolation and because chance, selection, and migration work as "filters" in each population to modify the relative frequencies of genetic variations in evolutionary time, different populations will have different combinations of DNA variations and hence, a different array of alleles and genotypes, for any particular susceptibility gene. This is a major fact that is most often ignored when developing strategies for understanding and predicting an individual's disease risk, as well as the development of therapies, by using knowledge about the phenotype(s) associated with a single-gene variation derived from studies of only a few populations.

Different subsets of genes will be influencing phenotypic variation in different subgroups of the same population. Because multigene genotypes will have a multinomial distribution, different combinations of susceptibility genes will be involved in determining disease risk in different individuals in different families. Therefore, every individual who has, or will develop, disease, even if they are drawn from the same population, is not expected to have the same genotype for all of the susceptibility genes. It follows that because the incidence of disease is a consequence of interactions of a population-specific distribution of susceptibility genotypes with the population-specific combinations of exposures to environmental agents over time, the cardinal genetic question must be this: which variations, in which genes, and in which populations are useful for understanding disease and predict-

ing which individuals will develop disease in which strata of environmental histories? Answering this question will establish the population-specific genetic architecture of disease and provide the basis for including genetic information into the practice of individualized medicine. Few genetic studies of common multifactorial diseases recognize the importance of this question.

Two Different Research Strategies for Studying Genetic Architecture

It is acknowledged by most researchers that information about genetic variation can be useful in the identification of presymptomatic individuals at increased risk of developing a common multifactorial disease and also in predicting progression and severity for those with disease. There is not universal accord, however, on the extent to which information about the complexity of the etiologies of a common multifactorial disease, such as CVD (illustrated in Figures 1 and 2), should be included in developing a strategy to use genetic information to diagnose and predict clinical end points in medicine and public health. The a priori importance assigned by researchers to the role that knowledge about the complexity of the etiologies of disease can play in developing genetic predictors is reflected by the research strategies that they employ. We next review two of these alternative strategies.

Disease Is a 'Simple' Consequence of Variations in Independent Causal Agents

The widely held belief that each case of disease is caused by a variation in a single agent follows from the medical successes that have been achieved for the infectious diseases.³¹ The industrialization of medicine has encouraged a search for particular agents that "control" the health of individuals and populations. It is taken to be axiomatic that knowledge about the nature of each causal agent can provide the power to prevent, or alter the course of, disease. The success of this reductionist research paradigm³² depends critically on three major simplifying assumptions: (1) it is possible to isolate each causal agent (genetic or environmental) without altering its role in producing the phenotype; ie, the role of an agent is not changed as a consequence of the process of measurement; (2) manipulation of an agent does not alter the behaviors of the other agents that influence the phenotype; and (3) the relationships between causal agents and outcomes are invariant; ie, they are static, not dynamic.

The validity of these assumptions is particularly critical for genotype-phenotype association studies. If they are false, then applications of single-gene analyses, one gene at a time, to unravel the genetic architecture of CVD are suspect and likely to be misleading. An analysis of studies reported in the leading cardiovascular research publications documents that the reductionist paradigm is accepted by most CVD genetic researchers without question, and its application is the rule rather than the exception. Much is being reported on the nature of the bits and parts of the genetic etiology of CVD, but too little attention is being paid to researching their integration into a model that predicts the emergent phenotypes that are measures of health.

It is apparent to many, but voiced by few, that focusing on the single genes with large marginal (independent) effects on disease risk might not produce the promised medical successes.^{3,12,15,33,34} As Morton³⁵ has emphasized, the genetic architecture of the continuously distributed phenotypes of health will not be revealed by the reductionist paradigm embraced by molecular biology. Currently, only a small fraction of the risk of CVD is attributable to the influences of variations in single genes with large phenotypic effects.^{36,37} Furthermore, if we do not accept and study the possibility that single genes, which have small, average genotypic effects in the population at large, can make a major contribution to understanding and predicting disease in particular individuals in particular strata of the population because of their interactions with other genes and environments, it will not be possible to adequately evaluate the utility of genetic information. As Morin³⁸ implores in a call for a paradigm shift, the very way we think about the problem prevents us from knowing. Cohen and Rice³⁹ call to attention the conundrum we all face: “The problem that afflicts all sciences is the fact that once you have defined the kind of answers that you expect to get, it is very difficult to know what you are missing.” Also, Popper, a prominent 20th-century philosopher of science, pointed out that the ability to formulate new questions is fundamental to initiating a paradigm shift.⁴⁰ This is clearly an issue faced by geneticists in their search for an understanding of the distribution of the common multifactorial diseases, such as CVD, among individuals, families, and populations in the post-genomic era. A plan for executing the prevailing research paradigm to identify independent, causal, genetic variations is laid out by Botstein and Risch.⁴¹ Morowitz⁴² presents chemical and biologic arguments that make clear the necessity for an alternative research paradigm for studying the genetics of human disease.

Disease Is a Consequence of the ‘Complex’ Organization of Interacting Agents

This view of the genetic analysis problem takes into account the networks of intermediate biochemical and physiological subsystems that connect genome variation with phenotypic variation illustrated in Figures 1 and 2. It embraces four fundamental aspects of disease etiology. First, the same network of interacting, quantitatively varying, intermediate biochemical and physiological agents that influences the so-called normal range of interindividual variability among the healthy also influences the development of disease. Individuals with disease are just in a different part of the multidimensional genotype-environment-intermediate biochemical and physiological state space, defined by variation and covariation of the agents, than are individuals who are healthy. Inferences about the role of molecular variation from studies that focus primarily on individuals who have CVD cannot provide unbiased information for prediction of disease risk among individuals in the population at large. The sampling issue is not widely appreciated among laboratory-based researchers. The probability of observing a particular genotype in a sample ascertained because they have a particular disease phenotype cannot be equal to the probability of encountering the disease phenotype in individuals with

a particular genotype. For instance, the cumulative risk of CVD death by age 65 years is 0.7 for individuals with familial hypercholesterolemia, a well-known lipid disorder associated with a defect in the LDL receptor gene.⁴³ However, only a small fraction of those with CVD have this gene defect.³⁶ It follows that genetic predictors that provide valuable information about selected patient groups seen in hospitals will be much less valuable in general practice, in which patients are drawn from the population at large.

Second, the biologic relationships between the network of interacting agents and cardiovascular health are nonlinear.^{44,45} Disease is a consequence of an individual’s homeodynamic mechanisms that do not compensate for disturbances in levels of, and the relationships among, the agents involved in causation.^{16,46} Changes in the level of one agent might influence disease risk by altering the relationship(s) among other agents. In most cases, the size of the marginal (independent) effect of a variation in an agent is inversely related to the dependence of its effect on the context defined by other interacting agents.¹⁵ Geneticists, in particular, must be aware that the context in which a molecular process takes place deserves as much study as the biochemical content of the process.^{24,39,47} We should be asking how do genetically influenced changes in relationships between agents influence risk of disease? The current disconnect between our knowledge about genomic variation and coordinated variation in intermediate traits that influence disease risk presents the challenge of reformulation of the questions being asked and new analytical skills yet to be developed that are necessary to address them.

Third, the genetic architecture of cardiovascular health is expected to be population-specific. Few populations will have the same relative frequency distributions of genetic variations.^{26–30,46,48} Differences among populations in the relative frequency of a susceptibility genotype or an environmental exposure will contribute to differences in the utility of a susceptibility genotype for prediction of disease within a particular population. Even for those rare instances when the relative genotype frequencies are the same across populations, the contribution of genetic variation to prediction will still be different, because its influence on variation in disease susceptibility depends on a particular combination of environmental exposures whose relative frequencies vary among populations.

Last, studies of genetic architecture can be guided by an understanding of the complexity of the etiologies of cardiovascular health. Studies to document the complexity of etiology and the influences of context defined by gender, age, and other measures of environmental effects should be a priority.^{49–51} We concur with Anderson⁵² that we must turn to nature to inform us about the type of model that should be used to describe nature. Even though it might never be possible to know everything about etiologies,^{52–55} the study of genetic architecture can be guided by our current knowledge, albeit incomplete. Recognition of the complexity of the organization of interacting agents can foster synergy between efforts to predict disease and efforts to understand the etiology of disease. For example, a particular gene might be a candidate for prediction because its product is involved in

the metabolism of intermediate biochemical and/or physiological agents that define disease. Also, etiologic relationships between agents involved in causation might suggest genetic studies of trait relationships and their role in prediction of disease. A research strategy to extract the full utility of genetic information that recognizes these considerations must begin by documenting the nature and extent of the complexities rather than seeking universal invariant, context-independent effects of single genes or genotypes. So, what steps should be taken to cope with these complexities? It seems imperative that geneticists consider the following.

Admit That Etiology Is Complex

We should ask questions and carry out research that reflect the reality of the problem. Pretending that the etiology of a common human disease like CVD is caused by the independent actions of multiple agents is deterring progress. Accepting the complexity of etiology provides a framework for organizing an immense amount of observations.⁵⁶ The result will bring clarity to the formulation of appropriate questions and selection of research methods.

Test Commonly Held Assumptions

The goal of CVD research must be to formulate a mathematical/statistical/computer model that summarizes the complexity of the etiology in manageable dimensions. What one is willing to assume guides the process and determines the validity of the applications of the model. The first step toward building such a model must be to test the validity of the assumptions that are being made in its formulation.

Ask Relevant Questions

Distinguishing between interesting questions and important questions is a function of social, economic, and cultural preoccupations of the community at large. The definition of an important question will suggest the appropriate model, measures, sampling design, and analytic tools necessary to interpret the data collected. A model counts as an explanation only if it meets the needs of an individual (in clinical medicine) or a community (in public health programs). Physicians and public health workers will ultimately decide whether a model has explanatory power or provides understanding.

Refocus on Measurement of the Environment

In the pregenome era, environmental factors were considered to be the major predictors of diseases. In the postgenomic era, genetic factors have supplanted rather than complemented the environmental approach. We must return to placing equal emphasis on the role of environment and the interactions with the new-found genetic factors. This will require new technologies to measure the environments, both internal and external to the individual, with the same precision as measurements of DNA.

Develop Nontraditional Analytic Methods

We need to explore new mathematical and statistical methods that consider the biological and genetic facts that have accumulated over the past 100 years and incorporate them into the analysis of the exponentially growing databases generated by contemporary molecular technologies. Tradi-

tional statistical approaches to modeling biologic data are inadequate and inappropriate for addressing questions of genotype–intermediate trait–disease phenotype relationships when hundreds to thousands of measurements are considered simultaneously. New methods for information handling, model pruning, and biological interpretation of research results are required. Identification of the context in which a genetic model is useful should be a high priority. Meta-analyses will be of less value in sorting out the genetic architecture. A meaningful strategy will integrate bottom-up, top-down, and hierarchical approaches to identify the subset of key variables for predicting and understanding gene-environment-disease end-point relationships.

Replicate to Sort Out Invariant and Context-Dependent Genetic Effects From Type I Errors

We expect that the effects of few genes will be invariant across populations and environmental strata; most will be context-dependent. A fraction of the context-dependent gene effects will be invariant across time within a particular stratum defined by age, gender, smoking habit, or other measures of exposures to environmental factors, both internal and external to the individual, which could help to distinguish them from Type I statistical errors (false-positives). A major research challenge will be to design and execute appropriate studies to distinguish between context-dependent effects and Type I errors. Inferences about genetic effects will be dependent on the coarseness of the graining of the interacting genetic and environmental agents that are involved in the etiology of the phenotype of interest.⁵⁷

Train Scientists for a Biocomplex Future

Connecting data with questions involves analytical skills that are in short supply. We have a plethora of data collectors and a dearth of qualified data analyzers. Turning the analytical step over to physicists, statisticians, and computer scientists who do not have basic training in biological research is likely to increase, rather than decrease, the disconnect between question and inferences in biology and medicine. The success of mining strategies being developed for very large data sets critically depends on the formulation of appropriate a priori hypotheses that follow from an understanding of the relevant questions.

Place Value on a Synthetic Mind Set in Promoting Young Scientists

The organization of the contemporary academic community is driven by the desire to deliver products and services according to management strategies adopted from industry. The genetic research community currently places highest value on finding and describing the bits and pieces of human health in promoting and rewarding research projects and individual scientists. Singularity of purpose, and a reductionist approach that has no interest in complexity, discourages imaginative solutions. We are in need of an academic environment that puts greater value on research projects and scientists committed to studying how the parts are put together. Such an environment should encourage a synthetic mind set among scientists espousing different disciplinary assumptions, foster communication that widens the context of

scientific research, and place a premium on collaboration in promoting and rewarding young investigators.

Summary

The Human Genome Project has revealed thousands of genes and millions of gene variations that might influence human health. We are now faced with reconciling a high-dimensional causal-state space of molecular networks that connect DNA variation and the well-established role of exposures to high-risk environmental agents with the emergent, discrete, clinical outcomes that are relevant to medicine and public health. We are entering an era of synthesis that will take advantage of the successes of reductionism in defining participating agents. Meaningful insights will depend on a fundamental change in how we use this information to model, measure, and analyze genotype-phenotype relationships. The fact that epigenetic feedback mechanisms and interactions of many agents from the genome through intermediate biochemical and physiological subsystems with exposures to environmental agents contribute to the emergence of an individual's clinical phenotype suggests that we will find heterogeneity in causation and predictability of agents among subsets of prevalent cases of disease. Embracing a more realistic biological model that incorporates the complexity of the interactions between agents as causations in a particular context indexed by time and space will be necessary to answer the three cardinal genetic questions about disease¹⁴: (1) where are the susceptibility genes located? (2) what are the functional DNA sequence variations in these genes? and (3) what are the statistical (for prediction) and biologic (for etiology) relationships between genotype variation and variation in onset, progression, and severity in which subsets of individuals? An unwillingness to adopt a realistic biological model for health when designing and analyzing studies of disease might be the greatest deterrent to answering these questions that are most relevant to the practice of medicine without prejudice.

Acknowledgments

This work was supported in part by National Institutes of Health (Bethesda, Md) grants HL39107, HL51021, and GM65509. We wish to thank our many friends for their criticisms and encouragement that helped to shape our perspective on common disease research.

References

- Baird PA, Anderson TW, Newcombe HB, Lowry RB. Genetic disorders in children and young adults: a population study. *Am J Hum Genet.* 1988;42:677–693.
- Baird PA. Genetics and health care: a paradigm shift. *Perspect Biol Med.* 1990;33:203–213.
- Strohman RC. Ancient genomes, wise bodies, unhealthy people: limits of a genetic paradigm in biology and medicine. *Perspect Biol Med.* 1993;37:112–145.
- Murray CJL, Lopez AD. Mortality by cause for eight regions of the world: global burden of disease study. *Lancet.* 1997;349:1269–1276.
- American Heart Association. Heart and stroke facts: 1994 statistical supplement. Dallas, Tex: American Heart Association; 1994.
- American Heart Association. Heart and stroke facts. Dallas, Tex: American Heart Association; 2002.
- American Heart Association. Heart and stroke statistical update. Dallas, Tex: American Heart Association; 2002b.
- Davignon J, DuFour R, Cantin M. *Hypertension: Physiopathology and Treatment.* 2nd ed. New York, NY: Mc Graw-Hill Book Co; 1983: 810–852.
- Sing CF, Moll PP. Genetics of atherosclerosis. *Annu Rev Genet.* 1990;24:171–187.
- Sing CF, Haviland MB, Templeton AR, Zerba KE, Reilly SL. Biological complexity and strategies for finding DNA variations responsible for inter-individual variation in risk of a common chronic disease. *Ann Med.* 1992;24:539–547.
- Sing CF, Reilly SL. Genetics of common diseases that aggregate, but do not segregate, in families. In: Sing CF, Hanis CL, eds. *Genetics of Cellular, Individual, Family, and Population Variability.* New York, NY: Oxford University Press; 1993:140–161.
- Sing CF, Zerba KE, Reilly SL. Traversing the biological complexity in the hierarchy between genome and CAD endpoints in the population at large. *Clin Genet.* 1994;46:6–14.
- Strohman RC. Epigenesis: the missing beat in biotechnology? *Biotechnology.* 1994;12:156–164.
- Sing CF, Haviland MB, Templeton AR, Reilly SL. Alternative genetic strategies for predicting risk of atherosclerosis. In: Woodford FP, Davignon J, Sniderman AD, eds. *Atherosclerosis X*, vol 10. Amsterdam, Netherlands: Elsevier Science Publishers BV; Excerpta Medica International Congress Series, 1995:638–644.
- Sing CF, Haviland MB, Reilly SL. Genetic architecture of common multifactorial diseases. In: Chadwick DJ, Cardew G, eds. *Variation in the Human Genome (Ciba Foundation Symposium 197).* Chichester, UK: John Wiley and Sons; 1996:211–232.
- Strohman RC. Maneuvering in the complex path from genotype to phenotype. *Science.* 2002;296:701–703.
- Dennis C. Altered states. *Nature.* 2003;42:686–688.
- Gottesman II, Gould TD. The endophenotype concept in psychiatry: etymology and strategic intentions. *Am J Psychiatry.* 2003;160:636–645.
- Zerba KE, Sing CF. The role of genome type-environment interaction and time in understanding the impact of genetic polymorphisms on lipid metabolism. *Curr Opin Lipidol.* 1993;4:152–162.
- Zerba KE, Ferrell RE, Sing CF. Genotype-environment interaction: apolipoprotein E (*Apo E*) gene effects and age as an index of time and spatial context in the human. *Genetics.* 1996;143:463–478.
- Zerba KE, Ferrell RE, Sing CF. Complex adaptive systems and human health: the influence of common genotypes of the *apolipoprotein E (ApoE)* gene polymorphism and age on the relational order within a field of lipid metabolism traits. *Hum Genet.* 2000;107:466–475.
- Lusis AJ. Atherosclerosis. *Nature.* 2000;407:233–241.
- Rose S. *Lifelines: Biology Beyond Determinism.* New York, NY: Oxford University Press; 1997.
- Lewontin RC. *The Triple Helix: Gene, Organism, and Environment.* Cambridge, Mass: Harvard University Press; 2000.
- Davidson EH, McClay DR, Hood L. Regulatory gene networks and the properties of the developmental process. *Proc Natl Acad Sci U S A.* 2003;100:1475–1480.
- Scriver CR, Byck S, Prevost L, Hoang L. PAH mutation analysis consortium. In: Chadwick DJ, Cardew G, eds. *Variation in the Human Genome (Ciba Foundation Symposium 197).* Chichester, UK: John Wiley and Sons; 1996:73–96.
- Weatherall D. The genetics of common diseases: the implications of population variability In: Chadwick DJ, Cardew G, eds. *Variation in the Human Genome (Ciba Foundation Symposium 197).* Chichester, UK: John Wiley and Sons; 1996:300–311.
- Clark AG, Weiss KM, Nickerson DA, Taylor SL, Buchanan A, Stengård J, Salomaa V, Vartiainen E, Perola M, Boerwinkle E, Sing CF. Haplotype structure and population genetic inferences from nucleotide sequence variation in human lipoprotein lipase. *Am J Hum Genet.* 1998;63:595–612.
- Fullerton SM, Clark AG, Weiss KM, Nickerson DA, Taylor SL, Stengård JH, Salomaa V, Vartiainen E, Perola M, Boerwinkle E, Sing CF. Apolipoprotein E variation at the sequence haplotype level: implications for the origin and maintenance of a major human polymorphism. *Am J Hum Genet.* 2000;67:881–900.
- Stengård JH, Clark AG, Weiss KM, Kardia S, Nickerson DA, Salomaa V, Ehnholm C, Boerwinkle E, Sing CF. Contributions of 18 additional DNA sequence variations in the gene for apolipoprotein E to explaining variation in quantitative measures of lipid metabolism. *Am J Hum Genet.* 2002;71:501–517.
- Golub ES. *The Limits of Medicine: How Science Shapes Our Hope for the Cure.* New York, NY: Times Books; 1994.
- Salthe SN. *Evolving Hierarchical Systems: Their Structure and Representation.* New York, NY: Columbia University Press; 1993.

33. Nelson MR, Kardia SLR, Ferrell RE, Sing CF. A combinatorial partitioning method (CPM) to identify multilocus genotypic classes that predict quantitative trait variation. *Genome Res.* 2001;11:458–470.
34. Chakravarti A, Little P. Nature, nurture and human disease. *Nature.* 2003;421:412–414.
35. Morton NE. Genetic epidemiology. *Annu Rev Genet.* 1993;27:523–538.
36. Hobbs HH, Russell DW, Brown MS, Goldstein JL. The LDL receptor locus in familial hypercholesterolemia: mutational analysis of a membrane protein. *Annu Rev Genet.* 1990;24:133–170.
37. Davignon J. *New Horizons in Coronary Heart Disease.* London, UK: Current Science; 1993:1–21.
38. Morin E. Method: *Towards a Study of Human Kind: The Nature of Nature (American University Studies Vol 111).* New York, NY: Peter Lang Publishing; 1992.
39. Cohen J, Rice SH. *Integrative Approaches to Molecular Biology.* Cambridge, Mass: MIT Press; 1995.
40. Popper KR. *A World of Propensities.* Bristol, UK: Thoemmes Press; 1995.
41. Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat Genet.* 2003;33:228–237.
42. Morowitz H. Phenetics: a born again science. *Complexity.* 2003;8:12–13.
43. Schuster H. High risk/high priority: familial hypercholesterolemia: a paradigm for molecular medicine. *Atherosclerosis.* 2002;2:27–32.
44. Goodwin B. *How the Leopard Changed Its Spots.* New York, NY: Scribner; 1994.
45. Sole RV, Goodwin B. *Signs of Life: How Complexity Pervades Biology.* New York, NY: Basic Books; 2001.
46. Somogyi R, Grellier LD. The dynamics of molecular networks: applications to therapeutic discovery. *Drug Discov Today.* 2001;6:1267–1277.
47. Lewontin RC. Genotype and phenotype. In: Keller EF, EA Lloyd, eds. *Keywords in Evolutionary Biology.* Cambridge, Mass: Harvard University Press; 1992:137–144.
48. Nickerson DA, Taylor S, Fullerton SM, Weiss KM, Clark AG, Stengård J, Salomaa V, Boerwinkle E, Sing CF. Sequence diversity and large-scale typing of SNPs in the human apolipoprotein E gene. *Genome Res.* 2002;10:1532–1545.
49. Kardia SLR, Haviland MB, Ferrell RE, Sing CF. The relationship between risk factor levels and presence of coronary artery calcification is dependent on *Apolipoprotein E* genotype. *Arterioscler Thromb Vasc Biol.* 1999;19:427–435.
50. Stengård JH, Kardia SL, Tervahauta M, Ehnholm C, Nissinen A, Sing CF. Utility of the predictors of coronary heart disease mortality in a longitudinal study of elderly Finnish men aged 65 to 84 years is dependent on context defined by *Apo E* genotype and area of residence. *Clin Genet.* 1999;56:367–377.
51. Fan J, Yang X, Wang W, Wood WH 3rd, Becker KG, Gorospe M. Global analysis of stress-regulated mRNA turnover by using cDNA arrays. *Proc Natl Acad Sci U S A.* 2002;99:10611–10616.
52. Anderson PW. More is different: broken symmetry and the nature of the hierarchical structure of science. *Science.* 1972;177:393–396.
53. Anderson PW. *Complexity: Metaphors, Models, and Reality* (Santa Fe Institute Studies of Complexity, vol 19). Reading, Mass: Addison Wesley; 1994.
54. Barrow JD. *Theories of Everything: the Quest for Ultimate Explanation.* New York, NY: Oxford University Press; 1991.
55. Laughlin RB, Pines D. The theory of everything. *Proc Natl Acad Sci U S A.* 2000;97:28–31.
56. Hoffmann R. Why buy that theory. *Am Sci.* 2003;91:9–11.
57. Kardia SLR, Haviland MB, Sing CF. Correlates of family history of coronary artery disease in children. *J Clin Epidemiol.* 1998;51:473–486.