Committee 2
Genetic Knowledge, Human Values
and Human Responbility

DRAFT--7/21/95
For Conference Distribution Only

LIMITATIONS OF LINKAGE STUDIES AND MULTIGENIC ORIGINS OF DISEASE

by

Richard C. Strohman
Professor Emeritus
Department of Cell and Molecular Biology
University of California
Berkeley, California, USA

The Twentieth International Conference on the Unity of the Sciences
Seoul, Korea August 21-26, 1995

I. Introduction

The Human Genome Project (HGP) is a superb technology capable of generating a high density map of perhaps all 100,000 human genes. As one of the leading architects of this projects has said, " ... this will also allow polygenic diseases and traits to be resolved into Mendelian components and thereby mapped" (1). However, while such a map is within the technical reach of the HGP it will not be possible to extract from a reading of the map information that, by itself, will be sufficient in the diagnosis or prediction of polygenic diseases and traits. It is not possible because the logic of polygenic diseases, as contrasted with monogenic diseases, is not to be found in the genome. Rather, that logic is encoded in a cellular epigenetic network of genes, gene products, and environmental signaling. This network is characterized by enormous complexity and informational redundancy from which is generated unexpected outcomes (phenotypes) driven by small changes in boundary conditions and environments (2,3). Continued emphasis on linear genetic logic, and discounting of epigenetic approaches presents serious problems for the future of biotechnology in the health/medical arena.

In what follows below I will try to convince you of the problems inherent in current directions within the Human Genome Project as well as in biotechnology in general. Analysis of these problems has been set out in prior publications which should be consulted for details and complete reference citations (2,3).

Informational redundancy is now well known at the level of the genome (4) and, in and of itself, constitutes a serious threat to the uniqueness equation underlying the HGP (5). This equation:

Unique Genes ----------> Unique Effects

defines saturation mutagenesis as the technological basis of efforts to associate specific genes with specific diseases (6). However, in the presence of redundant genes the uniqueness equation is severely compromised and is liable to serious error with equally serious consequences in clinical practice. In summary, the assumptions of the uniqueness equation are essential in the following areas: *Medical genetics,* which seeks isomorphic mapping of human diseases to Mendelian genes; *Molecular biology*, which seeks to identify unique genetically based mechanisms driving cellular processes; *Developmental biology*, which presupposes (1) genetic programs, (2) additivity of gene effects, and (3) the ability to map complex developmental stages to additive programmatic sequences in DNA.

These assumptions and presuppositions are being questioned at all levels of basic experimental biology (see Figure 2). New, and old, research findings are leading to a rejection of genetic determinism as the major paradigm of modern biology. Scientists are struggling with the apparent limits of genetic reductionism and are attempting to restructure genetic mechanisms within a larger context of "decision making" within cells and organisms (7). This context is sometimes seen to be ruled by principles of non linear dynamics and chaos theory, but for this presentation at least, it is described under the more traditional heading of epigenetic regulation. The main point here is that while aspects of basic biological research are moving away from linear genetic determinism and toward non linear complexity, applied biomedical technology remains unaffected by this fundamental shift in research emphasis. Thus, a rift is seen developing between basic and applied biology; a rift that could be dangerous to our public health.

Before we turn to a discussion of Figure 2 it is necessary to present a brief outline of epigenetic regulation.

II. Epigenetic aspects of cell regulation

What most biologists have assumed for years, but have never really formalized, is that every cell contains not one, but two, informational systems; the first is genetic and the second *epigenetic*.

The familiar genetic system of:

DNA----> RNA -----> Protein -----> Phenotype

is applicable to a small range of human phenotypes. In biomedicine it is restricted to monogenic diseases like Duchenne muscular dystrophy, hemophilia, and a host of other diseases (9). However, these diseases remain a small percentage of our disease load and account for less than 2% of the total (9,10 and below).

The epigenetic informational system in cells is depicted in Figure 1. The sets of (a) interactive genes (epistasis), (b) interactive genes and gene products, (epistasis, pleiotropy), and (c) interactive gene products and environment (polygenic and pleiotropic effects) define an unstable epigenetic system of great complexity inserted between unitary genetic elements and the final phenotype. In fact, this is a chaotic system with the major characteristic that, while a detailed map may be generated of all components, it will be impossible for mutational analysis alone to predict a unique outcome. As an example, a mutant gene may be redundant as seen in the case of an angiotensin II pathway in heart tissue (2). A single mutation in this pathway does not predict heart disease since the pathway contains many alternative genetic elements all of which perform identical functions. Alternatively, even without redundancy at the gene level, many

examples are found in cellular metabolic (epigenetic) networks where the network will simply be able to reset itself when given appropriate signals. The outcome of this resetting is often positive but unpredictable as found in adaptation of master runners to aerobic stress. These older men show epigenetic regulation of several key glycolytic and mitochondrial enzymes so that overall oxygen utilization is enhanced compared to younger men who achieve the same endpoint but through a different pathway (2). This response may even involve changes in gene expression since physical activity and electrical stimulation are known to repress and activate genes coding for isoenzymes in skeletal muscle cells (2). Thus, the context for patterns of gene expressions is found, not in the genome, but in interactive epigenetic networks.

Figure 1 here

An epigenetic system may be said to be chaotic in that, while it is impossible to predict which alternative pathway will be used, it will be possible to determine potential for adaptive change under precisely defined initial conditions(11,12). The system is thus a determinative chaotic system open to new approaches that combine linear genetic with non-linear complex system (epigenetic) analysis. Looked at in this way we may predict a new opportunity in biotechnology; viz., the definition of complex system parameters and specific environmental perturbations that elicit unique disease/health outcome(13).

We now look at specific instances where mainstream genetic determinism has found itself in conflict with new (and old) findings from basic research in genetics, and in other areas of molecular and cell biology.

Most of these conflicts may be resolved by recasting them in epigenetic terms.

Figure 2 here

III. Population biology conflicts with genetic determinism

Under the uniqueness equation described above equation the HGP has distilled a simplistic set of assumptions and goals:

a. Genes determine diseases

b. Genes determine aging

c. Genetic analysis provides diagnosis and therapy for disease and aging.

But fundamental rules governing population genetics stand in at least partial contradiction to the uniqueness equation and to the human genome project assumptions. This is a complex subject but essentially the unique relationship between genes and phenotypes is flawed because most complex phenotypes (including diseases) have no unique genetic basis. Rather the relationship between genome and phenome is characterized by great complexity involving interaction between many genes, gene products, and environmental signaling. This interaction may involve 10, 100, 1000 or more genes for any common disease like cancer or the heart diseases. We don't know exactly how many genes interact in these cases but as the number of interactive genes increases the effect of each gets smaller and more open to compensation (14). In addition, the interaction will be a function of an individual's natural history and environmental setting so that even in simplified cases where genetic connections may be traced genes will have different effects under different environments The basic message from population genetics is that a precipitating environment is

required to produce disease manifestation across the entire range of genetic variation (14,15). For problems like cardiovascular disease, most cancer, non insulin-dependent diabetes, most mental diseases, there is no evidence for single gene causality, and certainly none that would support the uniqueness equation fundamental to the human genome project.

Why is it that genetic diagnosis is predicted to fail in these cases? In brief, the argument is that the major statistical tool, analysis of variance, or ANOVA, as developed by Fisher is insensitive to heredity-environment interaction. This insensitivity is minimized in the agricultural breeding experiments for which ANOVA was designed because large sample size is normally the rule. In medical genetic studies (extended families) or in behavior genetics (twin studies), the sample sizes are small so that error is large in detecting lack of interaction between heredity and environment. As Wahlsten (14) points out, a newer statistical approach, multiple regression, is replacing ANOVA, but for the kinds of studies we are discussing the two procedures are essentially equivalent. Experts in agricultural genetics generally accept significant interaction between genes and environment and are extremely cautious in applying heritability coefficients or in assigning any significant numerical value to genetic cause when dealing with complex traits. Their position is that if gene effects are interactive (not additive) with environmental effects, it is incorrect to use ANOVA for assessing genetic contribution to a particular phenotype across a range of environments. Medical geneticists, however, using the same ANOVA but with significantly smaller sample size, not surprisingly do not find evidence for interaction and therefore assume that heredity and environment are additive. They then assign great significance to heritability coefficients and are confident these numbers describe

quantitatively the contribution of separate heredity and environment to any particular phenotype. We have a medical literature, then, that asserts with great confidence, but with serious theoretical reservations from sectors of population genetics, that this or that complex disease, while having an environmental component, also has a separate genetic component that can be discovered and utilized in pursuit of some hypothetical treatment strategy. It is beyond the scope of this review to enter this controversy fully; it is enough to state the minimum conclusion that medical genetics, with a linear view of gene-disease causality, finds itself in serious debate with a significant segment of its parent science of population genetics, which sees complex traits, including disease, as highly interactive and impossible to reduce to genetic elements alone (Figure 1).

IV. Disease natural history conflicts with genetic determinism

Here the issue is simple. There is a class of diseases for which the uniqueness equation is adequate. Diseases determined at fertilization, as Thomas Mckeown (15) has made clear, are based in genetic abnormalities of one kind or another. For simplicity we call these monogenic diseases. Examples are sickle cell anemia, cystic fibrosis, and Duchenne muscular dystrophy. There are literally thousands of these diseases but they occur within the human population at extremely low frequency and account for less than 2% of our total disease load. The basic message here is that 98% of the time our babies are well born with genetic constitutions capable of supporting a life span of over 100 years, an average life expectancy of about 85 years, and an old age relatively free of morbidity (16,17). In order for all this to happen the human genome needs to find itself in an

environment for which it has adequate representation...proper nutrition, housing, sanitation to name the obvious requirements.

There is a second level at which the biomedical paradigm is in conflict with actual disease distribution. It is assumed by the paradigm that complex traits like cancer and cardiovascular diseases have an important genetic component available not only to genetic analysis but also to genetic therapy. The reasoning is that these and other diseases attack people mostly in older (post-60) age groups. As such, the responsible genes would be beyond the reach of natural selection, which operates effectively at younger pre-reproductive ages. This being the case, it is argued that heart and cancer diseases are "old" entities, have always been with us (as have their genes), but show up significantly now because it is only recently that our population has aged sufficiently for them to become a problem. If this is true, then — so goes the argument — these are genetic diseases, pure and simple, and may be attacked as such (10,15).

But the natural history of our complex diseases shows that, in all probability, these are not genetic diseases, but are diseases of civilization. Of course, they have some genetic basis, but this basis is so broad as to be trivial with regard to providing precise genetic answers (Figure 1). Like all polygenic traits, genes are necessary but not sufficient. Evidence that diseases of civilization are not simply genetic includes the following (10). First, twin studies show extremely low concordance for most cancers and heart disease. Second, these same diseases show remarkable variation in identical populations over time and over geographical and migratory patterns. These variations disclose, for example, that diseases tend to be place-(environment) specific and that when people migrate, they tend to have those diseases common to their host population, not those which are

common to the genes they brought with them, i.e., not common to their native population. These variations are reversible. Finally, these diseases are rare in populations that have not come under Western habits. Natural history studies all indicate that our major premature killers are not genetic in any straightforward causal sense; they are diseases associated with changes in environment (10). That is the message from the past and present. That message, extended into the future, is that new diseases, their prevention and therapy, will also be associated with environmental change.

V. Evolutionary biology conflicts with genetic determinism

Most people, scientists included, are not aware of problems within evolutionary biology having to do with genetic mechanisms. These problems do not provide any weakening of the foundations supporting evolution; they do provide concern that we may have oversimplified the idea that evolution is to be explained by genetic mechanisms alone. Again, this is a complex area but we can state the following. Genetic change (adaptation) is seen as one end point of evolution, and change in genes (mutations) is seen as one element providing a basis for phenotypic variance that may be acted upon by natural selection. But gene changes alone will not and cannot explain evolution. For a complete explanation we require an understanding of how individual organisms generate their phenotypes in the presence (or absence) of gene changes in a variety of environmental settings (7,8). It is this understanding that we now do not have. Individual development is one missing link in our current theory of evolution; a link that is recognized, and one that the biological community is now struggling to supply and incorporate into a more complete picture of natural selection (18). We may provide a few examples of the conflict here.

First, there is the fact of absence of relationship between genetic and morphological complexity of species. Some closely related species cannot be seen by expert examination to be different (have different morphology) yet they show great variation in complexity at both genetic and protein sequence levels. Somehow organisms are able to take vastly different genomes and construct nearly identical phenomes, and this fact cannot presently be explained by a simple linear genetic paradigm. Second, humans and chimps have a very different morphology and yet humans do not differ genetically from chimps by more than 1-2%. Somehow we are able to construct very different organisms from very similar gnomes, and this is currently not explained by genetic theory (see below for further discussion).

VI. Developmental cell and molecular biology conflict with genetic determinism

There are many conflicts here. First, genetic determinism for complex traits has assumed the notion of "gene programs" to help explain the causal linkage between genes and phenotype. But this assumption has been found to be without experimental verification. There are no genetic programs (7,8). There are only genes that encode for proteins. Some of these genes, and their protein products, are extremely important. For example, some products bind to DNA and are involved in regulating development of morphological patterns (6). When they are mutated or missing the effects on a complex trait are profound. But these genes also exist in a cellular epigenetic context and depend on this context for the control of their expression in a species specific manner (2,7,8).

A second conflict comes with the realization of the fact of <u>informational redundancy</u> in organisms and especially within cells. The

uniqueness equation is undermined in the presence of a determination that more than one gene can bring about a unique effect. It is further weakened when we realize that not only is there informational redundancy at the gene level, but at the epigenetic level as well. There are many examples in the current literature of experimental biology testifying to the ability of the organism to get along without what were thought to be crucial genes. The organism, when a gene is missing, finds other genes or finds new ways (epigenetic controls) to interact vast numbers of remaining genes to produce the same or highly similar phenotypes (2). .

VII Conflict resolution

A major assumption of modern biomedical thinking is that genetic inheritance is the only inheritance. But biologists have always known this to be incomplete and we are now rediscovering the nature of our oversimplified paradigm (19). In modern developmental biology genetic the idea programs as a script for phenotype is being abandoned. There is no isomorphic mapping of complex phenotype to Mendelian factors (20), and the mechanism by which the organism elicits phenotypic variability from isogenic or near isogenic situations remains a profound mystery. The work on sibling species reveals that organisms may remain constant in morphology over millions of years even while they are enormously divergent at the level of DNA (21). Humans and chimps are shown to be nearly identical in genetic terms revealing that the organism is able to draw vastly different phenotypes from highly similar genotypes (22). Thus, profound questions are raised concerning the assumption of gene programming. First, that there appears to be less of a relationship between genetic and morphological complexity than we have thought. Second, if the

program is not in the genes, and organisms clearly are programmed, then where is the program? These and newer variations of complex, non-linear themes tend to be suppressed by our near monolithic commitment to molecular genetic mechanisms. This review has suggested that the molecular reductionist program to explain life has serious limits and that new epigenetic approaches to genetic regulation will be crucial. What might these new approaches be?

John Maddox, the editor of Nature, has written that modern biology, in concentrating on mechanism, has neglected theoretical approaches that might provide structure to the enormous data base accumulated by strictly molecular inquiry (23), and has suggested that such a conceptual structure might include a quantitative approach to dynamical cellular properties such as concentration fluxes of molecules which would control gene expression (24). Numerical characterization of these properties might then provide a basis for theory construction concerning regulation at levels higher than the gene. Theoretical physicist-cum- biologist, Walter Elsasser, has in fact laid out a basic description of a holistic theoretical biology in which dynamical properties play the role of higher order regulation (25). It is apparent that new research opportunities need to be created that will encourage work on these dynamical systems, and the theoretical structure hinted at by Maddox and Elsasser may lie, at least partially, in theory of complex systems .

One might begin the merger of genetic reductionism and epigenetic complexity with those areas where multigenic systems are know to be coordinated by higher order cellular responses to environmental conditions. Nobel laureate, Barbara McClintock, who described mobile genetic elements long before they were discovered by molecular biology, had always been

preoccupied with mechanisms that rapidly reorganize the genome. In one of her last reviews she wrote of the significance of responses of the genome to challenge. She ended that article as follows: "We know about the components of genomes .... We know nothing, however, about the how the cell senses danger and initiates responses to it that often are truly remarkable" (26).

At the cell level an interesting epigenetic approach to complex analysis of heart disease with multigenetic causality linked to interactive environments is the work of Sing and his group (13). At levels above the cell ... for complex physiological systems ... chaos theory builds on epigenetic thinking and already is providing new ways to think about complex systems. This is particularly true for cardiac function where sinus arrhythmia, long thought to be low level noise, or random fluctuation in heart rate, is now seen as high order chaos (11). Coupling of heart rate to brain function and thus to experience has long been appreciated as an observable patterned occurrence, but was mostly inexplicable through standard physiological experiment (27).. Chaos theory is able to provide a method of revealing generic pattern in what was thought to be random variation. Recognition of these patterns allows new insights into brain-heart physiology and may even allow prediction of sudden cardiac death among patients at risk (11).

It is here, at this interface between cell/organism and external world, that new research effort might be focused. Initial cellular responses are epigenetic in nature and involve selection of adaptive response from a bewildering array of molecular possibilities. At cellular and higher levels we expect that evolution has worked to select not just single genes but integrated behavior or generic patterns of response at all levels of

biological organization (12). These patterns can not be seen by linear analysis. It is at this level that theory of complex systems might prove useful. Generic patterns, with some ultimate basis in genomic reorganization, changes in gene expression, etc., would perhaps be open to theoretical structuring. Explanation and prediction of behavior leading to cancer or other cellular pathology, and to disease of heart and other complex organs would then not need to depend entirely on an apparently endless reductionistic analysis but could rely more on understanding rules of higher level organization; rules which, themselves, have been selected, and which control downstream mechanistic elements.

5/12/16

Selected References. For complete reference guide see reviews by R.C. Strohman cited below (#2,3).

1. Hood,L. 1992. In: The Code of Codes. D. J. Kevles and L. Hood. (Eds.) Harvard University Press, Cambridge, Massachusetts.

2. Strohman, R.C. 1993. Ancient Genomes, Wise Bodies, Unhealthy People: Limits of Genetic Thinking in Biology and Medicine. Perspectives in Biology & Medicine **37(1)**:112-145.

3. Strohman, R.C (1994) Epigenesis: The missing beat in biotechnology? Biotechnology **12(2)**: 156-164

4. Tautz, D. 1992. Redundancies, development and the flow of information. BioEssays **14**:263-266

5. Brenner, S., Dove, W., Hewrskowitz,I., and Thomas, R. 1990. Genes and development: molecular and logical themes. Genetics **126**: 479-486.

6. Wilkins, A.S. 1993. Genetic Analysis of Animal Development, 2nd Ed., Wiley-Liss Press, New York.

7. Nijhout, H. F. 1990. Metaphors and the Role of Genes in Development. *BioEssays* **12**:441-6.

8. Goodwin, B. C. 1985. What are the Causes of Morphogenesis? *BioEssays* **3**:32-6.

9. Weatherall, D. J. 1982. The new genetics and clinical practice. Nuffield Provincial Hospitals Trust, London.

10. McKeown, T. (1988). *The Origins of Human Disease.* Basil Blackwell, Inc., New York.

11. Skinner, J.E., Molnar,M., Vybiral,T., and Mitra,M. 1992. Application of chaos theory to biology and medicine. Integ. Physiol. & Behav. Sci. **27**:39-53

12. Kaufmann, S. 1993. The Origins of Order. Oxford University Press, New York.

13. Sing, C. F. and Reilley, S.L. 1993. Genetics of common diseases that aggregate but do not segregate in families. pp140-161 In: Sing, C.F.,

3/2-15

Hanis,C.L., (Eds.) Genetics of cellular, individual, family and population variability. Oxford Univ. Press, New York.

14. Wahlsten, D. 1990. Insensitivity of the analysis of variance to heredity-environment interaction. Behav. and Brain Sci. **13**:109-161.

15. Mckeown, T. 1979. The Role of Medicine: Dream, Mirage or Nemesis? Princeton University Press.Princeton, New Jersey.

16. Fries, J. F. and Crapo, L. M. (1981). *Vitality and Aging: Implications of the Rectangular Curve.* W. H. Freeman & Co., New York.

17. Tsai, S. P., Lee, E. S., and Hardy, R. J. 1978. The effect of reduction in Leading causes of death: potential gains in life expectancy. *AJPH* **68**:966-71.

18. Gottlieb, G. (1992). *Individual Development and Evolution: The Genesis of Novel Behavior.* Oxford University Press, Oxford.

19. Sapp, J. 1987) Beyond the Gene: Cytoplasmic Inheritance and the Struggle for Authority in Genetics. Oxford University Press, Oxford.

20. Stent, G. 1981. Strength and weakness of the genetic approach to the development of the nervous system. Ann. Rev. Neurosci. **4**:163-194.

21 Nanny, D.L. 1982. Genes and phenes in Tetrahymena. BioScience. **32(10)**:783-788.

22.. Wilson, A. C., Carlson, S. S., and White, T. J. 1977. Biochemical Evolution. Ann. Rev. Biochem. **46**:573-639.

23. Maddox,J. 1992. Is molecular biology yet a science? Nature **355**: 201

24. Maddox,J. 1992. Finding wood among the trees. Nature **333**:11

25. Elsasser, W 1987. Reflections on the Theory of Organisms, Orbis Publishing, Quebec.

26. McClintock,B. 1984. The significance of responses of the genome to challenge. Science **226**: 792-801.

27. Bond,W.C., Bohs,C., Ebey,J., & Wolf,S. 1973. Rhythmic heart rate variability (sinus arrhythmia) related to stages of sleep. Conditional Reflex 8(2):98-107.

Figure 2. Areas confounding genetic determinism in biomedicine.

Area	Confounding elements
A. Population Biology	Genetic effects are interactive and not additive
B. Disease natural history	Most common diseases are not inherited as gene mutations
C. Evolutionary Biology	Often no relationship between genetic and morphological complexity
D. Developmental Biology	There are no genetic programs
E. Molecular and Cell Biology	Informational redundancy and complex non linear organization inconsistent with linear genetics