# Traversing the biological complexity in the hierarchy between genome and CAD endpoints in the population at large

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. © Munksgaard, 1994

An emerging challenge facing those who are concerned about the efficacy of public health programs is to understand how information from the DNA revolution might be used to improve our ability to predict the initiation, progression and severity of a common disease having a complex multifactorial etiology. In the course of research to evaluate the role of information about DNA, combinations of genome types and environmental exposures that predispose to disease will be identified. Such information is expected to be useful in efforts to identify individuals and families at higher risk of disease and to predict their responses to a proposed therapy. This paper begins with a discussion of the features of a realistic biological model for the study of a common multifactorial disease. We present evidence for the complexity in the relationship between genome type variation and variation in risk of coronary artery disease (CAD) and review the preliminary results of our studies to determine whether information about genome type variation can improve our ability to predict the distribution of CAD among individuals in the population at large. Such studies make it apparent that new analytical strategies are necessary to deal with the plethora of genome type information available for the evaluation of risk of a common disease like CAD. This shift in the research paradigm will build upon new strategies to understand the organization of natural systems that are coming from outside the mainstream of genetic research.

**Charles F. Sing, Kim E. Zerba and Sharon L. Reilly**
Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI, USA

The dominant contemporary research paradigm for studying the genetics of human disease has its roots in the work of Garrod (Schull & Chakraborty 1979, Scriver & Childs 1992). This strategy seeks to identify and characterize the "inborn error" in a particular gene that is responsible for an individual's diseased state. It has been most successful for those monogenic diseases characterized by a direct one-to-one, Cartesian, mapping between allelic variation in a single "disease" gene and the Mendelian segregation of the disease phenotype among family members. For each in a long list of such diseases (McKusick 1990), an alteration in a single gene is necessary and sufficient to cause disease. These diseases and their causal mutations are rare in the population at large. The etiology of such single gene diseases, however, is not necessarily simple. Complexity arises most commonly when 1) different mutations in a disease gene may explain the same disease phenotype in different individuals, 2) mutations in different genes may be responsible for the Mendelian segregation of the disease in different families and 3) the genetic background and/or exposures to external environmental factors modify the severity of disease. Many extraordinary successes with the monogenic human diseases over the past 10 years have been realized as a consequence of combining molecular and physiological studies with the statistical tools of segregation and linkage analysis (Davies et al. 1983, Kan & Dozy 1978, Rommens et al. 1989). These successes have encouraged the selection of the

single gene model as the first choice for studies of the common human diseases having a complex multifactorial etiology.

Single gene models are not sufficient for studies of common multifactorial diseases, such as CAD, cancer, diabetes and the psychiatric disorders (Sing et al. 1992, Sing & Moll 1990, Sing & Reilly 1993, Weiss 1993). In every case, a particular manifestation of disease may aggregate in families but only in rare instances does it segregate according to Mendelian rules (Sing & Reilly 1993). In no case has the distribution of a common disease among individuals in the population at large been explained adequately by a single gene model. Disease is explained by rare allelic variations in a single gene in only a small fraction of those affected. In most cases, mutations in "susceptibility" genes contribute to small increases or decreases in risk of disease. A mutation in any one of these genes is not sufficient to cause disease.

Few molecular biologists recognize that the multifactorial diseases, which account for the major burden of public health costs in the human population (Baird et al. 1988), are a consequence of interactions of the effects of many genes with exposures to many environmental factors (Strohman 1993, Tauber & Sarkar 1992). For these common entities it is axiomatic that genome type by environment interaction effects determine risk of disease (Zerba & Sing 1993). This is documented by the observation that in the population at large many different combinations of genotypes and exposures to high risk environments are associated with a particular disease phenotype (Reilly et al. 1993a). Awareness of this fact has convinced a few medical researchers that the single gene, inborn error of metabolism paradigm that is appropriate for the study of phenotypes inherited in a Mendelian fashion, must be supplanted by a research strategy that accommodates the complex etiology of the common multifactorial diseases that aggregate but do not segregate (Gottlieb 1992, Sing et al. 1992, Sing & Reilly 1993).

The need for a shift in our working definition of the etiology of common multifactorial diseases may be clear to those of us working in the area, but the direction that should be taken is not. At this time no appropriate analytical strategy to tackle the genetic analysis of such diseases is available. Regardless, we believe that it is better to acknowledge the problem at hand and work toward developing the models and research strategies that are a realistic representation of nature than to redefine the questions of interest so that we can use well-established models and analytical tools that are not even a weak mimic of the etiology of disease.

In this paper we present the features of a realistic model for studying the biology of common multifactorial diseases, review our preliminary studies that address the three central questions defined by this model and discuss the implications of our findings for the design and interpretation of future genetic studies. We consider the study of the genetics of coronary artery disease (CAD) in the population at large to illustrate the issues.

## Modeling common multifactorial diseases

There are four features of CAD that must be considered in selecting a model for the joint distribution of causes and disease endpoints among individuals in the population at large (Sing & Moll 1990):

1) Many genes and many environmental factors influence the incidence of disease. Most genetic variations will be common, i.e. be polymorphic in the population under study, and be associated with small phenotypic deviations from the norm of the population.

2) Variations in different subsets of genes combine with different environmental exposures to predispose to CAD in different families (Sing & Reilly 1993). Most families with disease will be segregating for only a subset of the many genes involved in determining the distribution of risk of disease among individuals in the population at large.

3) The impact of a single- or multi-locus genotype on predisposition to CAD is conditioned by the genotype of other loci and the state of other risk factor traits, both internal and external to the individual. Only a fraction of those individuals with a particular common variation of a specific genetic or environmental cause will develop disease.

4) Quantitative variation in biochemical and physiological traits link discrete variations in genome type with interindividual variation in the probability of having, or acquiring, disease.

A visualization that illustrates important aspects of these features is presented in Fig. 1. We review elsewhere (Sing et al. 1992, Sing & Reilly 1993) the salient attributes of CAD that are shared by all complex adaptive systems (Salthe 1985, Simon 1962). These attributes are: 1) a hierarchial relationship between genome type variation and interindividual variation in the disease phenotype involving many agents organized into a complex network, 2) a coherency in the network of agents within and across levels in the hierarchy that has
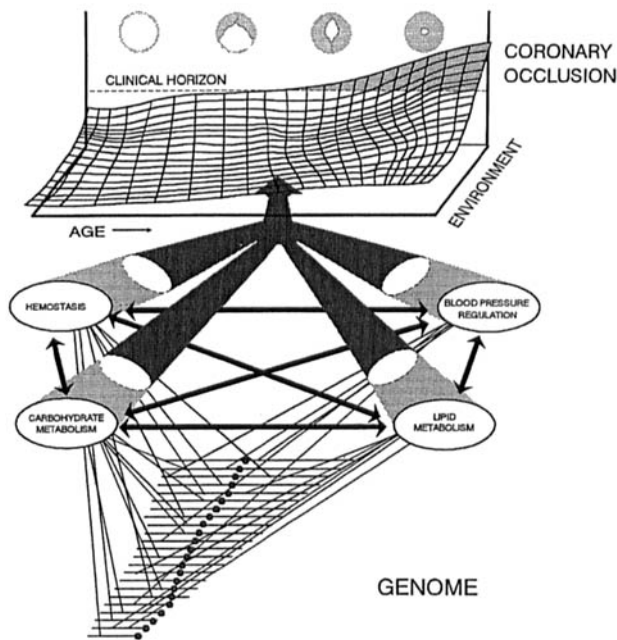
*Fig. 1.* A visual representation of the biological complexity of coronary artery disease.

been established by evolutionary and physiological constraints, 3) emergent properties, at each level in the hierarchy, that are a consequence of interactions among agents in the network and between these agents and the effects of exposures to environmental variations external to the network, and 4) dynamic changes in the relationships between agents, within and across levels in the hierarchy, over time. These attributes demand that the causes of CAD not be studied separately from the causes of health. The same coherent network of interrelated agents that define normal variability is also involved in defining the phenotypes of disease. Disease is a consequence of a disturbance of the levels of, and relationships between, these agents that is not compensated for by normal homeostasis (Childs 1977). Age of onset, rate of progression and degree of severity are emergent properties that can only be defined as an age-dependent probability function of the levels of, and relationships between, the agents in the pathways connecting the genome "in the basement" with clinical phenotypes at the upper level.

## Definition of fundamental research questions

We believe a research program that incorporates the characteristic features of complex adaptive systems can contribute to a more realistic understanding of the role of genetic factors in the etiology and prediction of common multifactorial disease like CAD. Specifically, research on three questions seems relevant at this time to help establish an

alternative to the Baconian-Cartesian, inborn errors of metabolism, paradigm. They are:

What are the coherent systems of biochemical and physiological agents at the intermediate level in the hierarchy that connect variation in genome type with variation in risk of disease? (Question 1),
What is the genetic architecture of interindividual variability in, and covariability between, these intermediate risk factor traits (i.e. coherency between genome and intermediate trait levels)? (Question 2), and
Does knowledge about genome type variation improve our ability to predict the emergent properties of onset, progression and severity of disease beyond that provided by the intermediate traits and measures of exposures to high risk environments? (Question 3)

The first two questions address the nature of the complexity of the mapping between genome type variation and variation in risk. The third question uses this information to address the public health question that motivates the consideration of measures of genome type variation. In the process of addressing these three questions, a more realistic model for the biology of multifactorial diseases like CAD will be revealed and the role of genetic information in prediction of endpoint phenotypes will necessarily take its appropriate position of importance among the plethora of agents involved in disease etiology.

## Evidence for complexity in the genetics of CAD

### Insights from studies addressing Question 1

There is a long list of agents that have been implicated in the etiology of CAD (Davignon et al. 1983, Ross 1993). Most intermediate risk factor traits involved in linking genome variation with variation in risk of disease have been measured in blood and plasma. Successes of the studies of lipid metabolism, hemostasis and blood pressure regulation and the inaccessibility of tissues that are involved have dampened incentive to develop new measures of relevant subsystems in the network of causation. Epidemiological studies that suggest only half of the cases of CAD are associated with the available risk factor traits have reinforced efforts to develop new methods of measurement in hopes of exposing the causes of the unexplained cases. Recent research on LDL oxidation (Esterbauer & Jürgens 1993), visualization of the disease (Hixson 1991, Stanford & Rumberger 1992) and the measurement of molecules associated with atheromatous tissue (Ross 1993) illustrate the po-

tential for further elucidation of the processes implied by Fig. 1. A sobering reality that is reinforced by all of these efforts is that there are tens, possibly hundreds, of intermediate biochemical, physiological and anatomical agents involved in determining risk of CAD. These intermediate traits are interrelated metabolically and cannot be considered independent risk factors for CAD. Epidemiological studies that evaluate the role of interindividual variation in coherency in this network of agents in determining risk of CAD could provide new insights.

### Insights from studies addressing Question 2

Studies to define the genetic architecture of CAD have focused on the genetic architecture of the quantitatively varying intermediate traits. Each of these many intermediate traits is likely to be influenced by many gene products. Two sorts of studies to identify and characterize the genes that influence interindividual variation in these traits have been carried out (Sing et al. 1988). The top-down approach attempts to identify the role of unmeasured allelic variations that have large effects on the level of one, or possibly two, of the traits identified in studies designed to answer question 1 (Moll 1993). The bottom-up approach relates measured allelic variations to phenotypic variation in one or more of these intermediate traits. Genes are selected for study because their products are involved in regulating the metabolism of the traits of interest.

Most investigations of measured genetic causes of variation in quantitative traits have focused on estimating the effects of single locus genotypes on the average level of one or more traits (Boerwinkle et al. 1987, Kaprio et al. 1991, Lusis 1988, Sing & Davignon 1985, Sing & Orr 1976, Xhignesse et al. 1991). However, the influence of such genetic variation on the mean level of a trait measures only one type of genetic influence on phenotypic variation. The ability of individuals with a particular genotype to adapt to perturbations from the environment is reflected in the intragenotypic phenotypic variances and covariances of the intermediate traits. The phenotypic variance of a trait may vary among genotypes because individuals with a particular genotype may be better buffered against exposures to environmental variations than individuals with a different genotype. Furthermore, many of the intermediate traits commonly associated with risk of CAD are interrelated metabolically. We must expect that genotypic variation determining a protein polymorphism will have effects on the levels of many traits. The resulting effects on coherency between traits will be expressed as intergenotypic differences in the covari-

ances between pairs of traits. Knowledge about the influences of genetic variation on intragenotypic variances and covariances between risk-factor traits and on the intragenotypic differences in relationships of these traits with concomitants, such as age and body size, may provide disease risk information that is not obtained by considering the trait means only (Murphy 1979, Murphy & Trojak 1986, Reilly et al. 1991, Reilly et al. 1992, Reilly et al. 1993c, Waddington 1957).

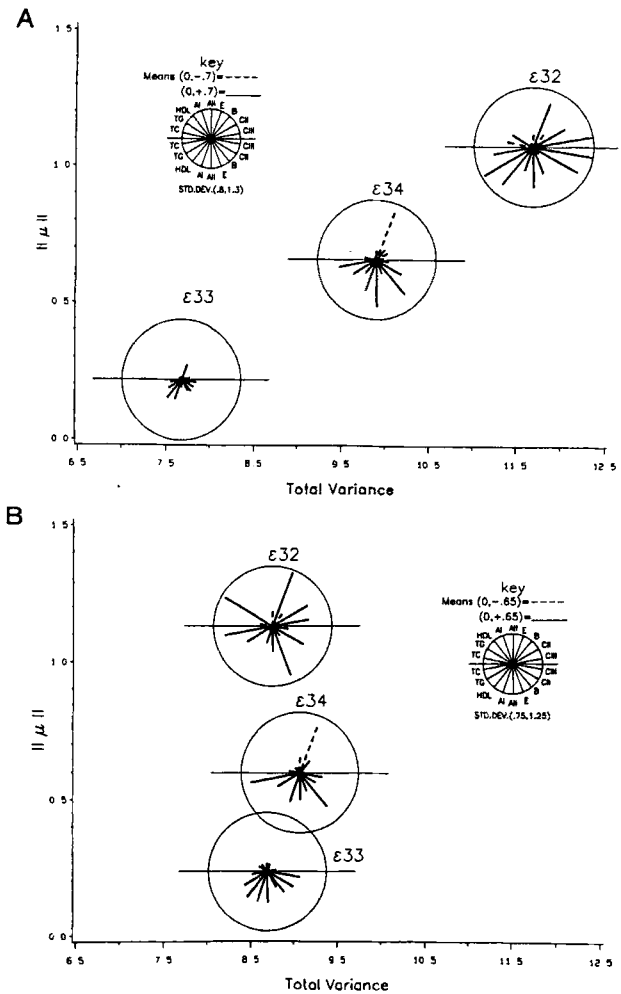We refer to the studies of allelic variation in the gene coding for the apolipoprotein (Apo) E



Fig. 2. Hyperspace of nine lipid and apolipoprotein traits in (A) females and (B) males, adapted from Reilly et al. (1991). To remove scale differences among traits, the gender-specific distribution of each trait was standardized to approximate a N(0,1) distribution. The genotype-specific mean and standard deviation of each standardized trait are represented by the length of the corresponding rays that are drawn from the intesection of the $\|\mu\|$ (the Euclidean distance) and Total Variance values for each genotype where $\|\mu\|_i = (\mu^2_{Total-Ci} + \mu^2_{lnTrigi} + \mu^2_{HDL-Ci} + \mu^2_{Apo\ Ali} + \mu^2_{Apo\ Alli} + \mu^2_{lnApo\ Ei} + \mu^2_{Apo\ Bi} + \mu^2_{Apo\ CIIi} + \mu^2_{lnApo\ CIIIi})^{1/2}$ and Total Variance $= \sigma^2_{Total-Ci} + \sigma^2_{lnTrigi} + \sigma^2_{HDL-Ci} + \sigma^2_{Apo\ Ali} + \sigma^2_{Apo\ Alli} + \sigma^2_{lnApo\ Ei} + \sigma^2_{Apo\ Bi} + \sigma^2_{Apo\ CIIi} + \sigma^2_{lnApo\ CIIIi}$ for $i = \varepsilon32, \varepsilon33,$ and $\varepsilon34$.

molecule to illustrate the nature of the complexity of the genetic architecture of the quantitative intermediate traits. Fig. 2 summarizes the complex pattern of influences of the three common Apo E genotypes on the means and intragenotypic variances of the univariate distributions of nine lipids and apolipoproteins in females and males reported by Reilly et al. (1991). To remove scale differences among traits, each trait was standardized to approximate a N(0,1) distribution. In this figure a multivariate measure of mean levels (i.e. Euclidean distance $||\mu||$) and trait variance (i.e. total variance adapted from Van Valen (1974)) are combined with each trait's contribution to these measures, depicted as rays on the upper and lower semi circles, respectively. The range of $||\mu||$ and the rank of the genotypes with respect to $||\mu||$ are approximately the same between genders, even though the contribution of each trait to $||\mu||$ is different across gender and genotype. With respect to the intragenotypic variances, both the range and the rank of total variance are different between genders. We see a corresponding difference in the contribution of each trait to the total variance of each genotype and gender. This figure illustrates that the gene coding for Apo E is both a level and variability gene with regard to multiple measures of lipid metabolism.

Fig. 3 presents an example from the work of Reilly et al. (1992) to explore the consequences of genotype-specific regression on concomitants for drawing inferences about the complex mapping function that must exist between the Apo E genotype and measures of lipid metabolism. These results were observed in males but not females. First we note that the marginal HDL-C means of the three common genotypes were not significantly different when variation in the waist-to-hip ratio
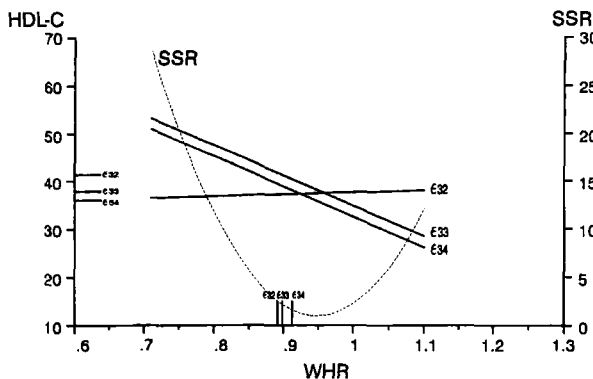
(WHR) was ignored. Likewise, the Apo E genotypes were not associated with significantly different mean WHR values when HDL-C values were ignored. However, as WHR varied the influence of the Apo E genotype on the expected value of HDL-C changed considerably. The ε32 genotype was associated with the lowest levels of HDL-C when WHR was below approximately 0.90 and was associated with the highest levels when WHR exceeded approximately 0.95. In the tails of the WHR distribution, we observed the greatest differences among genotypes in predicted HDL-C values while around the mean of the WHR distribution we observed the smallest differences among genotypes. This example illustrates that the component of phenotypic variance attributable to a genetic polymorphism may be dependent on the level of a second trait. In this case, the significant effect of the Apo E polymorphism on HDL-C levels was manifest only in the tails of the distribution of WHR; the marginal effects on HDL-C, when WHR was ignored, were trivial.

An added complexity that enters into the study of genetic architecture of CAD is the fact that the interactions between genome type and environments change over time. Different genes may play roles at different times throughout the life cycle. Different environments may be interacting with genome effects at different times. Few genetic studies are designed to investigate these possibilities. A recent report (Schwartz et al. 1993) that the contribution of genetic factors to interindividual variation in ambulatory blood pressure measurements varies throughout the day documents the reality of the problem. The contribution of genetic factors to determining interindividual differences was greatest during the part of the day when the subjects were active and virtually absent when subjects were in bed. It is patent that a realistic biological model must consider that the observed phenotype of a trait for an individual at a particular point in time, $P_{ijk}$, is the result of: 1) interactions between the effects of the entire constellation of genes that make up each individual's unique genome type and the effects of a unique suite of environmental experiences defined by time (and hence space) ($GE_{ij}$ in Fig. 4A (Dobzhansky 1960, Williams 1956, Zerba & Sing 1993)) and 2) chance events ($\varepsilon_{ijk}$ in Fig. 4B (Zerba & Sing 1993)). An individual i's functional phenotype ($GE_{ij}$) represents the deviation of his/her expected phenotype value at time j from the average of all functional phenotype values, $\mu$, that define the population of inference over all measurement times. Changes in an individual's functional phenotype value over time are a consequence of interactions of his/her genome type with environments indexed by time.



*Fig. 3.* Apo E genotype influence on HDL-C (mg/dl) regression on WHR in males, from Reilly et al. (1992). SSR=weighted sum of squared deviations, of the Apo E genotype-specific predicted values, from the average predicted value for a given level of WHR.
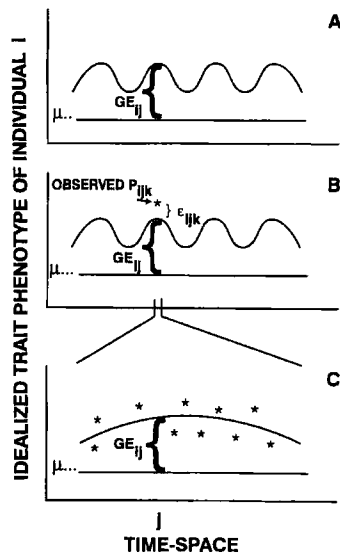
*Fig. 4.* From Zerba & Sing (1993). Relationships between the idealized observed phenotype of an individual, its functional phenotype and environmental perturbations. (A) The interaction of genome type with environmental perturbations determines a pattern of variation in the functional phenotype of an individual over time. The functional phenotype of a trait for an individual i at time j is represented as the deviation of the average pattern of its functional phenotype $(GE_{ij})$ from the average of the functional phenotype values $(\mu.)$ for all individuals that define the population of inference over all measurement times. The same population of individuals defines $\mu$ throughout the time period considered. (B) Chance events contribute to an individual's observed phenotype at any particular point in time. $P_{ijk}$ is the k th measurement of the trait phenotype of individual i at time j, $\mu_{..}$ is the population mean of the trait across all individuals, all environments at time j and all possible measures of i at time j, $GE_{ij}$ is the deviation of the phenotype from the population mean attributable to the interaction of the i th individual's genome type with all lasting environmental effects prior to time j and $\varepsilon_{ijk}$ is an independent and additive contribution of random variation in individual i's phenotype associated with time j. $\mu_{..}+GE_{ij}$ is the functional phenotype defined in panel A above. $\varepsilon_{ijk}$ includes random environmental perturbations with no lasting impact, i.e., environmental perturbations which do not become part of $GE_{ij}$ at time j+1 or later, and measurement error. (C) The effect of $GE_{ij}$ on the observed phenotype value at time j can be distinguished from that of $\varepsilon_{ijk}$, in theory, through repeated sampling of individual i across multiple times near j.

The etiology of an individual's functional phenotype at a particular time is a complex function of uncountable gene-gene and gene-environment interactions that influence physiological processes at all levels between genome and the traits of interest. Neither genes nor environments alone cause the particular deviation of an individual's functional phenotype from the mean of the population. The phenotype is a derivative of physiological and biochemical functions that are themselves emergent properties of interactions among genes and environments. The dynamic expression of the genome type is orchestrated by the environments experienced over time by the individual. The effects of $GE_{ij}$ and $\varepsilon_{ijk}$ are confounded at any particular point in time. Studies of variation in patterns of phenotypes as responses of genome types to experimentally defined changes in the environments are the key to understanding the etiology of genome type-environment interaction (Boerwinkle & Hallman 1993). Studies that sample at times before, during and after an environmental perturbation have the potential for separating the functional phenotype value, $GE_{ij}$, from the influence of chance events and for estimating the contribution of genome type variation to variation in response as well as variation in average level (Fig. 4C (Zerba & Sing 1993)).

The examples given above can only provide a small window into the true complexity of relationships between genome type variation and variation in intermediate risk factor traits. Only a fraction of the genes involved have been identified and described and hence most of the genome type variation is not accessible for the bottom-up studies. Furthermore, in addition to each gene being involved in determining variation in many traits, each of the intermediate traits involved in determining risk of disease will be influenced by variation in many genes (Kaprio et al. 1991, Kessling et al. 1992). These considerations of the genetic architecture of intermediate traits should send a clear message to those working on the genetics of common multifactorial diseases. The mapping across levels is many-to-many. Prediction of quantitative risk factor levels using genetic information is context dependent. Simple linear models of the relationships between genome type variation and interindividual variation in risk of disease destroy the potential for making meaningful inferences about how level, variability and covariability of risk factor traits contribute to determining risk of disease onset, progression and severity. The study of the distribution of genome type by environment interaction effects must take priority over the alternative of separate genetic or environmental studies.

## Information about DNA variation can improve the ability to predict CAD

Research on Questions 1 and 2 will identify genes involved in determining interindividual variation in intermediate traits that are predictors of CAD. Question 3 asks whether variation in these genes provides additional information about risk of disease that is not provided by the measures of intermediate traits. This would be the case if the measured genome type variation has pleiotropic effects on traits that have not been discovered, or cannot be measured *in vivo*, or if the ability of an inter-

mediate trait to predict disease was dependent on genome type. Therefore, research addressing Question 3 must also consider genes that may be involved in determining risk but were not identified as contributing to the genetic architecture of the measurable intermediate traits.

To illustrate the potential power of DNA information for predicting CAD we turn again to our work with the gene coding for Apo E (Reilly et al. 1993b). Fig. 5 gives the logistic regression analysis of selected predictors of parental CAD measured for a sample of healthy asymptomatic males ages 27 to 62 representative of the Rochester, MN, population. We begin our example by considering the relationship between the selected intermediate traits and probability of having a parent with definite CAD to address Question 1. Age and plasma Apo E level are evaluated in the four logistic regression models presented in the left panel of Fig. 5. Age and Apo E level each make a significant contribution to predicting parental CAD when considered separately, Models I and II, respectively. The contribution of these two traits is not additive, as indicated by the significant interaction effect revealed by comparing Models III and IV. The conclusion is that the ability to predict the probability of having a parent with CAD using the plasma Apo E level is not independent of the age of the asymptomatic adult child being measured. Question 2 has been addressed in a study of the Rochester population by Kaprio et al. (1991). Polymorphic variation in the gene coding for plasma Apo E explains only 10.64% (p= <0.001) of interindividual variation in plasma Apo E levels adjusted for variation in age, height, weight and smoking behavior.

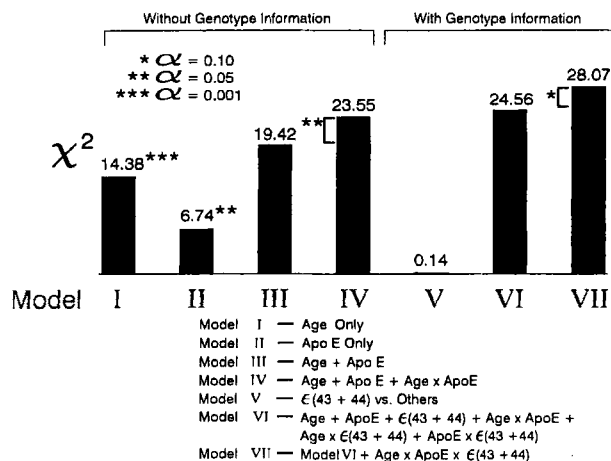Question 3 is addressed in the right panel of Fig.

5 which presents the analysis of the null hypothesis that adding information about allelic variation in the gene coding for plasma Apo E does not improve the ability to predict the probability of having a parent with CAD. Fitting restricted Apo E genotype effects only, Model V (the ε43+ε44 genotypes pooled compared to the other four genotypes pooled), did not result in a statistically significant result. When Apo E genotype information was added to the prediction model, and only first-order interactions with age and plasma Apo E level were considered, there was no evidence for improvement of prediction (Model VI compared to Model IV, Chi-Square=1.01). However, when the second-order interaction of Apo E genotype stratification with age and plasma apo E was considered in Model VII, there was statistically significant evidence, at the 0.10 level of probability, for improvement in prediction. This choice of significance level for the test of interaction is consistent with the arguments presented by Wahlsten (1990) and Zerba & Sing (1993). The surfaces of the predicted probabilities for the two groups of genotypes over the observed range of ages and Apo E levels are presented in Fig. 6.

Again, we believe the message is clear. Simple additive models for the prediction of CAD destroy the potential for making strong inferences. In most cases investigators would have stopped fitting
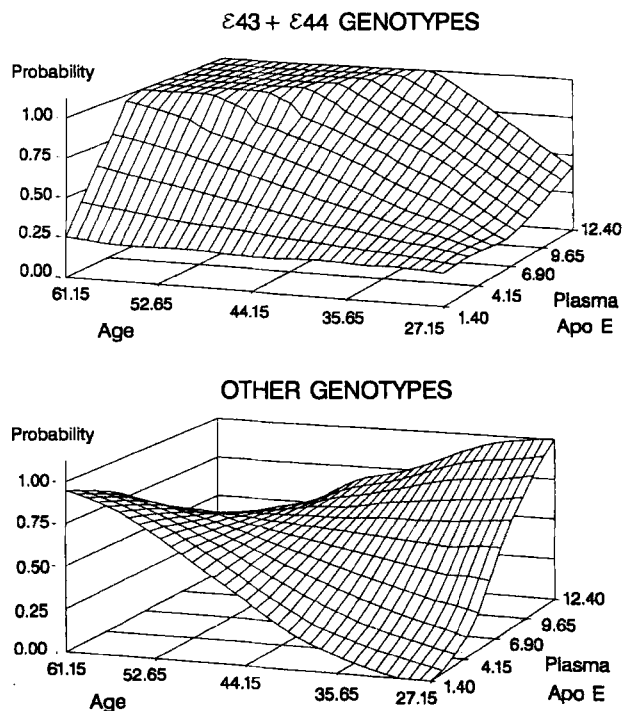


*Fig. 5.* Complexity in the prediction of parental CAD: Chi-square values (2[log $L_{model\ i}$ − log $L_{model\ with\ intercept\ only}$]) associated with logistic models of selected predictors with and without Apo E genotype information.



*Fig. 6.* Surfaces of the predicted probabilities of parental CAD illustrating the Apo E genotype by Age by plasma Apo E level interaction.

models with Model IV, missing Model VII and its information about the genetic contribution to prediction. The relatively simple consideration of all three levels in the hierarchy presented in the models given in Fig. 5 further emphasizes that prediction using intermediate quantitative traits can be context dependent. It is most likely that the stratifications of the population by Apo E genotype are associated with other (unknown) factors that influence the relationship between Apo E levels and age. It should be apparent to all but the most ardent reductionist that these genotype-dependent relationships between two risk factors and the probability of having CAD in a parent suggest a very difficult analytical problem when multiple genes, multiple measures of intermediate traits, multiple measures of exposure to the environment and the dynamics of change in environmental exposures over time are considered.

## Implications for future studies

We believe that the examples presented above document the complexity of the relationship between genome variation and variation in risk of CAD. The Baconian-Cartesian Mendelian method is not a valid model for the genetics of common multifactorial diseases. Genome types define the initial necessary conditions that place limits on the developing organism. Emergent disease properties associated with these genome types at a particular time in the life cycle are a function of interactions with exposures to past (since conception) and current environmental conditions. The simple single-gene models for diseases like CAD distort the multifactorial reality. A new research paradigm must recognize the primacy of interactions between genome types and exposures to environments. Neither genome types, nor environmental conditions are sufficient alone to determine onset, progression and severity of disease.

The utility of DNA information is context dependent. To accurately identify susceptible subgroups, screening for functional alterations in DNA will require auxiliary information about other genes and environmental factors. An understanding of potential effects of the interactions of an individual's initial biological conditions coded in the genome type with exposures to environments, and the epigenetic consequences of these interactions during the life cycle, will direct medicine to those changes in the environment that will minimize disease for the individual, the family and the population.

Modern molecular methods have revealed the involvement of many genetic, biochemical and physiological agents. Statistical association studies have documented the presence of coherence among these agents within and across levels. Genome type by environment studies are convincing evidence of the dominant influence of the dynamic aspects of disease etiology over the static effects of separate measures of the genome and the environment at a particular time. The causes of disease cannot be studied separately from the causes of health. It follows that there is a critical need for a shift in the research paradigm used in studying common multifactorial diseases. Models for complex adaptive systems (Waldrop 1992) and the methods of study that are being considered in computer science, economics, evolutionary biology and physics need to be explored. We must combine the knowledge gained in the past 20 years (information gathering phase) with more realistic biological models that are emerging to establish a new way of answering questions about balance, pattern and organization rather than continuing to only describe the agents that are involved.

## 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.

Boerwinkle E, Hallman DM. Genotype-by-environment interaction: It's a fact of life. In: Sing CF, Hanis CL, eds. Genetics of cellular, individual, family, and population variability. New York: Oxford University Press, 1993: 93–105.

Boerwinkle E, Visvikis S, Welsh D, Steinmetz MS, Hanash SM, Sing CF. The use of measured genotype information in the analysis of quantitative phenotypes in man. II. The role of the apolipoprotein E polymorphism in determining levels, variability and covariability of cholesterol, betalipoprotein and triglycerides in a sample of unrelated individuals. Am J Med Genet 1987: 27: 567–582.

Childs B. Persistent echos of the nature-nurture argument. Am J Hum Genet 1977: 29: 1–13.

Davies KE, Pearson PL, Harper PS, Murray JM, O'Brien T, Sarfarazi M, Williamson R. Linkage analysis of two cloned DNA sequences flanking the Duchenne muscular dystrophy locus on the short arm of the human X chromosome. Nucl Acids Res 1983: 11: 2303–2312.

Davignon J, DuFour R, Cantin M. Atherosclerosis and hypertension. In: Genest J, Kuchel O, Hamet P, Cantin M, eds. Hypertension, physiopathology and treatment, 2nd edn. New York: McGraw-Hill, 1983: 810–852.

Dobzhansky T. Individuality, gene recombination, and non-repeatability of evolution. Aust J Sci 1960: 23: 71–74.

Esterbauer H, Jürgens G. Mechanistic and genetic aspects of susceptibility of LDL to oxidation. Curr Opin Lipid 1993: 4: 114–124.

Gottlieb G. Individual development and evolution: the genesis of novel behavior. Oxford: Oxford University Press, 1992.

Hixson JE. Apolipoprotein E polymorphisms affect atherosclerosis in young males. Pathobiological Determinants of Atherosclerosis in Youth (PDAY) Research Group. Arterioscler Thromb 1991: 11: 1237–1244.

Kan YW, Dozy AM. Polymorphism of DNA sequence adjacent to the human β-globin structural gene: relation to sickle mutation. Proc Natl Acad Sci USA 1978: 75: 5631–5635.

Kaprio J, Ferrell RE, Kottke BA, Kamboh MI, Sing CF. Effects of polymorphisms in apolipoproteins E, A-IV, and H on quantitative traits related to risk for cardiovascular disease. Arterioscler Thromb 1991: 11: 1330–1348.

Kessling A, Ouellette S, Bouffard O, Chamberland A, Betard C, Selinger E, Xhignesse M, Lussier-Cacan S, Davignon J. Patterns of association between genetic variability in apolipoprotein (apo) B, apo AI-CIII-AIV, and cholesterol ester transfer protein gene regions and quantitative variation in lipid and lipoprotein traits: influence of gender and exogenous hormones. Am J Hum Genet 1992: 50: 92–106.

Lusis AJ. Genetic factors affecting blood lipoproteins: the candidate gene approach. J Lipid Res 1988: 29: 397–429.

McKusick VA. Mendelian inheritance in man. Catalogs of autosomal dominant, autosomal recessive and X-linked phenotype. Baltimore: The Johns Hopkins University Press, 1990.

Moll PP. Biometrical searches for new genes with effects on quantitative risk factors for atherosclerosis: problems and prospects. Curr Opin Lipid 1993: 4: 144–151.

Murphy EA. Quantitative genetics: a critique. Soc Biol 1979: 26: 126–141.

Murphy EA, Trojak JL. The genetics of quantifiable homeostasis. Am J Med Genet 1986: 24: 159–169.

Reilly SL, Congdon CB, Sing CF. Genetic algorithms as a method for studying risk factor combinations associated with common multifactorial diseases. Am J Hum Genet 1993a: submitted.

Reilly SL, Ferrell RE, Kottke BA, Kamboh MI, Sing CF. The gender-specific apolipoprotein E genotype influence on the distribution of lipids and apolipoproteins in the population of Rochester, MN. I. Pleiotropic effects on means and variances. Am J Hum Genet 1991: 49: 1155–1166.

Reilly SL, Ferrell RE, Kottke BA, Sing CF. The gender-specific apolipoprotein E genotype influence on the distribution of plasma lipids and apolipoproteins in the population of Rochester, Minnesota. II. Regression relationships with concomitants. Am J Hum Genet 1992: 51: 1311–1324.

Reilly SL, Ferrell RE, Sing CF. Prediction of family history of coronary artery disease (CAD) by apolipoprotein E genotype is context dependent. Am J Hum Genet 1993b: 53: 848A.

Reilly SL, Ferrell RE, Sing CF. The gender-specific apolipoprotein E genotype influence on the distribution of plasma lipids and apolipoproteins in the population of Rochester, Minnesota. III. Covariances and correlations. Am J Hum Genet 1993c: in press.

Rommens JM, Iannuzzi MC, Kerem B-S, Drumm ML, Melmer G, Dean M, Rozmahel R, Cole JL, Kennedy D, Hadaka N, Zsiga M, Buchwald M, Riordan JR, Tsui L-C, Collins FS. Identification of the cystic fibrosis gene: chromosome walking and jumping. Science 1989: 245: 1059–1065.

Ross R. The pathogenesis of atherosclerosis: a perspective for the 1990s. Nature 1993: 362: 801–809.

Salthe SN. Evolving hierarchical systems. New York: Columbia University Press, 1985.

Schull WJ, Chakraborty R (eds). Human genetics: A selection of insights. (Benchmark Papers in Genetics / 10). Stroudsburgh: Dowden, Hutchinson & Ross, Inc., 1979.

Schwartz GL, Turner ST, Sing CF. Contribution of genetic variation to interindividual differences in ambulatory blood pressure (ABP). American Heart Association 47th Annual Fall Conference and Scientific Sessions. September 28–October 1, 1993, San Francisco, CA.

Scriver CR, Childs B. Garrod's inborn factors of disease. New York: Oxford University Press, 1992.

Simon HA. The architecture of complexity. Proc Am Phil Soc 1962: 106: 467–482.

Sing CF, Boerwinkle E, Moll PP, Templeton AR. Characterization of genes affecting quantitative traits in humans. In: Weir BS, Eisen EJ, Goodman MM, Namkoong G, eds. Proceedings of the 2nd international conference on quantitative genetics. Sunderland, Sinauer, 1988: 250–269.

Sing CF, Davignon J. Role of the apolipoprotein E polymorphism in determining normal plasma lipid and lipoprotein variation. Am J Hum Genet 1985: 37: 268–285.

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, coronary artery disease. Ann Med 1992: 24: 539–547.

Sing CF, Moll PP. Genetics of atherosclerosis. Annu Rev Genet 1990: 24: 171–187.

Sing CF, Orr JD. Analysis of genetic and environmental sources of variation in serum cholesterol in Tecumseh, Michigan. III. Identification of genetic effects using 12 polymorphic genetic blood marker systems. Am J Hum Genet 1976: 28: 453–464.

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: Oxford University Press, 1993: 140–161.

Stanford W, Rumberger J. Ultrafast computed tomography in cardiac imaging: principles and practice. Mt. Kisco, NY: Futura, 1992.

Strohman RC. Ancient genomes, wise bodies, unhealthy people: limits of a genetic paradigm in biology and medicine. Perspect Biol Med 1993: in press.

Tauber AI, Sarkar S. The human genome project: has blind reductionism gone too far? Perspect Biol Med 1992: 35: 220–235.

Van Valen L. Multivariate structural statistics in natural history. J Theor Biol 1974: 45: 235–247.

Waddington CH. The strategy of genes. New York: Macmillan, 1957.

Wahlsten D. Insensitivity of the analysis of variance to heredity-environment interaction. Behavioral and Brain Sciences 1990: 13: 109–161.

Waldrop MM. Complexity: the emerging science at the edge of order and chaos. New York: Simon & Schuster, 1992.

Weiss KM. Genetic variation and human disease: principles and evolutionary approaches. New York: Cambridge University Press, 1993.

Williams RJ. Biochemical individuality: the basis of the genetotrophic concept. New York: John Wiley & Son, Inc., 1956.

Xhignesse M, Lussier-Cacan S, Sing CF, Kessling AM, Davignon J. Influences of common variants of apolipoprotein E on measures of lipid metabolism in a sample selected for health. Arterioscler Thromb 1991: 11: 1100–1110.

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 Lipid 1993: 4: 152–162.