Implications of the Human Genome Project for Medical Science

Francis S. Collins, MD, PhD
Victor A. McKusick, MD

The year 2000 marked both the start of the new millennium and the nouncement that the vast majority of the human genome had been quenced. Much work remains to understand how this "instruction book human biology" carries out its multitudes of functions. But the con quences for the practice of medicine are likely to be profound. Genetic p diction of individual risks of disease and responsiveness to drugs will re the medical mainstream in the next decade or so. The development of signer drugs, based on a genomic approach to targeting molecular pa ways that are disrupted in disease, will follow soon after. Potential misu of genetic information, such as discrimination in obtaining health insuran and in the workplace, will need to be dealt with swiftly and effectively. G nomic medicine holds the ultimate promise of revolutionizing the diagno and treatment of many illnesses.

JAMA. 2001;285:540-544 www.jama.

UNTIL RECENTLY, MANY PHYSICIANS and other health care professionals considered medical genetics as the province of specialists in tertiary care medical centers, who spent their time evaluating unusual cases of mendelian disorders, birth defect syndromes, or chromosomal anomalies. Asked whether genetics was a part of their everyday practice, most primary care practitioners would say no. That is all about to change.

To be sure, there are numerous medical conditions found in children and adults that have a strong, indeed predominant, genetic basis. The continuously updated Online Mendelian Inheritance in Man (OMIM) lists many thousands of such conditions,[1] but offers a far too narrow view of the contribution of genetics to medicine. Except for some cases of trauma, it is fair to say that virtually every human illness has a hereditary component.[2] While common diseases, such as diabetes mellitus, heart disease, cancer, and the major mental illnesses, do not follow mendelian inheritance patterns, there is ample evidence from twin and pedigree studies over many decades showing that all of these disorders have important hereditary influences. In fact, for many common illnesses of developed countries, the strongest predictor of risk is family history.

The role of heredity in most diseases is thus not in itself a new revelation. But in the past, it was considered unlikely that much could be done with this information other than to guide medical surveillance based on careful family history taking. A sea change is now underway, and it is likely that the molecular basis for these hereditary influences on common illnesses soon will be uncovered. Even though on average the quantitative contribution of heredity to the etiological characteristics of diseases like diabetes mellitus or hypertension may be modest, uncovering the pathways involved in disease pathogenesis will have broad consequences, pointing toward possible environmental triggers as well. The implications for diagnostics, preventive medicine, and therapeutics will be profound.

Genetics in the 20th Century

In the spring of 1900, 3 different investigators rediscovered Mendel's laws.[3] With Garrod's recognition of their application to human inborn errors of metabolism, the science of human genetics acquired a foundation. But it remained for Watson and Crick half a century later to uncover the chemical basis of heredity, with their elucidation of the double helical structure of DNA.[4] The role RNA as a messenger and the genetic co that allows RNA to be translated to pr tein emerged over the next 15 years. Th was followed by the advent of recomb nant DNA technology in the 1970s, o fering the ability to obtain pure prepa rations of a particular DNA segment However, sequencing of DNA was di ficult until Sanger and Gilbert indepen dently derived methods of sequencin DNA in 1977.[5,6] (It is remarkable in deed that the Sanger dideoxy method fo DNA sequencing remains the basic tech nology on which the genetic revolution is being built, albeit with major ad vances in automation of the analysis that have come along in the last 15 years.)

Author Affiliations: National Human Genome Research Institute, National Institutes of Health, Bethesda, Md (Dr Collins); Johns Hopkins University School of Medicine, Baltimore, Md (Dr McKusick).
Corresponding Author and Reprints: Francis S. Collins, MD, PhD, National Human Genome Research Institute, National Institutes of Health, 31 Center Dr, MSC 2152, Bldg 31 Room 4809, Bethesda, MD 20892-2152 (e-mail: fc23a@nih.gov).

use of variable DNA markers for analysis of human disorders was th in 1980.[7] Mapping of disor- linkage previously had been se- limited by the relatively small er of usable protein markers, such od groups. The notion that any elian disorder could be mapped chromosomal region caught the nation of geneticists. An early and ing success of this approach, the ing of the Huntington disease gene omosome 4 in 1983, gave a burst nfidence to this adventurous new ach.[8] But the difficulty of going linked marker to the actual dis- locus proved profoundly diffi- Years of work were required to candidate region and search for ual candidate genes, and many in- ators in the 1980s longed for a systematic approach to the ge-

the same time, potential ad- s in mapping and sequencing ology led certain scientific lead- articularly in the US Department gy, to propose the possibility of anized effort to sequence the en- uman genome. In the late 1980s controversy raged about such osals, with many in the scientific unity deeply concerned that this echnologically impossible and o consume vast amounts of fund- at might be taken away from other productive hypothesis-driven re- ch. But with the strong support of el of the National Academy of Sci- s,[9] and the enthusiasm of a few ers in the US Congress, the Hu- Genome Project (HGP) was ini- d in the United States by the Na- al Institutes of Health and the artment of Energy in 1990.[10]

Human Genome Project

m the outset, it was realized that a ailed set of plans and milestones uld be necessary for a project of this gnitude. The technology for carry- out actual large-scale sequencing had advanced to the point of being able tackle the 3 billion base pairs of the uman genome in 1990 nor were the necessary maps of the genome in hand to provide a scaffold for this effort.

Under the leadership of James Watson, it was decided to focus the first 5 years of the HGP on the development of genetic and physical maps of the human genome, which would themselves be of great value to scientists hunting for disease genes. The HGP also tackled mapping and sequencing of simpler model organisms, such as bacteria, yeast, the roundworm, and the fruit fly.[9-12] Considerable investments were made in improving technology. Perhaps the most unusual feature for a basic science enterprise, 3% to 5% of the budget was set aside from the outset for research on the ethical, legal, and social implications of this expected acceleration in obtaining genetic information about our species.[10] In the past, ethical, legal, and social analysis of the consequences of a scientific revolution often were relegated to other groups outside the scientific mainstream or lay dormant until a crisis developed. This time, the intention was to inspire a cohort of ethicists, social scientists, legal scholars, theologians, and others to address the coming dilemmas associated with increased knowledge about the genome, from social and legal discrimination on the basis of genetics to more philosophical issues such as genetic determinism.

The HGP has been international from the beginning. Although the United States made the largest investment, important contributions have been made by many countries, including Britain, France, Germany, Japan, China, and Canada. The original plan[9] called for completion of the sequence of the human genome by the year 2005, though there was limited confidence that this goal could be achieved. But one by one the intermediate milestones were accomplished. The HGP agreed at the outset to release all map and sequence data into the public domain. The availability of genetic and physical maps led to a considerable acceleration in the successful identification of genes involved in single gene disorders; while fewer than 10 such genes had been identified by positional cloning in 1990, that number grew to more than 100 by 1997.[13]

By 1996, the complete sequencing of several bacterial species and yeast led to the conclusion that it was time to attempt sequencing human DNA on a pilot scale. The introduction of capillary sequencing instruments and the formation of a company in the private sector promising to sequence the human genome for profitable purposes added further momentum to the effort. By 1999, confidence had gathered that acquiring the majority of the sequence of the 3 billion base pairs of the human genome could be attempted. In June 2000, both the private company and the international public sequencing consortium announced the completion of "working drafts" of the human genome sequence.

Current Research Focus

Though the working draft of the human sequence represents a major milestone, a vast amount of additional work remains to be done to understand its function.

It is necessary to complete the sequence analysis by closing the gaps and resolving ambiguities. This finishing process already has been accomplished for chromosomes 21[14] and 22[15] and will be carried out for the remainder of the genome during the next 2 years.

The genomes of other organisms also will need to be sequenced. Probably the most powerful tool to identify the coding exons, as well as the regulatory regions, is a comparison of the sequence across different genomes. For that purpose, full-scale sequencing of the laboratory mouse genome already has been initiated, and the sequencing of the rat and zebrafish genomes will not be far behind. In both the public and private sectors, serious consideration is being given to the sequencing of other large vertebrate genomes, including the pig, dog, cow, and chimpanzee.

An intense effort is under way to develop a catalog of human variation. While human DNA sequences are 99.9% identical to each other, the 0.1% of variation is expected to provide many of the clues to the genetic risk for common ill-

Research Opportunities and Forecast: Genomics	
Key Research Opportunities	**Forecast**
Define Complete List of All Human Genes and Proteins	Thousands of New Drug Targets for Heart Disease, Cancer, Diabetes, Asthma, etc
Define All Common Variants in the Genome, Determine Hereditary Factors in Virtually All Common Diseases and Refine Technology for Low-Cost Genotyping	Individualized Preventive Medicine Based on Genetic Risk
	Pharmacogenomics to Improve Outcome of Drug Therapy
	Environmental Risk Factor Assessment Becomes Individual-Specific
Determine Regulatory Signals That Affect Expression of All Human Genes in Normal or Abnormal State	Therapies for Developmental Defects
	Precise Molecular Analysis of Malignancies, Guiding Choice of Therapy
Determine Structures of All Human Proteins, Using a Combination of Experimental and Computational Methods	"Designer Drugs" Based on Precise 3-Dimensional Information About Targets
Develop Safe and Effective Gene Transfer Vectors for Many Different Tissues	Gene Therapy for Rare Single-Gene Disorders and Some Common Ones
Vigorously Explore the Ethical, Legal, and Social Implications of Genetic Research	Legal Safeguards Against Genetic Discrimination and Breaches of Privacy
	Effective Oversight of Clinical Application of Genetic Testing
	Mainstreaming of Genetics into the Practice of Medicine; Wide Achievement of "Genetic Literacy" Among Clinicians and Patients

nesses.[16] A public-private partnership has formed to build this catalog of variants as quickly as possible and has identified more than 2 million of these single nucleotide polymorphisms. Of particular interest are those common variants that influence gene function.

A powerful set of technologies for studying gene expression is being developed and explored.[17] These methodologies, which allow analysis of the transcription of as many as 10 000 genes in one experiment, make it possible to investigate the differences that occur between various tissue types and to explore the alterations in that expression pattern during disease. Such analyses have already been proved capable of identifying subtypes of certain malignancies that were identical by all other criteria.[18]

The same large-scale analysis strategies that have been applied so effectively to DNA and RNA also are being applied to proteins to characterize their structures, quantity, location in the cell, posttranslational modifications, and interaction partners.[19]

With the advent of these very large databases of information on sequence, variation, and expression, the field of computational biology is emerging as critically important to the future. Effective methods of sorting and analyzing the data will be required to glean biologically meaningful insights from the plethora of data.

The ethical, legal, and social implications research program has already fostered awareness of needs for intervention, particularly in the areas of privacy, genetic discrimination, guidelines for research, and education, and now focuses on the societal implications of increased information about human variation, in both medical and nonmedical situations.

The 21st Century: Critical Elements of the Medical Research Agenda

Obtaining the sequence of the human genome is the end of the beginning. As Knoppers has said, "As the radius of knowledge gets longer, the circumference of the unknown increases even more" (Bartha Knoppers, personal communication). For the full impact of advances in genetics to be felt in the practice of medicine, major challenges must be addressed.

Information about the human genome sequence and its variants must be applied to identify the particular genes that play a significant role in the hereditary contribution to common disease. This will be a daunting challenge. For a disease such as diabetes mellitus, 5 to 10 (or maybe more) genes are involved, each of which harbors a variant conferring a modest degree of increased risk. Those variants interact with each other and the environment in complex ways, rendering their identification orders of magnitude more difficult than for single gene defects. Nonetheless, with the combination of careful phenotyping (so that different disorders are not inadvertently lumped together) and sampling genetic variants at high density across the genome, it should be possible to identify disease gene associations for many common illnesses in the next 5 to 10 years.[2,16] One should not underestimate, however, the degree of sophistication in clinical investigation that will be necessary or the need for development of more efficient genotyping technology, such as the use of DNA chips or mass spectrometry, to make this kind of genome-wide survey a reality.

An understanding of the major pathways involved in normal homeostasis of the human organism must be developed along with how those pathways are deranged in illness. Identification of each gene that harbors a high-risk variant will point toward a critical pa

way for that illness. Many of those will come as a surprise, since the current molecular understanding of most common diseases is rather limited.

Efficient, high-volume methods will need to be developed and applied to the design of small-molecule drugs to modulate disease-related pathways in the desired direction. The pharmaceutical industry has been gearing up for this opportunity, and most companies now expect that the majority of future drug development will come from the field of genomics. With the application of methods that systematically combine chemical components into drugs and of high-volume assays for efficacy, it is expected that compounds can be efficiently identified that block or stimulate a particular pathway. A gratifying recent example is the development of the drug STI-571, which was designed to block the kinase activity of the bcr-abl kinase.[20] This protein is produced as a consequence of the translocation between chromosomes 9 and 22, a chromosome rearrangement that is characteristic of and central to the etiology of chronic myelogenous leukemia. STI-571 blocks the ability of the bcr-abl kinase to phosphorylate its unknown substrate and shows dramatic results in early clinical trials on patients with far advanced chronic myelogenous leukemia.

Along with the design of new drugs, genomics also will provide opportunities to predict responsiveness to drug interventions, since variation in those responses is often attributable to the genetic endowment of the individual. Examples have been identified where common variants in genes involved in drug metabolism or drug action are associated with the likelihood of a good or bad response. The expectation is that such correlations will be found for many drugs over the next 10 years, including agents that are already on the market. This field of pharmacogenomics promises to individualize prescribing practices.[21]

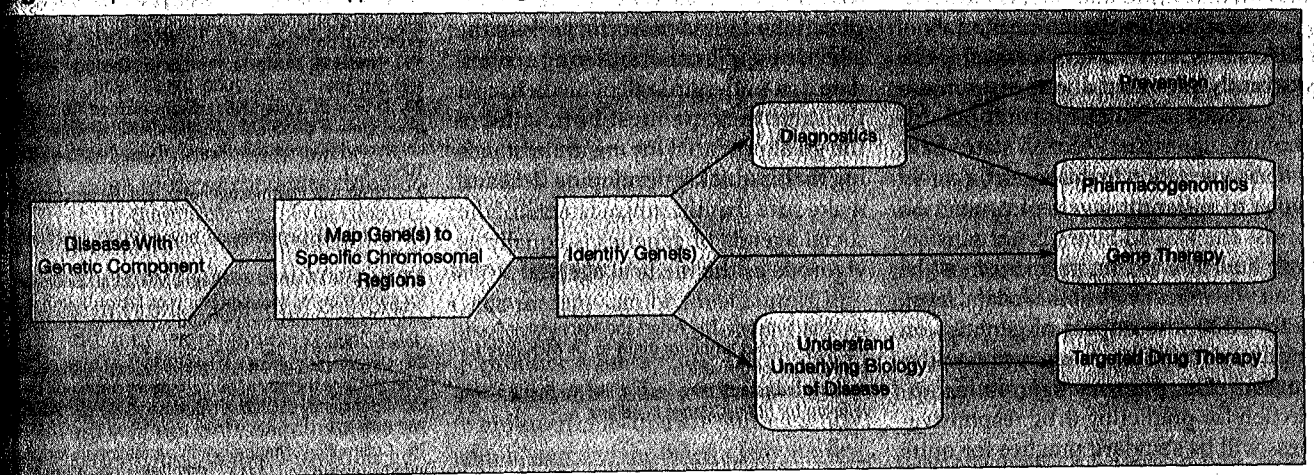
The field of gene therapy, having sustained a series of disappointments over the past few years, especially with the death of a volunteer in a gene therapy trial in the fall of 1999, has gone back to wrestling with the basic science questions of finding optimal methods for gene delivery.[22] While the optimism of the early 1990s about providing quick solutions to a long list of medical problems was probably never fully justified, it is likely that the development of safer and more effective vectors will ensure a significant role for gene therapy in the treatment of some diseases. There already have been promising reports of the application of gene therapy for hemophilia B[23] and severe combined immunodeficiency.[24]

Genetics in the Medical Mainstream

The power of the molecular genetic approach for answering questions in the research laboratory will catalyze a similar transformation of clinical medicine, although this will come gradually over the course of the next 25 years (FIGURE).

By the year 2010, it is expected that predictive genetic tests will be available for as many as a dozen common conditions, allowing individuals who wish to know this information to learn their individual susceptibilities and to take steps to reduce those risks for which interventions are or will be available. Such interventions could take the form of medical surveillance, lifestyle modifications, diet, or drug therapy. Identification of persons at highest risk for colon cancer, for example, could lead to targeted efforts to provide colo-

igure. Steps Involved in a Genetic Approach to the Diagnosis and Treatment of Disease



he rate of progress for applying a genetic approach to the diagnosis and treatment of each disease will be different depending on the research investment and the degree of biological complexity underlying the disease. First, the gene variants contributing increased disease risk must be identified by family studies and/or case-control studies. Diagnostic opportunities may then come along rather quickly, but will be of greatest clinical usefulness once prevention measures are developed that have proven benefit to those at high risk. Some gene variants will also show clinically useful associations with drug responsiveness (pharmacogenomics). In general, full-blown therapeutic benefits from identification of gene variants will take longer to reach mainstream medicine. In some instances, the gene itself will be the drug (gene therapy), while in others, a sophisticated knowledge of the underlying disease mechanism, built upon genetics, may allow the design of targeted and highly effective drug therapy.

noscopic screening to those individuals, with the likelihood of preventing many premature deaths.

Predictive genetic tests will become applicable first in situations where individuals have a strong family history of a particular condition; indeed, such testing is already available for several conditions, such as breast and colon cancers. But with increasing genetic information about common illnesses, this kind of risk assessment will become more generally available, and many primary care clinicians will become practitioners of genomic medicine, having to explain complex statistical risk information to healthy individuals who are seeking to enhance their chances of staying well. This will require substantial advances in the understanding of genetics by a wide range of clinicians.[25] The National Coalition for Health Professional Education in Genetics, an umbrella group of physicians, nurses, and other clinicians, has organized to help prepare for this coming era.

Another crucial step is the passage of effective federal legislation to outlaw the use of predictive genetic information in the workplace and in obtaining health insurance.[26,27] Numerous surveys have indicated that the public is deeply concerned about the potential for discrimination, and some individuals have forgone acquiring genetic information about themselves, since assurances cannot be currently provided about discriminatory misuse of the information. Although more than 2 dozen states have taken some action in this regard, a patchwork of different levels of protection across the United States is not satisfactory and this vexing problem must be dealt with effectively at the federal level.

By 2020, the impact of genetics on medicine will be even more widespread. The pharmacogenomics approach for predicting drug responsiveness will be standard practice for quite a number of disorders and drugs. New gene-based "designer drugs" will be introduced to the market for diabetes mellitus, hypertension, mental illness, and many other conditions. Improved diagnosis and treatment of cancer will likely be the most advanced of the clinical consequences of genetics, since a vast amount of molecular information already has been collected about the genetic basis of malignancy. By 2020, it is likely that every tumor will have a precise molecular fingerprint determined, cataloging the genes that have gone awry, and therapy will be individually targeted to that fingerprint.

Despite these exciting projections, certain tensions also will exist. Access to health care, already a major problem in the United States, will complicate these new advances, unless our medical care systems change in significant ways. Antitechnology movements, already active in the United States and elsewhere, are likely to gather momentum as the focus of genetics turns even more intensely on ourselves. Though the benefits of genetic medicine will be profound, there will be those who consider this advancement unnatural and dangerous. Efforts at public education need to start now to explain the potential benefits and to be honest about the risks.

In conclusion, this is a time of dramatic change in medicine. As we cross the threshold of the new millennium, we simultaneously cross a threshold into an era where the human genome sequence is largely known. We must commit ourselves to exploring the application of these powerful tools to the alleviation of human suffering, a mandate that undergirds all of medicine. At the same time, we must be mindful of the great potential for misunderstanding in this quickly developing field and make sure that the advancement of the social agenda of genetics is equally as vigorous as the medical agenda.

Funding/Support: The Albert and Mary Lasker Foundation provided an honorarium to Dr McKusick for preparation of this article.

Acknowledgment: We would like to thank Karin Jegalian for her assistance in editing.

REFERENCES

1. National Center for Biotechnology Information. Online Mendelian Inheritance in Man. Available at: http://www.ncbi.nlm.nih.gov/omim/. Accessed November 30, 2000.
2. Collins FS. Shattuck Lecture: medical and societal consequences of the Human Genome Project. *N Engl J Med*. 1999;341:28-37.
3. Henig RM. *The Monk in the Garden*. New York, NY: Houghton Mifflin; 2000.
4. Watson JD, Crick FHC. Molecular structure of nucleic acids. *Nature*. 1953;171:737-738.
5. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*. 1977;74:5463-5467.
6. Maxam AM, Gilbert W. A new method for sequencing DNA. *Proc Natl Acad Sci U S A*. 1977;74:560-564.
7. Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet*. 1980;32:314-331.
8. Gusella JF, Wexler NS, Conneally PM, et al. A polymorphic DNA marker genetically linked to Huntington's disease. *Nature*. 1983;306:234-238.
9. National Research Council, Committee on Mapping and Sequencing the Human Genome. *Mapping and Sequencing the Human Genome*. Washington, DC: National Academy Press; 2000.
10. US Department of Health and Human Services and Department of Energy. *Understanding Our Genetic Inheritance. The U.S. Human Genome Project: The First Five Years*. Washington, DC: US Dept of Health and Human Services; 1990.
11. Collins FS, Galas D. A new five-year plan for the U.S. Human Genome Project. *Science*. 1993;262:43-46.
12. Collins FS, Patrinos A, Jordan E, Chakravarti A, Gesteland R, Walters LR. New goals for the U.S. Human Genome Project: 1998-2003. *Science*. 1998;282:682-689.
13. Collins FS. Positional cloning moves from perditional to traditional. *Nat Genet*. 1995;9:347-350.
14. Hattori M, Fujiyama A, Taylor TD, et al. The DNA sequence of human chromosome 21. *Nature*. 2000;405:311-319.
15. Dunham I, Shimizu N, Roe BA, et al. The DNA sequence of human chromosome 22. *Nature*. 1999;402:489-495.
16. Collins FS, Guyer MS, Chakravarti A. Variations on a theme: cataloging human DNA sequence variation. *Science*. 1997;278:1580-1581.
17. Lockhart DJ, Winzeler EA. Genomics, gene expression and DNA arrays. *Nature*. 2000;405:827-836.
18. Bittner M, Meltzer P, Chen Y, et al. Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature*. 2000;406:536-540.
19. Pandey A, Mann M. Proteomics to study genes and genomes. *Nature*. 2000;405:837-846.
20. Druker BJ, Lydon NB. Lessons learned from the development of an abl tyrosine kinase inhibitor for chronic myelogenous leukemia. *J Clin Invest*. 2000;105:3-7.
21. Roses AD. Pharmacogenetics and the practice of medicine. *Nature*. 2000;405:857-865.
22. Verma IM. Gene therapy: beyond 2000. *Mol Ther*. 2000;1:493.
23. Kay MA, Manno CS, Ragni MV, et al. Evidence for gene transfer and expression of factor IX in haemophilia B patients treated with an AAV vector. *Nat Genet*. 2000;24:257-261.
24. Cavazzana-Calvo M, Hacein-Bey S, de Saint Basile G, et al. Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. *Science*. 2000;288:669-672.
25. Collins FS. Preparing health professionals for the genetic revolution. *JAMA*. 1997;278:1285-1286.
26. Hudson KL, Rothenberg KH, Andrews LB, Kahn MJE, Collins FS. Genetic discrimination and health insurance: an urgent need for reform. *Science*. 1995;270:391-393.
27. Rothenberg K, Fuller B, Rothstein M, et al. Genetic information and the workplace: legislative approaches and policy challenges. *Science*. 1997;275:1755-1757.