Genetics of Complex Disease
Approaches, Problems, and Solutions

NICHOLAS J. SCHORK

Department of Epidemiology and Biostatistics and Department of Genetics, Case Western Reserve University, Cleveland, Ohio; Program for Population Genetics and Department of Biostatistics, Harvard University School of Public Health, Boston, Massachusetts; and The Jackson Laboratory, Bar Harbor, Maine

Complex or multifactorial diseases are defined as diseases that are ultimately determined by a number of genetic and environmental factors. Although there are many technologies and strategies that can be used to detect genetic factors influencing complex diseases, these technologies and strategies have inherent limitations. In fact, the very name "complex disease" suggests that the results from relevant studies will not be simple to decipher. Ultimately, both the detection and precise characterization of a factor's contribution to a complex disease are difficult undertakings, because the effect of any one factor may be obscured or confounded by other factors. However, the genetic dissection of complex diseases can be greatly facilitated by paying heed to two very basic distinctions. The first distinction is between complexity at the level of individuals and complexity at the level of populations. The second distinction is between the two sequentially pursued components of gene discovery paradigms: gene identification and gene effect characterization. Although genetic epidemiology, as a research field, is oriented to both components of gene discovery for complex diseases, it is suited to gene effect characterization at the population level more than anything else. This paper reviews the origins of the genetic basis of complex traits, as well as the problems plaguing genetic epidemiologic analysis strategies, with the hope of showing how greater attention to these distinctions, as well as a greater integration of relevant knowledge, can alleviate confusion and shape future investigations. In addition, a new discipline, "phenomics" or "phenometrics," could be initiated that would complement genomic research as presently performed. Schork NJ. Genetics of complex disease: approaches, problems, and solutions.

AM J RESPIR CRIT CARE MED 1997;156:S103–S109.

The rapid pace at which molecular genetic technologies are being developed and applied is reflected in the multitude of research papers that fill the most eminent contemporary scientific journals. This pace has been somewhat dizzying and has created an aura of hope that finding genes that influence common complex diseases such as asthma will ultimately lead to the development of successful therapies and cures. Unfortunately, this hope must be tempered by the harsh reality that most common diseases are sufficiently complex to push the current array of molecular genetic analysis strategies to their limits. In fact, diseases with features that complicate detection of their contributing factors are generally referred to as "complex diseases" and are receiving a great deal of attention by contemporary geneticists and epidemiologists (1, 2).

This paper reviews the origins of the genetic basis of complex diseases, as well as the problems plaguing genetic epidemiologic analysis strategies for dissecting complex diseases. The focus, however, is on three items. First, two distinctions are elaborated that, if ignored, could contribute to overestimating the power of current genetic analysis strategies and obscure the role certain technologies and research paradigms have in the genetic analysis of complex traits. Second, general strategies for overcoming problems inherent in the current pool of genetic analysis strategies are discussed. Third, a new discipline, termed "phenomics" or "phenometrics," which attempts to unravel biochemical and physiological hierarchies leading from genes to clinical endpoints, is proposed and elaborated in the context of complex disease research.

ORIGINS OF COMPLEX GENETIC DISEASES

Physiological Complexity

The functioning human body is an extraordinarily complex biological phenomenon. A number of biochemical networks, physiologic systems, and organs must work together in intricately coordinated fashions for the human body to thrive. There are a great many features of the human body that require subtle control of a number of biochemical and physiological mechanisms. For example, respiratory control, blood pressure regulation, nutrient extraction and dispersion from food, and immune system surveillance all involve and implicate a number of biochemical and physiologic systems. These systems are typically under genetic control and often respond via feedback mechanisms to environmental stimuli, both to

Supported by National Institutes of Health grants HL94-011 (National Heart, Lung, and Blood Institute), HL54998-01 (National Heart, Lung, and Blood Institute), and RR03655-11 (National Center for Research Resources).

Correspondence and requests for reprints should be addressed to Nicholas J. Schork, Ph.D., Department of Epidemiology and Biostatistics, Case Western Reserve University, R215 Rammelkamp Building, MetroHealth Medical Center, 2500 MetroHealth Drive, Cleveland, Ohio 44109-1998.

ensure stability (i.e., homeostasis) and to allow adaptability. These systems are also often replete with redundancies and compensatory mechanisms such that if one component of the system is upset, other (nondysfunctional) systems can make up for it in some sense. One trivial but excellent example of a truly integrated physiologic phenomenon is the so-called "fight or flight" response, which implicates control of numerous physiologic systems by the central nervous system (3). Aspects of such systems can be thought of as hierarchies (4), with genes at the lowest level and clinical endpoints that define disease at the highest (Figure 1). What is important to consider about human physiologic complexity is that it suggests that (*1*) a number of things could go wrong that might ultimately influence a particular clinical endpoint taken as a sign or symptom of disease, and (*2*) different individuals could have different underlying pathologies that lead to similar phenotypic endpoints.

Population Complexity

Given the complexity of human physiology, and the fact that genes ultimately determine, in concert with the environment, phenotypic endpoints of the type evaluated in clinical settings, it is easy to see how complex the genetic structure of diseased populations can be. The human population is quite large and relatively old, has undergone a number of stratifications and divisions, has remixed people at one time within these subdivisions, and is settled in environments that vary widely. If, in fact, different genetically mediated physiologic mechanisms impact clinically relevant disease endpoints, then different genetic variants—each affecting different mechanisms—are likely to have distributed themselves throughout the human population. Some of these variants will have deleterious effects on individuals (possibly only within certain subpopulations), because they either exist in individuals with other disease-predisposing genes, or because the individuals possessing them are exposed to environments that exacerbate or bring out the deleterious effects of the gene (for interesting commentaries on the role of the environment, see references 3 and 6–9). Such shuffling of genes can create great genetic heterogeneity, to the point where individuals with similar clinical features of disease possess different dysfunctional genes as the root cause of those common clinical features. The forces that dictate population subdivisions, the (possibly growing) presence of a gene within a subpopulation, environmental changes, migrations, and other forces are often difficult to characterize and are therefore passed off as "random" or "stochastic" factors that affect populations. Sewall Wright proposed the notion that such factors not only influence the emergence, prevalence, and maintenance of novel phenotypes, but could also shape evolution in dramatic and important ways (5, 10, 11).

The Descriptive Complexity of Complex Diseases

The complexity of human physiology and the genetic complexity of human populations can cause a number of specific phenomena that, although amenable at times to mathematical modeling and experimental dissection, often make the determination of the factors influencing complex genetic diseases difficult. These phenomena can be summarized and categorized in the following ways:

1. *Classical polygenic or "threshold" inheritance*, in which a number of genotypes or mutations at different loci (which likely affect different physiologic systems) must be present before a system is sufficiently challenged to result in disease.

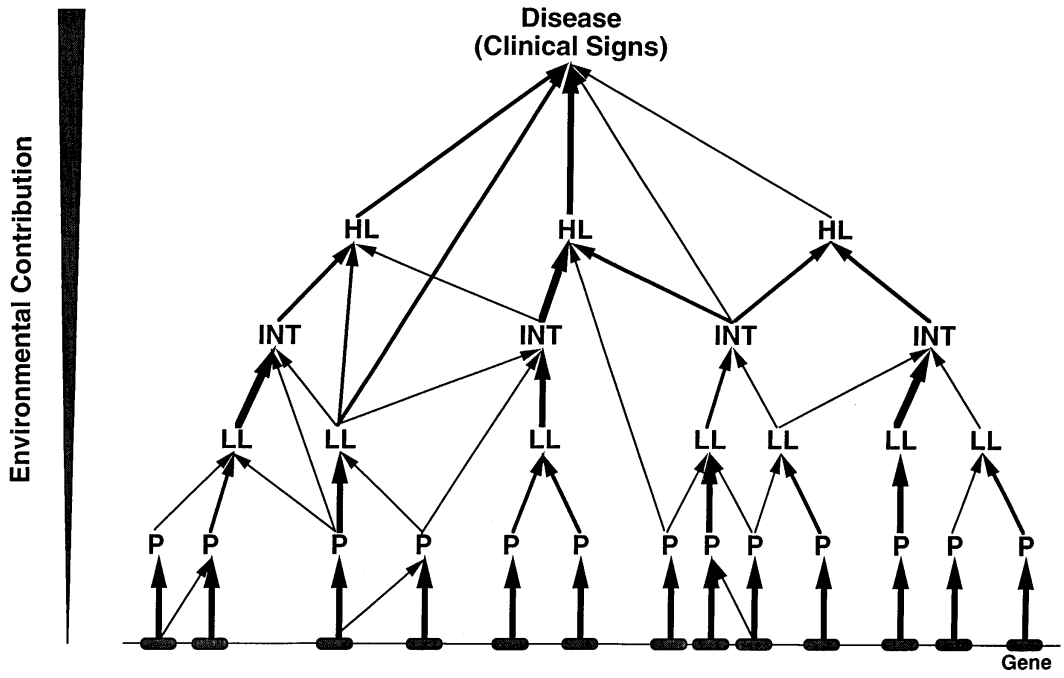

Figure 1. Schematic diagram of the relationship of genes and their products to intermediate phenotypes as well as the more overt clinical manifestations of a disease. The thickness of the arrows denotes the strength of the contribution of a lower-level factor to a higher-level factor. The inverted triangle on the left-hand side of the figure represents the (likely) diminishing effect of environmental conditions on factors integrated at lower and lower levels of a biochemical and physiological hierarchy. (Note: environmental mutagens that alter genetic structure should be recognized.)

2. *Locus heterogeneity*, in which defects in any of a number of genes or loci confer disease susceptibility independently of each other.
3. *Epistasis, or gene interaction*, in which the possession of a certain mutation or genotype will confer susceptibility to a degree dictated by the presence of other mutations or genotypes. Thus, epistasis reflects basic interactive effects of mutations, genotypes, and/or their biologic products (12).
4. *Environmental vulnerability*, in which gene products and the more remote phenotypes they affect are influenced by environmental stimuli.
5. *Gene × environment interactions*, in which a gene or genes have their deleterious effects only in the presence of a particular environmental stimulus.
6. *Developmental or time-dependent expression of genes*, in which a gene, whether in mutant form or not, has its most pronounced effect at a certain time or developmental stage (e.g., puberty).
7. *General aging of the system*, in which genes, either through programmatic senescence or general wear and tear, either mutate more rapidly (as may be the case for the mitochondrial genome) or do not produce vigorous products and therefore result in weakened or damaged physiologic mechanisms and phenotypic endpoints.

These factors make the isolation and characterization of genes influencing complex diseases difficult, since the contribution of a single gene to a disease may be obscured by the others. In addition, each factor that contributes to a complex disease will likely manifest a *context dependency*, in the sense that its deleterious effect may only be visible in the presence of other factors (f). To overcome these difficulties, geneticists must fashion studies that ultimately reduce the "noise" surrounding the "signal" (effect) of a gene in order to disclose or identify that gene. After a gene that is thought to influence a complex disease has been identified, one could then study its context dependency through laboratory manipulations of one sort or another or through large-scale epidemiologic investigations. These investigations could study (*1*) the frequency of the deleterious gene in the population, (*2*) the effect of the gene in the presence of other factors (e.g., environments), and (*3*) the impact on public health and health economics of the gene if its effects were ameliorated.

GENETIC ANALYSIS STRATEGIES

There are two basic strategies for disclosing and characterizing genes that influence complex diseases: candidate gene analysis and whole-genome searches or "genome scans." Each of these strategies has unique features, motivations, and problems associated with them and will therefore be discussed in isolation.

Candidate Gene Analyses

Methods and problems. Candidate gene analysis is very strightforward in that one merely seeks to test the association between a particular genic variant (i.e., allele) and a disease. If the variant is more frequent in persons with the disease than those without it, then one inference that could be made is that this is due to a causal relationship between that variant and the disease. Another inference could be that the gene is in linkage disequilibrium with a disease gene at a locus near the locus in question. Candidate gene analyses are therefore dependent on knowledge about a gene or variant, and the appropriateness of the analysis of a particular gene is only as good as the knowledge that makes the gene or variant a "candidate" in the first place (5). The knowledge that suggests the candidacy of a gene can come from a number of different sources: biological insights (e.g., the gene is known to be expressed in a certain tissue relevant to the disease), homology to other genes, guesswork, etc. The fundamental problem with candidate gene analysis given the complexity of human physiology is that there are likely to be an incredibly large number of candidate genes for any given disease. Analysis of each and every one of these candidates, in isolation of the others, may be prohibitive and difficult to interpret both statistically and biologically (1, 13). Heterogeneity, both with respect to the genes that contribute to a disease and the environments that study subjects are likely to live in, can cause further problems, although statistical strategies for overcoming aspects of these problems have been devised (14, 15).

Potential solutions. The use and abuse of candidate gene studies are likely to grow as more and more genes are identified whose ultimate functions are unknown but whose role in disease susceptibility is a possibility (16, 17). The obvious solution is simply to gain more compelling insight into the possible role a gene might have in human physiology and pathophysiology. Modern technological breakthroughs in the assessment of gene expression patterns (18), homology determination (19), functional studies in model organisms (20), and bioinformatics technologies such as the identification of gene families (21, 22) will all help solidify and narrow down candidate genes to be considered for a particular disease. Although better statistical methodologies will certainly help, greater biological understanding of the genes is likely to make more of an impact, unless one merely wants to search for associations using candidate polymorphisms—a task that is closer to whole genome searches than it is to true candidate gene analysis (13).

Genome Scan Technologies

Methods and problems. Whole genome searches make no assumptions about the candidacy of particular genes or genomic regions. They involve a true search for genetic effects at different locations along chromosomes. They typically involve gathering a large number of related individuals thought to be segregating for genes that influence a disease, and then tracing the putative parent-to-offspring cotransmission of variants (i.e., alleles or genotypes) at landmark spots along the genome (known as marker loci) with possible disease-influencing variants or alleles. If alleles at a particular marker locus appear to segregate (or be transmitted along with) genes seemingly influencing the presence of the disease in question, then an inference can be drawn that a gene actually influencing the disease resides near, or is "linked" to, the marker locus. Statistical methods used to draw inferences about the putative cotransmission of marker locus alleles and trait-influencing alleles have been termed "linkage analysis" methods, have received a great deal of recent attention, and are one of the primary building blocks in genome searches—the other being the DNA markers (1, 2, 23).

Linkage analysis methods are typically carried out in one of two ways: parametric pedigree analysis, which involves tracing cosegregation and recombination phenomena between observed marker alleles and unobserved putative trait-influencing alleles among members of large pedigrees; and allele-sharing methods, which assess the number of marker alleles shared at a particular locus among pairs of relatives manifesting the same trait. Schork and Xu have considered the relative advantages and disadvantages of each approach (24). Most linkage strategies are notoriously weak for detecting genes with small to moderate effects (13, 24) and rarely, if ever, are designed to simultaneously assess multiple gene and environmental effects

(5, 24–26). In addition, the collection of families necessary for conducting genetic linkage analyses and genome-wide searches may require finding a large number of families with individuals possessing the disease of interest. This may not only be extremely costly, especially at the stage of refining the location of disease genes (27), but it could create interpretive and power problems, because families with different environmental exposures and genetic or ethnic backgrounds could enter into the sample, create heterogeneity, and thereby increase the amount of noise obscuring the signal of a given gene effect.

Potential solutions. To overcome problems plaguing genome-wide searches for complex diseases, it is necessary to reduce the effects of other factors surrounding the effect of individual genes. This can be achieved through some simple measures, but not without a cost.

1. *Special sampling.* The easiest way to cut down noise for a genome scan study is to sample pedigree units or families that are likely to harbor and segregate genes of relevance and that are easy to ascertain and analyze. The affected sibling pair approach of linkage analysis is one example of this strategy (28). Other strategies include sequential sampling of family members (29), and merely ascertaining families on the basis of individuals with extreme or remarkable phenotypes. One strategy that is used often is to ascertain individuals with the disease that are relatively young, the idea being that early-onset forms of the disease are more likely to be genetic in origin as opposed to being merely a correlate of aging or exposure to environmental stimuli.
2. *Special clinical populations.* One way to create or ensure homogeneity is to work with populations whose clinical features are unique and likely to be under the influence of similar pathophysiologic mechanisms. An extremely good recent example of this strategy involved the linkage and candidate gene analysis of families showing mature-onset diabetes in the young (MODY), which was motivated by the hope of identifying candidate genes or genes that could be studied for the more common, garden-variety, noninsulin-dependent diabetes (30, 31).
3. *Special isolated populations.* Without recourse to unique clinical populations, another way to ensure homogeneity is to work with population isolates, or populations that do not undergo much immigration. Such groups are likely to have more restricted gene pools and therefore be more homogenous genetically. They are often very likely to be environmentally more homogenous as well. Consider religious isolates such as the Mennonites, the Amish, the Mormons, and the Hutterites, or island populations such as Icelanders. The people within such populations typically engage in similar lifestyles and have similar diets. These populations can also ease genetic analysis, because one can often rely on linkage disequilibrium analysis—a powerful gene mapping tool that will receive considerable attention in the future (32).
4. *Population assays.* In the absence of specific population isolates, one could try to obtain knowledge through genetic population assays of one sort or another that might suggest the existence of subpopulations. Thus, for example, one could survey the kinship or genetic relatedness of neighboring communities to see how genetically homogenous they might be—the idea being that if one can find evidence for greater genetic homogeneity within a community, then combining individuals or families from the different communities might introduce an unwanted heterogeneity. Techniques for assessing population structure in relevant ways are being developed and will likely receive greater attention in the future (33), as will the construction of genealogical data bases that could help link affected individuals genetically.
5. *Intermediate phenotypes.* Most physiologic systems have a hierarchical component to them, leading from the gene to its product, to ever more remote "intermediate" phenotypes, to the ultimate phenotypes used as signs and symptoms to diagnose disease (Figure 2). If one could identify a low-level physiologic mechanism known to be associated with a more remote, clinically relevant phenotype, then genetic analysis of this low level phenotype may be eased merely because its greater proximity to the genetic substrate suggests that a lesser number of genes and other factors influence it. It could be argued that such an emphasis on intermediate phenotypes is mis-

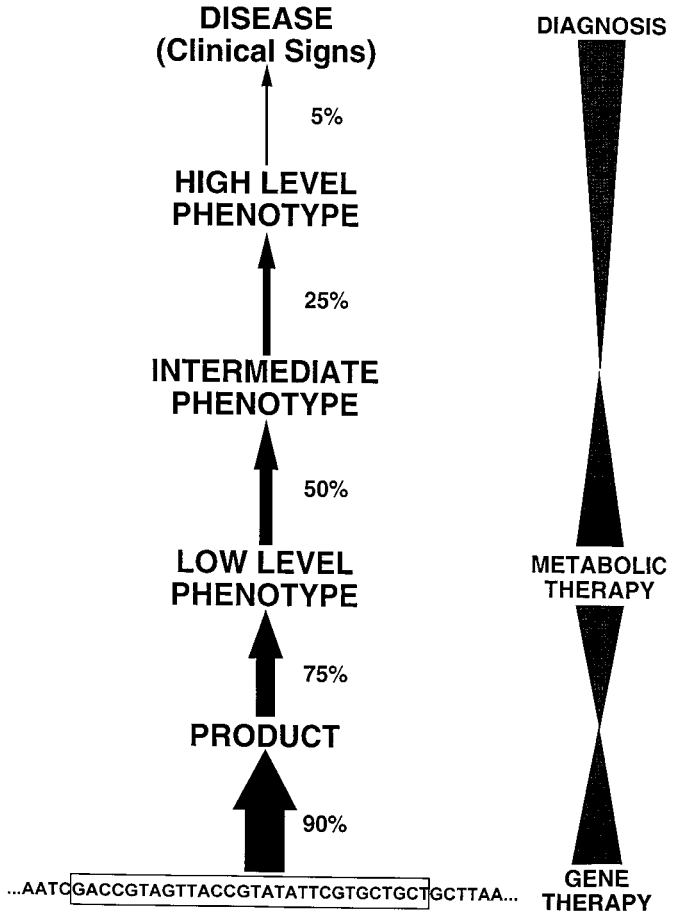

Figure 2. Schematic diagram of the relationship of a gene and its product to a single pathway that, when disrupted or dysfunctional, may contribute to disease. The thickness of the arrows between elements at different levels characterizes the strength of the contribution of a lower-level factor to a higher-level factor. The percentages next to an arrow give the hypothetical percentage of variation explained by the lower-level factor for the higher-level factor that they influence. The symbols on the right-hand side of the figure characterize the potential for diagnosis and therapeutic intervention at each level, with the size of the figures corresponding to the most realistic points for diagnosis or intervention. (Note that this figure is meant to characterize complex traits. For simple Mendelian disorders, knowledge of a gene variant or mutation could lead to easy diagnosis.)

placed, because one may have chosen an intermediate phenotype that is found, in the end, not to relate to the ultimate phenotype of interest. Such arguments are much like those directed at candidate gene analysis and suggest similar responses: the value of an intermediate phenotype is only as good as the physiologic insights that make the connection between it and the more remote phenotype in question. In addition, if one were to ignore intermediate phenotypes and rely on linking up the more remote phenotype of interest with the genes that ultimately influence it, then one would have to detect genes with very small effects. Intermediate phenotype analysis should receive greater attention—arguably, to the point of becoming a research area unto itself.

6. *New analytic methods.* Many traditional statistical analysis tools used in linkage analysis have acknowledged flaws. Overcoming these flaws will not be easy, but some analysis methods have emerged that warrant serious attention. For example, multipoint analysis and the transmission disequilibrium test (TDT) are now tools of choice for most statistical geneticists (15, 34). Other methods that should receive greater attention are variance component models (25, 35, 36), and haplotype sharing methods (37–40).

The Role of Genetic Epidemiology

Genetic epidemiology is a field whose purpose is to facilitate the design and implementation of studies meant to identify and characterize genes influencing traits of all sorts. However, it should be clear that studies designed to identify a gene are not necessarily optimal for characterizing the effect of that gene. Consider the often-used affected sibling pair strategy for linkage analysis. Merely collecting affected sibling pairs will not necessarily allow one to estimate the frequency of a disease-predisposing gene or estimate the contribution of other factors to the disease. As another example, consider the use of an isolated population. Such a population may make it easy to identify a particular genetic variant influencing disease among the people in that population, but the gene found may have a small (or possibly no) role in disease outside that population. This problem should not be a major concern, however, if one is trying to find a gene for the purposes of uncovering a pathophysiological mechanism (or pathway) influenced by that gene that could be studied in other contexts. Genetic epidemiologists should consider their field as one whose primary goal is to characterize the effect of genes at the population level; i.e., to explore the probable context dependency of complex disease-predisposing genes (4). In this light, once could say that true genetic epidemiology will come after the Human Genome Project has been completed. Such analysis will shed enormous light on public health and health economic issues, as well as pave the way for targeted environmental manipulations and therapeutic needs.

THE ULTIMATE SOLUTION: PHENOMICS, INTEGRATION, AND COORDINATION

Much of the material outlined in the preceding sections of this paper have focused on the origins of complex diseases, the types of problems that plague the analysis of complex diseases, and potential ways of making the analysis of complex diseases, in both abstract and concrete ways, easier.

Phenomics and Phenometrics: An Emerging Discipline

Comtemporary complex disease genetics research revolves around two basic paradigms: a "top-down" approach linking clinical endpoints or phenotypes with genotypes, and a "bottom-up" approach that takes genes as a starting point and then works back to the phenotype (4). Genome scans are basically a top-down approach, while many candidate gene studies are characteristic of the bottom-up approach. What is needed, then, are attempts to make sense of what is going on in-between the top and the bottom. Understanding events in-between is, in fact, what drives these two paradigms, because pathology and biochemical systems amenable to therapeutic intervention reside there. One could call the delineation of connections among various genes, gene products, intermediate phenotypes, and clinical endpoints "phenomics" or "phenometrics" to match "genomics" and "biometrics" associated with aspects of pure genetic research. Such a science could proceed quite naturally by mapping genes involved in very low-level phenotypes and activities such as gene product variation and hormone amounts (40), and then attempt to link the phenotypes studied with higher-level phenotypes.

Two simple strategies could be pursued for this purpose. First, one could take two strains of a model organism, such as mice or rats, that have been bred for opposite features (e.g., high blood pressure and low blood pressure) and then screen a number of possible intermediate phenotypes for differences between the two strains. Mapping genes that influence these phenotypes could then proceed by assuming pleiotropy; i.e., the breeding and selection for the remote phenotype during the creation of the strain brought along with it the intermediate phenotype because of the physiologic connection between it and the remote phenotype. Jacob and colleagues have, in fact, successfully undertaken studies in an effort to implement such a strategy (41). Second, one could use pharmacologic probes known to affect certain systems as a way of assaying possible genetically mediated mechanisms (*see* Schork and Weder for a relevant discussion [42]). The assignment of a role in pathogenesis to a particular intermediate phenotype could also appeal to those developing drug-based therapies, since many lower-level phenotypes are the targets of, or interface with, particular compounds.

Integration and Armchair Science

The amount of information relevant to researchers studying complex disease is almost overwhelming: molecular geneticists are characterizing possible candidate genes for disease, statistical geneticists are trying to find new genes through genome scan technologies; physiologists and anatomists are trying to determine mechanisms that, if upset, could result in disease; pathologists are attempting to understand pathways that mediate disease; epidemiologists are attempting to study the population-level context dependence of disease risk factors; pharmacologists are trying to find new compounds that may help ameliorate complex disease progression. This information needs synthesis not only to make sense of, or tie together, relevant bits and pieces of information strewn throughout the literature, but also to point specialists in productive directions. This type of synthesis could potentially be initiated and coordinated outside the lab, by "armchair" scientists akin to bioinformatics professionals, whose goal is to make sense of the tremendous amount of DNA sequence data obtained from various laboratories around the world (21, 22). Such scientists could attempt to devise crucial experiments linking mechanisms and genes of one sort or another, or integrate the knowledge in a way that can help put things in perspective.

DISCUSSION

This review has attempted to make clear some of the issues at the heart of genetic investigations of complex diseases. It has

focused on distinctions that, if taken seriously, could alleviate confusion and promote studies which, although they might not take on as comprehensive an orientation as some genetic epidemiologists would like, could provide information leading to a more rapid dissection of complex diseases when integrated properly with the results of other studies.

In addition to recognizing important distinctions that may help researchers put study designs into context, modern geneticists studying complex traits like asthma, diabetes, obesity, and hypertension must consider some words of caution. It seems that the prime motivation of a great deal of contemporary genetics research is drug discovery. This is a noble motivation, but not necessarily an easy one, and there are two sets of considerations that demand attention on this front. First, Lewontin and others have repeatedly stressed the notion that a focus purely on identifying genes and their unique DNA sequences for the purposes of drug discovery is limited and potentially wasteful:

> Many disorders can be explained by the failure of the organism to make a normal protein, a failure that is the consequence of a gene mutation. But intervention requires that the normal protein be provided at the right place in the right cells, at the right time and in the right amount, or else that an alternative way be found to provide normal cellular function. What is worse, it might even be necessary to keep the abnormal protein away from the cells at critical moments. None of these objectives is served by knowing the DNA sequence of the defective gene (43).

Second, without an appropriate understanding of a disease, particularly the details of just how physiologically integrated the systems involved in its pathogenesis are, one may not be able to gauge the effectiveness of potential therapeutic strategies. For example, if many of the systems implicated in a disease involve purely anatomical anomalies (for example, if a large component of asthmatic disease involves mechanical and structural deficiencies in the respiratory tract), then standard metabolic and gene therapies are not likely to help. In light of these considerations, it seems that greater integration of molecular, physiologic, genetic, pharmacologic, and epidemiologic studies of the type advocated in this paper is especially important.

References

1. Lander, E. S., and N. J. Schork. 1994. Genetic dissection of complex traits. *Science* 265:2037–2048.
2. Weeks, D. E., and G. M. Lathrop. 1995. Polygenic disease: methods for mapping complex disease traits. *Trends Genet.* 11:513–519.
3. Folkow, B. 1993. Physiological organization of neurohormonal responses to psychosocial stimuli: implications for health and disease. *Ann. Behav. Med.* 15:236–244.
4. Sing, C. F., M. B. Haviland, and S. L. Reilly. 1996. Genetic architecture of common multifactorial diseases. *In* C. F. Sing, editor. Variation in the Human Genome. Ciba Foundation, New York. 211–232.
5. Schork, N. J. 1997. Genetically complex cardiovascular traits: origins, problems, and potential solutions. *Hypertension* 29:145–149.
6. Boyd-Eaton, S., M. Konner, and M. Shostak. 1988. Stone-Agers in the fast lane: chronic degenerative diseases in evolutionary perspective. *Am. J. Med.* 84:739–749.
7. Turner, R. C., J. C. Levy, and A. Clark. 1993. Complex genetics of type 2 diabetes: thrifty genes and previously neutral polymorphisms. *Q. J. Med.* 86:413–417.
8. Ritenbaugh, C., and C. S. Goodby. 1989. Beyond the thrifty gene: metabolic implications of prehistoric migration into the New World. *Med. Anthropol.* 11:227–236.
9. Neel, J. V. 1962. Diabetes Mellitus: a "thrifty" genotype rendered detrimental by "progress"? *Am. J. Hum. Genet.*
10. Wright, S. 1931. Evolution in Mendelian populations. *Genetics* 16:97–159.
11. Wright, S. 1982. The shifting balance of evolution theory and macroevolution. *Annu. Rev. Genet.* 16:1–19.
12. Frankel, W. N., and N. J. Schork. 1996. Who's afraid of epistasis? *Nat. Genet.* 14:371–373.
13. Risch, N., and K. Merikangas. 1996. The future of genetic studies of complex human diseases. *Science* 272:1516–1517.
14. Speilman, R. C., R. E. McGinnis, and W. J. Ewens. 1993. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.* 52:506–516.
15. Spielman, R. S., and W. J. Ewens. 1996. The TDT and other family-based tests for linkage disequilibrium and association. *Am. J. Hum. Genet.* 59:983–989.
16. Schuler, G. D., M. S. Boguski, E. A. Stewart, L. D. Stein, and G. Gyapay. 1996. A gene map of the human genome. *Science* 274:540–546.
17. Collins, A., J. Frezal, J. Teague, and N. E. Morton. 1996. A metric map of humans: 23,500 loci in 850 bands. *Proc. Natl. Acad. Sci. U.S.A.* 93: 14771–14775.
18. Topper, J. M., J. Cai, D. Falb, and M. A. Gimbrone. 1996. Identification of vascular endothelial genes differentially responsive to fluid stimuli: cyclooxygenase-2, manganese superoxide dismutase, and endothelial cell nitric oxide synthase are selectively up-regulated by steady laminar shear stress. *Proc. Natl. Acad. U.S.A.* 93:10417–10422.
19. Eppig, J. T., and J. H. Nadeau. 1995. Comparative maps: the mammalian jigsaw puzzle. *Curr. Opin. Genet. Dev.* 5:708–716.
20. Smithies, O., and N. Maeda. 1995. Gene targeting approaches to complex genetic diseases: atherosclerosis and essential hypertension. *Proc. Natl. Acad. Sci. U.S.A.* 92:5266–5272.
21. Boguski, M. S. 1994. Bioinformatics. *Curr. Opin. Genet. Dev.* 4:383–388.
22. Fischman, J. 1996. Bioinformatics: Working the Web with a virtual lab and some java. *Science* 273:591–593.
23. Schork, N., and A. Chakravarti. 1996. A nonmathematical overview of modern gene mapping techniques applied to human diseases. *In* S. Mockrin, editor. Molecular Genetics and Gene Therapy of Cardiovascular Disease. New York, Marcel Dekker, Inc. 79–109.
24. Schork, N. J., and X. Xu. 1996. Sibpairs versus pedigrees: what are the advantages? *Diabetes Reviews* 5:1–7.
25. Schork, N. J. 1993. Extended multipoint identity-by-descent analysis of human quantative traits: efficiency, power, and modeling considerations. *Am. J. Hum. Genet.* 53:1306–1319.
26. Schork, N. J., S. P. Nath, K. Lindpaintner, and H. J. Jacob. 1996. Extensions of quantitative trait locus mapping in experimental organisms. *Hypertension* 28:1104–1111.
27. Kruglyak, L., and E. S. Lander. 1995. High-resolution genetic mapping of complex traits. *Am. J. Hum. Genet.* 56:1212–1223.
28. Weeks, D. E., and K. Lange. 1988. The affected pedigree member method of linkage analysis. *Am. J. Hum. Genet.* 42:315–326.
29. Cannings, C. A., and E. A. Thompson. 1977. Ascertainment in the sequential sampling of pedigrees. *Clin. Genet.* 12:208–212.
30. Yamagata, K., N. Oda, P. J. Kaisaki, S. Menzel, H. Furuta, M. Vaxillaire, L. Southam, R. D. Cox, G. M. Lathrop, V. V. Boriraj, X. Chen, N. J. Cox, Y. Oda, H. Yano, M. M. Le Beau, S. Yamada, H. Nishigori, J. Takeda, S. Fajans, A. T. Hattersley, N. Iwasaki, T. Hansen, O. Pedersen, K. S. Polonsky, R. C. Turner, G. Velho, J. C. Chevre, P. Froguel, and G. I. Bell. 1996. Mutations in the hepatocyte nuclear factor-1alpha gene in maturity-onset diabetes of the young. *Nature* 38:455–458.
31. Yamagata, K., H. Furuta, N. Oda, P. J. Kaisaki, S. Menzel, N. J. Cox, S. S. Fajans, S. Signorini, M. Stoffel, and G. I. Bell. 1996. Mutations in the hapatocyte nuclear factor-4alpha gene in maturity-onset diabetes of the young. *Nature* 38:458–460.
32. Jorde, L. B. 1995. Linkage disequilibrium as a gene-mapping tool. *Am. J. Hum. Genet.* 56:11–14.
33. Weir, B. S. 1990. Genetic Data Analysis. Sinauer Associates, Sunderland, MA.
34. Kruglyak, L., and E. S. Lander. 1995. Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Am. J. Hum. Genet.* 56:519–527.
35. Goldgar, D. E. 1990. Multipoint analysis of human quantitative variation. *Am. J. Hum. Genet.* 47:957–967.
36. Xu, S., and W. R. Atchley. 1995. A random model approach to interval mapping of quantitative trait loci. *Genetics* 141:1189–1197.
37. Te Meerman, G. J., M. A. van der Meulen, and L. A. Sandkuijl. 1995. Perspectives of identity by descent (IBD) mapping in founder populations. *Clin. Exp. Allergy* 25:97–102.
38. van der Meulen, M. A. 1996. Mutations and genetic disease: computational and population genetic approaches. *In* Medical Genetics. Groningen, Groningen, Netherlands. 152.
39. Houwen, R. H. J., S. Baharloo, K. Blankenship, P. Raeymaekers, J. Juyn, L. A. Sandkuijl, and N. B. Freimer. 1994. Genome screening by searching for shared segments: mapping a gene for benign recurrent intrahepatic cholestasis. *Nat. Genet.* 8:380–386.

40. Damerval, C., A. Maurice, J. M. Josse, and D. de Vienne. 1994. Quantitative trait loci underlying gene product variation: a novel perspective for analyzing regulation of genome expression. *Genetics* 137:289–301.
41. Brown, D. M., A. P. Provoost, M. J. Daly, E. S. Lander, and H. J. Jacob. 1996. Renal disease susceptibility and hypertension are under independent genetic control in the fawn-hooded rat. *Nat. Genet.* 12:44–51.
42. Schork, N. J., and A. B. Weder. 1996. The use of genetic information in large-scale clinical trials: applications to Alzheimer's disease. *Alzheimer Dis. Assoc. Disord.* 10:22–26.
43. Lewontin, R. C. 1997. Review of The Demon Haunted World: Science as a Candle in the Dark by Carl Sagan (Random House, New York; 1996). *New York Review of Books* XLIV:28–33.