

Special Article

SHATTUCK LECTURE — MEDICAL AND SOCIETAL CONSEQUENCES  
OF THE HUMAN GENOME PROJECT

FRANCIS S. COLLINS, M.D., PH.D.

**T**HE history of biology was forever altered a decade ago by the bold decision to launch a research program that would characterize in ultimate detail the complete set of genetic instructions of the human being. The idea captured the public imagination, perhaps less in the manner of America's wars on cancer and the acquired immunodeficiency syndrome than in the manner of the great expeditions — those of Lewis and Clark, Sir Edmund Hillary, and even Neil Armstrong. Scientists wanted to map the human genetic terrain, knowing it would lead them to previously unimaginable insights, and from there to the common good. That good would include a new understanding of genetic contributions to human disease and the development of rational strategies for minimizing or preventing disease phenotypes altogether.

The endeavor was both awesome and chancy. The instruction book — the human genome — was vastly larger than any genetic endowment tackled so far, and in 1990, the tools were not yet powerful enough to perform the task. Critical social questions, such as whether the new technologies for reading our genetic constitution would challenge our identities, our fundamental right to privacy, or our freedom from discrimination, loomed without answers.

Yet, a public science initiative focused so sharply on the molecular essence of humankind was too intriguing and too promising to forgo. Since the 1970s, nearly all avenues of biomedical research have led to the gene, for genes contain the basic information about how a human body carries out its duties from conception until death. In between, of course, our bodies struggle to survive in a challenging environment. Largely, but not entirely, at the behest of our genes, we fare better or worse.

Not surprisingly, disease researchers wanted to bag their leading gene suspects as soon as possible and at the least expense. This was no small order — the

80,000 or so human genes are scattered throughout the genome like stars in the galaxy, with genomic light-years of noncoding DNA in between. The billions and billions of uncharted DNA units (approximately 3 billion base pairs in humans) frustrated searches regularly in the late 1980s, often at great health and financial cost. If gene hunters were to mine miracles from the human genome, they needed more powerful tools and more ambitious strategies.

**ACCELERATED GOALS FOR SEQUENCING  
THE HUMAN GENOME**

In 1988, Congress appropriated funds to the Department of Energy (DOE) and the National Institutes of Health (NIH) to begin planning the Human Genome Project. Planners set a 15-year time frame, estimated that the price tag would be \$3 billion, and laid out formal goals to get the job done.<sup>1</sup> On October 1, 1990, the Human Genome Project officially began.<sup>2</sup> According to early plans, the human race would witness its own blueprint in fine detail in the year 2005.

In the fall of 1998, however, improvements in technology, success in achieving early mapping goals (see below), emerging research opportunities, and a growing demand for the human DNA sequence prompted project leaders in the United States and abroad to promise the blueprint — the complete DNA sequence of the human genome — two years ahead of schedule, in 2003.<sup>3</sup> The technology to accomplish such a task is at hand. Indeed, only six months later, in March 1999, 15 percent of the sequence was in a finished or nearly finished state. The largest centers participating in the Human Genome Project received new grants to begin full-scale sequencing of the human genome, and the timetable was moved up yet again. Pilot sequencing projects had been so successful that the planners of the Human Genome Project now felt confident that at least 90 percent of the human sequence could be completed in “working draft” form by the spring of 2000, considerably earlier than expected.

The NIH and DOE genome programs expect to contribute 60 to 70 percent of the sequence. Scientists funded by the Wellcome Trust at the Sanger Centre in Cambridge, England, along with other international partners, will produce the remainder.

Presented as the 109th Shattuck Lecture to the Annual Meeting of the Massachusetts Medical Society, Boston, May 8, 1999.

From the National Human Genome Research Institute, National Institutes of Health, 31 Center Dr., MSC 2152, Bldg. 31, Rm. 4B09, Bethesda, MD 20892-2152, where reprint requests should be addressed to Dr. Collins.

©1999, Massachusetts Medical Society.

Until the entire human genome sequence has been completed, in 2003 or perhaps earlier, the working-draft sequence will be very useful, especially for finding genes, exons, and other genomic features. But because the working draft will contain gaps and will not be entirely accurate, it will not be as useful as the finished sequence for studying DNA features that span large regions or require a high degree of accuracy over long stretches of the sequence.

The final product must have four characteristics — the four A's of the Human Genome Project. First, the sequence must be *accurate* — that is, the DNA spellings must have an accuracy of 99.99 percent or better. If this is the Book of Life, we should not settle for a rough draft over the long term but should remain committed to producing a final, highly accurate version. Second, large-scale sequencing requires that the shorter lengths of sequenced DNA be accurately *assembled* into longer, genomic-scale pieces that reflect the original genomic DNA. Third, the human DNA sequence must also be *affordable*, and technology development will aim to reduce the cost as much as possible. Finally, the high-quality, finished human DNA sequence should be *accessible* within 24 hours through public data bases.

With regard to this last point, it is imperative that the genome sequence, the fundamental building block of biomedical research, be made rapidly available at no cost to scientists around the world. The human DNA sequence arms scientists seeking to understand disease with new information and techniques to unravel the mysteries of human biology. This knowledge will dramatically accelerate the development of new strategies for the diagnosis, prevention, and treatment of disease, not just for single-gene disorders but for the host of more common complex diseases (e.g., diabetes, heart disease, schizophrenia, and cancer) for which genetic differences may contribute to the risk of contracting the disease and the response to particular therapies.

In 1996, at the Second International Strategy Meeting on Human Genome Sequencing, a group of scientists involved in large-scale genomic DNA sequencing passed a unanimous resolution that all sequence data produced by the Human Genome Project should be freely available and in the public domain, in order to encourage research and development and to maximize its benefit to society. These scientists believed that public availability of these data would encourage coordination and collaboration among scientists while preventing duplication of efforts.

As a result, planners of the Human Genome Project adopted a policy of releasing data every 24 hours to a free, publicly accessible data base. In the United States, GenBank (accessible at <http://www.ncbi.nlm.nih.gov>), run by the National Center for Biotechnology Information, serves as the public repository of sequence information. The value of this

data base as a resource for medical scientists around the world is incalculable. GenBank receives over 200,000 queries a day for information on gene sequences. It took 16 years to get the first billion bases into the data base, but it took only 15 months to add the second billion bases. Over 39,000 species are represented, and over 60,000 sequence-comparison searches are conducted each day.

Because researchers around the world participating in the public sequencing effort have agreed to submit their data to free, public data bases within one day after verification, any scientist, whether based at a university, a corporation, or a government laboratory, can log on by computer and have full and rapid access to the sequence data. The sooner the publicly funded effort to sequence the human genome is complete, with the data deposited in a free and publicly accessible data base, the sooner it will benefit all scientists working to understand and treat disease at a molecular level. In addition, scientists understand human genes by comparing human DNA with DNA from mice, flies, roundworms, and yeast. These comparisons make the human sequence much easier to understand.

As Human Genome Project plans for the period between 1998 and 2003 were being developed, two corporate ventures announced initiatives to sequence a major fraction of the human genome, using strategies that differ fundamentally from the publicly funded approach. The stated intention of one of these ventures to release its data publicly creates the possibility of synergy with the federal effort. If the corporate data and the public data can be merged, the depth of sequence coverage will be even greater. And because the public sequence data will be obtained from mapped DNA, it will provide critically needed anchoring to sequence data generated by private-sector efforts.

#### MORE THAN JUST THE SEQUENCE

Many people assume that since 1990 most of the work of the Human Genome Project has been devoted to large-scale sequencing. But this activity is actually a relatively recent arrival on the scene, since many other kinds of genome information were needed before full-scale sequencing could be undertaken. Before the 1998–2003 plan, the Human Genome Project's scientific goals mostly addressed genome mapping, technology development, and work to characterize the genomes of certain laboratory organisms. One type of map, known as a genetic map, is particularly helpful for following the inheritance of a disorder through several generations of a family. Genetic maps consist of a series of sequence-based markers that can be used to pinpoint the likely neighborhood of an altered gene responsible for a disease phenotype or other trait. Such maps have been invaluable in identifying and isolating highly penetrant

gene mutations with mendelian inheritance patterns. The goal was to establish markers close enough to give a gene hunter a high likelihood of placing a gene in a reasonably searchable interval. A year ahead of schedule, an international consortium of leading researchers in France and the United States published a genetic map containing almost 6000 markers, spaced less than 1 million bases apart.<sup>4</sup> Such detail was four to six times greater than the 1990 goals called for.

A second type of map, known as a physical map, provides cloned and ordered sets of contiguous DNA that represent regions of a chromosome, or even a whole chromosome. Once genetic markers define the region containing the sought-after gene, cloned pieces from the physical map provide a resource from which investigators can then isolate the gene. A copied replica of 98 percent of the human genome, consisting of thousands of linked pieces of DNA, is complete and meets the genome project's goal for physical mapping.<sup>5</sup> This map contains over 41,000 DNA markers, known as "sequence-tagged sites," or STSs, that properly align the pieces. With this density of markers, most genes in the human genome should lie within fewer than 100,000 bases of an STS.

The Human Genome Project also recognized from its inception its responsibility not only to develop gene-finding and analysis technology, but also to address the broader societal implications of these newfound abilities to decipher genetic information. So the project commits 5 percent of its annual research budget to a program that addresses the ethical, legal, and social implications of genome research (the ELSI program). This program has focused on four high-priority areas: the use and interpretation of genetic information, clinical integration of genetic technology, issues surrounding genetics research, and public and professional education about these issues.

The plan covering the years 1993 through 1998 expanded the mapping goals and explicitly included new objectives for identifying and mapping not just systems of anonymous markers but the genes themselves.<sup>6</sup> Scientists at the National Library of Medicine, at genome research centers, and in private industry later began a program to position expressed DNA from gene regions, or expressed sequence tags (ESTs), on the physical map. After two editions, this gene map represents the most extensive effort so far to locate and identify the 80,000 genes in the human genome. Over 38,000 gene tags have been placed on the map, giving disease-gene hunters a ready list of candidate genes residing in the chromosomal neighborhood they know is involved in a disease.<sup>5</sup>

Before 1996, the goals of DNA sequencing were largely targeted toward genomes of model organisms. From those efforts, scientists have sequenced the genomes of *Saccharomyces cerevisiae*, a species of yeast valuable to biologists and commonly used by bakers and brewers,<sup>7</sup> and *Escherichia coli*, a mainstay

of basic biology and the biotechnology industry.<sup>8</sup> The first complete genome sequence of a multicellular organism, the roundworm *Caenorhabditis elegans*, was completed in late 1998.

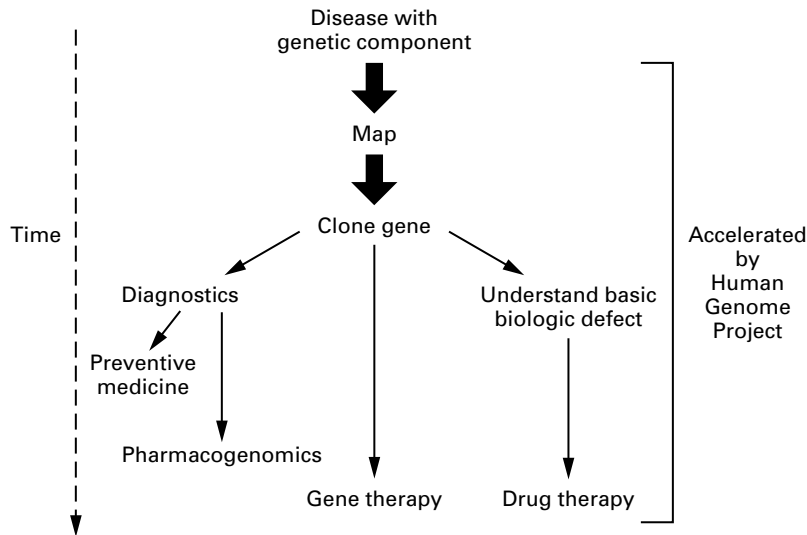
The yeast genome, which was the first genome of a eukaryote to be completely deciphered, contains 12,057,500 base pairs in its nuclear DNA. Containing some 6000 genes arranged on 16 chromosomes, yeast has already provided biologists with a valuable resource for determining the function of individual human genes involved in medical disorders such as cancer. Now, the complete sourcebook for that organism will allow scientists, for the first time, to piece together information that will provide a comprehensive look at how all the genes in a eukaryotic cell function as an integrated system.

The 100-million-base-pair *C. elegans* genome is distributed among six chromosomes and contains over 19,000 genes. Although barely visible with the naked eye, the tiny roundworm has become an invaluable tool for studying biologic processes such as development, neurobiology, and aging. The worm project began in 1990, as a collaboration between researchers at Washington University, in St. Louis, and at the Sanger Centre, in Cambridge, England. Computer comparisons between the worm sequence and the sequences of other organisms have shown that about 74 percent of named human genes have a definite homologue in *C. elegans*.

The 1998–2003 plan can be thought of as the development of a new and more diverse set of power tools, all of which are to be given away. These tools include the development of the catalogue of variations in the human DNA sequence; new forms of technology and new strategies for studying gene function on a whole-genome scale; and new areas of research in the ELSI program (the "safety goggles"), such as identifying and addressing issues that link genetics to personal identity and racial or ethnic background and examining the implications of these links for philosophical and religious traditions.

#### IMPLICATIONS FOR UNDERSTANDING GENETIC ILLNESS

Maps and other forms of genome technology provide the tools for a gene-isolation technique known as positional cloning.<sup>9</sup> This technique allows a researcher to confirm the genetic basis of a disease and identify the responsible gene, even when little is known about the gene's function. So far, over 100 disease-linked genes have been isolated with the use of the positional-cloning technique. Whereas gene discovery by this route once took years to decades, an investigator using these powerful tools can now sometimes map and isolate a gene in a matter of weeks. Increasingly, gene hunters are combining positional-cloning techniques with information in EST data bases to narrow their gene searches to rational



**Figure 1.** Steps Involved in the Genetic Revolution in Medicine.

Uncovering the genetic contributions to an illness is accomplished by cloning the gene for the disease, with the use of the tools of the Human Genome Project. Once the contributing genes and their disease-predisposing variants have been identified, diagnostic tests can be developed to predict future risk — but these tests are most effective when a preventive strategy is available to reduce the risk in persons found to be predisposed to a particular disease. Another rapidly developing application of diagnostics is pharmacogenomics, the prediction of responsiveness to drugs. Ultimately, the real payoff of genetic research will be the development of new gene therapies and drug therapies, but they will generally require many more years of intensive research.

candidates. This method, called positional candidate cloning,<sup>10</sup> has been used to isolate many altered genes associated with human disease.

Gene isolation provides the best hope for understanding human disease at its most fundamental level (Fig. 1). Knowledge about genetic control of cellular functions will underpin future strategies to prevent or treat disease phenotypes. The recent isolation of genes for Parkinson's disease,<sup>11-14</sup> for example, has greatly advanced molecular research on this baffling disease. In one study of families with early-onset Parkinson's disease, gene hunters mapped a suspect gene to a region of chromosome 4.<sup>11</sup> The region contained approximately 100 genes, among which was 1 known to encode the protein  $\alpha$ -synuclein. Earlier research had shown that  $\alpha$ -synuclein accumulates in brain cells of people with Alzheimer's disease,<sup>15</sup> and people with Parkinson's disease have similar deposits in the substantia nigra. In just a few months, the researchers showed conclusively that a missense mutation in the  $\alpha$ -synuclein gene caused Parkinson's disease in the study families. Further research has shown that a mutation in a gene encoding a protein critical to the breakdown of  $\alpha$ -synuclein and other proteins also results in the Parkinson's disease phenotype in a different family.<sup>14</sup> Although most cases of Parkinson's disease appear to have limited heritability, studying rare families of

this sort provides crucial clues to the pathways involved —  $\alpha$ -synuclein is found in the Lewy bodies in virtually all cases of Parkinson's disease. An understanding of the genetic control of the proteolytic processes of brain proteins may provide new targets for interventions in a number of related neurodegenerative disorders characterized by the accumulation of protein deposits, including Alzheimer's disease, Huntington's disease, and spinocerebellar ataxia.

Even before a gene's role in disease is fully understood, diagnostic applications can be useful in preventing or minimizing the development of health consequences (Fig. 1). DNA tests that look for the presence of disease-linked mutations, for example, are proving to be the most immediate commercial application of gene discovery and the one now used most frequently by clinicians. These tests may help establish the diagnosis of a genetic disease, foreshadow the development of disease later in life, or identify healthy heterozygote carriers of recessive diseases. Genetic tests can be performed at any stage of the human life cycle, and the sampling procedures are becoming less invasive. Whereas genetic testing was once sought almost exclusively by couples with a family history of early-onset disease, for the purpose of family planning, information about genetic status is increasingly sought by persons who wish to learn about their own predisposition to adult-onset illness.

In a growing number of instances, strategies can be implemented to reduce or prevent illness when a genetic cause or predisposition is known. Successes in reducing disease through treatment have been achieved for the hereditary disorders hemochromatosis, phenylketonuria, and familial hypercholesterolemia, among others. Risk reduction through early detection and lifestyle changes may be possible in the case of disorders associated with predisposing mutations, such as some cancers. As therapies build on knowledge gained about the molecular basis of disease, increasing numbers of illnesses that are now refractory to treatment may yield to molecular medicine in the future.

The recent discovery of an altered gene (*HFE*) that leads to hereditary hemochromatosis,<sup>16</sup> a common disorder of iron metabolism, provides an interesting example of the potential for using information about mutations to prevent an adult-onset disease phenotype. A recessive condition, hereditary hemochromatosis affects about 1 in 300 persons of northern European descent and is easily treatable if diagnosed early. Its major symptoms — liver cirrhosis, heart failure, diabetes, arthritis, and other organ damage — do not occur until midlife and are easily misdiagnosed. Untreated, the disease causes early death, but treatment by phlebotomy to remove excess iron allows people with hereditary hemochromatosis to live a normal life span. A single substitution of the amino acid tyrosine for cysteine at codon 282 accounts for the majority of cases.<sup>17</sup>

At first glance, hereditary hemochromatosis seems to be an ideal target for public health approaches to the prevention of hereditary disease: the disorder is common, the number of disease-linked mutations in the gene are few, and effective treatment can minimize or eliminate the effects of the disease. But closer examination reveals a number of complexities that have so far militated against the rapid introduction of this genetic test as a tool for disease prevention.<sup>16</sup> Because the penetrance of the altered *HFE* gene is reduced, especially in females, clinical signs can range from none that are detectable to severe organ damage from iron overload. At the moment, simple detection of the mutation does not predict the most likely clinical course. Before population testing for *HFE* mutations is considered, further research is needed to explain the variations in phenotype among mutation carriers and to correlate the genotype more closely with health outcomes.

#### IMPLICATIONS FOR THE STUDY OF COMMON DISORDERS

The rather straightforward mendelian rules that govern the inheritance of disease traits have been worked out for many rare disorders that result from highly penetrant changes in a single gene. But teasing out the genetic components of the so-called complex disorders — diabetes, heart disease, most com-

mon cancers, autoimmune disorders, and psychiatric disorders — that result from the interplay of environment, lifestyle, and the small effects of many genes remains a formidable task. Most of the successful efforts to identify genes associated with common diseases have focused on highly heritable subgroups, including the *BRCA1*<sup>18</sup> and *BRCA2*<sup>19</sup> genes in breast cancer, the gene for hepatocyte nuclear factor 4 $\alpha$  (*HNF-4 $\alpha$* ) in maturity-onset diabetes of the young (MODY) type 1,<sup>20</sup> the gene for glucokinase (*GCK*) in MODY type 2,<sup>21</sup> the gene for hepatocyte nuclear factor 1 $\alpha$  (*HNF-1 $\alpha$* ) in MODY type 3,<sup>22</sup> the gene for human Mut S homologue 2 (*hMSH2*)<sup>23,24</sup> and the gene for human Mut L homologue 1 (*hMLH1*)<sup>25</sup> in hereditary nonpolyposis colon cancer, and the gene for  $\alpha$ -synuclein in Parkinson's disease.<sup>11</sup> Linkage analysis and positional-cloning techniques are well suited to discovering genes with such strong influences. But these strategies are not as easily applied to the multiple, low-penetrance variants, which in the aggregate account for a larger percentage of illnesses. Identification of weakly penetrant alleles that contribute to common disorders requires new and more powerful approaches.

To assist in these efforts, the Human Genome Project is initiating new studies of genetic variation in the human population to provide a dense map of common DNA variants. DNA sequence variations include insertions and deletions of nucleotides, differences in the copy number of repeated sequences, and single-nucleotide polymorphisms, or SNPs (pronounced "snips"), which occur most frequently throughout the human genome. About 1 in every 300 to 500 bases in human DNA may be a SNP.

SNPs can be used as markers in whole-genome linkage analysis of families with affected members, as well as in association studies of individuals in a population. Association studies may directly test a variant with potential functional importance or may take advantage of the phenomenon of linkage disequilibrium — in which a marker and a gene are inherited together — to map gene variants associated with disease. Because the human species consists of relatively few generations, recombination events have not disrupted linkage disequilibrium over distances of 3000 to 100,000 bases in most populations. Consequently, association studies view large human populations as evolutionary families and do not rely on studies of nuclear families for gene mapping.<sup>26,27</sup>

Some SNPs may contribute directly to a trait or disease phenotype by altering function. Though most SNPs are located outside protein-coding sequences, those within coding sequences, called cSNPs, are of particular interest because they are more likely to affect gene function. A large, well-characterized collection of SNPs will become increasingly important for the discovery of DNA sequence variations that affect biologic function. Work is already under way

with NIH support to develop a catalogue of 60,000 or more SNPs. A recently formed pharmaceutical consortium will support the production of 300,000 more, with the work being done at the publicly funded genome centers and all the data deposited in the public domain. This is a wonderful example of a public-private partnership to develop a powerful set of research tools that all can use.

#### NEW FORMS OF TECHNOLOGY FOR GENETIC ANALYSIS AND RISK ASSESSMENT

The transition from genetics to genomics marks the evolution from an understanding of single genes and their individual functions to an understanding of the actions of multiple genes and their control of biologic systems. Whereas the tools of the Human Genome Project initially advanced research on single genes, they are now forming the basis for genomic-scale analysis of the human organism.

The so-called DNA chip currently provides one promising approach to genome-scale studies of genetic variation,<sup>28</sup> detection of heterogeneous gene mutations,<sup>29</sup> and gene expression.<sup>30</sup> The result of an adaptation of dot blot hybridization techniques, DNA chips, also called microarrays, generally consist of a thin slice of glass or silicon about the size of a postage stamp on which threads of synthetic nucleic acids are arrayed. Sample probes are added to the chip, and matches are read by an electronic scanner. As with semiconductors, the capacity of DNA chips has doubled about every two years, so chips that held a few hundred arrays not so long ago now hold hundreds of thousands.

Microarray technology has been applied to the detection of DNA variations as well as expression of messenger RNA in individual cells and tissues. Microarrays are used clinically to detect human immunodeficiency virus sequence variations, p53 gene mutations in breast tissue, and expression of cytochrome P450 genes. In the laboratory, microarray technology has also been applied to genomic comparisons across species,<sup>31</sup> genetic recombination,<sup>32</sup> and large-scale analysis of gene copy number and expression, as well as protein expression, in cancerous tissues.<sup>33</sup>

Use of microarrays and other new technologies to detect DNA variations holds promise, along with family histories and data from large population studies, for establishing a person's risk of contracting common, adult-onset disorders. A base-line genome scan could provide helpful information about a person's risk profile and point to the prevention strategies — if available — that should be used.

#### GENETIC KNOWLEDGE AND INDIVIDUALIZED MEDICINE

Identifying human genetic variations will eventually allow clinicians to subclassify diseases and adapt

therapies to the individual patient. There may be large differences in the effectiveness of medicines from one person to the next. Toxic reactions can also occur and in many instances are likely to be a consequence of genetically encoded host factors. That basic observation has spawned the burgeoning new field of pharmacogenomics, which attempts to use information about genetic variation to predict responses to drug therapies.

For example, researchers discovered that patients with Alzheimer's who have the  $\epsilon 4$  subtype of the gene for apolipoprotein E (*APOE*  $\epsilon 4$ ), which affects cholinergic function in the brain, are less likely to benefit from the cholinomimetic drug tacrine than are patients without this subtype.<sup>34</sup> This finding will help in the analysis of data from clinical trials of Alzheimer's therapies and will promote the development of new therapies specifically designed for *APOE*  $\epsilon 4$  carriers.

In another example, cholesteryl ester transfer protein (CETP) plays an important part in the metabolism of high-density lipoprotein, a lipoprotein associated with lowered susceptibility to atherosclerosis. A certain genetic variant of the *CETP* gene is correlated with higher plasma CETP levels and lower levels of plasma high-density lipoprotein. One study showed that in men who carried this genetic variant, treatment with pravastatin slowed the progression of coronary atherosclerosis.<sup>35</sup> This finding may allow physicians to predict which patients with coronary artery disease will benefit from treatment with pravastatin.

In a third example, the formation of venous blood clots in the brain and legs is a rare but serious side effect of birth-control pills. One study has shown a dramatically increased risk of cerebral-vein thrombosis among women taking oral contraceptives who also carry the blood-clotting variant factor V Leiden.<sup>36</sup> The risk of other venous thrombotic events is also increased in this group. Foreknowledge of the presence of this variant and consideration of alternative forms of birth control might be useful in minimizing the risk of thrombosis in these women.

Not only will genetic tests predict responsiveness to drugs on the market today, but also genetic approaches to disease prevention and treatment will include an expanding array of gene products for use in developing tomorrow's drug therapies. Since the Food and Drug Administration's approval of recombinant human insulin in 1982, over 50 additional gene-based drugs have become available for clinical use. These include drugs for the treatment of cancer, heart attack, stroke, and diabetes, as well as many vaccines.<sup>37</sup>

Not all therapeutic advances for gene discovery will be genes or gene products. In other instances, molecular insights into a disorder, derived from gene discovery, will suggest a new treatment. Sodium phen-

ylbutyrate, for example, which is approved for the regulation of blood ammonia levels, is being tested in clinical trials for the treatment of cystic fibrosis.<sup>38</sup> The main clinical phenotype in people with cystic fibrosis results from a mutation in the gene for cystic fibrosis transmembrane conductance regulator (CFTR) protein. The mutation prevents normal amounts of CFTR protein from crossing the cell membrane, diminishing the ability of chloride and water to enter and exit the cell. Sodium phenylbutyrate apparently stimulates expression of CFTR protein, allowing more of it to reach the correct location.

### ETHICAL, LEGAL, AND SOCIAL IMPLICATIONS

One of the most active areas of the ELSI program has been policy development related to the privacy and fair use of genetic information, particularly in health insurance, employment, and medical research. Debates in this area focus largely on the potential of genetic information to predict an increased likelihood of the eventual development of a disease phenotype in a currently healthy person.

Although many states have attempted to address "genetic discrimination" in health insurance and employment, federal legislation would provide the most comprehensive protection. Concern about the confidentiality of genetic information may make people reluctant to volunteer for studies involving disease-linked gene mutations or genetic therapy, for fear that the results could result in the loss of a job or the loss of insurance coverage.

Largely on the basis of recommendations formulated in workshops held by the Human Genome Project and the National Action Plan on Breast Cancer,<sup>39,40</sup> the Clinton administration endorsed the need for congressional action to protect against genetic discrimination in health insurance and employment. In 1996, Congress enacted the Health Insurance Portability and Accountability Act, which represented a large step toward protecting access to health insurance in the group-insurance market but left several serious gaps in the individual-insurance market that must still be closed.

In the area of workplace discrimination, the Equal Employment Opportunity Commission has interpreted the Americans with Disabilities Act as covering on-the-job discrimination based on "genetic information relating to illness, disease or other disorders."<sup>41</sup> But no claims of genetic discrimination have been brought to the commission, and the guidance has yet to be tested in court, so the degree of protection actually provided by the act remains uncertain.

In the area of privacy, as part of the partnership between the National Action Plan on Breast Cancer and the Human Genome Project, medical researchers, policy makers, and representatives of law, government, the insurance industry, and public health

have recently assessed current policies and practices designed to protect confidentiality in genetics research and have identified areas where new or modified policies or practices might enhance the protection of privacy and promote the conduct of research. The group is developing a set of principles for researchers, research institutions, state and federal agencies, and policy makers to consider in formulating measures to protect the privacy of research information.

Other important steps have been taken to ensure the responsible integration of genetic tests into clinical practice. For the most part, genetic testing in the United States has developed successfully, providing options for avoiding, preventing, and treating inherited disorders. But the rapid pace of test development combined with the rush to market new products may create an environment in which the tests are made available before they have been adequately validated. On the recommendation of the Task Force on Genetic Testing,<sup>42</sup> assembled by the Human Genome Project's NIH-DOE ELSI Working Group, the Secretary of the Department of Health and Human Services has established an advisory panel to ensure the safe introduction of genetic tests into clinical practice.

Completion of the first human-genome sequence and the expansion of human genetic research to include studies of genetic variation among subpopulations have raised new questions about ethical, legal, and social issues. The 1998-2003 plan includes an examination of these issues as well as the integration of genetic technology and information into health care and public health activities; the use of knowledge about genomics and gene-environment interactions in nonclinical settings; examination of a variety of philosophical, theological, and ethical perspectives on new genetic knowledge; and consideration of the ways in which racial, ethnic, and socioeconomic factors affect the use, understanding, and interpretation of genetic information, the use of genetic services, and the development of policy.<sup>3</sup>

### A HYPOTHETICAL CASE IN 2010

General visions of gene-based medicine in the future are useful, but many health care providers are probably still puzzled by how it will affect the daily practice of medicine in a primary care setting. A hypothetical clinical encounter in 2010 is described here.

John, a 23-year-old college graduate, is referred to his physician because a serum cholesterol level of 255 mg per deciliter was detected in the course of a medical examination required for employment. He is in good health but has smoked one pack of cigarettes per day for six years. Aided by an interactive computer program that takes John's family history, his physician notes that there is a strong paternal history of myocardial infarction and that John's father died at the age of 48 years.

To obtain more precise information about his risks of contracting coronary artery disease and other illnesses in the future, John agrees to consider a battery of genetic tests that are available in 2010. After working through an interactive computer program that explains the benefits and risks of such tests, John agrees (and signs informed consent) to undergo 15 genetic tests that provide risk information for illnesses for which preventive strategies are available. He decides against an additional 10 tests involving disorders for which no clinically validated preventive interventions are yet available.

A cheek-swab DNA specimen is sent off for testing, and the results are returned in one week (Table 1). John's subsequent counseling session with the physician and a genetic nurse specialist focuses on the conditions for which his risk differs substantially (by a factor of more than two) from that of the general population. Like most patients, John is interested in both his relative risk and his absolute risk.

John is pleased to learn that genetic testing does not always give bad news — his risks of contracting prostate cancer and Alzheimer's disease are reduced, because he carries low-risk variants of the several genes known in 2010 to contribute to these illnesses. But John is sobered by the evidence of his increased risks of contracting coronary artery disease, colon cancer, and lung cancer. Confronted with the reality of his own genetic data, he arrives at that crucial "teachable moment" when a lifelong change in health-related behavior, focused on reducing specific risks, is possible. And there is much to offer. By 2010, the field of pharmacogenomics has blossomed, and a prophylactic drug regimen based on the knowledge of John's personal genetic data can be precisely prescribed to reduce his cholesterol level and the risk of coronary artery disease to normal levels. His risk of colon cancer can be addressed by beginning a program of annual colonoscopy at the age of 45, which in his situation is a very cost-effective way to avoid colon cancer. His substantial risk of contracting lung cancer provides the key motivation for him to join a support group of persons at genetically high risk for serious complications of smoking, and he successfully kicks the habit.

This vision of genetically based, individualized preventive medicine is exciting, and it could make a profound contribution to human health. For this vision to be realized, however, protections against the misuse of genetic information will need to be firmly in place. And another critical challenge must be met: physicians, nurses, and other health care providers will need to become familiar with the emerging field of genetic medicine. The need for medical genetic specialists who can sort out the most complex cases will be considerable, but there will not be enough of them to go around, and genetic medicine will be practiced for the most part by primary care provid-

**TABLE 1.** RESULTS OF GENETIC TESTING IN A HYPOTHETICAL PATIENT IN 2010.

CONDITION	GENES INVOLVED*	RELATIVE RISK	LIFETIME RISK (%)
Reduced risk			
Prostate cancer	<i>HPC1, HPC2, HPC3</i>	0.4	7
Alzheimer's disease	<i>APOE, FAD3, XAD</i>	0.3	10
Elevated risk			
Coronary artery disease	<i>APOB, CETP</i>	2.5	70
Colon cancer	<i>FCC4, APC</i>	4	23
Lung cancer	<i>NAT2</i>	6	40

\**HPC1, HPC2, and HPC3* are the three genes for hereditary prostate cancer; *APOE* is the gene for apolipoprotein E; *FAD3* and *XAD* are hypothetical genes for familial Alzheimer's dementia; *APOB* is the gene for apolipoprotein B; *CETP* is the gene for cholesteryl ester transfer protein; *FCC4* is the hypothetical gene for familial colon cancer; *APC* is the gene for adenomatous polyposis coli; and *NAT2* is the gene for *N*-acetyltransferase 2.

ers. Numerous surveys have shown that we are not currently prepared for this — most practitioners in primary care have not had a single hour of instruction in genetics as part of their formal training.<sup>43,44</sup>

To meet this urgent need for education, the National Human Genome Research Institute, the American Medical Association, and the American Nurses Association have recently banded together to form the National Coalition for Health Professional Education in Genetics (NCHPEG). NCHPEG (accessible at <http://www.nchpeg.org>) is a national effort to promote professional education and access to information about advances in human genetics. An interdisciplinary group, NCHPEG comprises the leaders of approximately 100 diverse health professional organizations, consumer and voluntary groups, government agencies, private industry, managed-care organizations, and professional genetics societies. By facilitating frequent and open communication, NCHPEG seeks to capitalize on the collective expertise and experience of its members and to reduce duplication of effort.

## CONCLUSIONS

Since the beginning of the Human Genome Project, the increasing detail and quality of genome maps have reduced the time it takes to find a gene from years to months to weeks. We have learned the location of approximately half the 80,000 or so genes packaged on human chromosomes. Genome researchers are at the forefront of cross-disciplinary technological development for the identification and analysis of not just single genes but also whole genomes.

Most recently, we have begun to get the first glimpses of the human genome at its most detailed level: about 15 percent of the 3 billion bits of DNA code that spell out instructions for every function a human body carries out is now available in public

data bases. In just 12 months, a highly useful working draft of 90 percent of the genome will be available, and by 2003, the full DNA sequence of the human will give us unprecedented opportunities to observe and understand the literal Book of Life.

As hoped, genome maps, sequence data, and analytic tools are providing a robust technological infrastructure for biomedical research well into the next century. The Human Genome Project's commitment to make these powerful tools freely available means that any scientist with access to the Internet can already apply genomic approaches to individual research problems. High-technology research tools developed by the Human Genome Project will offer unprecedented opportunities to study human biology and disease in entirely new ways.

As genome technology moves from the laboratory to the health care setting, new methods will make it possible to read the instructions contained in an individual person's DNA. Such knowledge may foretell future disease and alert patients and their health care providers to undertake better preventive strategies. In the wrong hands, however, that same information could be used to discriminate against or stigmatize a person. In response to this concern, the Human Genome Project has catalyzed the development of policy options for lawmakers to consider in their efforts to prohibit genetic discrimination and to protect the privacy of genetic information. The stage is set to solve these vexing problems with effective federal legislation, but this window of opportunity will not stay open indefinitely.

Writing 97 years ago, Sir William Osler described the goals of medicine this way: "To wrest from nature the secrets which have perplexed philosophers in all ages, to track to their sources the causes of disease, to correlate the vast stores of knowledge, that they may be quickly available for the prevention and cure of disease — these are our ambitions."<sup>45</sup> The Human Genome Project, with its audacious goal of providing the tools to uncover the hereditary factors in virtually every disease, has become a major modern component of Osler's vision. The genetic revolution in medicine is under way.

*I am indebted to Leslie Fink and Dr. Margaret Bouvier for their assistance in the preparation of the manuscript.*

## REFERENCES

1. National Research Council Committee on Mapping and Sequencing the Human Genome. Mapping and sequencing the human genome. Washington, D.C.: National Academy Press, 1988.
2. Department of Health and Human Services, Department of Energy. Understanding our genetic inheritance: the U.S. Human Genome Project: the first five years: FY 1991-1995. Washington, D.C.: Government Printing Office, 1990. (NIH publication no. 90-1590.)
3. Collins FS, Patrinos A, Jordan E, Chakravarti A, Gesteland R, Walters I. New goals for the U.S. Human Genome Project: 1998-2003. *Science* 1998;282:682-9.
4. Murray JC, Buetow KH, Weber JL, et al. A comprehensive human linkage map with centimorgan density. *Science* 1994;265:2049-54.
5. Deloukas P, Schuler GD, Gyapay G, et al. A physical map of 30,000 human genes. *Science* 1998;282:744-6.
6. Collins FS, Galas D. A new five-year plan for the U.S. Human Genome Project. *Science* 1993;262:43-6.
7. The yeast genome directory. *Nature* 1997;387:Suppl:5.
8. Blattner FR, Plunkett G III, Bloch CA, et al. The complete genome sequence of *Escherichia coli* K-12. *Science* 1997;277:1453-74.
9. Collins FS. Positional cloning: let's not call it reverse anymore. *Nat Genet* 1992;1:3-6.
10. *Idem*. Positional cloning moves from perditional to traditional. *Nat Genet* 1995;9:347-50. [Erratum, *Nat Genet* 1995;11:104.]
11. Polymeropoulos MH, Lavedan C, Leroy E, et al. Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science* 1997;276:2045-7.
12. Gasser T, Muller-Myhsok B, Wszolek ZK, et al. A susceptibility locus for Parkinson's disease maps to chromosome 2p13. *Nat Genet* 1998;18:262-5.
13. Jones AC, Yamamura Y, Almasy L, et al. Autosomal recessive juvenile parkinsonism maps to 6q25.2-q27 in four ethnic groups: detailed genetic mapping of the linked region. *Am J Hum Genet* 1998;63:80-7.
14. Leroy E, Boyer R, Auburger G, et al. The ubiquitin pathway in Parkinson's disease. *Nature* 1998;395:451-2.
15. Ueda K, Fukushima H, Masliah E, et al. Molecular cloning of cDNA encoding an unrecognized component of amyloid in Alzheimer disease. *Proc Natl Acad Sci U S A* 1993;90:11282-6.
16. Burke W, Thomson E, Khoury MJ, et al. Hereditary hemochromatosis: gene discovery and its implications for population-based screening. *JAMA* 1998;280:172-8.
17. Feder JN, Gnirke A, Thomas W, et al. A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. *Nat Genet* 1996;13:399-408.
18. Miki Y, Swensen J, Shattuck-Eidens D, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 1994;266:66-71.
19. Wooster R, Bignell G, Lancaster J, et al. Identification of the breast cancer susceptibility gene BRCA2. *Nature* 1995;378:789-92. [Erratum, *Nature* 1996;379:749.]
20. Yamagata K, Furuta H, Oda O, et al. Mutations in the hepatocyte nuclear factor-4 $\alpha$  gene in maturity-onset diabetes of the young (MODY 1). *Nature* 1996;384:458-60.
21. Froguel P, Zouali H, Vionnet N, et al. Familial hyperglycemia due to mutations in glucokinase: definition of a subtype of diabetes mellitus. *N Engl J Med* 1993;328:697-702.
22. Yamagata K, Oda N, Kaisaki PJ, et al. Mutations in the hepatocyte nuclear factor-1 $\alpha$  gene in maturity-onset diabetes of the young (MODY3). *Nature* 1996;384:455-8.
23. Fishel R, Lescoe MK, Rao MR, et al. The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell* 1993;75:1027-38. [Erratum, *Cell* 1994;77:167.]
24. Leach FS, Nicolaidis NC, Papadopoulos N, et al. Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell* 1993;75:1215-25.
25. Papadopoulos N, Nicolaidis NC, Wei YF, et al. Mutation of a mutL homolog in hereditary colon cancer. *Science* 1994;263:1625-9.
26. Schafer AJ, Hawkins JR. DNA variation and the future of human genetics. *Nat Biotechnol* 1998;16:33-9.
27. Collins FS, Guyer MS, Chakravarti A. Variations on a theme: cataloging human DNA sequence variation. *Science* 1997;278:1580-1.
28. Wang DG, Fan J-B, Siao C-J, et al. Large-scale identification, mapping and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 1998;280:1077-82.
29. Hacia JG, Brody LC, Chee MS, Fodor SP, Collins FS. Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two-colour fluorescence analysis. *Nat Genet* 1996;14:441-7.
30. DeRisi J, Penland L, Brown PO, et al. Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat Genet* 1996;14:457-60.
31. Hacia JG, Makalowski W, Edgemon K, et al. Evolutionary sequence comparisons using high-density oligonucleotide arrays. *Nat Genet* 1998;18:155-8.
32. Winzeler EA, Richards DR, Conway AR, et al. Direct allelic variation scanning of the yeast genome. *Science* 1998;281:1194-7.
33. Kononen J, Bubendorf L, Kallioniemi A, et al. Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat Med* 1998;4:844-7.
34. Poirier J, Delisle M-C, Quirion R, et al. Apolipoprotein E4 allele as a predictor of cholinergic deficits and treatment outcome in Alzheimer disease. *Proc Natl Acad Sci U S A* 1995;92:12260-4.

35. Kuivenhoven JA, Jukema JW, Zwinderman AH, et al. The role of a common variant of the cholesteryl ester transfer protein gene in the progression of coronary atherosclerosis. *N Engl J Med* 1998;338:86-93.
36. Martinelli I, Sacchi E, Landi G, Taioli E, Duca F, Mannucci PM. High risk of cerebral-vein thrombosis in carriers of a prothrombin-gene mutation and in users of oral contraceptives. *N Engl J Med* 1998;338:1793-7.
37. *New medicines in development: biotechnology 1998*. Washington, D.C.: Pharmaceutical Research and Manufacturers of America, 1998:35.
38. Rubenstein RC, Zeitlin PL. A pilot clinical trial of oral sodium 4-phenylbutyrate (Buphenyl) in deltaF508-homozygous cystic fibrosis patients: partial restoration of nasal epithelial CFTR function. *Am J Respir Crit Care Med* 1998;157:484-90.
39. Hudson KL, Rothenberg KH, Andrews LB, Kahn MJ, Collins FS. Genetic discrimination and health insurance: an urgent need for reform. *Science* 1995;270:391-3.
40. Rothenberg K, Fuller B, Rothstein M, et al. Genetic information and the workplace: legislative approaches and policy challenges. *Science* 1997; 275:1755-7.
41. Compliance manual. Section 902. Washington, D.C.: Equal Employment Opportunity Commission, March 15, 1995.
42. Holtzman NA, Watson MS, eds. Promoting safe and effective genetic testing in the United States: final report of the Task Force on Genetic Testing. Baltimore: Johns Hopkins University Press, 1998.
43. Scanlon C, Fibison W. Managing genetic information: implications for nursing practice. Washington, D.C.: American Nurses Publishing, 1995:1-50.
44. National Center for Genome Resources. National survey of public and stakeholders' attitudes and awareness of genetic issues. Washington, D.C.: Schulman, Ronca, and Bucuvalas, 1996.
45. Osler W. Chauvinism in medicine. *Montreal Med J* 1902;31:684-99.