FY03-95ER62116

0571

DOE/ER/62116--T1

Univ. of Houston

GENETIC SECRETS: PROTECTING PRIVACY AND CONFIDENTIALITY IN THE GENETIC ERA

Mark A. Rothstein, editor

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

MASTER

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible electronic image products. Images are produced from the best available original document.

GENETIC SECRETS: PROTECTING PRIVACY AND CONFIDENTIALITY IN THE GENETIC ERA

TABLE OF CONTENTS

PREFACE

The United States Department of Energy sponsored a highly successful workshop on Medical Information and the Right to Privacy at the National Academy of Sciences in Washington, D.C. on June 9-10, 1994. The idea to produce a volume exploring the full range of issues related to genetic privacy arose from that meeting. I was pleased to accept the Department of Energy's invitation to organize, solicit, and edit the manuscripts contained in this volume. Any opinions, findings, conclusions, or recommendations expressed in the book are solely those of the authors and do not necessarily reflect the views of the Department of Energy.

Several individuals were instrumental in compiling this work. Dan Drell and John Peeters of the Department of Energy gave me their unqualified support, as well as the independence to pursue my vision of the structure and content of the book. I am indebted to my chapter authors who permitted me to intrude into their busy lives to produce a work for me. They also were willing to revise their work several times to integrate the book chapters more closely. Several of the authors also reviewed drafts of the concluding chapter and offered valuable criticism.

At the Health Law and Policy Institute at the University of Houston, I am indebted to Cathy Rupf, who coordinated the publisher's and authors' agreements, and to Diana Huezo, who processed all the manuscripts. Harriet Richman, Faculty Services Librarian at the University of Houston Law Library, supplied essential reference support. Laura F. Rothstein not only authored

an excellent chapter on Genetics and Schools, but supplied much-needed encouragement in marshalling the talents of thirty-two colleagues.

Mark A. Rothstein

October 1996

Foreword

Arthur C. Upton

Few developments are likely to affect human beings more profoundly in the long run than the discoveries resulting from advances in modern genetics. The increasingly powerful diagnostic, predictive, and life-enhancing tools generated by molecular genetics and biotechnology have already begun to revolutionize medicine, science, agriculture, animal husbandry, and a growing number of industries.[1]

Exemplifying the power of the new technologies are their uses to: (1) identify the specific strains and sources of microorganisms responsible for certain outbreaks of tuberculosis[2] and Legionnaire's disease;[3] (2) implicate the human papilloma virus in causation of the majority of cancers of the uterine cervix;[4] (3) elucidate many other aspects of carcinogenesis;[5] (4) clarify the causal mechanisms of certain allergic reactions;[6] and give rise to increasing numbers of new and improved varieties of disease-and pest-resistant animal and plant species.[7]

Although the developments in genetic technology promise to provide many additional benefits to mankind in the years to come, their application to genetic screening poses ethical, social, and legal questions, many of which are rooted in issues of privacy and confidentiality. Still to be resolved, for example, is the extent to which the highly personal information contained in

one's genome differs in kind from other medical and legal information and, consequently, deserves greater protection against disclosure to one's employer, health insurer, family members, or others.[8]

Concerns about the disclosure of genetic information are prompted in large part by the fear that it could stigmatize the affected person and also, perhaps, even members of his or her family, causing such persons to be barred from employment, denied insurance, or subjected to other forms of discrimination.[9] Such concerns are heightened, moreover, by the fact that protection of the confidentiality of genetic information is being rendered increasingly difficult by the computerization and electronic transfer of medical records, coupled with the rapid growth of managed care and other sweeping changes in the organization of the health care delivery system.[10]

Also complicating the issue is the tension that exists under certain circumstances between the desire to respect the confidentiality of genetic information and the competing need and/or responsibility to share the information: e.g., (1) a parent who possesses a disease-causing gene may be under the moral obligation to share the information with his or her child if the health of the child would otherwise be jeopardized;[11] (2) newborn infants in most states are required by law to undergo genetic screening for phenylketonurea;[12] (3) members of the military are required to contribute specimens of their DNA to a central armed forces repository, in order to facilitate their

identification if killed in the line of duty;[13] (4) criminal offenders are required in many states to contribute DNA to databases maintained for forensic purposes by law enforcement agencies;[14] and (5) persons in all walks of life are being called upon increasingly to contribute to DNA data banks for research purposes.[15] While most DNA banks and DNA databases are generally acknowledged to serve important and beneficial purposes, the adequacy of existing safeguards for protecting the confidentiality of the genetic information they contain is not without question.[16]

Also subject to question are the circumstances under which genetic information should, or should not, be disclosed to the affected individual himself.[17] For example, should a person who is found on genetic testing to carry a gene that may predispose him or her to a disease of uncertain likelihood, for which no methods of treatment or prevention are known, be told of the condition, even if the disclosure under such circumstances might possibly do the person more harm than good? Also, by extension, if the same gene might pose a risk to other members of the person's family, who also happened to be carriers of the same mutation, should they too be notified?[18]

The ethical, practical, and legal ramifications of these and related questions -- which are at the forefront of contemporary medicine and medical research -- are explored in depth in the chapters to follow. The broad range of topics dealt with in these chapters includes: the privacy and confidentiality of

genetic information, considered from an ethical standpoint and also in the framework of the patient-physician relationship, public health, the family, and society at large; the challenges to privacy and confidentiality that may be projected to result from the emerging genetic technologies and from the application of such technologies to exposure surveillance, population screening, and forensic problems; the role of informed consent in protecting the confidentiality of genetic information in the clinical setting, including the issues surrounding the right to know, and/or not to know; the potential uses of genetic information by third parties, including employers, insurers, and schools; the implications of changes in the health care delivery system for privacy and confidentiality; relevant national and international developments in public policies, professional standards, and laws; recommendations for addressing problems in each of these subjects areas; and the identification of research needs. The chapters that follow address the privacy and confidentiality of genetic information of all types, considering the full range of their social, ethical, and legal ramifications.

1. Philip Kitcher. The Lives to Come: The Genetic Revolution and Human Possibilities. New York: Simon & Schuster, 1996; U.S. Congress, Office of Technology Assessment. New Developments in Biotechnology: Patenting Life---Special Report. Washington, D.C.: U.S. Government Printing Office, 1989.

2. Agnes Genewein, et al. "Molecular Approach to Identifying Route of Transmission of Tuberculosis in the Community." Lancet 342 (1993): 841-4.

3. W. Gary Hlady, et al. "Outbreak of Legionnaire's Disease Linked to a Decorative Fountain by Molecular Epidemiology." American Journal of Epidemiology 138 (1993): 555-62.

4. Mark Schiffman, et al. "Epidemiologic Evidence Showing that Human Papillomavirus Causes Most Cervical Intraepithelial Neoplasia." Journal of the National Cancer Institute 85 (1993): 958-64.

5. John Mendelsohn, et al., eds. The Molecular Basis of Cancer. Philadelphia: W.B. Saunders, 1995; I. Bernard Weinstein. "The Contribution of Molecular Biology to Cancer Epidemiology." Annals of the New York Academy of Sciences 768 (1995): 30-40.

6. Delores Graham and Hillel Koren. "Biomarkers of Inflammation in Ozone-Exposed Humans." American Review of Respiratory Disease 142 (1990): 152-6.

7. Kitcher, op. cit.

8. Institute of Medicine, Committee on Assessing Genetic Risks: Assessing Genetic Risks: Implications for Health and Social

Policy, Lori Andrews, et al., eds. Washington, D.C.: National Academy Press, 1994.

9. Arthur L. Frank. "Scientific and Ethical Aspects of Human Monitoring." Environmental Health Perspectives 104 (Suppl.3) (1996): 659-62; Harry Ostrer, et al. "Insurance and Genetic Testing: Where Are We Now?" American Journal of Human Genetics 52 (1993): 565-77; Walter C. Zimmerli. "Who Has the Right to Know the Genetic Constitution of a Particular Person?" Human Genetic Information: Science, Law, and Ethics, Chadwick, et al., eds. Chichester and New York: John Wiley & Sons, 1990.

10. Committee on Regional Health Data Networks, Institute of Medicine. Health Data in the Information Age: Use, Disclosure, and Privacy. Molla Donaldson and Kathleen Lohr, eds. Washington, D.C.: U.S. Government Printing Office, 1994; John K. Iglehart. "Physicians and the Growth of Managed Care." New England Journal of Medicine 331 (1994): 1167-8; Institute of Medicine, National Academy of Sciences. Health Data in the Information Age. Washington, D.C.: National Academy Press, 1994.

11. Sonia M. Suter. "Whose Genes Are These Anyway? Familial Conflicts Over Access to Genetic Information." Michigan Law Review 91 (1993): 1854-908.

12. Jean McEwen and Philip R. Reilly. "Stored Guthrie Cards as DNA 'Banks.'" American Journal of Human Genetics 55 (1994): 196-200.

13. Jean E. McEwen. "DNA Data Banks." Chapter 15 in this volume.

14. Jean E. McEwen. "Forensic DNA Data Banking by State Crime Laboratories." American Journal of Human Genetics 56 (1995): 1487-92.

15. McEwen (ch. 15), op. cit.

16. Barry Scheck. "DNA Data Banking: A Cautionary Tale." American Journal of Human Genetics 54 (1994): 931-3.

17. Committee on Assessing Genetic Risks, op. cit.

18. Zimmerli, op. cit.

Chapter 1

Genes, Genomes and Society

Leroy Hood and Lee Rowen

Deciphering human heredity will allow one to glimpse into the innermost workings of ourselves. Two pioneering scientific endeavors laid the framework for this venture, perhaps the most far-reaching scientific exploration ever undertaken. In 1953, Jim Watson and Francis Crick elucidated the structure of DNA, the informational molecule of human heredity.[1] From this work came the fundamental insight that DNA employs a digital code, similar to that used by computers, except that a four-letter language, G, C, A and T is employed rather than the two-letter language of computers, O and 1. Thirty-seven years later, in 1990, the Human Genome Project was initiated, a 15-year program to decipher the human DNA digital code by mapping and sequencing the 23 pairs of human chromosomes that reside in the nucleus of every human cell (Figure 1). These chromosomes contain the DNA code that directs the marvelous process of human development wherein each of us goes from one cell (the fertilized egg) at conception to 10^{14} cells as an adult.

Three Types of Biological Information

There are three types of biological information that function in living organisms. The first type is the digital or linear information of the DNA backbone of our chromosomes (Figure 2). Its unit of information is the gene; it is estimated that

human chromosomes contain perhaps 100,000 units of information or genes. Genes are expressed in a differential manner; that is, in different cells (e.g. muscle and brain) different combinations of the genes are expressed and this leads to the different appearances and behaviors or phenotypes of these cells. The DNA molecules are made up of two strands oriented in opposite directions where the G/C and A/T letters always pair or exhibit molecular complementarity across the strands (Figure 2). If chromosomes are broken into small pieces and the two strands are separated from one another, even in a complex mixture of DNA fragments, the correct partners can find one another through molecular complementarity and "zipper" back together. This molecular complementarity is the basis for the DNA diagnostics that will be discussed later. Each gene is expressed as messenger RNA, also exhibiting a four-letter language closely related to that of DNA. This mRNA molecule is processed by a specialized complex cellular machine, the ribosome, to generate the second type of biological information--the protein molecule--initially formed as a linear string of protein letters (Figure 2). The genetic code dictionary connects the DNA and protein languages (e.g. three adjacent DNA letters encode one protein letter).

The molecular language of proteins is more complex than that of DNA, with 20 different letters rather than four. The particular order of these letters in each protein string directs it to fold into a unique three-dimensional shape (Figure 3). Each protein is a three-dimensional molecular machine; these machines

catalyze the chemistry of life and give the body shape and form. As we shall see later, deciphering the DNA code may give us new insights into two of the most fundamental problems of proteins. (1) How does the order of the protein letters direct the three-dimensional folding into a precise shape (the protein folding problem)? (2) How does the three-dimensional structure of a protein permit it to execute its function (the structure-function problem)? Proteins and other biological macromolecules assemble together to create the functional units of living organisms, their cells (Figure 1).

The third type of biological information resides in the complex systems and networks arising from complex cellular interactions. For example, the human brain is composed of a 3xmillion million (10^{12}) nerve cells that form 10^{15} connections (synapes) to create an incredibly complex network (Figure 4). The interactions of these nerve cells lead to the so-called emergent properties of the brain (e.g. memory, consciousness, and the ability to learn). One could study one particular nerve cell for 20 years to learn everything it could do. Yet, this study would provide no insights into these emergent properties because they arise as a consequence of the network interactions of many different cells. The information of complex systems and networks, or biological complexity, is, in a sense, four-dimensional--it changes both in time and space.

Studying complex systems and networks is very difficult: (1) the components and their connections must be defined; (2)

biological experiments must probe how emergent properties arise from the network; (3) mathematical modeling will probably be necessary to thoroughly define complex systems; and (4) these models must ultimately be tested against biological reality by experimentation. Interestingly enough, many of the powerful new tools that scientists need to analyze biological complexity are emerging from technologies developed by the human genome project, as we shall see shortly.

Deciphering Biological Information

The Human Genome Project has catalyzed a quantum jump in our ability to decipher the one-dimensional biological information of DNA. However, the deciphering of biological information actually has two different meanings for each of the three types of information. For DNA, it is one thing to determine the order of DNA letters across each of the different chromosomes (e.g. the DNA sequence) and quite another to decipher the biological meaning that 3.7 billion years of evolution has inscribed in our DNA. For protein, it is one thing to identify a threedimensional structure and quite another to understand how this structure carries out its function. For a network, it is one thing to define the components and their connections and something different to understand how the emergent properties arise from these biological networks. Applied mathematicians and computer scientists will play a critical role in deciphering each of these types of biological information because they will create the powerful tools needed for complex analyses of large data sets.

Deciphering the biological information of complex systems and networks will be the major challenge in biology and medicine as we move into the 21st century. Analyzing biological complexity will require breaking the systems down into more experimentally tractable subsystems whose properties still reflect those of the system as a whole. One will also have to identify key bottlenecks or control points in the complex systems, both to understand their biology and manipulate the system for 21st century medicine. The tools of the Human Genome Project or genomics are beginning to allow us to tackle biological complexity.

The Human Genome Project

The Human Genome Project is the enterprise to map and sequence the 24 different human chromosomes (22 autosomes and the two sex chromosomes, X and Y). Humans have 46 chromosomes, half come from one's mother, and the other half from one's father. Homologous chromosomes differ on average by one in 1,000 letters of the DNA sequence. These variations within the human population are called polymorphisms. Since humans have three billion ($3x10^9$) DNA letters in the maternal or paternal complement of chromosomes, typically three million ($3x10^6$) polymorphisms distinguish the maternal and paternal chromosome sets. However, the genes may occupy only 3-5 percent of the DNA; hence, most of the polymorphisms will lie outside genes and, presumably, have little effect on the functioning or appearance (phenotype) of the organism. However, a few polymorphisms will predispose to human genetic diseases like cystic fibrosis or certain kinds of cancer

and are, therefore, medically important.[2]

The Human Genome Project is creating three types of maps for each chromosome (Figure 5). The genetic map has identified approximately 6,000 polymorphic markers spread evenly across all human chromosomes (except the Y).[3] A polymorphic marker is, typically, a particular site on an individual chromosome where a single DNA letter or small group of letters varies among members of the human population (Figure 6). The genetic map can then be used to identify genes that predispose to disease. The perfect co-segregation or passage of adjacent pairs of genetic markers through families together with a disease trait allow the localization of the disease-predisposing gene between the two polymorphic markers. Genetic markers further away from the disease gene do not co-segregate in families because their association is lost by chromosomal recombination, that is, apparently random breakage and reunion between the paternal and maternal chromosomes which scrambles the associations between particular forms of genetic markers. (In other words, the further the markers are away from the disease gene, the more likely it is that recombination will have unlinked an association). Genes predisposing to disease are crudely localized by first analyzing them against ~400 genetic markers scattered across the genome. Once the general location is identified, more genetic markers in that region can be studied to further narrow the disease gene location. This process is termed genome-wide genetic mapping. It provides an approximate localization of disease-predisposing

genes (to a region of perhaps one million DNA letters; the gene may only be 50,000 letters long). The actual gene must then be localized by other methods (e.g. DNA sequencing).

The physical map is made up of overlapping human DNA fragments that, taken together, span the length of the chromosomes (Figure 5). These DNA fragments can be used to physically localize disease genes. These fragments are also the source of material for the final map, the sequence map.

The sequence map for each chromosome represents the order of the letters of the DNA language all the way along each chromosome--this is termed the DNA sequence of the chromosome. The average human chromosome contains 130,000,000 DNA letters. The current DNA sequencing machines can only read about 500 letters in each DNA fragment at one time.[4] Hence, the DNA sequence maps constitute by far the largest challenge the Human Genome Project faces. Indeed, the genetic and physical maps are nearing completion; the next ten years of the project will be spent on the sequence maps. As the sequence maps are completed, computational approaches and biological experiments will allow the identification of the 100,000 human genes.

Model Genomes

The Human Genome Project also proposes to map and sequence the genomes of five model organisms (Table 1). Four are simple organisms with significantly smaller genomes than humans. The bacterium, yeast, simple roundworm (nematode) and the fly (Drosophila) all have genes that are similar to a subset of human

genes. Hence, one can use these simple model organisms to gain insights into how the evolutionarily related counterpart genes in humans may function. The sequence of the yeast genome has recently been completed,[5] as have the sequences of three prokaryotic genomes.[6]

The mouse genome is as complex as that of human. Accordingly, the mouse can serve as a valid model organism to study the function of many genes that control complexities not found in the simple model organisms. The mouse can also serve as a model for studying disease genes--how they cause pathology and as a vehicle to search for drugs to prevent the disease. Experiments can be done in mouse that are impossible or impermissible in humans.

The Human Genome Project will have a profound impact on biotechnology, biology and medicine as we move into the 21st century.

Technology Development for Genomics: Implications for Biology and Medicine

The study of genomes has, necessarily, led to the development of technologies that have the capacity to decipher large amounts of biological information from DNA. In the past, biologists tended to focus on the analysis of one gene or protein for extended periods of time. Today, the tools of genomics permit the analysis of thousands to billions of units of information per day (Table 2). For example, the Genome Center in the Department of Molecular Biotechnology at the University of Washington has 10

DNA sequencers that provide the capacity to sequence 360,000 (10 x 36,000) DNA letters per day. Likewise, our four genetic mappers can analyze more than 4,000 genetic markers per day.

Our large-scale DNA arrayer can place 20,000 human DNA fragment clones in one hour on a filter about the size of this page. DNA fragments can be obtained from two different sources. First, DNA from chromosomes can be fragmented and cloned into a recombinant DNA vector which can be grown in an appropriate host (e.g., bacteria). This is called a genomic library. Second, mRNA can be copied into DNA to make copies of all the genes expressed in a tissue, cell type or even tumor. This is called a copy DNA (cDNA) library. The presence of an mRNA in a tissue indicates that the gene coding for that mRNA is expressed in that tissue, that is, is used to produce a protein. Hence, for example, to examine the differences between the genes expressed in normal and tumor cells, 20,000 cDNA clones from a normal prostate gland can be arrayed on a filter and used to analyze the cDNA information present in hundreds of prostate tumors by molecular complementarity or hybridization. Ten identical normal cDNA filters can easily be prepared in one day and compared against the cDNA libraries from 10 tumors, thus making 200,000 comparisons of informational units (e.g., 20,000 x 10 hybridizations). A second approach to DNA arrays is the synthesis of a 100,000 oligonucleotide (e.g., a string 20 DNA letters long) arrays on a glass or silicon chips the size of your thumb nail (Figure 7).[7] In time, the expression patterns of all 100,000

human genes can be studied with these DNA chips. From these studies, insights into which proteins play key roles in cellular development, both normal and cancerous, will be obtained on a scale not heretofore possible.

Very powerful computational analyses can also be carried out on DNA sequences. For example, the 360,000 DNA letters per day coming from 10 DNA sequencers can be matched against the more than 600,000,000 letters in the genome data base to determine whether any of the new sequences match the preexisting sequences. The DNA sequence in the data base comes mostly from very short stretches of experimental human genes or from the genomes of other organisms, e.g., the model organisms. Only about 0.39% of the human genome has been sequenced to date.

The important point about these large-scale instruments is they can be used to study complex biological systems and networks. Indeed, the Human Genome Project is already beginning to revolutionize the practice of biology and medicine--and will have even more of an impact as we move toward completion of the human and model organism genomes early in the 21st century.

Genomics and Biology

The major challenge genomics presents to biology is the identification of the functions of all of the 100,000 human genes (Figure 8). This is very hard. The functions of a few of these genes are understood to varying levels of sophistication. The functions of some others can be guessed at because they resemble genes whose functions are known. The functions of many genes are

unknown. In the past, biologists would study a function, develop an assay for it (e.g. a way to measure it), through the assay purify the protein, and through the protein obtain the gene.[8] Thus, function was tied to gene identification. Genomics has inverted this pathway (Figure 8). There are three discrete challenges: (1) correlating genes with their proteins; (2) determing the three-dimensional structures of proteins; and (3) understanding how particular three-dimensional protein structures execture their functions. There may be shortcuts in correlating genes with presumptive functions. For example, computational methods can be used to determine whether the gene (or its protein translation via the genetic code dictionary) is similar to a gene (or protein) previously studied. If so, this may give a due as to function. If not, one may use large-scale DNA arrays to identify the cells or tissues in which the gene is expressed.

The localization of the gene product to particular cells may also give additional clues as to function. Then biological experiments must be done to elucidate the function. For example, the gene can be rendered non-functional ("knodced out") in a mouse to determine whether it has any noticeable effect on the phenotype (appearance or behavior) of the mouse. Indeed, experiments are now underway to knock out each of the 6,000 different yeast genes to determine their effect on yeast biology. Genes and their proteins can now be readily linked in yeast (because the entire genome is sequenced). For example, two-dimensional protein separation gels, those which separate

complex protein mixtures in one dimension by size and in a second dimension by charge, isolate relatively pure yeast protein spots (Figure 9). Individual proteins can be extracted from gels, cut with protein cutting enzymes, the fragment sizes analyzed in a mass spectrometer and, because the entire yeast genome has been sequenced, the corresponding gene can be identified from computational comparisons of predicted and experimental protein fragment sizes. The behavior of particular proteins (levels of expression, chemical modifications that alter function) can be followed on two-dimensional gels over the time necessary for a cell or organism in order to carry out a complex function to correlate protein behavior with particular functions. Thus, the worlds of DNA and protein can be joined. It will be some time before we can use similar tools to analyze human genes (at least until most of the genome is sequenced).

There are several computational approaches that may facilitate understanding the gene/protein/function relationships. The regulatory sequences (usually lying immediately to one side or even within the gene) determine when in development (time), where in the tissues (space) and how much of the gene is to be expressed (magnitude). As we develop systems analyses for the problems of gene regulation (studying, for example, the interactions of DNA regulatory sequences and the proteins that operate on these sequences to trigger the control decisions for gene expression of time, space, and magnitude), we will begin to decipher the regulatory code. Perhaps there will be a time when

this code can be deciphered directly from the gene sequence to predict for every gene these three parameters of gene expression (Figure 10).

A second computational approach will be to attempt to identify the lexicon of motifs that are the fundamental building blocks of genes and proteins (Figure 11). A motif is a segment of protein sequence that causes a particular fold and/or facilitates a particular function. An analogy is a train. The train is made up of many different cars that have discrete functions (e.g. caboose, engine, box car). So a protein is often made up of several domains each with a discrete function. Each car in the train has smaller components that facilitate function (e.g. the windows, stove, chimney, doors, and walls of the caboose). So a protein domain has as its building block motifs which may vary in size from a few letters to 100 or more letters (Figure 11). Perhaps a few hundred motifs out of a possible 10^3-10^4 have been identified (e.g. the zinc finger motif found in proteins which bind DNA).[9] These motifs correlate with defined structure and sometimes can actually facilitate a function. Motifs can be difficult to identify because many of them are highly degenerate; that is, out of 10-30 amino acid letters, perhaps only a few are conserved or partly conserved.

Two advances will facilitate the identification of the entire lexicon of motifs. The first is finishing the sequences of the genomes of the human and other model organisms. The cross-species comparisons can be useful in identifying motifs as

can the identification and cross comparisons of all the individual members of gene families within a species. Gene families arise when one gene has been very successful. Often many copies of that gene are made at the same chromosomal site and the individual genes diverge to carry out distinct but related functions. This group of genes is termed a gene family. Really successful gene families can make copies of themselves which move to different chromosomal sites (Figure 12). Second, the determination of many more three-dimensional structures for proteins will permit the cross comparisons of one-dimensional patterns and threedimensional structures to facilitate motif identification. This lexicon of protein motifs could play a key role in solving the protein-folding problem and in linking three-dimensional structures of proteins to their functions.

Thus, the tools of genomics will let us approach in new and powerful ways the analysis of complex biological systems and networks. Areas such as immunity, development, and nervous system function can all be approached from the systems viewpoint using many of the powerful tools of genomics. Virtually every area of biology can be attacked with these new tools and approaches.

Finally, it has been said that the history of much of our past evolution is buried in our genomes. The complete genome sequence will, indeed, let us identify all of the families of related genes and delineate the nature of their molecular archeology (Figure 12). Comparisons with the genome sequences of the model organisms will enormously enrich our understanding of

molecular evolution. As a product of evolution, the digital information of human chromosomes actually contains many different languages, some discrete and others overlapping. For example, the coding regions of genes represent one language; the regulatory code a second; the major features of genome evolution a third; the chromosomal machinery necessary for rapid DNA replication from many sites a fourth; etc. The initial efforts to decipher the multiplicity of languages present in human chromosomes have proved challenging. For example, Figure 13 is a schematic illustration of the 700,000 letters of the DNA alphabet spanning one important gene family of immune receptors. The vertical bars on the first line indicate the 94 gene elements found in this family. The lower colored bars all represent distinct types of digital information present in the longest contiguous stretch of human sequence analyzed to date.[10] Knowing all the members of this gene family permits one to interrogate and manipulate the immune system with striking new strategies. This is moving us toward a preventive medicine of the 21st century.

Genomics and Medicine

The tools of genomics provide the large-scale capacity to study human polymorphisms to determine which are irrelevant, which cause variations within the range of normal physiology (e.g. height and longevity variations), and which correlate with diseases or the predisposition to diseases. As we identify all 100,000 human genes, we will have the tools to study variation in large human populations (using large-scale DNA sequencing or even

large-scale DNA arrays or chips to detect polymorphisms) (Figure 7). Indeed, with the use of DNA chips to study polymorphisms, it should be possible to increase the throughput of genetic marker analyses one hundred-fold. Thus, genes that cause or predispose to disease could, given appropriate numbers of families with the disease, rapidly be identified.

Some sequence variations (polymorphisms) within genes invariably cause disease. For example, some defects in the structure of collagen, a protein required for building bones, have been traced to small deletions in the DNA coding for the protein.[11] A person who inherits this mutation from either parent will suffer from osteogenesis imperfecta, a bone disease. Such a situation of so-called autosomal dominance is rare. More commonly, a defective version of a gene inherited from one parent can be fully or partially compensated by the normal version of the gene inherited from the other parent Thus, many diseases will be caused only if the same defective gene is inherited from both parents.

Alternatively, inheritance of a defective gene may result in a continuous gradient of phenotype ranging from no effect to explicit disease. For example, the severity of some diseases such as Huntington's disease or fragile X syndrome, which causes mental retardation, has been correlated with an increase in the number of consecutive repeats found in a group of three DNA letters tied to a specific location within the gene.[12] Below a certain number of consecutive repeats, no disease symptoms are

manifested. Above this, as the number of repeats increase, so does the severity of the disease.

Finally, some polymorphisms are associated with a probability of getting a disease. In these cases, terms such as 'susceptibility' or 'predisposition' may be used to describe the propensity to disease. For example, when the mutant (altered) form of the breast cancer 1 gene (BRCA1) is present in one defective copy, this gene predisposes a 60 year-old woman to a 70 percent chance of getting breast cancer.[13] The 70 percent probability figure may arise from one or both of two possibilities. First, women with the defective BRCA1 gene may require one or more environmental factors to trigger the disease process. Presumably, the overall probability that women with the defective gene will experience those factors is 70 percent. Second, perhaps other genes have the ability to modify the expression (or function) of the BRCA1 gene so as to enhance or limit its ability to cause cancer. By this alternative, the overall probability of having the requisite "bad" set of genes, without an offsetting collection of "good" genes, is 70 percent.

Complicating matters even further, when larger numbers of families with a particular disease phenotype are studied by genetic analyses, it often turns out that multiple genes can predispose to the same apparent disease. For example, the BRCA1 and BRCA2 genes both can cause breast cancer[14] and four different genes have been identified as predisposing to Alzheimer's disease.[15] In essence, some diseases, such as

cancers and dementias, appear to result from any number of different defective genes, possibly in combination with environmental triggers. Along this same line, some diseases, such as multiple sclerosis, are likely to be multigenic in origin, requiring that two or more separate genes be defective, in order for the disease to occur.

Thus, sequence variations in genes can lead to diseases that have an all-or-none symptamatology, a degree in the severity of symptoms, or a likelihood of causing symptoms if other genetic or environmental factors exacerbate or fail to ameliorate the effects of the defective genes.

If we look 25-30 years into the future (or perhaps less), we can imagine a time when perhaps one hundred (or more) polymorphic variations in genes will have been identified as predisposing to common diseases--cancer, cardiovascular diseases, autoimmune diseases, etc. It will certainly be possible to identify these defective genes in individuals and deduce a future "probable health history" complete with numerical assignments to the probabilities. Many of these "probable health histories" can be quite complex and, thus, difficult to interpret. In the future, there will be therapeutic interventions or preventive measures that will circumvent the effects of many of these disease predisposing genes.

Today, however, there exists a gap between the ability to diagnose the predisposition to diseases such as breast cancer and the ability to prevent the disease. The debate on whether to use

DNA diagnostic tests to identify susceptible women is framed in terms of this gap between diagnostic capacities and the ability to intervene therapeutically. For complex diseases, this gap may span decades. The preventive measures will, in all likelihood, employ the manipulation of the three types of biological information (Table 3). In this scenario, one of the major functions of medicine would be to keep people well (today medicine generally treats the sick). That is, preventive intervention would be given before the disease symptoms manifest in persons whose genes conduce to a high probability of causing a disease.

Thus, genomics will provide powerful tools for correlating DNA polymorphisms with disease followed by subsequent manipulations of biologic information and/or environmental factors (such as diet) to prevent disease. This approach is not likely to change the natural life span. Rather, it will reduce the toll of chronic illnesses which often strike in middle age. It will presumably let individuals live weli into their 70s and 80s mentally alert and physically healthy. Society will then, as it is now, be challenged to deal with an expanding population of 70 and 80 year-old people capable of contributing to society in a productive and creative manner.

Ethical and Social Implications of Human Variation

The tools of genomics will provide powerful and large-scale means for deciphering human polymorphisms that predispose to disease. This biological information must be acquired, stored,

analyzed, and distributed from computers. As we learn more about how human polymorphisms correlate with disease--increasingly comprehensive knowledge can be gained about the predicted health history of each individual if appropriate DNA tests are carried out. If preventive or therapeutic measures were available to circumvent the deleterious effects of diseasepredisposing genes (readily available to all), then the question of privacy of genetic information would not be quite so compelling. However, we do not now have preventive measures or therapies for most genetic diseases, nor will we in the near future. DNA testing has begun and presumably will continue. Databases are now accumulating DNA testing information. Accordingly, the issue of genetic privacy presents a compelling challenge. Since other chapters of this book consider these and related issues in considerable detail, after the following reflection on values, we will briefly outline a few of the major issues surrounding the deciphering and use of biological information.

Values and Policy Making

How does one, either an individual or a society, decide whether a particular action (to test or not?; to abort or not?; to inform or not?) is the right thing to do? Typically, the rightness or wrongness of actions is evaluated in light of probable consequences. Will the action cause the most good and/or the least harm to self and others? Will the action cause immediate good but long term harm, or vice-versa? Will the action benefit many or only a privileged few? What might be the

unintended consequences of a course of action? How are benefits and costs to be defined and assessed? What happens when different assumptions about what is best for individuals and/or society dictate different and conflicting courses of action? What is more important -- the benefits to individuals or the benefits to society, if there is a conflict between the two? What benefits are most important to the long term health of a society -- fairness?; optimal distribution of resources?; protection of the individual's rights?; prevention of harm to others? Judgments such as these are difficult to make for highly informed and reflective individuals. They are even harder for persons who lack the necessary background of information required for sensibly evaluating contrasting viewpoints or who lack the will or ability to think critically about ethical issues. Nonetheless, decisions regarding the uses of genetic information must and will be made, and many of these decisions will be embodied in public policy.

Policy decisions about genetic information are complicated by the following factors: legal precedents already in place, to which analogies will be made; current and future social, economic, and political realities that play into the content of various policies, and into the processes by which consensus regarding the content of policies is achieved; and the inherent complexity of the issues posed by the future explosion of genetic data, techniques, and concepts.

The precedents in place surrounding issues such as testing and consent (e.g. testing newborns for phenylketonuria, a disease

whose deleterious effects can be largely ameliorated by dietary modifications, or testing blood donations for hepatitis and HIV) may apply well to some genetic disease-causing polymorphisms but not others. Determining the validity of the analogical reasoning that ties precedents to new situations requires both an understanding of the ethical, political, and social issues at stake and an understanding of the scientific and technical nuances that pertain to a particular disease and its predisposing genes. Attaining the required depth and breadth of understanding is difficult, especially when such understanding is frequently not held or agreed upon even by the experts in the field. Because of an urgent need to make policy (so that industries such as insurance companies can be regulated, and laws can be consistently applied to individual cases) decisions will be made in the absence of complete information and social reflection about the underlying values. As a result, new precedents may be set that turn out to be shortsighted or deleterious but which will be extremely difficult to overturn.

Factors inherent to social, economic, and political reality affect policy formation. For example, while American citizens may all be created equal in terms of possessing certain rights, we are not all created equal in terms of life circumstances, and our genes comprise an important and irreversible component of our life circumstances. In the arena of genetic information, many policy decisions will require a clarification of the application of the concept of a 'right.' Do citizens have a right to privacy

of information about their genes even if the withholding of that information causes harm to others? Do citizens have a right to health care, even if the care is expensive? Who is obligated to pay for the care? Do fetuses have a right to life, even if the life they have will be miserable, due to some genetically caused disease? Does society have the right to say that certain fetuses must be aborted, because the care of individuals with debilitating genetic diseases is too great a burden for its limited resources, resources which should instead be placed where they will do the most good? Addressing these questions and others will force an examination of our social values and commitments. Ideally, a consensus will emerge that optimizes benefits for both individuals and society over the long term. There will be costs, however, perhaps to individual liberty, perhaps to a conception of the sanctity of life, perhaps to our resources in the form of increased obligations towards those who, through no fault of their own, bear the burden of genetic diseases.

Finally, the data and concepts emerging from the Human Genome Project will challenge many of our beliefs about human nature. We will discover the ways in which we humans are special, gifted with abilities not had by other species. We will also discover how very similar we are to other species inhabiting our planet. We believe that two fascinating issues will dominate social policy discussions in the 21st century. One is our capacity to shape our own evolution as a species. Decisions may be made about who should live to adulthood and who should die,

either through permissible or mandatory abortion or through the withholding of therapeutic intervention for life-threatening genetic diseases in the interest of rationing the precious resources of health care. Policies may be considered regarding the regulation of who should be allowed to bear children. Lastly, it may become possible to develop and implement techniques aimed at altering the DNA in the chromosomes passed on through inheritance, thus potentially eliminating some disease-causing genes from the gene pool, at least for several generations.

The second issue concerns the age-old free-will versus determinism and nature versus nurture debates. A vastly increased understanding of the relationships between genes and behavior will bring these debates to the fore, with social implications for education, therapeutic intervention, and legal adjudications. Distinctions between normal and deviant behavior might be drawn, in part, along the lines of genetic polymorphisms. Just as with genetic diseases, some polymorphisms may promote an all-or-none phenotype with regard to a particular behavioral trait, and some may affect degrees of expression of a behavioral trait. Decisions may be made that certain behaviors are intolerable for society and, thus, that therapeutic intervention modifying the behavior will be mandatory. It will take will, resourcefulness, and soul-searching on the part of society's policy makers to find an ethical path through the thorny issues created by the notion that there might be identifiable genetic predispositions to certain behaviors.

We will turn now to a brief listing of specific issues arising from the deciphering and use of biological information.

Screening. Who should get genetic tests? When is it advisable to screen the members of a family with genetic disease? When is it appropriate to use genetic screening for prenatal diagnostics? When is it appropriate to carry out populationwide genetic screening?

Privacy. Who should or will have access to genetic information beyond the patient and his or her physician? Does your insurance company or employer? Does your family or future husband or wife? There are complex questions. It appears a priori obvious that only an individual and his or her physician should know about one's genetic information. Yet, this approach creates complications. For example, if the entire population knew who had good and bad genes, but the insurance companies did not, would it not be possible for those with more bad genes to carry far more insurance than those with few bad genes, thus placing insurance companies at a disadvantage? Alternatively, suppose an individual knowingly chooses to work for a company with an environment unhealthy for him or her because of a defective gene. If the individual becomes sick, is the company responsible if it had no right to deny employment based on defective genes? This relates to the complex question of how responsible an individual is for his or her choices. Several bills are now in Congress addressing the issue of genetic privacy.

Counseling. How does one explain to the lay public genetics

and probability? Where are we going to get the trained experts to handle the volume of potential future patients?

Physicians. How can one effectively train all physicians about the complexities of human genetic disease and biological information? The training of physicians will have to change dramatically as we move into preventive medicine of the 21st century; it will need to be more analytic and conceptual; training in the use of computers will be critical; and training in how to educate and communicate with patients will become important.

Abortion. As the 100,000 human genes are identified, it will be possible to screen in utero for increasing numbers of human genetic diseases. How can the boundary conditions for permissible therapeutic abortions be determined? Will wrongful life suits be permissible if a fetus with a genetic defect is not aborted? In some cases, most reasonable individuals could agree that abortion is not appropriate (e.g., to obtain the desired sex). In other cases, most reasonable individuals, apart from those with religious convictions against abortion, would agree that abortion is appropriate when a severe, untreatable disease is involved (e.g., Tay-Sachs disease is a rapid and progressive neurological disease that generally kills infants within the first year of life). How can society set the boundary conditions between these extremes (assuming that abortion is legal)?

Genetic Engineering. Genetic engineering or gene therapy involves the replacement of defective genes by good ones. There

are two types of genetic engineering--gene replacement in the cells of the body (somatic cell engineering) and gene replacement in the sex cells, the sperm and egg cells (germ cell engineering). Somatic cell engineering is, in one sense, just an extension of contemporary medical therapies. Changes made in genes die with the individual. Limited somatic cell engineering has been carried out, but many technical problems remain to be solved.

Germ cell engineering modifies the human gene pool. Consequently, these changes can alter human heredity. Germ cell genetic engineering is unlikely to be practiced (in humans) for a long while, if ever. First, enormous technical difficulties must be solved before it is safe in just the technical sense. Second, most interesting human traits (e.g., intelligence, emotional stability, and physical attractiveness) are complex multigenic traits that will probably not be completely understood in our lifetimes. Hence, they could not be engineered. However, there may come a time when society could engineer human heredity to modify fundamentally human traits. Society will then have to determine whether this is appropriate and, if so, establish reasonable rules and guidelines.

Genes and behavior. Genes do appear to influence behavior. For example, genes appear to contribute to homosexual behavior[16] and thrill seeking.[17] A Johns Hopkins scientist recently created a strain of mice in which the gene synthesizing an important neural transmitter (the signal molecules brain cells

use to communicate with one another) was destroyed.[18] Mutant male mice, if placed in a cage with normal males, killed them. If placed in a cage with females, the male mice brutally attacked them. It would appear that the loss of this neural transmitter caused mice to become extremely violent. A similar observation has been made for humans. Several years ago, a Dutch geneticist reported on a large family with eight males predisposed to extremely violent behavior (armed robberies, brutal assaults, rapes, etc.).[19] The violent males all had defects in a gene which breaks down a particular neural transmitter. None of the normal family members tested had this defect. Accordingly, it does appear likely that genes may influence some aspects of behavior. This poses an interesting challenge for our society and its judicial system. Since our system of law is based on free will and individual responsibility, could a future criminal argue extenuating circumstances because his genes made him commit the criminal act?

Forbidden Science. Are there some types of biological research that are considered so dangerous (e.g. connections between genes and behavior) or so socially inappropriate (to some) such as the use of fetuses for investigation, that the research should be banned? We would argue that the fundamental knowledge of how our genes and human development work is so important to dealing with some of humanity's most deadly and devastating diseases that few, if any, restrictions should be placed on fundamental research. Society should control the

application of this knowledge in the form of technologies, not the acquisition of this basic knowledge.

Scientists and Society

Never have the research opportunities been greater in the biological and medical sciences. Yet, scientists face a skeptical general public. The public wonders whether science has really brought benefits as they are surrounded by pollution, disease (cancer and AIDS), and poorly understood new technologies that appear to have a science fiction cast (e.g., Jurassic Park). They are vaguely aware of the ethical issues emerging from human genetics often without sufficient knowledge of this science to think rationally about them. We believe the fundamental contract between scientists and society has changed markedly, even in the last five to ten years. Scientists must reach out to society and educate them as to the opportunities (wonders) and benefits of science, as well as the ethical challenges.

When we moved to the University of Washington to create the first Department of Molecular Biotechnology, one of us (LH) had two objectives: (1) to create an interdisciplinary environment for developing and applying tools to study systems complexity to biology and medicine; and (2) to create an environment to encourage scientists to spend five to ten percent of their time bringing science to society. The most effective way we have found to do this is to catalyze system change in K-12 schools in Seattle. For example, we have an elementary program, recently funded by a $4.25 million grant from the National Science

Foundation to bring hands-on, inquiry-based science through 100 hours of instruction to each of the 1,400 elementary teachers in the Seattle Public School District over the next five years. This effort is a collaboration including the School District, Boeing Co., the Fred Hutchinson Cancer Research Center, and our Department, together with nine other departments at the University of Washington. In addition, we are also teaching high school students and teachers how to sequence the human genome. Twenty schools are participating in an endeavor to sequence an unknown gene causing deafness in a large Costa Rican family. We also have the students break up into groups of four and imagine that they are a family with Huntingtons disease. The students are taught how to analyze the situation ethically and then they are asked to decide whether they want to know if they (hypothetically) have the defective gene. Needless to say, the experience is a challenging and educational adventure. These students, we hope, will realize that science is not about answers, but rather about asking questions. We hope they will be excited by challenges, curious about the world, and aware that learning is a life-long commitment. As such children become citizens, they will be uniquely capable of dealing with the complexities of the world in which they live.

We would argue that scientists (and other academicians) should make a commitment to bring science (and the benefits of education) to the public. It is perhaps the only way we can make our case to society about the fundamental importance of science

to society's future. We can, at the same time, prepare tomorrow's citizens to appreciate and deal with the opportunities and challenges coming from the recent and exponentially increasing explosion in deciphering biological information.

Table 1.

Genome Sizes of Model Organisms

	Megabases (Millions of Bases)
E coli	5
Yeast	15
Nematode (Worm)	100
Drosophila (Fly)	180
Mouse	3,000
Human	3,000

Table 2

Tools of Genomics

Tools	Throughput
Large-scale DNA sequencer	36,000 DNA letters per day
Genome-wide genetic mapping	1,200 genetic markers per day
Large-scale DNA arrays	2000, hybridizations per day
Computational (similarity anlyses)	3×10^{12} DNA letters per day

Table 3

Molecular Therapies*

1°	Anitsense Gene Therapy
3°	Protein Engineering Applied Molecular Evolution Hormones Neurotransmitters
4°	Stem Cells Immunomanipulation

* 1°, 3°, and 4° dimensional indicate the three types of biological information (see text).

Figure 1. A drawing of the cell, its nucleus and chromosomal strand extending from the nucleus. (From *The Human Genome Project: From Maps to Medicine*. Department of Health and Human Services, Public Health Service National Institutes of Health, NIH Publication No. 96-3897).

Figure 2. A schematic illustration of the flow of biological information from DNA to messenger RNA to protein.

Figure 3. The three-dimensional structure of an enzyme, lysozyme, that cleaves sugar molecules.

Figure 4. A photograph of stained nerve cells and their communicating extensions.

Figure 5. A schematic illustration of the three types of maps being determined by the Human Genome Project.

Figure 6. An illustration of a hypothetical polymorphic site or genetic marker on a human chromosome. Similar portions of the same chromosome are given for the maternal and paternal chromosomes for two individuals. One of these four chromosomes has a single letter substitution or polymorphism.

Figure 7. A schematic of a DNA chip or oligonucleotide array. Different short DNA sequences (e.g. ~20 letters) can be synthesized on a glass or silicon chip and then used to detect messenger RNA (or their DNA copies) or DNA fragments that are complementary in sequence by hybridization (9).

Figure 8. A schematic diagram illustrating the challenge presented by the Human Genome Project through the identification of the 100,000 or so human genes. The challenges include correlating genes with their proteins, proteins with their structures and protein structures with their functions.

Figure 9. A two-dimensional protein gel. The proteins (dark spots) are separated in one dimension by size and in a second dimension by electrical charge.

Figure 10. A schematic illustration of the DNA regulatory code governing the expression of particular genes in different tissues. The long rectangles represent genes and the squares, triangles, and circles various regulatory elements. (Adapted from Figure 19, page 150, in **The Code of Codes, Scientific and Social Issues in the Human Genome Project**. Eds. Kevles, D.J. and L. Hood. Harvard University Press, Cambridge, MA, 1992).

Figure 11. A schematic illustration of the domains and motifs of a hypothetical protein (see text). (Adapted from

Figure 20, 154, in **The Code of Codes, Scientific and Social Issues in the Human Genome Project.** Eds. Kevles, D.J. and L. Hood. Harvard University Press, Cambridge, MA, 1992).

Figure 12. Members of a related set (the immunoglobulin super family) of a very successful proteins that are encoded by genes and gene families scattered across the human genome. (From Hunkapiller, T. and L. Hood. Diversity of the Immunoglobulin Gene Superfamily. *Advances in Immunology* 44, 1-63, 1989).

Figure 13. A schematic illustration of the human ß T cell receptor gene family. The vertical bars represent genes. The colored patterns represent various other types of biological information. Adapted from L. Rowen et al., *Science*, in press).

1. James D. Watson and Francis C. Crick. "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acids." Nature 171 (1953): 737-8.

2. John R. Riordan, et al. "Identification of the Cystic Fibrosis Gene: Cloning and Characterization of Complementary DNA." Science 245 (1989): 1066-73.

3. Colette Dib, et al. "A Comprehensive Genetic Map of the Human Genome Based on 5,264 Microsatellites." Nature 380 (1996): 152-54.

4. Michael W. Hunkapiller, et al. "Large-Scale and Automated DNA Sequence Determination." Science 254 (1991): 59-67.

5. Nigel William. "Yeast Genome Sequence Ferments New Research." [News]. Science 272 (1996): 481.

6. Robert D. Fleischmann, et al. "Whole-Genome Random Sequencing and Assembly of Haemophilus Influenzae Rd." Science 269 (1995): 496-512; Claire M. Fraser, et al. "The Mycoplasma Genitalium Genome Sequence Reveals a Minimal Gene Complement." Science 270 (1995): 397-403; Carol Bult, et al. "Insights into the Origins of Cellular Life from the Complete Genome Sequence of the Methanogenic Archeon. Methanoccoccus jannaschii." In press.

7. Stephen P. Fodor, et al. "Light-Directed, Spatially Addressable Parallel Chemical Synthesis. Science 251 (1991): 767-73.

8. Michael W. Hunkapiller, et al. "A Microchemical Facility for the Analysis and Synthesis of Genes and Proteins." Nature 310 (1984): 105-11.

9. Frank Eisenhaber, Bengt Persson, and Patrick Argos. "Protein Structure Prediction: Recognition of Primary, Secondary, and Tertiary Structural Features from Amino Acid Sequence." Critical Reviews in Biochemistry and Molecular Biology 30 (1995): 1-94.

10. Lee Rowen, B.F. Koop and Leroy Hood. "The Complete 685 kb DNA Sequence of the Human ß T Cell Receptor Locus." Science. In press.

11. Darwin J. Prockop and Kari I. Kivirikko. "Heritable Diseases of Collagen." New England Journal of Medicine 311 (1984): 376-86.

12. Neal G. Ranen, et al. "Anticipation and Instability of IT-15 (CAG)n Repeats in Parent-Offspring Pairs with Huntington Disease." American Journal of Human Genetics 57 (1995): 593-602; E. Pintado, et al. "Instability of the CGG Repeat at the FRAXA Locus and Variable Phenotype Expression in a Large Fragile X Pedigree." Journal of Medical Genetics 32 (1995): 907-08.

13. Mary-Claire King, Sarah Rowell, and Susan M. Love. "Inherited Breast and Ovarian Cancer. What are the Risks? What are the Choices?" Journal of the American Medical Association 269 (1993): 1975-80.

14. Oshio Miki, et al. "A Strong Candidate for the Breast and Ovarian Cancer Susceptibility Gene BRCA1." Science 266 (1994): 66-71; S.V. Tavtigian, et al. "The Complete BRCA2 Gene and Mutations in Chromosome 13q-Linked Kindreds." Nature Genetics 12 (1996): 333-37.

15. Marcia Barinaga. "Missing Alzheimer's Gene Found." [News]. Science 269 (1995): 917-18.

16. Dean H. Hamer, et al. "A Linkage Between DNA Markers on the X Chromosome and Male Sexual Orientation." Science 261 (1993): 321-27; Stella Hu, et al., "Linkage Between Sexual Orientation and Chromosome Xq28 in Males but no in Females." Nature Genetics 11 (1995): 248-56.

17. Richard P. Ebstein, et al. "Dopamine D4 Receptor (D4DR) Exon III Polymorphism Associated with the Human Personality Trait of Novelty Seeking." Nature Genetics 12 (1996):78-80; Jonathan Benjamin, et al. "Population and Familial Association Between the D4 Dopamine Receptor Gene and Measure of Novelty Seeking." Nature Genetics 12 (1996): 81-84.

18. Olivier Cases, et al. "Aggressive Behavior and Altered Amounts of Brain Serotonin and Norepinephrine in Mice Lacking MAOA." Science 268 (1995): 1763-1766; Randy Nelson, et al. "Behavioural Abnormalities in Male Mice Lacking Neuronal Nitric Oxide Synthase." Nature 378 (1995): 383-86.

19. H.G. Brunner, et al. "X-Linked Borderline Mental Retardation with Prominent Behavioral Disturbance: Phenotype, Genetic Localization, and Evidence for Disturbed Monoamine Metabolism." American Journal of Human Genetics 52 (1993): 1032-39.