



Published in final edited form as:

Perspect Biol Med. 2006 ; 49(4): 490–503.

Complex Adaptive System Models and the Genetic Analysis of Plasma HDL-Cholesterol Concentration

Thomas J. Rea, Christine M. Brown, and Charles F. Sing

Department of Human Genetics, University of Michigan, Ann Arbor.

Abstract

Despite remarkable advances in diagnosis and therapy, ischemic heart disease (IHD) remains a leading cause of morbidity and mortality in industrialized countries. Recent efforts to estimate the influence of genetic variation on IHD risk have focused on predicting individual plasma high-density lipoprotein cholesterol (HDL-C) concentration. Plasma HDL-C concentration (mg/dl), a quantitative risk factor for IHD, has a complex multifactorial etiology that involves the actions of many genes. Single gene variations may be necessary but are not individually sufficient to predict a statistically significant increase in risk of disease. The complexity of phenotype-genotype-environment relationships involved in determining plasma HDL-C concentration has challenged commonly held assumptions about genetic causation and has led to the question of which combination of variations, in which subset of genes, in which environmental strata of a particular population significantly improves our ability to predict high or low risk phenotypes. We document the limitations of inferences from genetic research based on commonly accepted biological models, consider how evidence for real-world dynamical interactions between HDL-C determinants challenges the simplifying assumptions implicit in traditional linear statistical genetic models, and conclude by considering research options for evaluating the utility of genetic information in predicting traits with complex etiologies.

We are moving from an Age of Reductionism to an Age of Emergence, a time when the search for ultimate causes of things shifts from the behavior of parts to the behavior of the collective.

—R. B. Laughlin (2005)

A WIDELY ACCEPTED MODEL of living organisms, which has its roots in 19th-century thought, emphasizes a deterministic and programmatic basis of biological structure and function (Mayr 1982, Woese 2004). This model has led to a research strategy whose goals have been to reduce the organism and its characteristic functional properties to presumed essential component agents, to model the relationships between these components and biological function by physical, chemical, or statistical methods, and to test the utility of models obtained for predicting, controlling, and understanding the contributions of the individual components to functional properties by comparing their impact on function in well-controlled experimental situations with their influence in intact, free-living organisms. Significant progress has been achieved in resolving the essential component agents to higher and higher resolution (e.g., metabolic pathways and products, enzymes, protein structures, encoding genes, polynucleotide sequences, nucleotide sequence variations). Modeling approaches have extended the limits of the mechanistic metaphor.

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

The authors wish to thank Ole Faergeman, Vinod Misra, Stuart Newman, and Günter Wagner for stimulating questions and discussions, and Lynn Illeck and Deborah Theodore for manuscript assistance. This work was supported by NIH grants HL072905, HL039107, and GM065509.

On the other hand, living organisms are better understood as complex adaptive systems characterized by multiple participating agents, hierarchical organization, extensive interactions among genetic and environmental effects, nonlinear responses to perturbation, temporal dynamics of structure and function, distributed control, redundancy, compensatory mechanisms, and emergent properties (Anderson 1972; Cowan, Pines, and Meltzer 1994; Kauffman 1995; Salthe 1993). It is increasingly clear from empirical evidence that our ability to predict the function of complex adaptive systems from the properties of single components is inversely related to the perceived “organized complexity” of the etiology of the emergent phenotypes of the system of interest, and that the determination of such emergent phenotypes involves properties of the whole system that cannot be understood simply in terms of the properties of individual component agents (Weaver 1948). The inadequacies of considering only the additive and independent contributions of single agents, exemplified by the enigma of emergent property prediction, were anticipated as a consequence of evolutionary interpretations of biological observations and analogies to “simple” nonlinear physical systems (Anderson 1972; Mayr 1982). Indeed, rigorous but controversial thought has posited that the properties of organisms cannot be reduced to physicochemical causality beyond a certain scale (Elsasser 1987), a reality that has faced the physical sciences and that has been a topic of great interest from the origins of genetics. If true, expectations for prediction of disease risk warrant qualification. The consequences for understanding causality are sobering.

The onus of medical genetics is to demonstrate whether genomic variation can supplement or supplant other predictors of disease risk, especially given assertions that most of the risk is environmentally determined (Reddy 2004). To what extent—if at all—can the prediction of risk be reduced to genes? In the specific case of ischemic heart disease (IHD) discussed here, this would entail evidence that genetic data can improve upon disease prediction that is afforded by gender, obesity, hypertension, smoking, diabetes, or dyslipidemia.

One fundamental shortcoming of studies of the role of genes in prediction of disease risk is the false proposition that knowing the genetic program is sufficient to describe both composition and spatial form that are characteristics of function when, in fact, it informs only the former (Goodwin 1994; Schrödinger 1945). This fact is pertinent to formulating biologically relevant research questions and designing studies to address those questions. Simply stated, detailed information about microscopic-scale molecular components has not provided an adequate basis for understanding the etiology of larger macroscopic features in physics, and certainly, it follows, not in biology (Laughlin and Pines 2000; Platt 1961). For biological systems, the question of whether genome variation is or is not within the hypothetical Elsasser limit of physicochemical causality that is beyond reduction has yet to be resolved. Many scholars query if, and how, we may achieve a stage of critical insight that will provide a synthetic approach to the prediction, control, and understanding of the emergence of biological form and function that includes both genetic and environmental determinants. Observations from three different perspectives suggest ways we might pursue this problem. The first couples evidence for the ubiquity of interactions between agents within and between multiple levels of biological organization with the contribution of those interactions to emergent phenotypic variations at the organismal level. The second recognizes that life-course changes in the relationships between causal agents, and between these agents and the emergent phenotype, are influenced by interactions between genotypic and environmental effects that vary across time and space (Sing, Stengård, and Kardia 2004). The third follows from theoretical studies of the architecture of complex adaptive systems designed to identify and characterize those properties of the interacting subsystems that define biologically meaningful functions. The confluence of these perspectives forms the conceptual basis for biological research options (Beattie 2004; Simon 1996; Sing, Stengård, and Kardia 2004).

Consistent with the goal of determining the value of genetic information in human health, the importance of identifying genetic predictors of inter-individual variation in risk of disease has become evident, and researchers have sought to determine how genetic variation may impact the properties of particular complex adaptive systems that make up the living organism. One such complex system of interacting agents is reverse cholesterol transport (RCT), in which plasma high-density lipoprotein cholesterol (HDL-C) concentration is considered to be a relevant biological property. Attention to the RCT system is motivated by a need to understand and control HDL-C, which is a major risk factor for IHD, a leading cause of morbidity and mortality in Western societies. Despite the widespread recognition that RCT has the properties of a complex adaptive system, characterization of the contribution of genetic variation to inter-individual variation in plasma HDL-C (or any other IHD risk factor) has proceeded consistently with the deterministic paradigm by seeking single gene–single variant predictors of phenotypic effects. Over 1,500 publications in the past 30 years document that the persistent devotion to explaining the genetic component of plasma HDL-C variation by single gene effects has resulted in inconsistent and irreproducible inferences. The biological reality that genes are only one part of the interacting system that determines a particular phenotype, and that the effect of a particular genetic variation depends on the effects of other genes and exposures to environmental agents, has been largely ignored (Newman 2003). Given these considerations, the question can no longer be which gene variation causes the phenotype of interest, but, rather, which combination of variations in which subset of genes in individuals exposed to a particular combination of environmental agents in a particular population contribute to the propensity for developing the phenotype.

In this article we use RCT as a prototypic example to address issues that geneticists face in demonstrating the value of genomic information in predicting human health, with the realization that a complete understanding of causation is unattainable (Goldstein 2005; Popper 1990). First, we identify and characterize RCT as a complex adaptive system. Second, we consider how evidence for real-world dynamical biological interactions between component agents challenges assumptions underlying current statistical approaches to evaluating the role of genomic information for predicting a phenotype that has a complex etiology. Finally, we examine research options for addressing the challenge of connecting inter-individual variation in HDL-C, an emergent property of RCT, to genetic variation.

RCT as a Complex Adaptive System

The world is made up of natural little bits and pieces that fit together in some natural way and bring to whole objects their own properties. The properties of the bits and pieces are properties they acquire in actually being parts of the wholes.

—R. Lewontin (1998)

Basic and clinical studies suggest that mechanisms mobilizing cholesterol from peripheral tissues can reduce the burden of atherosclerosis plaque in the arterial wall, maintain lumen patency, and reduce the risk of cardiovascular disease. The central role of plasma HDL-C concentration in this process derives from the properties of the HDL particle and the inverse correlation between HDL-C concentration and risk of IHD observed in many studies. In addition to its RCT function as a carrier of cholesterol to the liver, HDL-C plays multiple roles in other processes considered important in IHD pathogenesis, including platelet function, anti-inflammatory mechanisms, and endothelial cell responses. Interest in predicting disease risk using HDL-C and predicting inter-individual variation in HDL-C from genomic data derive from these fundamental observations.

The RCT system is described as a network of reactions inferred from the characterization of the key plasma lipid chemistries. The sequence of reactions, primary components of the system,

expression and regulation of gene products, and some of the environmental agents that influence functions of the system have been defined (Fielding and Fielding 1995; Tall 1992). These observations have fostered the construction of various mathematical models of lipid metabolism and have led to the genetic engineering of experimental animal models to better characterize the relationships between component agents and their contributions to the functional properties of the RCT system (Breslow 1996; Knoblauch et al. 2000). Data from all of these approaches indicate that plasma HDL-C concentration is influenced by a great number of factors.

Mammalian somatic cells require and synthesize, but are unable to catabolize, cholesterol. RCT removes cholesterol from these cells and transports it to the liver for redistribution or secretion, thereby providing a mechanism for cholesterol homeostasis. Close inspection of the RCT process reveals several fundamental features of a complex adaptive system. It is composed of many agents (apolipoproteins, receptor and membrane proteins, the genes that encode them, and various lipid classes), each of which may have multiple forms. There are specific organizational domains (protein-lipid interfaces, lipoprotein particle populations, and membranes) and nonlinear responses to interventions such as drug therapy. The relationships between the agents involved in RCT, and between these agents and the emergent plasma HDL-C phenotype, are dynamic at multiple time scales (short-term postprandial responsiveness to lipemia and age-dependent changes in the metabolism of lipids and lipoproteins). There are nonspecific effects of environmental agents (dietary, pharmacological, and pathological perturbations) that simultaneously affect many agents that influence the etiology of plasma HDL-C concentration. Finally, redundancy and compensatory processes operate at multiple steps in the RCT pathway (multiple lipases with overlapping activities and at least three hepatic cholesterol uptake mechanisms). The collective, coordinated activity of these features of RCT determines the emergence of the plasma HDL-C concentration, which in turn interacts with other pathways of lipid metabolism to influence membrane composition, biliary secretion of lipids, hormone production, and other cholesterol-dependent processes.

Characteristically, and similar to observations of other complex metabolic systems, analysis and engineering of individual components do not produce invariant predictions of the emergent functional properties of HDL-C. The ability of a particular agent to predict is dependent on context. Ignoring the complex adaptive system properties of RCT results in the inability of any particular component agent to accurately predict variation in plasma concentration of HDL-C across strata defined by time, genotype, gender, and exposure to environmental factors.

The potential role of genes in contributing to HDL-C variation is supported by the participation of an extensive number of gene products in regulating RCT (Ghazalpour et al. 2004). The genetic contribution to inter-individual variation of RCT components in humans using linkage and association studies has been quantified (reviewed in Sing and Boerwinkle 1987). Single gene-based linear models have been used to test for evidence of association between inter-individual variability in plasma HDL-C concentrations and variation in candidate genes. Such models are inherently limited in their ability to predict inter-individual variation of a trait that is a consequence of a complex etiology. The nature of these limitations is manifold. Single-locus strategies to predict inter-individual variation in human populations do not address the multifactorial interactions characteristic of complex biological traits, something that is especially critical when estimating and testing the role of gene-gene interaction, when the individual participating genes do not have separate, independent effects (Culverhouse et al. 2002). Furthermore, most studies ignore interactions of the particular gene of interest with other genetic and environmental agents that act dynamically through the time course of the development of the shape and function of the emergent phenotype. In summary, current models and statistical methodologies for predicting plasma HDL-C concentration are fundamentally

flawed because they do not consider extensive interactions between agents, and their dynamics at scales inherent to biological systems, that influence RCT.

Interactions and Dynamics Among Agents in the RCT System

“But wait,” the exasperated reader cries, “everyone nowadays knows that development is a matter of interaction. You’re beating a dead horse.”

I reply, “I would like nothing better than to stop beating him, but every time I think I am free of him he kicks me and does rude things to the intellectual and political environment. He seems to be a phantom horse with a thousand incarnations, and he gets more and more subtle each time around. ... What we need here, to switch metaphors in midstream, is the stake-in-the-heart move, and the heart is the notion that some influences are more equal than others.”

—S. Oyama (1985)

Most lexicons state that interaction consists of mutual or reciprocal action or influence. However, this obvious definition does not imply or reveal the mechanisms, causative power or meaning, range of effects, or specifications for measurement. Significantly, it also does not specify whether or not the individual agents that allegedly interact have actions or influences independent of their relationships with other agents. Despite these ambiguities, categories of interactions can be defined and prove to be informative with respect to system functions. In biological sciences, the languages of interaction have historically fallen into two general categories, biophysical and statistical. Among many important distinctions between the two is the realization that measurement of interaction by one does not necessarily imply measurability by the other. Failure to recognize this possibility frequently leads to inappropriate and misleading inferences about the interactive mechanisms of gene products from statistical studies of the relationships between genetic and phenotypic variation.

Interactions in RCT

It is reasonable to expect to find biophysical interactions at multiple component levels of RCT. Cholesterol flux through the RCT system is the result of an evolution-selected network of complementary agents; variation in plasma HDL-C is influenced by variation in many of them. Some of the genes involved (e.g., apolipoprotein genes) give rise to protein products that function in multiple roles such as lipoprotein assembly, receptor-ligand specificity, and enzyme activation, that are critical to RCT. Specific and dynamic biochemical interactions, commonly at membrane interfaces, are components of the physical basis for RCT. These interactions are dependent upon both the quality and quantity of many gene products and a broad spectrum of specific lipid moieties, including many of environmental origin, that are critical to protein and apolipoprotein conformation, orientation, and activity.

Biophysical and statistical studies of the relationships among and between the effects of genes, proteins, lipids, and exposures to environments document the ubiquity of the role of interactions in RCT. Sing, Stengård, and Kardia (2004) describe the types of interactions (genetic, biochemical, and environmental) known to contribute to structural or functional properties of RCT that are prototypical of the interactions among agents ubiquitous to lipid metabolism. Statistical analyses of gene-gene interactions in studies of model organisms anticipate the RCT situation: the majority of statistically significant contributions of genetic variations to phenotypic variation appear to be attributable to genotypes defined by combinations of DNA sequence variants that do not have separate and independent effects (Dolinski and Botstein 2005; Wolf et al. 2005). One practical consequence for statistical strategies is that failure to test for interaction among gene loci that do not have significant individual genotypic effects will likely result in false negative findings.

Complexity of Interactions Imposed by Dynamics

We use the term *dynamics* here in reference to changes or modification of relationships between component agents over time. Genetic variation influences these dynamical relationships at multiple time scales. DNA and protein-sequence variations may affect protein-protein interactions and transcription regulation on the microsecond to millisecond time scale, with resulting changes in RNA synthesis occurring within minutes to hours. Genotypic variation may change the dynamics of biochemical reactions and metabolic pathways that occur on the order of seconds to days. For genotypic effects on developmental or life course changes, time scales are on the order of years or decades. Finally, the impact of genotype changes on the dynamics of agent relationships transpires over the evolutionary time course of generations.

Human lifespan data provide compelling evidence that there is complexity of interactions imposed by dynamics on the RCT system. For example, in a study of 1,876 individuals from Rochester, Minnesota, ranging in age from five to 90, Zerba, Ferrell, and Sing (1996, 2000) demonstrated that relationships among a number of plasma RCT traits are dependent upon genotypes of the apolipoprotein E gene (*APOE*), gender, and age. Correlations among plasma concentrations of apolipoproteins B and E, TC, triglycerides, and HDL-C all varied in a statistically significant manner across the human lifespan and, importantly, the time course of the correlations differed according to gender and *APOE* genotype. This study suggests that hormonal and other cumulative environmental exposures (such as diet) impact HDL-C metabolism and RCT in a genotype-dependent manner. Consistent with this suggestion is the observation that elevated estrogen levels in premenopausal women and hormone replacement therapy in postmenopausal women are associated with increased plasma HDL-C concentration (Erberich et al. 2002). These results reinforce the proposition that context-dependent effects on relationships between components of RCT metabolism may account for a large fraction of inter-individual variation in plasma HDL-C not explained by invariant gene and environmental effects.

Given the documented role of the influence of dynamic interactions between participating agents on RCT, it is not surprising that the literature confirms that simple linear models and single agents poorly predict inter-individual variation in plasma HDL-C concentration. Advances in research design and analysis that employ realistic biological models must incorporate measures of network interactions and dynamics (Cork and Purugganan 2004). Conversely, ignorance of these dynamical interactions will ensure failure to ascribe context dependencies to specific disease-risk genotypes and disregard for Simpson's paradox (the idea that associations found in subpopulations may be different from the population considered as a whole; Simpson 1951) in interpreting statistical relationships. The irreproducibility often seen in association and linkage studies of phenotype-genotype relationships may well be attributable, in large part, to failure of the analysis to deal with the complexity inherent in the dynamics of interactions between participating agents.

Research Options

It is high time to back up assertion with argument.

—P. Kitcher (2001)

As a surviving and dominant research strategy from a prolific age of biological reductionism, studies that have sought to understand the genetic contribution to the etiology of plasma HDL-C concentration have made fundamental contributions to our knowledge of RCT. However, linear models, static hierarchical definitions of relationships between participating agents, and cross-sectional population study designs serve a paradigm that has lost its logical foundation with consequences analogous to the epidemiological “thought-tormented world” and reveals a reductionist paradox (Beattie 2004; Susser 1989). Because the RCT pathway is influenced

by a multitude of dynamic interactions, mathematical modeling cannot fully characterize the etiological connection between variation in HDL-C and genome variation. The challenge is to incorporate what can be understood and what can be measured into models for prediction of inter-individual variation in plasma HDL-C, so as to take full advantage of variation in the genetic substrate.

We begin this undertaking, perhaps unconventionally, by considering reactions to an orthologous problem identified by observations and data specifically derived from quantum mechanics experiments. The fundamental problem is well known as the quantum measurement paradox, and it may be communicated generally as the impossibility of defining the causal pathway from microscopic scale to macroscopic outcomes based upon quantum mechanical analyses. Leggett (2005) divided reactions of physicists to this paradox into three viewpoints, and we use his reaction-sorting schema as a starting point for the choice of research options we shall propose. One may view the status of information from the genome sequence as: (1) the necessary and complete basis (we intentionally choose not to use Leggett's use of the word *truth* here) of the biological world, at all levels, that sufficiently defines biological causation; (2) a necessary and complete basis of the biological world for reliable prediction purposes, but one that is insufficient for understanding causation; or (3) a necessary but incomplete basis of the biological world, one in which—at one or more levels between genotype and phenotype—unknown variables and processing rules mediate biological prediction and etiology.

The preceding sections of this paper are most plausible and consistent with the third of these possibilities. Occasionally, highly penetrant alleles may be sufficient for prediction (scenario 2): genome-sequence information is sometimes “sufficient” to predict rare HDL-C phenotypes to the extent that variations in phenotypes are statistically correlated with variations in the encoded sequence. However, the frequency of these mutations and scale of their effects have not been carefully analyzed in large populations. At population scales, such correlations are inevitably inadequate for unraveling the causal pathways from genotype to phenotype that involve a hierarchy of subsystems that are each open and dynamic and causally embedded in a number of interactions. Two complementary research strategies to address these problems may be considered: the first is designed to advance etiological understanding from detailed measurement, description, and analysis within specific biological levels of organization and the second is expected to enhance predictive utility across levels of biological organization that take advantage of knowledge about the agents and their etiological relationships in each of the participating subsystems. These two strategies may be used synergistically to advance our knowledge of intra-subsystem etiology (first strategy) and inter-subsystem correlations (second strategy) that measure the collective organization that takes form and effect across levels in the hierarchy that connect the small world with the macro world (Holland 1999; Morowitz 2002).

The first strategy, pursuing ever more detailed characterizations (physiology, metabolism, structure definition, information encoding, and variation) of intra-level agents for each of the multiple levels of biological organization, is a long-established experimental strategy that operates with the assumption that causation can be determined. Although this descriptive approach is essential, and compelling because of progress in characterizing intra-level behavior under controlled experimental conditions, it is, nonetheless, demonstrably insufficient to address the challenges of system-wide synthetic integration—in other words, it does not enable us to use information about variation in properties of a lower-level system in order to predict variation in higher-order, system-level properties. A suitable example arises from our nascent understanding of human genetic and epigenetic variation based on knowledge about variation in the nuclear genome sequence, DNA methylation, large-scale copy number polymorphisms, insertions/deletions, inversions, relative allele frequency and linkage disequilibrium patterns in diverse populations, and the contribution of the mitochondrial genome. The scale of possible

interactions among these variables within and across levels, and their connections with emergent phenotypic outcomes, is intractable (consistent with Elsasser 1987). The inferential barrier for this research strategy is characteristic of the problem of traversing the boundary between the statistical representation of the quantum mechanical world and the deterministic rules of the gravitational world in physics research (Laughlin 2005).

The problem of identifying variables and their combinatorial statistical relationships in prediction models that relate the outputs of one subsystem to the inputs of a second system is precisely one that the second complementary research strategy might address. The goal of this strategy is not to obtain etiological understanding within a hierarchical level per se, but rather to use the information about etiology from the subsystems in the selection of inter-level predictive measures. As such, this approach will entail identifying ways of reducing the high dimensionality (very large number of variable agents at each level)—for example, by extracting the variables from each level necessary for building predictive models and by testing hypotheses about relationships of variables between subsystems that may reveal the higher-order collective organizing principles. Current analytical strategies for accomplishing this goal are woefully inadequate. It is probable that such an approach will require a type of global genome metric that integrates multiple locus associations that are just beginning to emerge (Schaid et al. 2005). It will rely on methods of dimension reduction, as suggested by the sum stat test, the combinatorial partitioning method, and the tree scan method (Nelson et al. 2001; Templeton et al. 2004; Wille, Hoh, and Ott 2003). This selective reduction process will require the ability to identify context-dependent statistical effects in different subpopulations and must be capable of identifying combinatorial effects of predictor variables with or without separate, independent statistical effects. Concurrently, the theoretical basis for computational methods appropriate for such approaches needs substantial research attention. To refine this statistical strategy, it will be necessary to develop methods for selecting predictor variables within and between levels of the hierarchy between genome and phenotype, and for testing their contribution to predicting disease risk in a range of genetic and environmental backgrounds.

With respect to this strategy for prediction, very large human data sets representative of carefully defined populations will be required—both training sets for model building, and test sets for model validation. The case for these studies is well documented, and planning for national or international projects based on large-scale longitudinal designs is well under way (Check 2004; Collins 2004; Khoury 2004; Pembrey 2004). Such large population studies will underscore the challenges of incorporating heterogeneity of genetic variation and phenotype-genotype relationships within and between populations into statistical models and methods of analysis (Clark et al. 2005). These studies will require resources for the creation of regulatory-compliant infrastructures for recruitment and retention, sample handling, data acquisition, and statistical analyses, resources that few research groups will be able to generate (Kruglyak 2005). If this effort is successful, we fully expect it will sort out a subset of genomic variations that influence RCT and are candidates for predicting IHD risk, identify those variations that have utility for predicting IHD beyond traditional risk factors in specific genetic and environmental strata in particular populations, and reveal ever-increasing paradoxes and challenges to developing personalized medicine, diagnostic methods, and therapeutic interventions. Given the promises of the past three decades, investment in such large-scale genetic studies is essential.

A variety of contextual factors will assist our understanding of the genetic components of human diseases with complex multifactorial etiologies (Sing, Stengård, and Kardia 2004). We propose four. First, we suggest a broader discussion of the meaning, utility, and role of causation, prediction, and emergence in complex disease research and a reconnection to a natural philosophy that cultivated relationships between science and humanity in a more integrative era (Harpham 2006). Second, we propose a return to research intent on falsifying

hypotheses as a means of rigorous testing, rather than research devoted to proving hypotheses, which suppresses new ideas while perpetuating oversimplified or false predictive relationships. Third, we urge the constant monitoring of how experimental conditions and designs for the collection of population-based data we employ in our studies are representative of, and consistent with, the synthetic biological realities they purport to sample. Last, we highlight the need to encourage research communities that foster communication, inclusivity, and improved resource utilization that are better equipped to address questions relevant to phenotype-genotype relationships across numerous multifaceted contexts.

Summation

It is almost an intrinsic part of our concept of science that we never know enough. At all times one could almost say: we can explain it all, but understand only very little.

—E. Chargaff (1971)

The study of plasma HDL-C concentrations exemplifies the challenges faced in biology and medicine of utilizing the plethora of information that has become available as a consequence of the genome revolution. Because of the lack of a theory connecting the small world to the large world, recognized by generations of physicists but denied by the reductionist strategy that dominates biology, we should redefine the goal of “understanding” emergent phenotypes in medicine. Experimental reductionist science may be used to characterize the causal agents and the etiological relationships between them within subsystems, solely for the objective of generating hypothesized elements for consideration in building statistical models for predicting “emergent” phenotypic outcomes in free-living populations. However, only a small fraction of these many hypothesized agents (or relationships between agents) will prove to be useful in building a predictive model. Progress will be measured in terms of the simplicity and large-population applicability of these models for predicting complex phenotypes such as plasma HDL-C for individuals with differing genetic and environmental contexts that vary across time and space. Simultaneously sorting among large numbers of genetic variations and environmental strata, to find which combination predicts emergent measures of health in which subset of individuals at which time in the lifecycle, will be the challenge. Those with the imagination and creativity equal to this challenge will have the greatest opportunities for contributing to the prediction and understanding of human diseases that have a complex multifactorial etiology.

References

- Anderson PW. More is different. *Science* 1972;177:393–96. [PubMed: 17796623]
- Beattie, A. Figures in an epigenetic landscape: competing paradigms of complexity in epidemiology. 2004. <http://www.lancaster.ac.uk/ias/documents/complexity%20workshop/a%20b%20figures%20in%20an%20epigenetic%20lands>
- Breslow JL. Mouse models of atherosclerosis. *Science* 1996;272(5259):685–88. [PubMed: 8614828]
- Chargaff E. Preface to a grammar of biology. *Science* 1971;172:637–42. [PubMed: 4929532]
- Check E. Huge study of children aims to get the dirt on development. *Nature* 2004;432:425. [PubMed: 15565117]
- Clark AG, et al. Determinants of the success of whole genome association testing. *Genome Res* 2005;15(11):1463–67. [PubMed: 16251455]
- Collins FS. The case for a U.S. prospective cohort study of genes and environment. *Nature* 2004;429(6990):475–77. [PubMed: 15164074]
- Cork JM, Purugganan MD. The evolution of molecular genetic pathways and networks. *BioEssays* 2004;26(5):479–84. [PubMed: 15112228]
- Cowan, GA.; Pines, D.; Meltzer, D. *Complexity: Metaphors, models, and reality*. Boston: Addison Wesley Longman; 1994.

- Culverhouse R, et al. A perspective on epistasis: Limits of models displaying no main effect. *Am J Hum Genet* 2002;70(2):461–71. [PubMed: 11791213]
- Dolinski K, Botstein D. Changing perspectives in yeast research nearly a decade after the genome sequence. *Genome Res* 2005;15(12):1611–19. [PubMed: 16339358]
- Elsasser, W. Reflections on a theory of organisms. Frelighsburg, Quebec: Orbis Publishing; 1987.
- Erberich LC, et al. Hormone replacement therapy in postmenopausal women and its effects on plasma lipid levels. *Clin Chem Lab Med* 2002;40(5):446–51. [PubMed: 12113285]
- Fielding CJ, Fielding PE. Molecular physiology of reverse cholesterol transport. *J Lipid Res* 1995;36(2): 211–28. [PubMed: 7751809]
- Ghazalpour A, et al. The pathogenesis of atherosclerosis: Toward a biological network for atherosclerosis. *J Lipid Res* 2004;45(10):1793–1805. [PubMed: 15292376]
- Goldstein, R. Incompleteness: The proof and paradox of Kurt Gödel. New York: Norton; 2005.
- Goodwin, BC. How the leopard changed its spots: The evolution of complexity. New York: Scribners; 1994.
- Harpham G. Science and the theft of humanity. *Am Sci* 2006;94:296–98.
- Holland, JH. Emergence: From chaos to order. New York: Perseus; 1999.
- Kauffman, SA. At home in the universe: The search for the laws of self-organization and complexity. Oxford: Oxford Univ. Press; 1995.
- Khoury MJ. The case for a global human genome epidemiology initiative. *Nat Genet* 2004;36(10):1027–28. [PubMed: 15454932]
- Kitcher, P. Battling the undead: How (and how not) to resist genetic determinism. In: Singh, RS., et al., editors. Thinking about evolution: Historical, philosophical, and political perspectives. 2. New York: Cambridge Univ. Press; 2001. p. 396-414.
- Knoblauch H, et al. A pathway model of lipid metabolism to predict the effect of genetic variability on lipid levels. *J Mol Med* 2000;78(9):507–15. [PubMed: 11140376]
- Kruglyak L. Power tools for human genetics. *Nat Genet* 2005;37(12):1299–1300. [PubMed: 16314858]
- Laughlin, RB. A different universe: Reinventing physics from the bottom down. New York: Basic Books; 2005.
- Laughlin RB, Pines D. The theory of everything. *Proc Natl Acad Sci USA* 2000;97(1):28–31. [PubMed: 10618365]
- Leggett AJ. The quantum measurement problem. *Science* 2005;307:871–72. [PubMed: 15705838]
- Lewontin, R. Not all in the genes. In: Wolpert, L.; Richards, A., editors. Passionate minds. Oxford: Oxford Univ. Press; 1998. p. 103-12.
- Mayr, E. The growth of biological thought. Cambridge: Belknap Press; 1982.
- Morowitz, HJ. The emergence of everything: How the world became complex. Oxford: Oxford Univ. Press; 2002.
- Nelson MR, et al. A combinatorial partitioning method to identify multilocus genotypic partitions that predict quantitative trait variation. *Genome Res* 2001;11(3):458–70. [PubMed: 11230170]
- Newman SA. The fall and rise of systems biology. *GeneWatch* 2003;16(4):8–12.
- Oyama, S. Ontogeny of information: Developmental systems and evolution. Cambridge: Cambridge Univ. Press; 1985.
- Pembrey M. Genetic epidemiology: Some special contributions of birth cohorts. *Paed Perin Epidemiol* 2004;18(1):3–7.
- Platt JR. Properties of large molecules that go beyond the properties of their chemical subgroups. *J Theor Biol* 1961;1:342–58. [PubMed: 13736280]
- Popper, KR. A world of propensities. Bristol: Thoemmes; 1990.
- Reddy KS. Cardiovascular disease in non-western countries. *N Engl J Med* 2004;350(24):2438–40. [PubMed: 15190135]
- Salthe, SN. Development and evolution: Complexity and change in biology. Cambridge: MIT Press; 1993.
- Schaid DJ, et al. Nonparametric tests of association of multiple genes with human disease. *Am J Hum Genet* 2005;76(5):780–93. [PubMed: 15786018]

- Schrödinger, E. *What is life?*. Cambridge: Cambridge Univ. Press; 1945.
- Simon, HA. *The sciences of the artificial*. Cambridge: MIT Press; 1996.
- Simpson EH. The interpretation of interaction in contingency tables. *J Royal Stat Soc (Ser B)* 1951;13(2):238–41.
- Sing CF, Boerwinkle EA. Genetic architecture of inter-individual variability in apolipoprotein, lipoprotein and lipid phenotypes. *Ciba Found Symp* 1987;130:99–127. [PubMed: 3327665]
- Sing, CF.; Stengård, JH.; Kardia, SLR. Dynamic relationships between the genome and exposures to environments as causes of common human diseases. In: Simopoulos, AP.; Ordovas, JM., editors. *Nutrigenetics and nutrigenomics: World review of nutrition and dietetics*. Basel: Karger; 2004. p. 77-91.
- Susser M. Epidemiology today: “A thought-tormented world.”. *Int J Epidemiol* 1989;18(3):481–88. [PubMed: 2807649]
- Tall AR. Metabolic and genetic control of HDL cholesterol levels. *J Intern Med* 1992;231(6):661–68. [PubMed: 1619389]
- Templeton AR, et al. Tree scanning: A method for using haplotype trees in phenotype/genotype association studies. *Genetics* 2004;169(1):441–53. [PubMed: 15371364]
- Weaver W. Science and complexity. *Amer Sci* 1948;36:536–44.
- Wille A, Hoh J, Ott J. Sum statistics for the joint detection of multiple disease loci in case-control association studies with SNP markers. *Genet Epidemiol* 2003;25(4):350–59. [PubMed: 14639704]
- Woese CR. A new biology for a new century. *Microbiol Mol Bio Rev* 2004;68(2):173–86. [PubMed: 15187180]
- Wolf JB, et al. Epistatic pleiotropy and the genetic architecture of covariation within early and late-developing skull trait complexes in mice. *Genetics* 2005;171(2):683–94. [PubMed: 16020793]
- Zerba KE, Ferrell RE, Sing CF. Genotype-environment interaction: Apolipoprotein E (ApoE) gene effects and age as an index of time and spatial context in the human. *Genetics* 1996;143(1):463–78. [PubMed: 8722796]
- 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(5):466–75. [PubMed: 11140944]