University of Pennsylvania
ScholarlyCommons

Departmental Papers (HSS)

Department of History and Sociology of Science (HSS)

1993

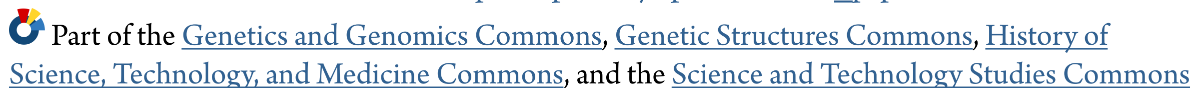
A Guide to the Human Genome Project: Technologies, People, and Institutions

Susan L. Speaker
University of Pennsylvania

Susan M. Lindee
University of Pennsylvannia, mlindee@sas.upenn.edu

Elizabeth Anne Hanson
University of Pennsylvania

Follow this and additional works at: http://repository.upenn.edu/hss_papers


Part of the Genetics and Genomics Commons, Genetic Structures Commons, History of Science, Technology, and Medicine Commons, and the Science and Technology Studies Commons

Recommended Citation (OVERRIDE)
Speaker, S.L., Lindee, M.S., & Hanson, E. (1993). *A Guide to the Human Genome Project*. Philadelphia: The Chemical Heritage Foundation.

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/hss_papers/28
For more information, please contact repository@pobox.upenn.edu.

A Guide to the Human Genome Project: Technologies, People, and Institutions

Abstract
There are many scientific reports and full-length books dealing with the Human Genome Project in all its facets; this simple, concise guide is intended for those who need a broad overview and a quick reference.

The information presented here is drawn from such journals as *Cell*, *FASEB Journal*, and *Science*; from official publications, in particular *Human Genome News*, the Office of Technology Assessment's *Mapping Our Genes*, and the Department of Energy's program reports; and from several secondary sources, including the prepublication draft of *Gene Quest*, by Robert Mullan Cook-Deegan, an insider's historical account, and the exploration of techniques in *Exons, Introns, and Talking Genes*, by Christopher Wills. The report focuses almost entirely on the genome project in the United States. We have shortened and simplified whenever possibly, providing citations and a bibliography for those who would like more detailed information.

We begin by exploring the origins of the genome project and the questions and criticisms it has provoked in the scientific community. Then we explore important techniques; the isntitutions connected with the prokect, including designated genome centers, important suppliers of resources, and corporations; systems of communication; and the ethical, legal, and social issues raised by the project. After two appendixes—lists of key personnel and of the disease associated with each chromosome—the report closes with a bibliography, a glossary (including a list of acronyms), and a timeline.

Disciplines
Genetics and Genomics | Genetic Structures | History of Science, Technology, and Medicine | Science and Technology Studies

This book is available at ScholarlyCommons: http://repository.upenn.edu/hss_papers/28

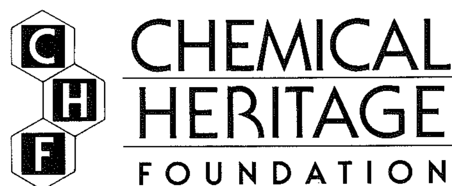
BIMOSI

A GUIDE TO THE HUMAN GENOME PROJECT

Technologies, People, and Institutions

By Susan L. Speaker and M. Susan Lindee
With Elizabeth Hanson

The Biomolecular Sciences Initiative of the
Beckman Center for the History of Chemistry
Philadelphia 1993



Chemical Heritage Foundation Publication No. 11

QUARTO
QH
447
.S631
1993
C.1

ISBN 0-941901-10-6
©1993 by the Chemical Heritage Foundation.
All rights reserved.
Copyright is not claimed for the illustrations credited to the publications listed below or for the glossary and acronyms, reprinted from *Human Genome: 1991–92 Program Report*.

A Guide to the Human Genome Project was published as part of the Biomolecular Sciences Initiative (BIMOSI), a project of the Beckman Center for the History of Chemistry, an academic unit of the University of Pennsylvania that is sponsored by the Chemical Heritage Foundation. BIMOSI is supported in part by the Andrew W. Mellon Foundation.

All illustrations credited to *HGN* are from *Human Genome News;* to *HGPR 1991–92* from *Human Genome: 1991–92 Program Report* (Washington, D.C.: Department of Energy, 1992); and to OTA, *Mapping Our Genes,* from Office of Technology Assessment, *Mapping Our Genes: Genome Projects—How Big, How Fast?* (Washington, D.C.: USGPO, 1988).

THE JOHNS HOPKINS
THE
MILTON S.
EISENHOWER
LIBRARY

For further information, including copies of *A Guide to The Human Genome Project,* please contact the

Chemical Heritage Foundation
3401 Walnut Street, Suite 460B
Philadelphia, PA 19104-6228 USA
Telephone (215) 898-4896
Fax (215) 898-3327

On the cover: A computer-generated image of the DNA double helix, looking down the helical axis. Computer graphic modeling and photography by Arthur J. Olson. Courtesy Arthur J. Olson and the Scripps Research Institute. ©1985 TSPI

AGD 4011
65
2 MY 95

CONTENTS

PREFACE 5

OVERVIEW

Initial Stages of the HGP in the United States 6

Genome Research Abroad 7

HUGO: A United Nations for the Human Genome 7

Pros and Cons 8

- Big Bad Science 8
- Patenting 10

TECHNOLOGIES AND TECHNIQUES

DNA, Chromosomes, and Genes 11

How To Map a Chromosome 12

- Genetic Maps 12
- Physical Maps: Low Resolution 13
- Physical Maps: High Resolution 14
- Amplifying and Ordering DNA Fragments 14
- Sequencing Technologies 16
- Keeping Results Comparable 16

Model Organisms 17

WHERE THE HGP IS CARRIED OUT

Designated Genome Centers 18

- DOE Centers 18
- NIH Centers 18

Genome Resource Centers 20

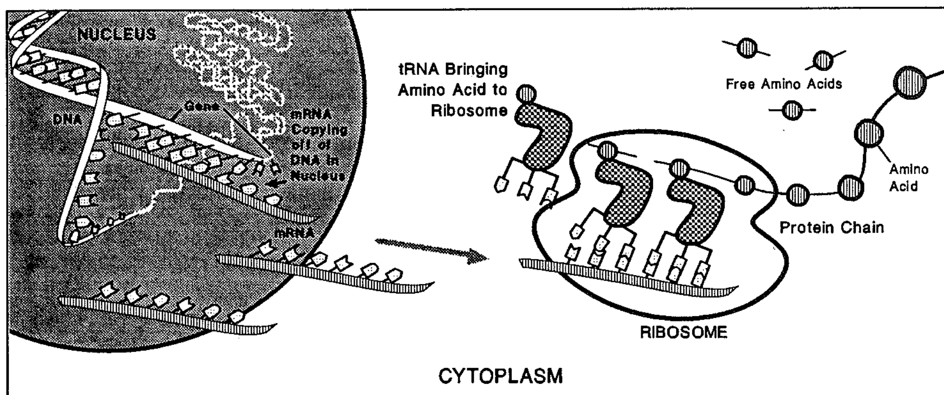
- International Resources 20
- U.S. Repositories 21

Corporate Involvement 22

- Early Corporate Interest 22
- Corporate Collaboration 22
- Industry and Gene Therapy 23
- Technology Development and Transfer: DOE 23
- Corporate Competition—and Managing It 24

6-12-95 MAC

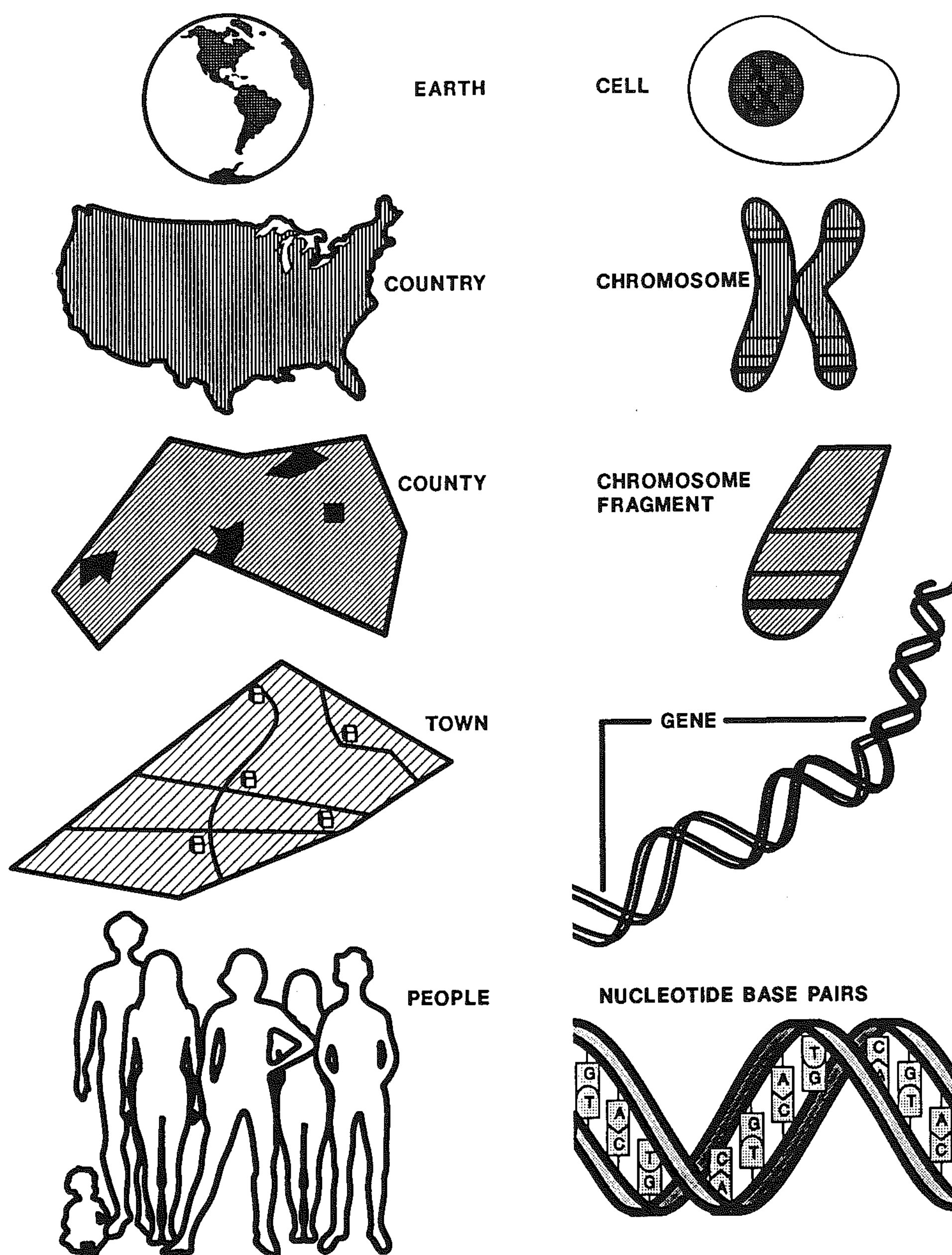
COMMUNICATIONS
Print Media 25
Newsletters 25
Journals and Series 25
Other Forums 26
Electronic Publications 26
Conferences 27
ETHICAL, LEGAL, AND SOCIAL IMPLICATIONS 28
APPENDIXES
A. Key Personnel 29
B. Chromosomes and Their Associated Diseases 31
SELECTED BIBLIOGRAPHY 32
GLOSSARY AND ACRONYMS 34
A GENOME PROJECT TIMELINE 38



In the first step of gene expression, messenger RNA (mRNA) is synthesized, or transcribed, from genes by a process somewhat similar to DNA replication. In higher organisms, this process takes place in the nucleus of a cell. In response to certain signals (e.g., association with a particular protein), sequences of DNA adjacent to, or sometimes within, genes control the synthesis of mRNA. Protein synthesis, or translation, is the second major step in gene expression. Messenger RNA molecules are known as such because they carry messages specific to each of the 20 different amino acids that make up proteins. Once synthesized, mRNAs leave the nucleus of the cell and go to another cellular compartment, the cytoplasm, where their messages are translated into the chains of amino acids that make up proteins. A single amino acid is coded by a sequence of three nucleotides in the mRNA, called a codon. The main component of the translation machinery is the ribosome—a structure composed of proteins and another class of RNAs, ribosomal RNAs. The ribosome reads the genetic code of the mRNA, while a third kind of RNA molecule, transfer RNA (tRNA), mediates protein synthesis by bringing amino acids to the ribosome for attachment to the growing amino acid chain. Transfer RNAs have three nucleotide bases that are complementary to the codons in the mRNA

Gene expression. From OTA, Mapping Our Genes, *p. 23.*

PREFACE



The number of base pairs of DNA in human cells is roughly comparable to the number of people on Earth. The scale of genetic mapping efforts can be compared to population maps, with chromosomes (50 to 250 million base pairs) analogous to nations, and genes (thousands to millions of base pairs) to towns.

A comparative scale of mapping. From OTA, Mapping Our Genes, *p. 5.*

There are many scientific reports and full-length books dealing with the Human Genome Project in all its facets; this simple, concise guide is intended for those who need a broad overview and a quick reference.

The information presented here is drawn from such journals as *Cell, FASEB Journal,* and *Science;* from official publications, in particular *Human Genome News,* the Office of Technology Assessment's *Mapping Our Genes,* and the Department of Energy's program reports; and from several secondary sources, including the prepublication draft of *Gene Quest,* by Robert Mullan Cook-Deegan, an insider's historical account, and the exploration of techniques in *Exons, Introns, and Talking Genes,* by Christopher Wills. The report focuses almost entirely on the genome project in the United States. We have shortened and simplified whenever possible, providing citations and a bibliography for those who would like more detailed information.

We begin by exploring the origins of the genome project and the questions and criticisms it has provoked in the scientific community. Then we explore important technologies and techniques; the institutions connected with the project, including designated genome centers, important suppliers of resources, and corporations; systems of communication; and the ethical, legal, and social issues raised by the project. After two appendixes—lists of key personnel and of the diseases associated with each chromosome—the report closes with a bibliography, a glossary (including a list of acronyms), and a timeline.

The Human Genome Project is an endeavor of sufficient complexity at all levels to require far more than this brief report to explain it. The information compiled here, however, can help those interested in learning more.

OVERVIEW

The Human Genome Project (HGP) is a scientific research program focused on mapping (including sequencing) the entire human genome (the total genetic material in the human cell). To map the genome is to place genes or other markers in statistically calculated or absolute positions on a conventionalized depiction of the chromosomes that make up the genome (see Technologies and Techniques, page 11).

The genomes of many different individuals will be mapped in the course of the project. While it is common to refer to "the" human genome, there is no such singular thing. Every individual human genome is different, and indeed even a single individual's genetic material may differ in different cells as a consequence of mosaicism or mutation. Ideally the project will be able to produce a map that reflects both the variety in human populations and the stability of some sections of the genome. The genomes of several model organisms, including mice and flies, are also being mapped as part of the project.

The HGP is a government project, generally funded by either the National Institutes of Health (NIH) or the Department of Energy (DOE). Human genome mapping is under way in the United States in a wide variety of institutions, including universities, private industry, national laboratories, and clinical laboratories. But the HGP is only one part of a large international effort. According to a recent analysis of genome research worldwide, "the percentage of the world share for 1990 from European Community countries was just under 30 percent, compared with 50 percent for the United States. The U.K. provided about a third of the total, followed by France and Germany."[1] Important genome work is also being done in Japan, Australia, Canada, Latin America, the former Soviet Union, and South Africa (see Genome Research Abroad, below).

The HGP can be described as a collaborative program that will create a readily accessible database useful to clinical geneticists, molecular biologists, evolutionary biologists, and other scientists who need information about the location of human genes. The project is not cutting-edge science in itself: the intellectual tools needed to map the human genome were worked out before the project began. It could thus be compared to the production of star maps in astronomy. But spinoffs from the project will include new technologies for manipulating DNA and new therapies that draw on the information in gene maps. Another hope is that mapping the genome will lead to dramatic advances in medicine. The long-term legacy of the project, then, can be expected to include not only the actual map(s) but the technological innovations that made the maps possible, and diagnostic, therapeutic, and commercial applications.

Initial Stages of the HGP in the United States

Mapping has been an important activity in genetics since the first linkage maps of fruit fly (*Drosophila*) genes produced by T. H. Morgan and his coworkers at Columbia University during 1910–1920. Since then, the genomes of model organisms such as mice and flies and easily manipulated experimental organisms such as *Escherichia coli* and yeast have been partially or completely mapped. But mapping the large and complex human genome did not seem technologically feasible until the last decade.

In June 1985 Robert Sinsheimer, then chancellor of the University of California at Santa Cruz, convened the first meeting to discuss the technical prospects of a particular type of mapping the human genome: sequencing its DNA (see pages 8, 16). This meeting did not lead to its intended result—a genome mapping center at the Santa Cruz campus—but it did plant the idea for such a project in the mind of the Nobel laureate Walter Gilbert. Gilbert, a professor of biology at Harvard, began promoting large-scale sequencing at professional meetings in the late summer and fall of 1986.[2]

About the same time, the Department of Energy, still wrestling with the difficulties of detecting radiation mutation in human populations, began to take an interest in constructing a map of the human genome. At a meeting at Alta, Utah, in December 1984, participants proposed that a direct analysis of DNA might be used to detect heritable mutations in the survivors at Hiroshima and Nagasaki. While these survivors have been the focus of almost fifty years of epidemiological and molecular research, assessing mutation rates among them has proved extraordinarily difficult. Late in 1985 Charles DeLisi of the DOE proposed a three-part approach involving DNA sequencing, computer analysis, and methods to order DNA fragments cloned from the human genome. DeLisi began a research program based on these objectives in 1987. The DOE Human Genome Initiative was the first government program of human genome research.

The DOE program sparked an active public debate among scientists about the idea of sequencing the entire genome. At first many were highly skeptical of a directed approach (see Big Bad Science, page 8). Many were also skeptical of the DOE as sponsor. The National Academy of Sciences (NAS) issued a report in February 1988 that argued strongly for a broader human genome project, not funded solely by the DOE. In April 1988 the Office of Technology Assessment reported that the debate was not about whether the genome should be sequenced, but about how to organize the effort. In the same period, Congress appropriated $17.3 million to the National Institutes of Health to fund genome research and $11.8 million to the DOE. These budgets were increased to $28 million for the NIH and $18 million for the DOE in 1989, $58.5 million and $26 million in 1990, $87.5 million and $47.7 million in 1991, and $104.8 million and $59 million in 1992.

NIH and DOE coordinated their planning and early in 1990 submitted a five-year joint research plan. The agencies established joint working groups on mapping, informatics (the hardware, databases, and computational analysis that support mapping), and the social, ethical, and legal implications of human genome research. Other sources of financial support for the HGP include the National Science Foundation (NSF), which does not have a genome program per se, but provides grants for research on instrumentation efforts and for genomic research on plants and nonhuman organisms, and the Department of Agriculture, which supports extramural and

1. Walter Bodmer, "Genome Research in Europe," *Science* 256 (1992), 480–481.

2. Much of the following history is drawn from a short account by James Dewey Watson and Robert Mullan Cook-Deegan, "Origins of the Human Genome Project," *FASEB Journal* 5 (1991), 8–11.

intramural studies of the genomes of plants of agricultural significance. (The Howard Hughes Medical Institute also initially funded genomics databases and conferences; it still gives grants to individual investigators.) These sponsors shape the wide variety of activities carried on in universities, government labs and agencies, nonprofit research institutes, and industry that are known collectively as the Human Genome Project.

Genome Research Abroad

The United Kingdom, Denmark, France, Germany, and Italy all have established national genome programs. Of these the United Kingdom's national program, funded by the Medical Research Council (MRC), is probably the most highly developed. The main support is for a resource center that provides specialized services—reagents, materials, and data—to the community. Nongovernmental sources in the U.K. also make significant contributions to the national and international genome efforts: the Imperial Cancer Research Fund has a substantial involvement in human genome analysis, and the Wellcome Trust provides the major support for the European office of the Human Genome Organization (HUGO), based in London.[3]

In France the Centre d'Etude du Polymorphisme Humain (CEPH), which is supported by a mixture of government and private sources, has been a major contributor to international collaboration in the construction of human gene maps (see Genome Resource Centers, page 20). More recently the French Muscular Dystrophy Association funded a large related project called Généthon, designed to coordinate disease-gene mapping on a large scale and carry out organized genomic sequencing.[4]

Several organizations in Europe support genome research and help coordinate efforts in different countries. These include the European Economic Community (EEC), the European Molecular Biology Organization (EMBO), and the European Molecular Biology Laboratory (EMBL) in Heidelberg. Projects supported by the EEC require participation of two or more EEC countries. Such projects must define clear work plans and targets and emphasize infrastructure, the provision of resources through centers, and networking.[5] EMBO is funded by seventeen European countries; it supports fellowships, workshops and training courses, occasional scientific meetings, and a journal. It does not fund research directly, but rather helps strengthen the training of European molecular biologists. EMBL "sponsors research in instrumentation, biocomputing, and gene mapping and sequencing as well as other areas of biology. . . . EMBL also operates the major European database of nucleotide sequences, which works in cooperation with GENBANK [GenBank, the DOE's database of nucleotide sequences], to gather and disseminate sequence data."[6]

In Japan several government agencies provide genome project support. Most mapping and sequencing research falls under the domain of the Ministry of Education, Science, and Culture, which is analogous to the NIH in the United States. The Science and Technology Agency primarily supports mission-oriented basic research. It has expedited the development of automated sequencing technology. The Ministry of Health and Welfare has integrated human genome research into its Research Project for Aging and Health.[7]

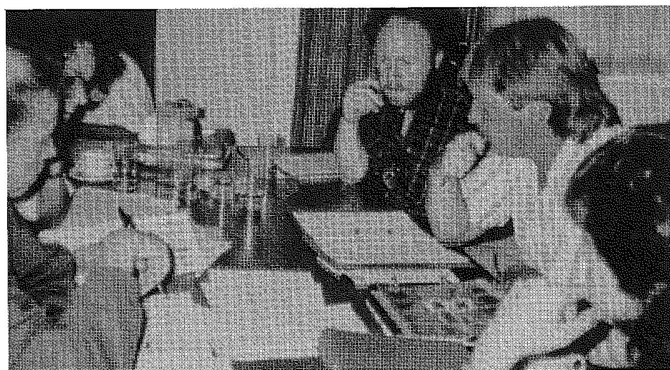
HUGO: A United Nations for the Human Genome

The proliferation of mapping and sequencing efforts worldwide created the need for an international coordinating scientific body. The Human Genome Organization was incorporated in 1989 to address this need, serving (in the words of Norton Zinder) as a "U.N. for the human genome." Its purposes are

- to help coordinate research on the human genome and in particular to foster collaboration between scientists so as to avoid unnecessary competition or duplication of effort, and to coordinate this research with parallel studies in model organisms;
- to coordinate and facilitate the exchange of data and biomaterials relevant to human genome research, and, through a training program, to encourage the spread of the related technologies;
- to encourage public debate and provide information and advice on the scientific, ethical, social, legal, and commercial implications of human genome projects.

To these ends, HUGO's activities were to include some or all of the following:

- promoting the international exchange of knowledge and research techniques by providing training fellowships, instructional courses, and workshops, as well as by organizing and funding international meetings;
- offering expert advice to governmental and nongovernmental agencies on the support of human and other genome research;
- gathering and distributing information on human-genome-related programs and projects.[8]



An international effort. *Participants at the First International Workshop on Human Chromosome 5, held in London, 1990. From* HGN 2(4) *(Nov. 1990), 13.*

3. Bodmer, "Genome Research in Europe" (see note 1).
4. *Ibid.*
5. *Ibid.*
6. Office of Technology Assessment, *Mapping Our Genes: The Genome Projects—How Big, How Fast?* (Washington, D.C., USGPO, 1988), 141–142.
7. *Ibid.*; and Yoji Ikawa, "Human Genome Efforts in Japan," *FASEB Journal* 5 (1991), 66–69.
8. Walter Bodmer, "HUGO, The Human Genome Organization," *FASEB Journal* 5 (1991), 73.

Much has been demanded of HUGO, and it has taken some time to get things under way. But recently the EEC awarded HUGO a contract to help coordinate workshops on the individual chromosomes, which lie at the heart of future international collaboration.

Pros and Cons

Some members of the molecular biology community have viewed the Human Genome Project with skepticism. Their criticisms tend to fall into four main categories. First, some consider the project bad science. Others deplore the use of "big science" in biology. Some critics say that the HGP creates an improper milieu for doctoral training in science. Finally, many believe that the HGP takes funding away from other worthy projects. In addition to these basic criticisms of the project itself, problems related to the patenting of HGP discoveries have recently prompted controversy and merit discussion here.

The project has also raised wide concerns both within and beyond the scientific community. These ethical, legal, and social issues, treated in the section on page 28, led to the formation of the ELSI program of the HGP.

Big Bad Science

The "bad science" charge derives mainly from the project's initial emphasis on sequencing. The focus of the project has now shifted to more general mapping, but it began as a program to sequence the entire genome. Sequencing involves establishing the order of all nucleotides in a nucleic acid, while mapping locates the genes on the chromosomes. Only parts of each DNA molecule act as genes; other parts seem to have no known function in heredity, though they may have regulatory roles. Because much of the genome material seems to be "junk," a crash program to sequence it appeared to promise a questionable return on the investment of funds and human effort. Many scientists argued that "function-based" approaches already in use (that is, moving from messenger RNA or a cell protein, via complementary DNA, to its gene and regulatory regions) presented the best chance of finding out about the genome. Nobody disputed that a complete sequence would be a very nice thing to have on hand, but many asserted that a physical map would be more immediately useful.[9] Finally, some have questioned the wisdom of devoting large-scale funding to a definite goal like sequencing when so many of the important discoveries in molecular biology have been adventitious.[10]

The argument against biological "big science" was well summarized by the biologist Bruce M. Alberts in an essay written in 1985, before the HGP existed. Alberts said that the best biological research was rarely done in a large laboratory, because large laboratories are both inefficient and poor environments for younger scientists. First, the bigger the research group, the more time a group leader must spend on helping with job applications, finding and accounting for funds, cranking out publications, and so on. There is much less time for thinking about science or keeping up with the research literature. The leader may tend to encourage associates to do obvious rather than innovative experiments, and will also be less familiar with techniques being used—and therefore less able to judge their potential or limitations. In such circumstances, Alberts said, graduate and postdoctoral students often function more as factory workers and less as young scientists. He concluded by stating, "Science is not a business and bigger is not better. . . . Any value system based on acquiring the largest research team, or on maximizing either total grant support or publications, is counterproductive to good science and should be viewed with alarm."[11] (In 1986–87 Alberts, now president of the NAS, chaired the National Research Council panel that drafted the strategy for the HGP; he was chosen specifically as a counterbalance to more enthusiastic panel members.)

Other critics note that once the technical and organizational problems of the HGP are resolved, much of the "research" will be rather dull, technician-oriented work far beneath the abilities of postdocs and graduate students. What provisions will be made for these younger colleagues?[12] Scientific journals no longer consider sequences worthy of

The Asilomar Conference (1975). *Early concerns over the possible risks of recominant DNA led to the drafting of guidelines adopted by the National Institutes of Health in 1976. Concerns over related ethical, legal, and social issues led to creation of the HGP's ELSI program (see page 28 below). Shown here are four Asilomar Conference organizers: Maxine Singer (NIH), Norton Zinder (Rockefeller University), Sidney Brenner (Medical Research Council, U.K.), and Paul Berg, chair of the organizing committee. Courtesy Andrew A. Stern and the National Academy of Sciences.*

9. See Roger Lewin, "Proposal to Sequence the Human Genome Stirs Debate," *Science* 232 (1986), 1598–1600; Lewin, "Shifting Sentiments over Sequencing the Human Genome," *Science* 233 (1986), 620–621; and Victor A. McKusick, "Current Trends in Mapping Human Genes," *FASEB Journal* 5 (1991), 12–20.

10. Bernard D. Davis, "Some Problems with a Crash Program," *FASEB Journal* 5 (1991), 76.

11. Bruce M. Alberts, "Limits to Growth: In Biology, Small Science is Good Science," *Cell* 41 (1985), 337–338.

12. Leslie Roberts, "Plan for Genome Centers Sparks Controversy," *Science* 246 (1989), 204–205.

publication, and so young scientists engaged in genomics research might be unable to publish unless their results have some fundamental biological importance.

The geneticist Victor McKusick, founding president of HUGO, has responded to these criticisms by saying that the HGP is not so much "big science" as coordinated, interdisciplinary science. HGP funding (ca. $164 million in 1992) is relatively modest in the world of big science: AIDS funding for 1991, for example, was $800 million. Yet the organization of the project is unusual in biology. HGP looks like big science because of the way it is institutionalized, not because of its funding. McKusick notes that scientists in the future, given a full map and sequence, would be spared the drudgery of cloning and sequencing and could get on to more interesting work. This defense underscores the technical utility of the HGP—as a resource or tool with future value (a map that will stand for all time)—rather than its potential for generating innovative or basic science.[13]

Critics of big-science biology generally do not dispute the importance of the genome mapping project, but question the benefits of singling out genome science and subjecting it to top-down administration. As one biologist wrote to *Science*, "It is budgeted directly from Congress like a separate institute, with its own administration and council and with its own study sections. As a result, it is overbudgeted, overtargeted, overprioritized, overadministered, and micromanaged." If, he continued, the main goals of genome research are to identify genes that influence human health and to learn a lot of interesting biology about gene organization, expression, and evolution, this will be better accomplished by the independent research already in progress.[14]

The big-science issue surfaced again in 1989 when the NIH announced that it would create twenty-five genome centers. James Watson, responding to complaints about the general quality of work at big science centers, admitted that they often did not carry out their mandates: "We all know how fraudulent most centers are." But he stated that the cottage-industry approach was not practical for the genome map—with some historical justification: major *geographical* mapping projects have generally required a coordinated effort in many different locations.[15] Norton Zinder of the NIH Program Advisory Committee on the Human Genome commented that the issue was how to avoid creating a monster—and how to kill any monsters created. "In the past, centers were like werewolves—you couldn't kill them. And a lot of them go bad." Genome centers would therefore need strong leaders, definite missions, and tough peer review and accountability standards.[16] Perhaps in response to these concerns, the NIH initiated an accountability system that uses physical maps of a certain resolution as a way to measure progress at genome centers (see Keeping Results Comparable, page 16).

Related to the big-science theme was considerable apprehension about the DOE's involvement in the HGP. Would biological research be likely to flourish within the kind of bureaucratic structures that supported "big physics"? Is the DOE equipped to make decisions about biological research, that is, will the agency demand the same quality from funding applications that the NIH or NSF would? DOE supporters noted that the DOE (or its predecessor agency, the Atomic Energy Commission) has been involved in genetic research since 1945. It has also established gene libraries and the database GenBank. In addition, it is accustomed to handling big projects, and the HGP is as much an organizational as a technical challenge.[17]

Finally, many scientists have become convinced that the HGP has been funded at the expense of other projects, because NIH cut back its biological research funding at the same time the HGP appropriations were made. In 1989 two letter-writing campaigns were organized to protest the HGP, by Martin Rechsteiner of University of Utah and by Michael Syvanen of the University of California at Davis. They urged Congress to kill or at least curtail the project because, as currently conceived, it was a wasteful use of scientific funding and detrimental to the training of young scientists.[18] By 1990 the project's budget requests were being scrutinized much more closely by Congress. Supporters of

Two Key HGP personnel. *Norton Zinder, left, receives a certificate in recognition of his contributions as first chair of the NIH Program Advisory Committee on the Human Genome. James Watson, then director of the NIH National Center for Human Genome Research is the presenter. From* HGN *3(3) (Sept. 1991), 10.*

13. McKusick, "Current Trends in Mapping Human Genes" (see note 9).

14. Donald D. Brown, letter, in "Two Views of the Genome Project," *Science* 251 (1991), 854–855. See also Leslie Roberts, "A Meeting of the Minds on the Genome Project?" *Science* 250 (1990), 756–757.

15. See Stephen S. Hall, *Mapping the Next Millenium: The Discovery of New Geographies* (New York: Random House, 1992), 3–28.

16. Roberts, "Plan for Genome Centers Sparks Controversy" (see note 12).

17. Lewin, "Proposal to Sequence Human Genome" (see note 9); and Roger Lewin, "Politics of the Genome," *Science* 235 (1987), 1453.

18. Martin Rechsteiner's letter in *FASEB Journal* 4 (1990), 2941–2942, summarizes his position.

the project argue that the HGP appropriations are separate from the NIH's regular budget, and that canceling the program would not alleviate the current funding crunch. They also worry that if HGP budgets are cut just as the project is hitting its stride, researchers will find something else to do, and biologists may never get the genome maps.[19]

Patenting

Perhaps more threatening to the project than any of these criticisms has been the NIH's controversial efforts to patent complementary DNA (cDNA). In June 1991 Craig Venter, a biologist at the National Institute of Neurological Disorders and Stroke, filed for patents on 350 unique cDNA clones representing genes not seen before. He had been advised to do this by Reid Adler at NIH's technology transfer office, even though the functions of the genes were unknown and the discovery may prove to have little utility. The genomics communities in the United States and abroad reacted to this move with horror, for several reasons. First, sequencing of short pieces of unidentified clones is very basic work. To patent one early element of a process could undercut patent protection for workers who finally elaborate the functions of proteins encoded by genes. It would also discourage investment in inventions based on that work. Second, critics believe that patenting such basic pieces of the genomics puzzle would hinder the free flow of information in the genomics community. They raise the specter that the whole national and international effort may collapse into scientific fiefdoms as patenting of basic data increases costs and competition.[20]

Patenting per se is not the issue for the critics. Patenting has long played a role in the development of biotechnology, from recombinant processes to genetically engineered bacteria. The U.S. government encourages patents because they facilitate technology transfer—the conversion of scientific knowledge into useful products. The reasoning is that if individuals and industries can expect to benefit from their discoveries and innovations (i.e., be compensated for their time and effort), they will be both more likely to invest further time, effort, and resources and in fact more willing to share discoveries and innovations. Scientists often publish their data soon after filing patent applications, so that information flow is not impeded, but the investigators retain control over inventions and knowledge and are assured of any financial rewards. Nor are patents typically intended to confer exclusivity, but rather to generate licensing revenue.

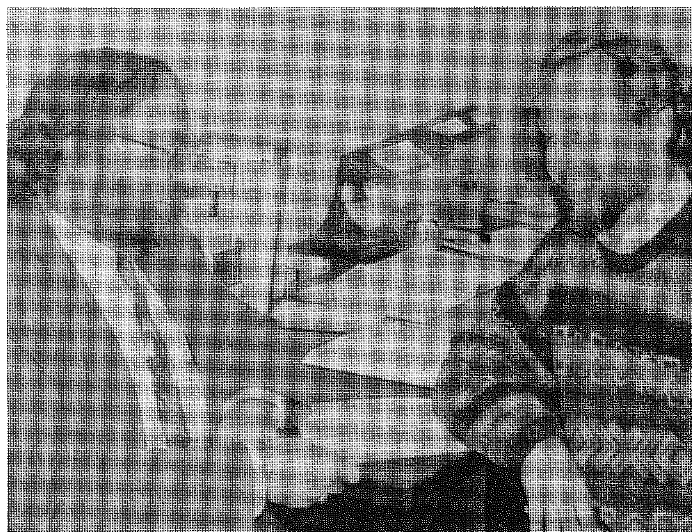
Researchers who do not file for patent protection before publishing risk losing control of their discoveries. The failure to file also inhibits technology transfer, since any firm investing in a new product or process wants a guarantee that this investment will be protected. The process or product might also be patented by a foreign company or individual, thereby inviting "foreign exploitation of research funded at U.S. . . . expense." For this reason Federal statutes require that recipients of Federal funds report all patentable inventions.[21]

Yet the patenting process may inhibit the free exchange of information because for a patent to be issued the invention must be novel, that is, not used by anyone other than the inventor. Publication of method cannot precede the patent application by more than one year. Scientists might be reluctant to publish information about their processes or products if doing so might endanger patentability. If information about such basic things as sequenced sections of DNA is routinely delayed and contested, collaborative research will be difficult. So far this seems to be a hypothetical rather than a real problem, but it does invite consideration of alternatives to existing patent laws, which would address the special requirements of biotechnology and the Human Genome Project.

19. Leslie Roberts, "Genome Backlash Going Full Force," *Science* 248 (1990), 804; Roberts, "Tough Times Ahead for the Genome Project," *ibid.*, 1600–1601; and J. H. Weis, "Usefulness of the Human Genome Project" (letter), *ibid.*, 1595.

20. See Leslie Roberts, "Genome Patent Fight Erupts," *Science* 254 (1991), 184–186; and letters in Nov. 29 and Dec. 20 issues, *ibid.*

21. OTA, *Mapping our Genes,* (see note 6) 165–171, contains an excellent discussion of patents and copyrights as they apply to the genome project.



ELSI. *Daniel Drell (DOE Human Genome Program) and Eric Juensgt (NIH National Center for Human Genome Research) discuss their agencies' respective programs in the ethical, legal, and social issues that may arise from data produced by human genome research. From* HGPR 1991–92, *p. 22 (photo provided by Leslie Fink, NIH NCHGR).*

TECHNOLOGIES AND TECHNIQUES

One unique aspect of the Human Genome Project is the degree to which technologies and techniques drive the work. Indeed, the project's success depends on developing technologies that can speed up mapping and sequencing and thus reduce costs. Understanding these techniques requires some basic background in genetics.

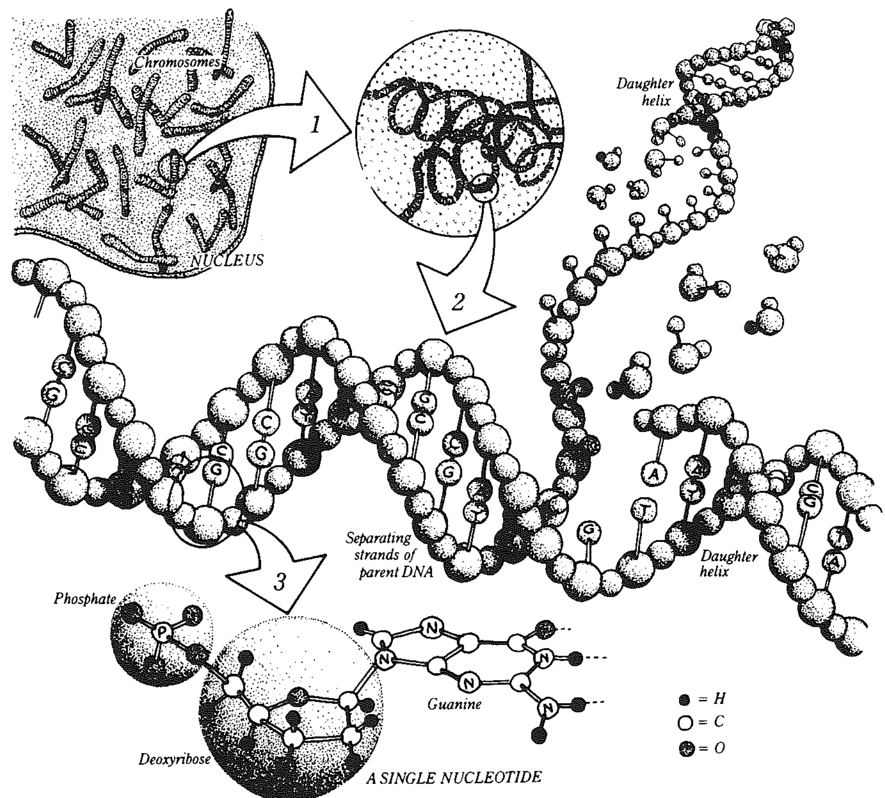
What follows is a brief primer on the structure of chromosomes and genes. Then the two basic types of genome maps—genetic and physical—are explained, along with some of the steps involved in constructing maps. Next, sequencing technologies are discussed. The section concludes with a brief account of model organisms. Refining mapping and sequencing techniques and creating new ones are major activities in genomics research. Because the state of the art changes so rapidly, this discussion is necessarily incomplete and covers only the most commonly used techniques.

DNA, Chromosomes, and Genes

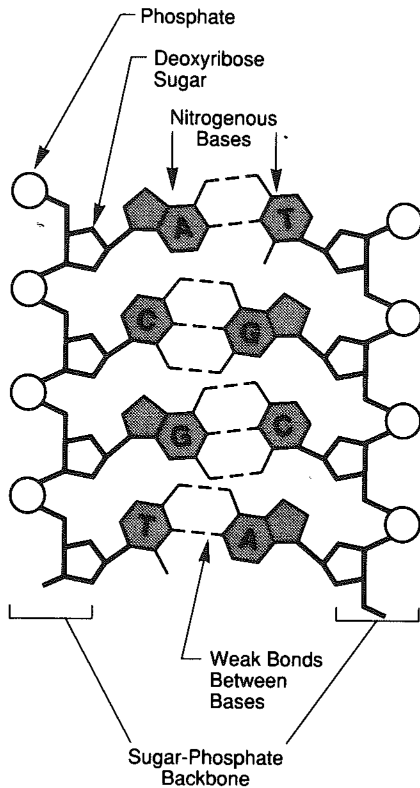
The genome is the total complement of genetic material present in a single cell. This genetic material is deoxyribonucleic acid—DNA—and it is found in the nuclei of human cells. DNA contains a set of coded instructions that guide all cellular activities for the lifetime of the cell or organism. Each of a person's 10 trillion cells contains essentially the same DNA.

The human genome is divided into distinctively shaped, physically separate units called chromosomes. The number of chromosomes in a cell varies dramatically from species to species. Dogs have 78 chromosomes, the most of any mammal; the fruit fly, a favorite research organism, has 8 chromosomes. Humans have 46 chromosomes, arranged in 23 pairs. One chromosome in each pair is inherited from the mother, the other is inherited from the father. Twenty-two of these pairs consist of matched chromosomes, known as autosomes, which are identified by number in order of decreasing size (chromosome number 1 is the longest). The remaining pair determines sex. In females this pair is matched (XX); in males it is unmatched—the two chromosomes have different shapes (XY). The total number of different types of human chromosomes is thus 24: the 22 autosomes, plus the X and Y chromosomes.

Each chromosome is a single molecule of DNA tightly coiled into a compact shape. Stretched out, a DNA molecule would be about two inches long—it is the largest type of molecule known. The shape of the DNA molecule is called a double helix. It can be imagined as a twisted ladder. Sugar-phosphate groups make up the sides of the ladder. These are connected by "rungs" consisting of molecules called *bases* (compounds that react with acids). Each rung of the DNA ladder is made up of two bases, one protruding from each sugar-phosphate group on the side. (The combination of base, phosphate, and sugar is known as a nucleotide.) DNA contains four kinds of base: adenine (A), thymine (T), guanine (G), and cytosine (C). Each base of the rung matches up with, or is *complementary* to, a specific base from the opposite side: adenine always pairs with



The human genome at four levels of detail. *Apart from reproductive cells (gametes) and mature red blood cells, every cell in the human body contains 23 pairs of chromosomes, each a packet of compressed and entwined DNA (1, 2). Each strand of DNA consists of repeating nucleotide units composed of a phosphate group, a sugar (deoxyribose), and a base (guanine, cytosine, thymine, or adenine) (3). DNA ordinarily takes the form of a highly regular double-stranded helix, the strands of which are linked by hydrogen bonds between guanine and cytosine and between thymine and adenine. Each such linkage is a base pair; some 3 billion base pairs constitute the human genome. The specificity of these base-pair linkages underlies the mechanism of DNA replication illustrated here. Each strand of the double helix serves as a template for the synthesis of a new strand; the nucleotide sequence (i.e., linear order of bases) of each strand is strictly determined. Each new double helix is a twin, an exact replica, of its parent. Figure and caption text from* HGPR 1991–92, *p. 193.*



DNA structure. *The four nitrogenous bases of DNA are arranged along the twisted sugar-phosphate backbone or ladder in a particular order (the DNA sequence), encoding all genetic instructions for an organism. Adenine (A) pairs with thymine (T), while cytosine (C) pairs with guanine (G). The two DNA strands are held together by weak bonds between the bases. A gene is a segment of a DNA molecule, located in a particular position on a specific chromosome, whose base sequence contains the information necessary for protein synthesis. Figure and caption text from* HGPR 1991–92, *p. 194.*

thymine, guanine with cytosine. The sequence of bases "GGATCC," for example, always pairs with (is complementary to) "CCTAGG" on the opposite side of the ladder. The number of base pairs in a human chromosome averages 150 million; the human genome contains approximately 3 billion base pairs. These bases can be thought of as the alphabet whose order, or *sequence,* encodes genetic information (see the figure on page 38).

When a cell divides, its DNA breaks apart along the rungs of the ladder, separating the base pairs. Each half, or single strand, of the DNA can then serve as a *template* for the formation of a new DNA molecule. Complementary nucleotides assemble along the single-stranded DNA and stick to it, so that two complete, double-stranded DNA molecules are synthesized, one for each of the "daughters" of the dividing cell. DNA can also be treated with chemicals that cause its two strands to separate; researchers take advantage of this trick in mapping chromosomes.

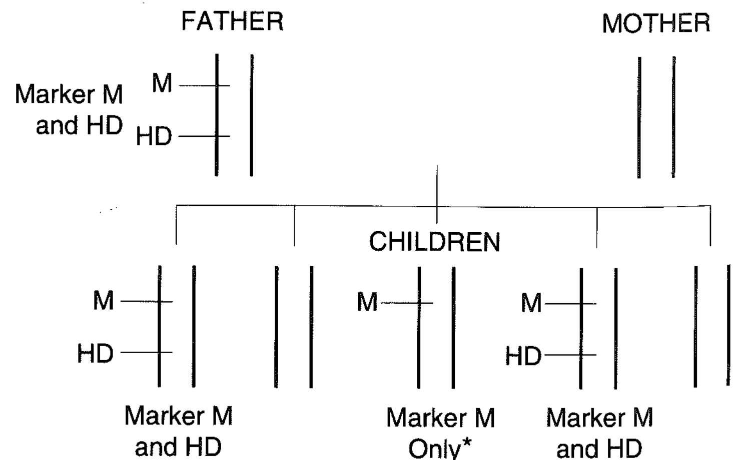
A specific sequence of bases that contains instructions for making a protein is called a gene. Because all genes share universal "start" and "finish" sequences, researchers are able to locate genes on a chromosome by looking for these sequences, even if they do not know what proteins are coded for. But much of the DNA sequence—even within a gene—has no known biological function. In fact, a typical gene is about 90 percent "stuffer"—stretches of base pairs with no known function. These sequences get edited out in the process of making proteins. Such "stuffer" sequences are called *introns;* the functional sequences are termed *exons.* A human gene might contain up to 30,000 base pairs, but only about 10 percent of these base pairs contain information now known to be of use to the cell. This still represents about 3,000 base pairs—an enormous number to analyze in order to sequence even a single gene.

How To Map a Chromosome

There are many methods of mapping chromosomes. Genomics researchers choose a method based on the level of detail in which they want to analyze a chromosome or portion of chromosome. All chromosome maps basically show the order of genes along the chromosome. The ultimate chromosome map is the complete base sequence. With the technology available today, however, sequencing each chromosome is prohibitively expensive. One of the goals of the Human Genome Project is to create different types of maps, at varying levels of detail, for each human chromosome. These maps fall into two basic categories: genetic maps, also called linkage maps, and physical maps.

Genetic Maps

Like all organisms that reproduce sexually, humans inherit half their chromosomes from each parent. But chromosomes do not pass from generation to generation unadulterated. The processes that produce egg and sperm cells also shuffle genes. Like all cells, the cells that give rise to egg and sperm *replicate,* or copy, their chromosomes as they di-



Constructing a genetic linkage map. *Genetic linkage maps of each chromosome are made by determining how frequently two markers are passed together from parent to child. Because genetic material is sometimes exchanged during the production of sperm and egg cells, groups of traits (or markers) originally together on one chromosome may not be inherited together. Closely linked markers are less likely to be separated by spontaneous chromosome rearrangements. In this diagram, the vertical lines represent chromosome 4 pairs for each individual in a family. The father has two traits that can be detected in any child who inherits them: a short known DNA sequence used as a genetic marker (M) and Huntington's disease (HD). The fact that one child received only a single trait (M) from that particular chromosome indicates that the father's genetic material recombined during the process of sperm production. The frequency of this event helps determine the distance between the two DNA sequences on a genetic map. From* HGPR 1991–92, *p. 201.*

vide, in this case in the special process called *meiosis.* But in the first division of meiosis the matching chromosomes pair up and then separate, to end up in separate cells; during this time a process called *crossing over,* a form of *recombination,* occurs. Sections of chromosomes are transposed; the copies of chromosome 11, for example, may swap sections. Crossing over may also occur between different chromosomes so that, for example, chromosome 11 ends up with a small piece of chromosome 4 and vice versa. During recombination, genes that are close together on a chromosome are likely to stay together, whereas genes that are far apart are more likely to be split up.

A genetic, or linkage, map estimates the "distance" between two genes or two markers on a chromosome in terms of the likelihood that they will be inherited together. The distance is measured in units called centimorgans. Two genes that are one centimorgan apart have a one percent chance of being separated during recombination. On genetic maps no genes can be farther apart than 50 centimorgans because any gene in a particular genome has at least a 50 percent chance of being inherited with any other. (See also RFLPs, page 14.)

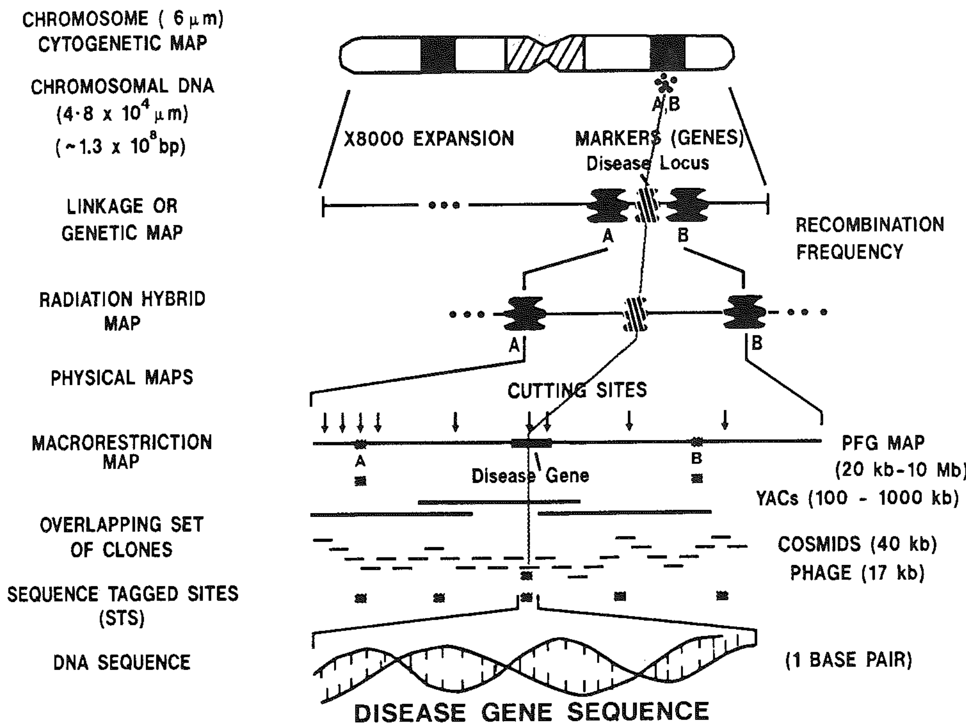


Diagram relating the multiple levels at which the human chromosome can be mapped. *The line running vertically through the diagram represents the tracking of markers A and B through progressively more precise levels of mapping. Tracking markers allows investigators to follow a candidate disease gene from the coarsest to the finest map resolution, the* DNA sequence. *The* cytogenetic map *provides the lowest level of resolution, measuring the distance between chromosomal features (i.e., bands or breakpoints) visible under the light microscope. Chromosome banding can resolve features to about 5 megabases (Mb). The* linkage or genetic map *measures the recombination frequency between two linked markers (A and B in this diagram), which can be genes or polymorphisms (see RFLPs, page 14).* Radiation hybrid maps *are produced by breaking chromosomes with radiation and then identifying the fragment carrying the marker (the breakpoint); the resolution of these maps is comparable to that of linkage maps. At the next resolution level,* macrorestriction *fragments of 1 to 2 Mb are separated and the markers localized and mapped. Finer mapping resolution is provided by ordered libraries of* yeast artificial chromosomes *(YACs), which have insert sizes from 100 to 1000 kilobases (kb). Ordered libraries of* cosmids *have smaller insert sizes, usually about 40 kb, and produce higher-resolution maps. The* DNA base sequence *is the highest-resolution map, with sequence tagged sites (STSs) used as unique reference points. Figure and caption text from* HGPR 1991–92, *p. 73 (figure provided by C. E. Hildebrand).*

Physical Maps: Low Resolution

Physical maps, as the name implies, show the physical distance between genes or markers on a chromosome. This distance can be measured, for example, in terms of base pairs. For the human genome the coarsest physical map essentially shows what stained chromosomes look like under a light microscope. Cells are isolated at a stage in the cell cycle when the chromosomes become short and thick in preparation for cell division. When stained, the chromosomes then show a pattern of up to 1,000 light and dark bands. These bands do not correspond to a particular DNA sequence; in fact, any particular band may contain as many as 100 genes. Such maps can nonetheless sometimes be used to identify the region in which to search for a gene—if the gene causes disease when defective, for example, and the appearance of the band differs in normal and afflicted individuals. For most genes, however, this level of resolution is too low to detect differences between individuals.

A technique called *somatic cell hybridization* provides another low-resolution means of telling which genes are on which chromosomes. It also allows researchers to isolate a particular chromosome from the others so it can be studied alone. In this method researchers create hybrid cells, cells that have both human and mouse chromosomes. To do this they fuse human cells growing in culture (usually a type of skin cell) with mouse tumor cells, also grown in culture. Cells like this, with too many chromosomes—too many instructions for carrying on their daily lives—get rid of some chromosomes with each new generation, and they tend to throw out the human chromosomes. Over time the hybrid cell line looks more and more like a mouse cell line: all of the mouse chromosomes replicate and are passed on when the cell divides, but only one or a few of the human chromosomes persist. Researchers have created lines of hybrid cells that contain single copies of human chromosomes 2 to 9, 11 to 14, 16 to 19, 21, 22, X, and Y. If a human protein is produced by a given cell line, say the line with chromosome 7, then researchers know that the gene coding for that protein lies on chromosome 7.

The technique of *in situ hybridization* allows researchers to find a known base sequence on a chromosome. It requires first having isolated a bit of DNA

known to be a gene, part of a gene, or some other marker, and knowing its base sequence. This bit of DNA is called a *probe.* It is a single strand, and it can be labeled (marked so it is distinguishable from other such strands) with either a radioactive isotope or a fluorescent dye. Researchers can then take a slide with a complete set of chromosomes and heat it or treat it chemically so that the chromosome's DNA strands separate. When the probe is washed over the slide, it will bind to sequences complementary to it on the chromosomes. If the probe is labeled with a fluorescent dye, the region of the chromosomes it binds to lights up under a fluorescence microscope; if it is labeled radioactively, the slide can be exposed to a photographic emulsion to produce pictures to be examined under a microscope.

A well-prepared *in situ* hybridization can provide direct evidence that a particular gene or DNA sequence is present on a specific region of a chromosome. It has been a principal method of mapping human genes to chromosomes; it is useful in the initial effort to trace specific genes or disease markers to a particular chromosome. Once a general region has been identified by means of *in situ* hybridization, higher-resolution mapping methods can be used to characterize the location in more detail.

Physical Maps: High Resolution

To make detailed, high-resolution chromosome maps, researchers first need purified samples of chromosomes. Human chromosomes can be extracted from cells growing in culture. Then they must be separated. When the chromosomes have been labeled with a fluorescent dye, an instrument called a *flow sorter* can separate them by size by detecting how brightly they glow: since more dye sticks to the larger chromosomes than to the smaller ones, the larger ones fluoresce more brightly. Hybrid cell lines like the mouse-human lines described above can be used to manufacture copies of a single human chromosome; a flow sorter then separates the human chromosome from the mouse chromosomes.

With 150 million bases, however, the average chromosome is too unwieldy to sequence. So researchers cut the chromosome into smaller, more manageable chunks using enzymes called *restriction enzymes.* Physical mapping is the process of reassembling those chunks in their original order. Restriction enzymes recognize specific base sequences and cut DNA at those sequences. Some restriction enzymes recognize sequences that are repeated frequently along DNA. Treating DNA with such enzymes results in many small fragments. The restriction enzyme called *Alu* I is one of these frequent-cutting enzymes; it would cut chromosome 11, for example, into about 250,000 pieces. Other restriction enzymes cut DNA at less frequent intervals: treating chromosome 11 with the enzyme *Hind* III would result in about 17,000 DNA fragments. The length of the fragments can vary among individuals because the patterns of repeating sequences in their DNA differ. These variations are known as *restriction fragment length polymorphisms,* or RFLPs ("riflips)," and the use of probes to find these variations is known as *DNA fingerprinting.* RFLPs are used as *markers* (identifiable sites on a chromosome) for both genetic and physical maps.

In practice it is useful to stop the chemical reaction before the enzyme has cut the DNA at every place it can. This results in fragments with overlapping sequences. Detecting the overlaps helps researchers put the fragments in order. First, however, the pieces must be separated by size using *gel electrophoresis.* An electric current is passed through a gel in which the fragments are suspended; it sets them moving at different rates, depending on their size and electric charge. Larger fragments tend to move all at the same rate, but a new technique, pulsed-field gel electrophoresis, can separate even these: the rate at which they respond to changes in the direction of the electric field sorts them out.

Amplifying and Ordering DNA

Once fragments of DNA have been obtained and sorted by size, genome researchers need to prepare the large amounts required for physical mapping—to *amplify* them. They must also package the fragments so they can be analyzed. Researchers may *clone* the fragments, incorporating them into the DNA of organisms like bacteria and yeast that can be grown in large quantities, thus producing lots of DNA. The fragments can also be grown in cultures, using the *polymerase chain reaction* (PCR).

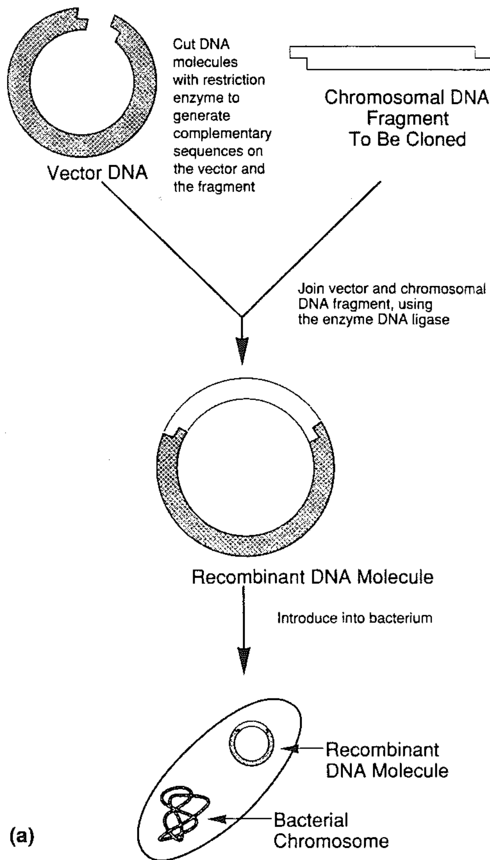
In PCR the DNA to be copied (replicated) is heated to separate it into its two strands, then two *primers*—short chains composed of nucleotide sequences complementary to those at one end of each strand—are added, along with a heat-resistant enzyme to catalyze synthesis of new strands. The new strands act as templates in their turn. Since PCR does not require a host cell, it is very fast: a fragment of DNA can be copied more than a billion times in three hours. PCR can also be used to detect the existence of a particular sequence in a DNA sample.

Cloning relies on *vectors*—DNA molecules that can accept foreign DNA, be inserted into a host cell, and still replicate themselves. If the target DNA fragments are relatively small (less than 12,000 base pairs), they can be spliced into circular bits of bacterial DNA called *plasmids.* The bacteria can then be grown in culture so that the bacteria in each culture have a plasmid containing a different fragment of human DNA.

Bacteria can also be made to carry larger fragments. A type of vector called a *cosmid,* containing part of a virus that normally infects bacteria, can carry DNA inserts of up to about 45,000 base pairs. Relatively huge DNA fragments—up to a million base pairs long—can be inserted into a vector containing the portions of yeast chromosomal DNA that cause it to replicate. This is known as a *yeast artificial chromosome,* or YAC.

Researchers refer to a collection of clones, like those grown in bacteria, as a *library* or *random library.* There is no obvious order to the library—putting the clones in order is the job of physical mapping. The clones in a library are ordered by subdividing the DNA fragments into even smaller fragments and identifying which clones have some common, overlapping, subfragments. Groupings of clones representing overlapping or contiguous regions of the genome are known as *contigs.* On an incomplete map contigs are separated by gaps where not enough clones have been mapped to allow the connection of neighboring contigs. Of all the steps in physical mapping, connecting all contigs poses the most technical problems.

Genome researchers refer to mapping with contigs as mapping from the "bottom up." Physical mapping that begins with large DNA fragments, as in yeast artificial chromosomes, is called "top-down" mapping. Although top-down mapping is of lower resolution than bottom-up mapping, it produces maps with fewer gaps.

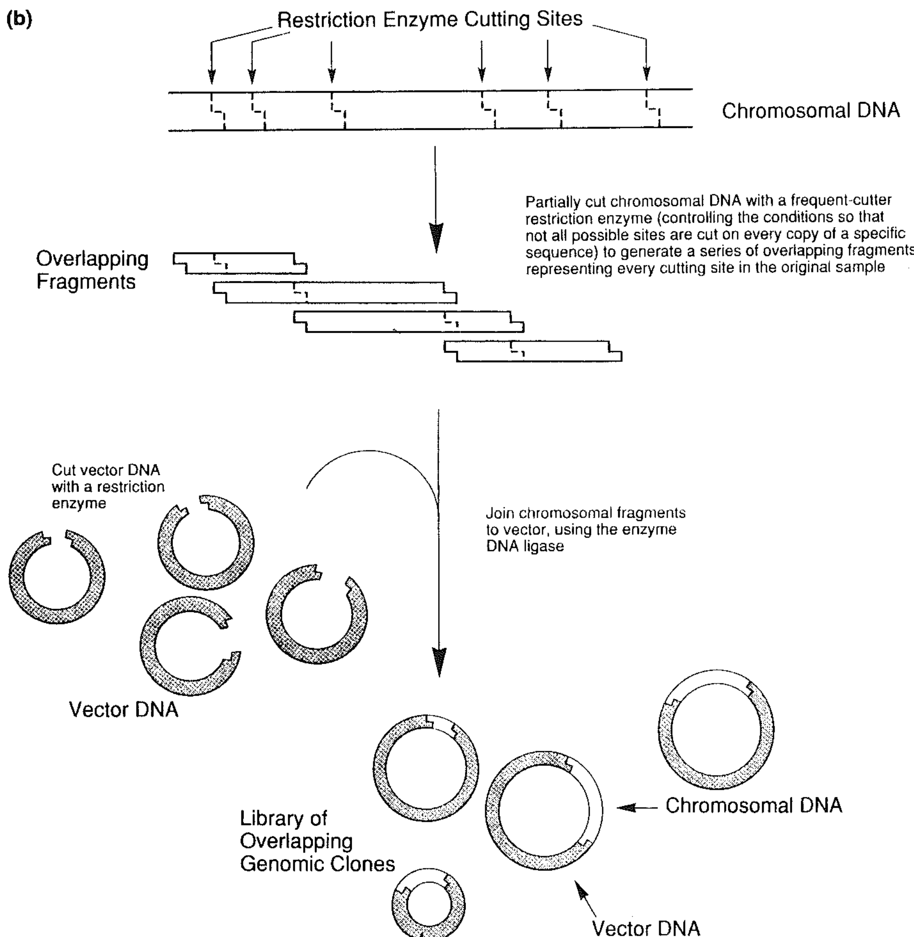


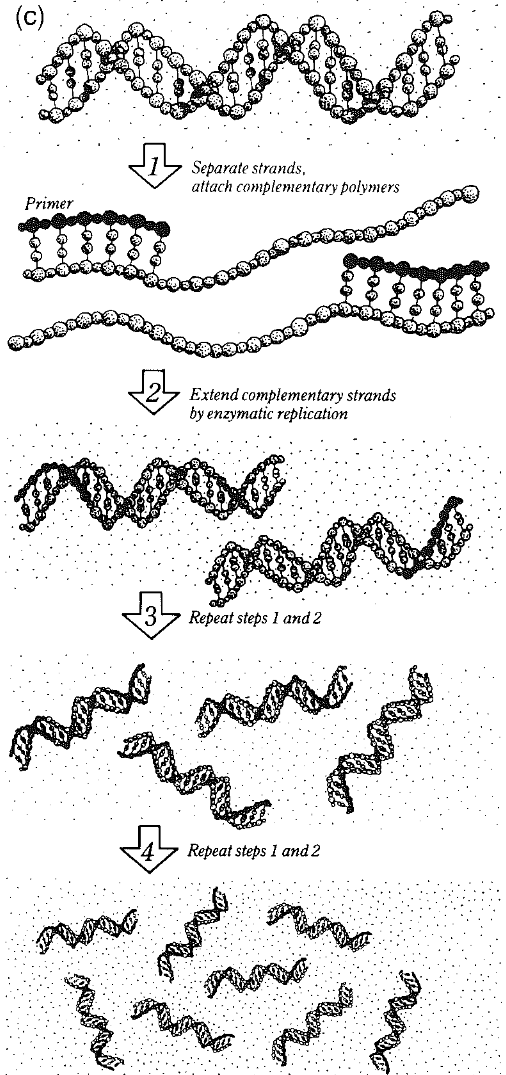
Amplifying and ordering DNA.

***(a) Cloning DNA in plasmids.** If DNA of any origin (human, animal, or plant) is fragmented and inserted in the DNA of rapidly reproducing foreign cells, billions of copies of a single gene or DNA segment can be produced in a very short time. Here the DNA to be cloned is inserted into a plasmid (a small, self-replicating circular molecule of DNA) that is separate from chromosomal DNA. When the recombinant plasmid is introduced into bacteria, the newly inserted segment will be replicated along with the rest of the plasmid. From* HGPR 1991–92, *p. 208.*

***(b) Constructing an overlapping clone library.** A collection of clones of chromosomal DNA, called a library, has no obvious order to show where the cloned pieces originally sat on the uncut chromosome. To establish that two given clones are next to each other in the genome, libraries of clones that partly overlap must be constructed. They are ordered by dividing the inserts into smaller fragments and determining which share common DNA sequences. From* HGPR 1991–92, *p. 209.*

***(c) Polymerase chain reaction (PCR).** In vitro DNA amplification: PCR can amplify a target DNA sequence of any origin (virus, bacteria, plant, or human) hundreds of millions of times in a matter of hours—a task that takes days with recombinant technology. The method relies on a specialized enzyme, a polymerase that can synthesize a strand complementary to a given DNA strand in a mixture containing the four DNA bases (A, T, C, G), and two DNA fragments about 20 bases long, known as primers. Repeated heating and cooling separates the DNA strands, then allows the primers to find and bind to their complementary sequences and the polymerase to extend the primers into new complementary strands. From* HGPR, 1989–90, *p. 128.*





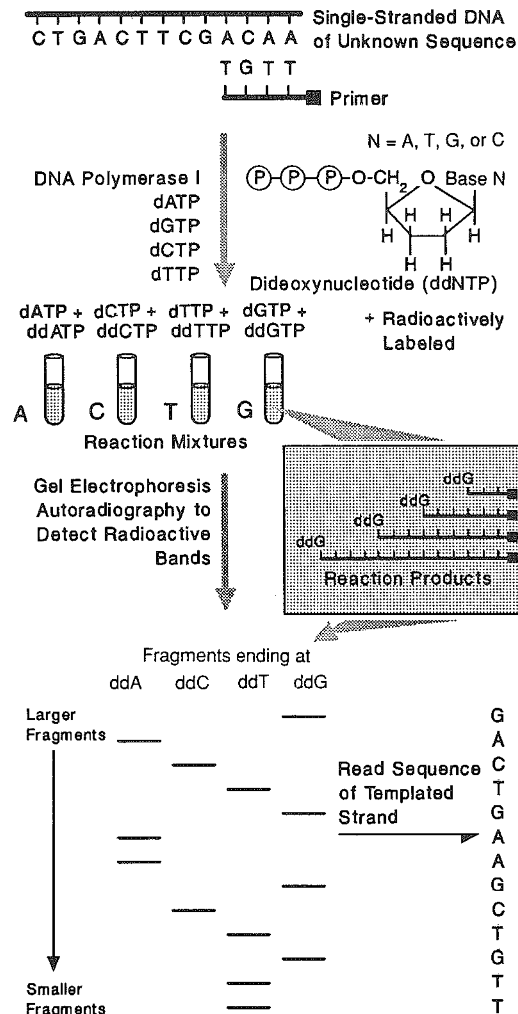
Sequencing Technologies

Sequencing is the highest-resolution mapping technique. Strategies for sequencing the human genome are much more controversial than other approaches to mapping. Some scientists favor sequencing only genes that are known to be expressed (to contain instructions for, e.g., making a given protein). Others propose that sequencing should be targeted at specific regions of interest, as is currently done. Still others hold that the whole genome should be sequenced because it could reveal sections of the genome with important functions that would otherwise go unidentified. Some geneticists believe that all or most of the DNA has some function, even though the function may be unknown right now.

Two methods for sequencing DNA are standard in laboratories today. The Sanger method, also called the dideoxy or chain-termination method, uses an enzymatic procedure to synthesize DNA chains of varying lengths along a single-stranded DNA template, much as in PCR, ending in a predictable base. Four reactions are run simultaneously, one for each base. The Maxam-Gilbert method uses chemicals that degrade DNA in a controlled way. In both methods the lengths of the resulting chains (how fast they move) are read using gel electrophoresis. Most steps in these methods are now automated.

Technologies to make sequencing faster and more sensitive, accurate, and economical are constantly under development. One adaptation of the Maxam-Gilbert method, called multiplex sequencing, enables a researcher to analyze a large set of cloned DNA fragments—about 40—as a mixture on a single gel throughout most of the sequencing steps. As in traditional sequencing, each clone in the mixture is tagged with short, unique sequences of DNA in the first step, and its nucleotide is deciphered in the final step. Multiplex sequencing is ten times as efficient as the standard procedure.

Other techniques promise to bring "visual" inspection down to the molecular level. Scanning tunneling microscopy uses a tip close to atomic dimensions to scan the surface of a specimen without destroying it; the makeup of simple surfaces has been distinguished at close to the atomic level with this approach. With DNA specimens the objective, still not achieved, is to distinguish the four bases on the sugar-phosphate backbone of DNA. The recent development of intense, coherent (laser) X-ray sources and high-quality X-ray optics may also allow resolution fine enough to define the sequence of DNA bases on a chromosome. In principle, a single strand could provide sufficient data to reconstruct a holographic image of the DNA molecule.



DNA sequencing. In the Sanger method four reactions are prepared. All contain the target DNA to be sequenced, the four nucleotides needed for DNA synthesis, DNA polymerase to catalyze the synthesis, a primer that will begin the reaction at one *end of the chain (see the figure for PCR), and an artificial* (dideoxy) *nucleotide that will halt the reaction at the base it resembles once it is incorporated in the chain; all four reactions will result in chains of different lengths. Each reaction tube, however, contains a different dideoxy nucleotide, and its corresponding nucleotide is radioactively labeled. Each chain in, e.g., tube G will end where a guanine should be taken up. When each reaction is placed in its own lane on a gel and subjected to an electric current, the chains of different lengths will move at different rates; because all four reactions begin at the same base, no chains that end at, e.g., A will be the same length as any that end at, e.g., T. The resulting four-lane radiogram can then be read as the sequence of bases in the genome fragment. Adapted from* OTA, Mapping Our Genes, *p. 45.*

Keeping Results Comparable

Constructing the human genome map requires considerable sharing of data. Several of the individual chromosomes are being mapped at more than one location. The researchers at these different labs often use different markers, cells from different stricken families, different techniques and strategies. How can all this information about any given chromosome be coordinated in a single map?

The current solution is the sequence-tagged site, or STS, strategy for mapping. This strategy requires that a researcher or research team find a short DNA sequence that occurs only once in the genome in the element (individual clone, contig, or sequenced region) being mapped. This unique STS defines the element. The researchers would use whatever mapping techniques they chose but would always report the results in terms of the STS markers, that is, in the same language. Eventually a crude map of the entire genome, showing the order and the spacing of STSs, could be constructed. The STS system has many other benefits: permitting cross-referencing of physical and genetic maps; eliminating the need to obtain probes and other reagents from the original investigator; being affordable for small as well as large labs; and generally saving money, effort, and time.[22]

22. "STS—New Strategy May Provide Common Link for Mapping," *Human Genome News* 2(3) (Sept. 1990), 1–2, 16.

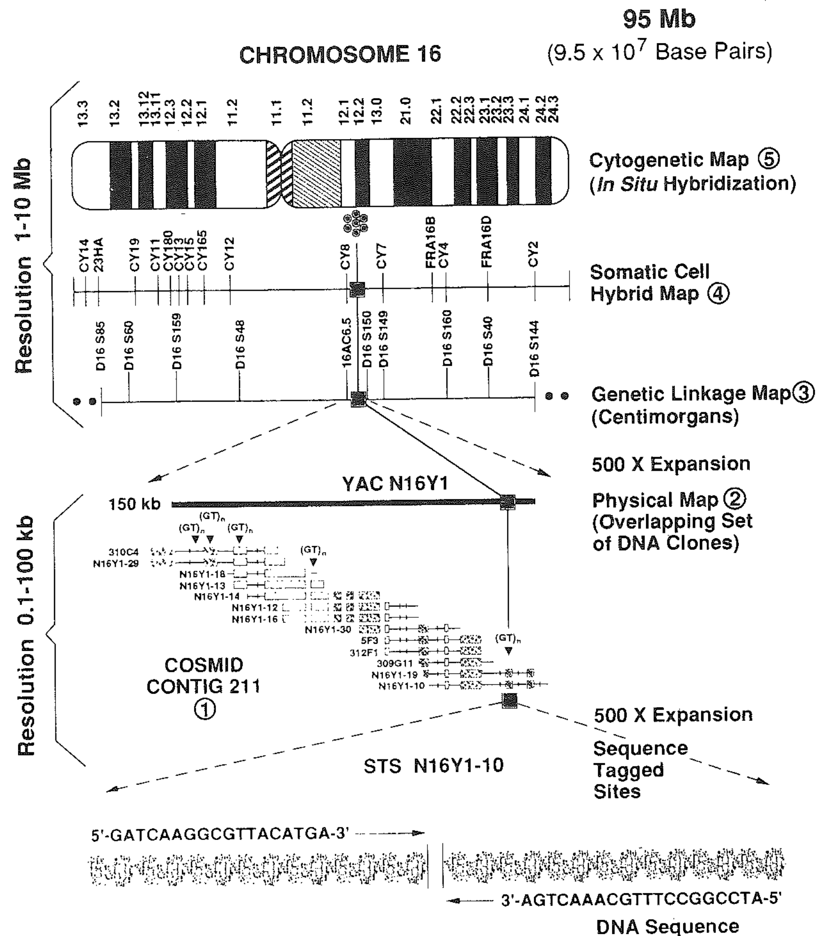
Model Organisms

While not precisely a technology, model organisms constitute one technical resource of the Human Genome Project. The genomes of such organisms as roundworms, yeasts, fruit flies, and mice are similar to the human genome in various ways, and researchers can use them as models to study gene identity, organization, and function, to examine the processes and diseases that have counterparts in humans, and to search for homologous genes.[23]

The roundworm *Caenorhabditis elegans* is the simplest organism possessed of a nervous system; it has only six chromosomes. Extensive study over the past thirty years has revealed much about its cellular development and physiology, especially the relation between specific behaviors and particular cells. Nearly complete physical and genetic linkage maps have been generated for its six chromosomes, and a pilot project has begun to determine the entire base sequence of its genome.[24]

Single-cell organisms also have their uses. *Saccharomyces cerevisiae*—the common brewer's yeast—is valuable because it is a *eukaryote:* unlike bacteria but like human cells, yeast cells have a well-defined subcellular structure, including a discrete nucleus bound by a membrane. Many major proteins fill the same functions in yeast and in the higher eukaryotes. More is known about the relation between genetics and biochemistry and between structure and function than for any other eukaryote. Work with the *prokaryotes* (whose DNA is contained in a single circular chromosome)—especially *Escherichia coli* and *Mycoplasma capricolum*—is also yielding considerable information about biochemical mechanisms responsible for gene expression.

Among multicellular organisms, fruit flies, especially *Drosophila melanogaster,* have long been used in genetic research. The genetics of fruit flies is among the best characterized; morever, there are many human counterparts to fruit fly genes (over 400 described to date), and information about how fruit fly genes work can illuminate the function of human genes. *Drosophila* studies are being coordinated at a new genome center based at the University of California, Berkeley. Mice are also widely used in genomics research. Mice share up to 80 percent of their DNA with humans, thus the mouse genome can serve as a model for human hereditary disease.



Integrating genetic and physical maps. *A cloned DNA fragment is related to the cytogenetic map using sequence-tagged site (STS) markers (a unique short segment of DNA of known sequence). The location of STS N16Y1-10 is shown from the bottom up in (1) an ordered set of cosmid clones (cosmid contig 211); (2) a 150-kb YAC insert (YAC N16Y1); (3) a genetic linkage map of chromosome 16 between two genetic markers (16AC6.5 and D16S150); (4) a somatic cell hybrid map between two markers (CY8 and CY7); and (5) a cytogenetic map between 16q12.1 and 16q12.2, as determined by in situ hybridization. From* HGPR 1991–92, *p. 52 (figure provided by Monica Fink, LANL).*

Large numbers of genome scientists are involved in generating the physical and genetic maps of the nineteen mouse chromosomes for comparison with human maps, most notably at the designated genome center at MIT (See under NIH Centers, page 18).

23. This section is based on Leslie Fink and Anne E. Adamson, "NCHGR Conducts Model Organism Studies," *Human Genome News* 2(5) (Jan. 1991), 1–2; and Verne Chapman, Peter D'Eustachio, and Joseph Nadeau, "Workshop on Mouse Genome Mapping," *ibid.*, 3(1) (May 1991), 11–12. See also OTA, *Mapping Our Genes* (see note 6), 41–42.

24. Leslie Fink, "Successful Worm Studies Yield Much Data," *Human Genome News* 4(1) (May 1992), 1–2.

WHERE THE HGP IS CONDUCTED

Dozens of institutions—national, university, and corporate laboratories—are engaged in research that relates to the Human Genome Project. In the next two sections we examine first those institutions formally designated as centers, then the repositories that provide the standardized cultures and samples without which the HGP could not proceed. In a third section we examine the relationship between science (including science conducted at the centers listed here) and industry. These tidy divisions give some idea of the scope of the project, although they gloss over its extent and the complexity of its interaction.

Designated Genome Centers in the U.S.

Two major funding agencies, DOE and NIH, which joined forces in early 1989, support the official genome centers in the United States. Most centers are funded through one or the other, though several have joint funding. The majority of DOE's work is conducted at the department's multidisciplinary national laboratories. The focus of its research and development program is to construct linearly ordered maps of DNA clones (prepared by the National Laboratory Gene Library Project) specific to a given chromosome; to improve the efficiency of sequencing DNA; and to upgrade the computer capabilities needed to organize, disseminate, and interpret the sequence of the human genome—that is, to improve *informatics*. The DOE also gives technology transfer a high priority. In contrast, the NIH funds research largely through individual grants to research teams working at universities and similar not-for-profit centers. The agency's effort focuses on comparative genetic studies of human and model organisms, promotes both predoctoral and postdoctoral training programs, and emphasizes the study of disease genes, with an eye to future applications. Several NIH centers also work on informatics—the algorithms, software, and computer hardware that analyze newly collected data and guide future research.[25]

DOE Centers

There are three major DOE centers, all working on physical mapping of an individual chromosome. The Center for Human Genome Studies at Los Alamos National Laboratory (LANL), of which Robert K. Moyzis is director (established 1988), is mapping and fingerprinting chromosome 16, which controls the synthesis of hemoglobin. Using repeat sequences as nucleation sites (starting points), the LANL center seeks to produce contig maps with landmarks that are useful for integrating genetic and physical maps rapidly. The center's informatics project has developed the computer and statistical techniques needed to predict the probability overlap pairs of cosmid clones when mapping fingerprint data. The center also has an ELSI program (see Ethical, Legal, and Social Issues, page 28).

The Human Genome Center at Lawrence Berkeley Laboratory (LBL), of which Jasper Rine is now director (established 1988), is mapping and sequencing chromosome 21, connected with Down syndrome and with several degenerative diseases. Work includes physical mapping with single-copy DNA probes as anchor points; informatics; and the implementation of a comprehensive chromosome-21 information system, sequencing, and DNA manipulation. The center is determining the conditions necessary for selective cleavage of single-stranded DNA and RNA adjacent to hybridization sites.

The Human Genome Center at Lawrence Livermore National Laboratory (LLNL), where Anthony V. Carrano is director (established 1990), is mapping chromosome 19, which contains several genes important to the repair of damaged DNA, and exploring new approaches to physical mapping of chromosomes. LLNL participates with LANL in the National Laboratory Gene Library Project. In informatics LLNL is developing new computer software to analyze physical maps and display the resulting data graphically. It is developing new instruments as well, including an image-analysis system used in fluorescent *in situ* hybridization (FISH) studies; a general robotic tool to automate the handling of DNA and relieve researchers of highly repetitive tasks; and scanning-tunneling microscopy or spectroscopy for cost-effective, high-volume DNA sequencing.[26] (See also Sequencing Technologies, page 16; and Corporate Involvement, page 22.)

Three other DOE labs—Argonne National Laboratory, Brookhaven National Laboratory, and Oak Ridge National Laboratory—also participate in genomics research, mainly DNA sequencing.

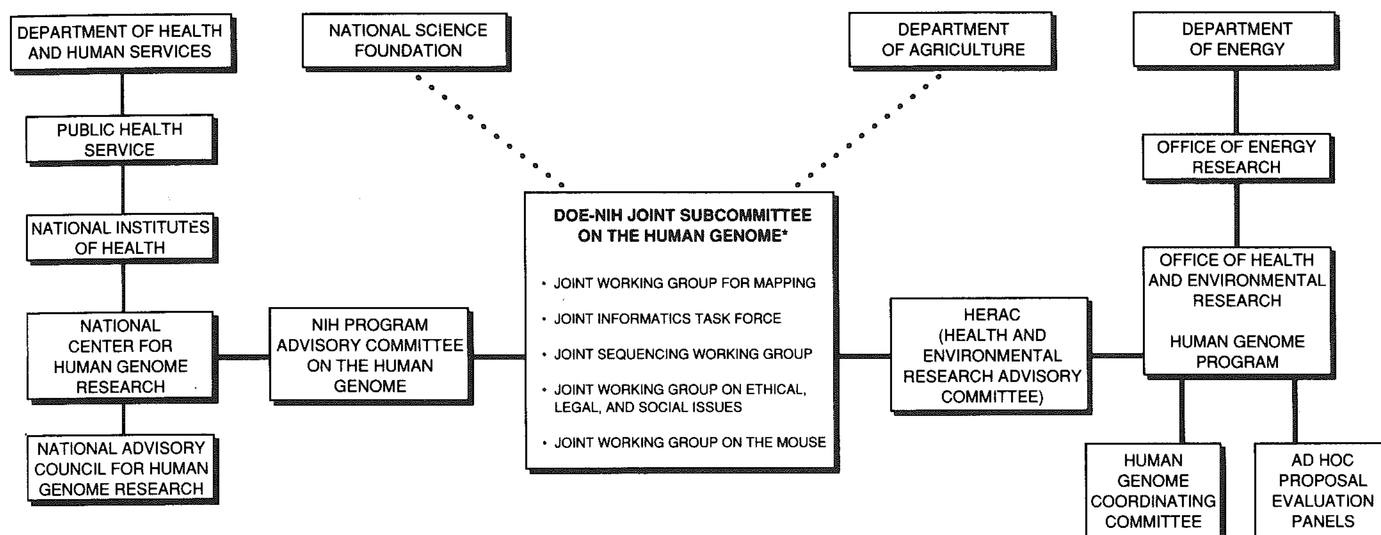
NIH Centers

Centers funded by NIH engage in a variety of programs. One, based at the Whitehead Institute for Biomedical Research and the Massachusetts Institute of Technology (established 1990), focuses on the mouse genome. Its director, Eric S. Lander, leads a consortium of twelve principal researchers at MIT, Whitehead, Harvard University, Princeton University, and the Jackson Laboratory (see Genome Resource Centers, below). The researchers are constructing highly detailed maps of mouse chromosomes 1, 11, and X, with the long-term goal of constructing both a high-resolution genetic map and a low-resolution physical map of the entire genome. Because the mouse is a well-studied mammalian model, scientists use mouse mutations to study the effects of mutations on cell function, immunology, neurobiology, reproduction, and behavior (see

---

25. Much of the material on genome research centers is based on the table in *Human Genome News* 4(6) (March 1993), 2–8. See also Betty K. Mansfield and Judy M. Wyrick, "DOE Holds First Human Genome Contractor-Grantee Workshop," *Human Genome Quarterly* 1(3) (Winter 1990), 2–4; and Leslie Fink, "Baylor, Utah, Receive NCHGR Center Grants," *Human Genome News* 2(6) (March 1991), 12.

26. "New NIH Genome Centers Signal Milestone," *Human Genome News* 2(4) (Nov. 1990), 3–4.



Organizational chart of joint DOE–NIH activities. From HGPR 1991–92, *p. 65.*

Model Organisms, page 17). The MIT center is also constructing a physical map of the human genome and preparing a YAC library as a resource for other scientists interested in studying the mouse genome.

Another model organism, the fruit fly, *Drosophila melanogaster,* is the focus of a relatively new center at the University of California, Berkeley, directed by Gerald Rubin (established 1992). Researchers connected with this center will map the complete genome of the fruit fly. The center will also serve as a centralized source of technologies and materials for scientists studying fruit fly genetics (see Model Organisms, above). This center includes researchers at Harvard, the Carnegie Institution of Washington, and Lawrence Berkeley Laboratory, as well as Berkeley.

Several NIH centers, like the DOE centers, focus on a single human chromosome. One is funded jointly by DOE and NIH: the Salk Institute at La Jolla, California, where Glen A. Evans is director. The Salk center (established 1990) is working on human chromosome 11, to which 133 genes have been mapped, including those for Wilms' tumor, genitourinary defects, and mental retardation. Genes that play a role in several forms of cancer and allergies are also believed to be located on chromosome 11. One of the center's priorities is rapid identification and isolation of disease genes.

The center at the University of California, San Francisco, directed by Richard M. Myers and David R. Cox (established 1990), recently moved to Stanford University (March 1993). The center is working on chromosome 4, one of the largest and potentially most important chromosomes in terms of human disease genes. Chromosome 4 is believed to contain genes for Huntington's disease and one form of Alzheimer's disease. Researchers are constructing a rough map of the chromosome using *in situ* hybridization and then filling in details with landmarks prepared by other methods.

Children's Hospital of Philadelphia (CHOP) is the first hospital in the nation to be designated a genome center site. The center (established 1991), directed by Beverly S. Emanuel, operates in collaboration with the University of Pennsylvania School of Medicine; it also has collaborative arrangements with the Fox Chase Cancer Center in Philadelphia and the Du Pont–Merck Pharmaceutical Company. The CHOP-Penn center is mapping human chromosome 22, the "Philadelphia chromosome," associated with eight types of cancer and a number of birth defects, including DiGeorge syndrome. Investigators are using a variety of techniques to locate some 300 anchor markers at regular intervals along the chromosome. They are also using yeast artificial chromosomes (YACs) as cloning vectors to subdivide the chromosome into an ordered collection of DNA fragments. Once isolated, this set of fragments spanning the entire chromosome will be available to the scientific community for use in locating and isolating important genes.

Chromosome 3 is the focus of the recently established center (1992) at the University of Texas Health Science Center at San Antonio, directed by Susan L. Naylor. Besides constructing radiation hybrid, genetic linkage, and contig maps, the center is developing a chromosome 3–specific database with the Utah center described below.

At Washington University, in St. Louis, Missouri, researchers under the direction of David Schlessinger are using the YAC technology developed at their laboratory to construct complete maps of human chromosomes 7 and X. Chromosome 7 is believed to contain a total of about 5,000 genes, including the cystic fibrosis gene and genes that control immune response. The X chromosome has also been the target of intensive study; genes for hemophilia A and B, diseases of the adrenal gland, fragile X syndrome, and color blindness are among those located on it. The YAC library developed here is a resource for other genomics researchers.

Among the centers studying a number of chromosomes, that at the University of Michigan in Ann Arbor, directed until recently by Francis S. Collins, is particularly focused on disease genes as opposed to strict mapping or sequencing. This group identified genes that cause cystic fibrosis (chromosome 7) and

neurofibromatosis (17, 22). As a genome center (established 1990), the group has focused on improving technologies and speeding up the process of identifying disease genes "from clinic to base pair," with special emphasis on the Huntington's disease gene (chromosome 4). (The center may be moving to Bethesda, Maryland, now that Collins is director of the NIH National Center for Human Genome Research.)[27]

The genome center at Baylor College of Medicine in Houston, Texas, though also focusing on genes responsible for diseases, is working on both physical and genetic mapping. The goal of the center, directed by C. Thomas Caskey (established 1990), is to improve DNA-sequencing technology while developing a physical map of human chromosomes X and 17 and a genetic-linkage map of chromosome 6. Investigators collect DNA samples from patients with inherited diseases and, with special computer programs, attempt to locate disease genes on cloned YACs.

Genetic mapping is the focus of the genome center at the University of Utah in Salt Lake City, directed by Raymond Gesteland and Ray White (established 1990). This center is developing high-quality DNA markers to add to the genetic-linkage map of human chromosomes 16 and 17 and part of 5 and to help connect the genetic and physical maps. Investigators expect to generate about 640 markers each year. The Utah center collaborates with a group at the University of Alberta in Edmonton, Canada, on more rapid DNA-sequencing methods. It performs computerized genetic and statistical analyses to link information about inherited diseases to specific chromosomes and genes. Mapping technology developed at the Utah center will be available to other gene hunters through collaborations.

Finally, the new center at the University of Iowa, Iowa City, where Jeffrey Murray is director (established 1992), has an ambitious three-part program. Its goals include generating a high-resolution genetic linkage map of the entire human genome to help in locating disease-causing genes; addressing ethical, legal, and social issues raised by research in genetics; and educating high school science teachers about genetics and the HGP. Research teams at Fox Chase Cancer Center in Philadelphia, the Marshfield Medical Research Foundation in Marshfield, Wisconsin, and Harvard are also involved in the project.

Genome Resource Centers

In order to produce results that can be compared from one lab to another, molecular geneticists need to work with starting materials—cell lines and DNA samples—that have known characteristics. Researchers can obtain some source materials from private companies that sell them. But a variety of nonprofit institutions around the world also maintain such cell lines. Investigators in the Human Genome Project rely on these genome resource centers for the biological materials they manipulate and map. Repositories of biological materials supply researchers with standardized cell cultures and DNA for a nominal fee. They also collect, authenticate, amplify, and store these materials. Some of these repositories also maintain databases on results obtained using their resources.[28]

International Resources

The Centre d'Etude du Polymorphisme Humaine (CEPH) in Paris is one of the most widely used resources of this type. CEPH is a private nonprofit institution established in 1983 to facilitate construction of genetic linkage maps of each human chromosome by using DNA polymorphisms, that is, by comparing variations in the DNA sequences among related individuals. CEPH is a partner in Généthon in Paris (see page 7) but collaborates with researchers in over 150 laboratories in North America, Europe, South Africa, Japan, and Australia.[29]

The main premise of CEPH is that collaborative research on DNA from the same families will result in earlier completion of the human genetic linkage map. CEPH provides its collaborators with samples from a reference panel of DNA from three to four generations of about 60 large (with at least 6 children) families. About half of these families are drawn from the Utah Mormons, studied since the 1960s for various traits, including a gene for colon cancer. Some Venezuelan pedigrees are also included in the CEPH collection, as are other families from France and Denmark. Most of these families have no known genetic diseases. In 44 of these families cells have been obtained from all of the grandparents.[29]

In order to obtain material from CEPH, collaborating investigators must first possess DNA probes (see *in situ hybridization*, pages 13–14) that detect genetic markers; the markers targeted are generally *restriction fragment length polymorphisms* (RFLPs; page 14). The collaborators must agree to use the probes to test the entire panel and to provide CEPH with all of their data. There are no enforcement mechanisms, but so far researchers have cooperated. The collaboration requires the sharing of data but not of the actual probes, which could prove to be patentable, to help avoid potential competition.

The CEPH database, which contains genotypes for all tested genetic markers, has two components, one collaborative, one public. The collaborative database is available only to CEPH investigators; it includes unpublished and published data. Published data are later moved to the CEPH public database, available to the general scientific community. Unpublished data also can be released to the public database, after two years.

CEPH is also collaborating with the American Type Culture Collection to construct probe kits for the mapped genetic markers on CEPH consortium maps. Sponsored by NIH National Center for Research Resources, this project will enhance use of the CEPH consortium genetic linkage map by enabling researchers to localize disease-determining and other interesting genes.

Another important European resource is the United Kingdom's DNA Probe Bank. Funded by the Medical Research Council as part of the U.K. Human Genome Mapping Project Resource Centre, the probe bank offers some 650 DNA probes free to the research community of the United Kingdom. Foreign investigators may be subject to fees and restricted access. The European Cell Bank also maintains cell lines of interest to

27. See Larry Thompson, "Healy and Collins Strike a Deal," *Science* 259 (1993), 22–24.

28. Much of the material on genome resource centers is based on Denise Casey, "Nonprofit Resource Centers Facilitate Mapping," *Human Genome News* 3(1) (May 1991), 14–17.

29. Much of the material on CEPH is drawn from OTA, *Mapping Our Genes* (see note 6), 146. See also Alexander Dorozynski, "Gene Mapping the Industrial Way," *Science* 256 (1993) 463.

30. Christopher Wills, *Exons, Introns, and Talking Genes* (New York: Basic Books, 1991), 278–280, discusses the CEPH families and their importance to genomics research.

human genome mappers. And the Japanese Cancer Research Resources Bank, although established to facilitate cancer research, includes cell and gene repositories.

U.S. Repositories

Resource centers in the United States also are generally private, nonprofit organizations that house projects funded by various government agencies, especially sections of the NIH. The American Type Culture Collection and the Coriell Institute for Medical Research are the two major facilities. The National Cell Culture Center and the Jackson Laboratory provide similar services.

The American Type Culture Collection (ATCC) in Rockville, Maryland, has maintained and distributed biological resource material to researchers since 1925, including such traditional material as viruses, bacteria, fungi, protozoa, and metazoan cell lines (i.e, cells from animals with differentiated tissues and organs). In 1985 the NIH National Institute of Child Health and Human Development established an international Repository of Human and Mouse DNA Probes and Libraries within ATCC. ATCC staff obtain, amplify, and distribute over 100 probes that detect RFLPs. They also maintain clone and genomic repositories from known genes. The clone repository (over 1,640 human and 155 mouse clones) contains discrete gene sequences from various parts of the human and the mouse genomes. Clones specific to each human chromosome and most mouse chromosomes are available.

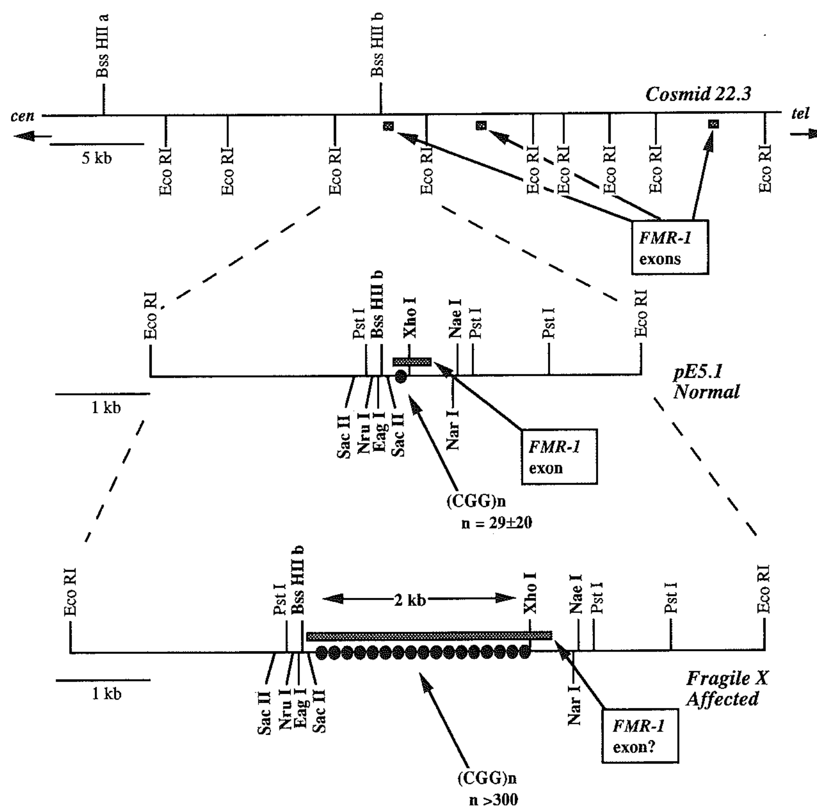
The ATCC also manages the National Laboratory Gene Library Project (NLGLP), preserving and distributing sixty gene libraries specific to given chromosomes, which were constructed at Lawrence Livermore National Laboratory and Los Alamos National Laboratory. In addition, ATCC offers predesigned primers—labeled single strands of DNA or RNA with a known sequence—for in vitro amplification of human DNA; it also puts out product sheets with information on the primers, including allele sizes, gene name, cytogenetic location, and sequence.

The Coriell Institute for Medical Research in Camden, New Jersey, establishes, characterizes, stores, and distributes more than 7,000 cell lines. Coriell began offering DNA samples in 1990.

Three major NIH cell repositories housed at Coriell constitute the world's largest collection of human cells. These are the Human Genetic Mutant Cell Repository of the National Institute of General Medical Sciences, founded in 1972; the Aging Cell Repository of the National Institute on Aging, created in 1974; and the National Cell Repository of the National Institute of Mental Health, established in 1990. The National Cell Repository facilitates research on the genetic components of manic depression, Alzheimer's disease, and schizophrenia. These three repositories, along with smaller collections on cancer, diabetes, and other diseases, are collectively known as the John T. Dorrance, Jr., International Cell Science Center.

Coriell's cell cultures are derived from human fibroblasts, lymphoblasts, and amniotic fluid; they represent more than 400 genetic diseases and 800 chromosomal aberrations. For research on the links between given genes and diseases Coriell offers cultures from multigenerational family groups, with samples from affected individuals and families (including nonaffected members) with cystic fibrosis, fragile X–linked mental retardation, Huntington's disease, retinitis pigmentosa, and major affective disorder. Coriell also has human-rodent somatic cell hybrids (see page 13), available both as cell cultures and as purified DNA, plus DNA from single-chromosome hybrids for human chromosomes 2–9, 11–14, 16–19, 21, 22, X, and Y. Coriell also maintains an extensive bibliographic database and abstracts of literature citations on its various cultures.

The National Cell Culture Center in Minneapolis, Minnesota, makes animal cells and secreted proteins available in large quantities for small research laboratories and larger collaborative groups. Sponsored by NIH National Center for Research Resources, this service is available to researchers in the United States and Canada, though preference is given to NIH-sponsored projects.



Map of the fragile X site in cosmid 22.3 (a cloning vector; see page 14). *The top line shows the general location; the second line, an enlarged view of the site in a normal individual; and the third line, a similar view of the site in an individual affected by the syndrome. The greatly increased repetition of the base sequence CGG at the site (see the* n *given below the site) presumably causes the partial detachment of a fragment of the X chromosome for which the syndrome is known. From* HGPR 1991–92, *p. 8 (figure provided by David Nelson).*

The Jackson Laboratory, in Bar Harbor, Maine, supplies genetically defined mice. Jackson, a participant in the MIT center working on the mouse genome (see page 18), maintains a diverse collection of mouse strains for basic research and for studies of the genetic and developmental factors that underlie a variety of disorders. Over 700 different mutations and 1,000 different strains and sets of strains are regularly used for mapping studies. The Jackson Laboratory's Genetic Information Resource compiles and distributes information on the lab's numerous mouse strains and mutants.

Corporate Involvement

The genome project depends on private industry as well as on government agencies and the nonprofit sector. While much of the high-visibility science takes place in academic or national laboratories, this activity depends on corporate suppliers of data analysis systems, innovative technology, routine materials, and assistance of various kinds. Conversely, corporations both profit from the HGP's current laboratory needs and have a stake in future commercial applications. Many private companies have adopted the so-called gold-rush strategy of producing machines and supplies for the growing genomics community—from sequencing machines to computer programs. Biotechnology and pharmaceutical firms are also helping to translate the scientific findings of genomics research into diagnostic tools or medical therapies—potential medical applications often given as the primary public justification for government funding of HGP—and will market them.

This section explores the various roles of private industry in genomics research, suggesting some of the problems raised by the collaborations between university researchers and pharmaceutical or biotechnology firms, and some of the ways that industrial involvement has been productive for HGP. (For many of the processes, see Techniques and Technologies, above.)

Early Corporate Interest

Corporate interest in developing instrumentation and methods for sequencing genetic materials existed before the HGP was conceived. The recombinant DNA research of the 1970s made clear the potential of this kind of work in drug development and medical care. The case of the group that worked under Leroy Hood at the California Institute of Technology in the 1980s illustrates the various levels at which academic science, government, and industry can cooperate. These researchers began working on automating DNA sequencing in the early 1980s, collaborating first with the Beckman Corporation and then with Applied Biosystems, Inc., a company formed to sell the DNA sequences that Hood's lab invented. These efforts were aided by donations from Monsanto and Upjohn, as well as by an NSF grant. Caltech was later the designated site for the Center for Molecular Biotechnology (1988), with federal funding of $3.5 million a year. It was housed in the newly constructed Beckman Institute, just donated to Caltech by Arnold Beckman.

Technology already in place helped make the HGP feasible. Applied Biosystems had a fluorescent-dye sequencer on the market by 1987. Du Pont produced a similar product the same year, and EG&G Biomolecular marketed a machine based on detecting radioactive phosphorus. This last was cheaper and intended for use by small conventional labs, rather than the mega-sequencing facilities of most HGP sites.

Another HGP milestone, the polymerase chain reaction (PCR), was actually produced by industry rather than academe. (See Constructing Libraries of DNA Fragments, under Techniques and Technologies, above.) Kary Mullis was working for Cetus Corporation when he developed the technique. The rights to PCR were later sold to Hoffmann–La Roche. More recently companies such as Genmap have been developing laboratory techniques for mapping, while others such as Transkaryotic Therapies are working on so-called speed-walking techniques, which help researchers proceed from known chromosomal markers to disease-causing genes.[31]

Corporate Collaboration

Today private companies collaborate with genome researchers in academia and government in a number of ways. At the simplest level, hundreds of small companies supply biotechnological materials and services to genome labs. In the journal *Science* their advertisements run from full-page spreads to small items tucked into the classified section. Takara offers several DNA-cloning kits, while Operon Technologies ("World's Leading Supplier of Synthetic DNA") sells DNA primers. Research Genetics promises "Custom DNA Purified and Delivered in 48 Hours" and gives customers a toll-free number for orders. Lofstrand Labs provides "research support services"—DNA sequencing, plasmid isolation, oligonucleotide synthesis—to larger labs that want to increase their productivity without increasing their staff.

Government and university labs sometimes contract out a portion of their research to private companies. Bios Laboratories, for example, not only sells cell cultures and DNA, as well as "kits" that make it easier for researchers to manipulate them, but will also carry out research using these products under contract. Bios has a contract with the National Cancer Institute to characterize a gene associated with lung cancer susceptibility and has entered into two other government research agreements. Private companies market products developed in research labs and often handle commercial applications from the start: Bert Vogelstein, an oncologist at Johns Hopkins University, is working with Hoffmann–La Roche to develop diagnostic tests for cancer based on genetics.[32]

Companies may also both fund research at universities and benefit from it. William Gates III, CEO of Microsoft Corporation, recently donated $12 million to the University of Washington to set up a new department of molecular biotechnology and recruit Leroy Hood from Caltech to head it. Hood anticipates further collaboration with Microsoft in the future. His lab will develop machines to sequence DNA; researchers at Microsoft will help design the software needed to run the machines and to analyze results.[33]

31. "Gene Rush," *Scientific American*, Jan. 1991, 112–113.

32. Scott Veggeberg, "Biotech Industry Gearing Up for More Science, Less Hype," *Scientist*, 22 June 1992, 1, 15; and Fred Gebhart, "Bios Labs Aims to Be Top Supplier to HGP Researchers," *Genetic Engineering News*, 1 June 1992, 20.

33. Doris Jones Yang, "Lighting a Fire at 'Camp DNA,' " *Business Week*, 16 Nov. 1992, 73–76.

Industry and Gene Therapy

One major role that corporations play in genomics research is as developers of gene therapy and screening tests for genetic diseases.

There are many approaches to gene therapy, but generally it involves introducing certain genes into the cells of a patient whose copies of those genes do not function. The first gene therapy trial was undertaken at the National Institutes of Health in 1990, to treat an enzyme-deficiency disease. Worldwide, nearly twenty gene-therapy trials are now under way in university or government laboratories.

In February 1992 a Federal review panel gave approval for the first clinical trial of gene therapy to be run by a company. That company was Targeted Genetics Corporation, which has devised a gene-therapy technique for treating AIDS patients. Several other companies are developing such techniques. Genetic Therapy, Inc., is aiming its therapies at cancer, hemophilia, and cystic fibrosis. Somatix Therapy has targeted cancer, neurological diseases, and hemophilia. Viagene is developing gene therapies for AIDS, cancer, and hepatitis B. Additional therapies are being developed by Cell Genesys for AIDS and eye disease; by Targetech for hemophilia and high cholesterol; by Transkaryotic Therapies for hemophilia, short stature, and anemia; and by Vical for AIDS and muscular dystrophy.[34]

Crude genetic tests that detect chromosomal abnormalities, in cells from amniotic fluid, for example, have been available for years. Genome research, however, has provided the basis for genetic tests for diseases caused by a mutation in a single gene. Once a gene responsible for a disease, say, cystic fibrosis, has been isolated, mapped, and sequenced, a screening test using a probe for that gene can be developed. Such single-gene defects are known to cause about 4,000 inherited diseases.

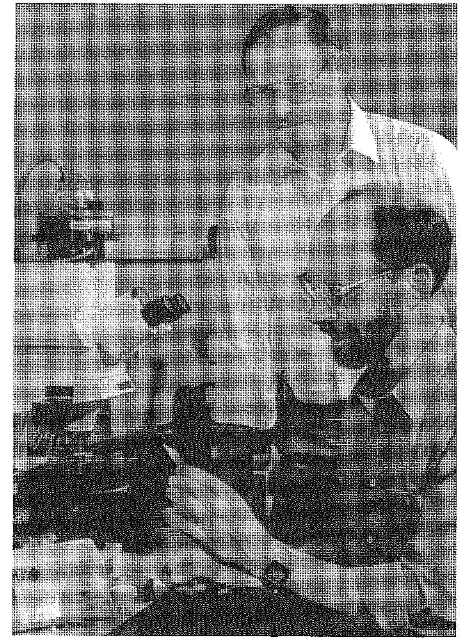
Genetic screening services provided by corporations account for about half the genetic testing carried out in the United States. (The rest is done in academic laboratories.) Vivigen, for example, performs genetic testing for cystic fibrosis and sickle cell anemia. Collaborative Research, Inc., offers testing for cystic fibrosis and adult polycystic kidney disease. GeneScreen offers a test for Duchenne muscular dystrophy. In addition to tests for disease detection, for example in prenatal diagnosis, genetic tests are marketed for determining paternity and for forensic analysis.

Technology Development and Transfer: DOE

The Department of Energy's national laboratories have made technology transfer and collaboration with industry a major focus of their operations.[35] Collaborations occur in all activities pertinent to the HGP. In biological materials Lawrence Livermore National Laboratory pioneered chromosome painting, a new technology to identify chromosomal abnormalities. The technique uses fluorescent dyes market by Life Technologies, Inc. (LTI), it is being manufactured and funded by the medical diagnostic firm Imagenetics, Inc. LLNL also developed the vectors used in chromosome painting; Amoco is now marketing these and funding their further development. LTI developed new bacterial host strains that give greater stability to several new cosmid cloning vectors produced by LLNL; the firm is marketing these vectors. Dynal Corporation and Lawrence Berkeley Laboratory are working together on better ways to isolate mRNA, chromosomes, and restriction fragments.

In instrumentation, LLNL and Autogen have been working to extend the capabilities of Autogen's automated plasmid DNA extractor so that it can extract cosmid DNA as well. Applied Biosystems and LLNL have developed chemistries and software to automate clone fingerprinting. Los Alamos National Laboratory and Bio-Rad Corp are developing Bio-Rad's pulsed-field gel electrophoresis technology so that it can be used to enhance resolution of large DNA fragments. LBL and Hewlett-Packard are devising new applications for HP's robotic systems, with HP providing technical assistance and lending some robotic hardware. LBL is also generating DNA templates for Applied Biosystems' automated DNA sequencer and working with Cruachem on a low-cost high-volume DNA synthesizer. A laser-excited fluorescence gel scanner (for reading DNA sequencing and fingerprinting gels) designed at LBL is being licensed by Molecular Dynamics, which will develop a commercial version.

In informatics—software and database designs—IBM and LANL have jointly tested computer-developed programs for efficiently calculating the probability of clone overlap, based upon fingerprint characteristics that the clones have in common. LANL has negotiated with Sybase and Servio Corporation to make available to HGP investigators Sybase's relational and Servio's object-oriented database management systems, at greatly reduced cost. Feedback from HGP researchers using the products will in turn show originators where modifications need to be made. Allied Biosystems is marketing the Contig Browser, a prototype version of software developed at LLNL, used to analyze overlapping clones—contigs—and display them graphically.



Technology to be transferred. *The co-inventors of chromosome painting prepare to examine a sample. From* HGPR 1991–92, *p. 29 (photo provided by Joe Gray).*

34. Andrew Pollack, "Commercial Test of Gene Therapy," *New York Times*, 14 Feb. 1992, D1, D6.

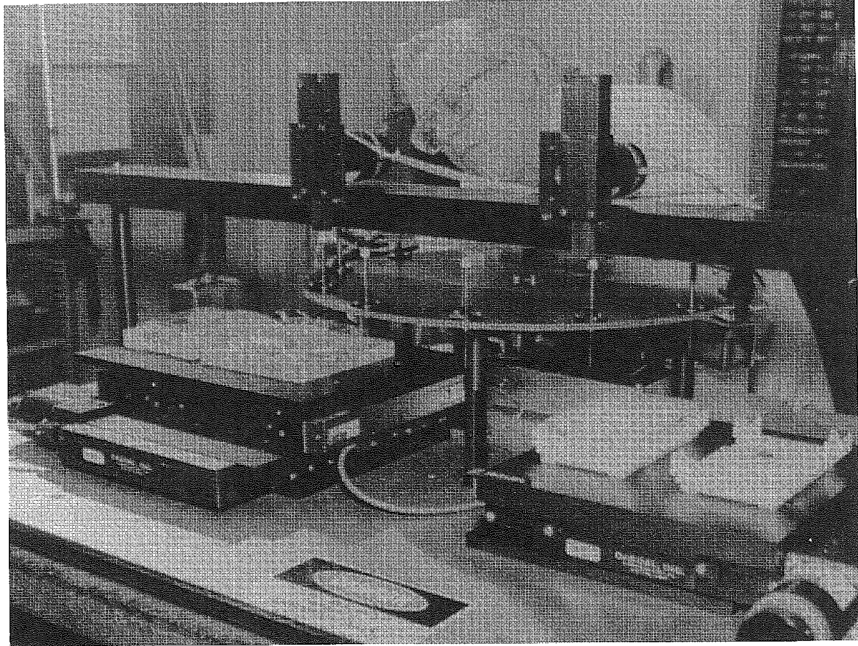
35. Much of the information in this section comes from Denise K. Casey, "DOE Centers Facilitate Technology Transfer," *Human Genome News* 2(5) (Jan. 1991), 4–6. See also "Technology Transfer," in *Human Genome: 1991–92 Program Report* (Washington D.C.: DOE, 1992), 24–29

Corporate Competition—and Managing It

The proliferation of companies that carry out tasks traditionally done in university or government laboratories suggests that mapping and sequencing—especially sequencing—could and should become industrial activities. Some researchers believe that industrial-scale involvement will be necessary in order to sequence the entire human genome. Leaving the tedious task of sequencing to industry could also free academic researchers to study fundamental biological problems, such as how genes function. Yet the relationship between government or academic researchers and commercial labs does not always run smoothly. From the beginning, projects such as building DNA sequencers generated friction over patent rights and royalties, as well as over scientific credit. This friction continues to be a problem as new techniques and genetic products are developed.

The movement of scientists between academia and industry may contribute to the difficulty. Most genome scientists have some connection to private industry, through licensing agreements. With a notable lack of historical perspective, the *Economist* suggested in 1988 that biotechnology was "the first business with enough glamour to persuade eminent scientists that the entrepreneurial spirit and academic respectability are not mutually exclusive."[36] A number of genome researchers have gone so far as to leave academic labs to join industrial ranks. Some began as consultants to a company, among them Mike Hunkapiller, who left Leroy Hood's group at Caltech for Applied Biosystems in 1983 and later became a company officer. Others founded their own companies, among them Walter Gilbert. Hood himself heads an academic center, but he is planning a corporate venture into genome research, a Seattle-based company backed by a $50 million investment from the entrepreneur Frederick Bourke. This company will focus on sequencing genes important in diagnosing and treating disease.

The largest private gene-sequencing laboratory, the nonprofit Institute for Genomic Research, located in Maryland near NIH, has been under the direction of former NIH researcher J. Craig Venter since the summer of 1992. The institute started with a $70 million grant from the Healthcare Investment Corporation; the initial thrust of its research, however, will be to scale up Venter's NIH research on sequencing fragments of DNA known as expressed sequence tags. The institute eventually plans to carry out much the same research as the academic and government genome centers: sequencing, mapping, and the biology of gene expression. It will thus go beyond scaling up to compete with the original center.[37]



Automation and the HGP. *The HGP relies heavily on automated instrumentation and invests in its development. This high-speed colony picker developed at Lawrence Berkeley Laboratory picks and arrays yeast or bacterial colonies containing target DNA sequences; it can be serviced by a robot. From* HGPR *1991–92, p. 1.*

Concern over patent rights to gene sequences has tempered academe's enthusiasm for industrial gene-sequencing ventures.Gilbert was widely criticized for the plans his short-lived Genome Corporation had to copyright sequence information and sell it to companies or researchers who wanted it. And before Venter left NIH, his attempt to patent "anonymous" cDNA—bits of expressed genes whose function was not yet known—sparked a controversy that illuminates how complex are the issues involved in patenting and the industrial development of knowledge and public resources (see Patenting, under Overview, above). The NIH in fact defended its action by maintaining that patenting turned the DNA into private property—and thus made it more likely to be the focus of expensive corporate research and development.

The DOE has started to address some of these issues, formally. In March 1991 Los Alamos National Laboratory signed a Cooperative Research and Development Agreement (CRADA) with Life Technologies, Inc. Under the three-year agreement LANL and LTI will cooperate in developing faster, cheaper techniques for determining base sequences of the human genome in much longer DNA fragments than now possible. LTI will have the first opportunity to license any products resulting from the effort and would pay royalties to LANL under such a license. The DOE plans to arrange more CRADAs, and such arrangements may become a standard part of the collaborative genome efforts.[38]

36. "Biotechnology: Inherited Wealth," *The Economist*, 3 April 1988, S3.

37. Christopher Anderson, "Controversial NIH Researcher Leaves," *Science* 258 (1992), 95.

38. Anne Adamson, "LANL, Life Technologies Approve CRADA," *Human Genome News* 3(1) (May 1991), 5–7.

COMMUNICATIONS

The size and complexity of the Human Genome Project have made improved scientific communication imperative. A number of newsletters, journals, computer nets, and other resources have been developed to fill this need. Many older journals such as *Genetics, Human Genetics,* and *Cytogenetics and Cell Genetics* (see Conferences, below) have devoted increasing amounts of space to genome research during the past ten years, and it is likely that journals and other forums would have developed even without the HGP. The HGP has, however, accelerated this process, especially the establishment of databases and computer nets.

Print Media

Newsletters

One excellent source of news on the HGP, especially for the uninitiated, is the *Human Genome News* (formerly *Human Genome Quarterly*), published by the National Center for Human Genome Research and DOE. It carries current news of progress in all areas of human genome research as well as schedules of upcoming meetings and funding news. It also features an acronym list on the last page of each issue, which is helpful to readers who cannot tell a FISH from a YAC. This newsletter was first published in spring 1989 by the DOE, but in May 1990 the NIH became copublisher and contributor, so that a more consolidated source of information on the genome project would be available. At this time *HGN* became a bimonthly instead of a quarterly publication.

HGN is a fine resource. Managing editor Betty K. Mansfield and her staff deserve much applause for collecting and coordinating a huge and extremely diverse body of information. Indeed, *this* Guide would have been much more difficult to compile without access to *HGN.* It is true that in the tradition of official government publications, the newsletter tends to be long on reports of plans and progress, but rather short on discussions of delays, debates, and controversies. Consequently, it is not always clear how well different parts of the project are working.

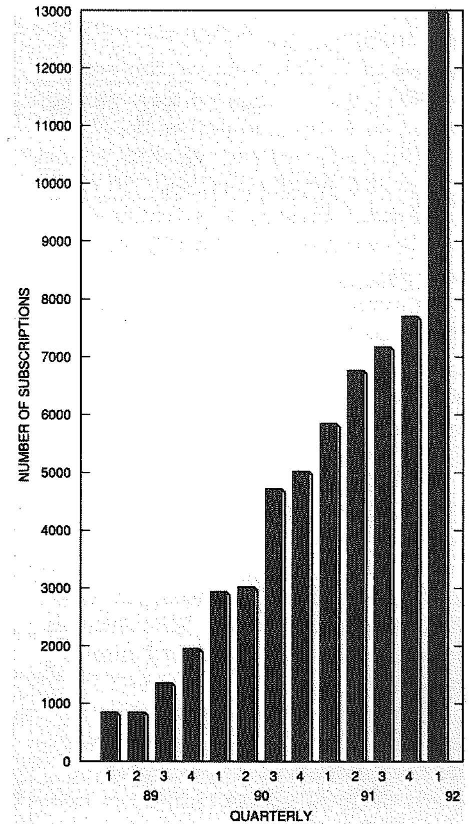
HGN frequently lists publications, including periodicals, often under the general rubric "Resources." Its May 1992 issue, for example, lists, describes, and gives contacts for seven newsletters: *Baylor Genome Center News, Bioinformatics, CELLS, Genome* (Michigan), *Linkage Newsletter* (Columbia), *NCBI News,* and *Probe* (USDA).

Journals and Series

A number of publications have been generated primarily by genome projects and the needs of genome scientists. The early issues of these periodicals are important parts of HGP history, because they reflect the changing interests and needs of the genomics community during the mid 1980s and early 1990s. The earliest specialized genome journal was *Genetic Analysis: Techniques and Applications,* which began publication as *Gene Analysis Techniques* in 1984. Published bimonthly by Elsevier Science Publications, it focuses on new methods, materials, and instruments for molecular biology, cell biology, biochemistry, and genetics, with an emphasis on recent developments in gene cloning and nucleic acid analysis. It changed its name in 1990 because the editors wished to broaden the journal's focus to include more applications. They added a review section and a section of "very brief techniques articles (in 'cookbook' format)."

The premier journal of genome research as it is embodied in the human genome initiative is *Genomics,* published monthly by Academic Press. In the first issue, September 1987, the editors explained that the term *genomics* had been adopted for the new discipline of gene mapping and sequencing—a field "born from a marriage of molecular and cell biology with classical genetics and . . . fostered by computational science." They continued:

> *Genomics* will not only report new data concerning genome maps and improved methods for mapping and sequencing . . . but also will publish analyses of the information, methods for those analyses, methods for storage, retrieval, searching, pattern recognition, comparisons, etc., as well as interpretation of structural findings in light of their biologic significance and biomedical applications. . . . *Genomics* will be a meeting ground for molecular biologists and biochemists, human and somatic cell geneticists, cytogeneticists, population and evolutionary biologists, genetic epidemiologists, clinical geneticists, theoretical biologists, and computational scientists, all interested in the biology and genetics of the human and other complex genomes.



Growth of the **Human Genome News** *Mailing List, April 1989 to April 1992. From* HGPR 1991–92, *p. 60.*

The journal publishes mainly basic research communications, but it also includes book reviews and commentaries, as well as a section titled "Genomics Update," which gives reference to new or confirmatory gene assignment information reported elsewhere. *Genomics* is able to coordinate a wide variety of research reports largely through its interaction with the major extant genetics databases, such as the Human Gene Map Library in New Haven, GenBank, and the European Molecular Biology Laboratory.

Earlier in 1987 an older journal, *Canadian Journal of Genetics and Cytology,*

changed its name to *Genome*. The editors explained that "when the Genetics Society of Canada founded the Journal in 1959 it was anticipated that cytology would be an integral part of the Journal's contents. Instead, genetics, cytogenetics, and evolution became the main substance." Pressed by scientific information services for a shorter, more relevant title, the editorial board "chose *Genome* to recognize that the Journal has traditionally reported extensively on the structure, function, and evolution of entire genomes, frequently in an agricultural, entomological, or vertebrate context. . . . The basic science of genomes in terms of molecular, genetic, and selection mechanisms is becoming a significant portion of the Journal, thereby justifying the new title." This journal has begun to include more articles on human genetics, but continues to focus mainly on plant and animal genomes.

Two journals began publication in 1991. *PCR Methods and Applications* is devoted exclusively to amplification methods and their use; it intends to keep PCR (polymerase chain reaction; see Technologies and Techniques, above) users up to date through research reports, review articles, letters, and news of other items of interest. It is published quarterly by the Cold Spring Harbor Laboratory Press. *Mammalian Genome*, the official journal of the International Mammalian Genome Society, is devoted to molecular studies of mammalian genomes, with special emphasis on genetic and physical maps, analysis of gene complexes and complex traits, and analysis of human genetic disorder and animal models. It is published quarterly by Springer-Verlag New York.[39]

A recent addition to the genome-related literature is *Nature Genetics*, a monthly journal designed to supplement *Nature*. *Nature Genetics* will focus on the link between human genome structure and disease, publishing a broad spectrum of papers on the latest findings in gene mapping, linkage analysis, candidate genes, positional cloning of important chromosomal regions, clinical genetics, aspects of developmental biology, imprinting, and fundamental advances in the Human Genome Project's research for humans and other organisms.[40]

Work on the HGP may seem to move too fast to make publishing books a high priority for its researchers. In 1990, however, Cold Spring Harbor Laboratory Press began publishing a series of short, single-theme surveys under the title *Genome Analysis*. The books consist of invited papers that review data, methods, and emerging ideas in the study of genetics both in humans and in other species. The first five volumes are titled *Genetic and Physical Mapping; Gene Expression and Its Control; Genes and Phenotypes; Strategies for Physical Mapping;* and *Regional Physical Mapping*.

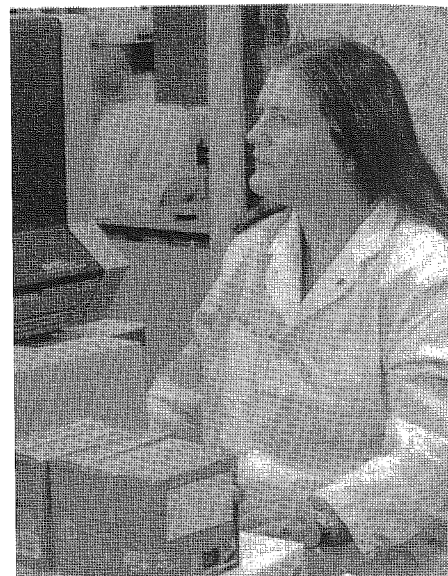
To facilitate the access of genome scientists and other researchers to the periodical literature, *Human Genome Abstracts* began bimonthly publication in February 1990. Each issue contains about 600 abstracts, mainly from published literature relating to the Human Genome Project, selected from over 5,000 sources. Cambridge Scientific Abstracts publishes the printed version; the online version is available in Dialog File 76, Life Sciences Collection.[41]

Other Forums

Electronic Publications

The vast amount of data generated by genome researchers, as well as the continual changes occurring in genome technology and techniques, makes rapid communication essential. This need has led to the creation of various databases and electronic bulletin boards. The online version of *Human Genome Abstracts* has been mentioned above. A more specialized database of bibliographic references on computing (C) and mathematical aspects of molecular biology (MB) and genetics (G) is the *CMBG Bibliography*, which contains nearly 5,000 citations and is expanding. It is available to the public and maintained as part of the GenTools Project at the University of Texas System Center for High Performance Computing, based in Austin.[42]

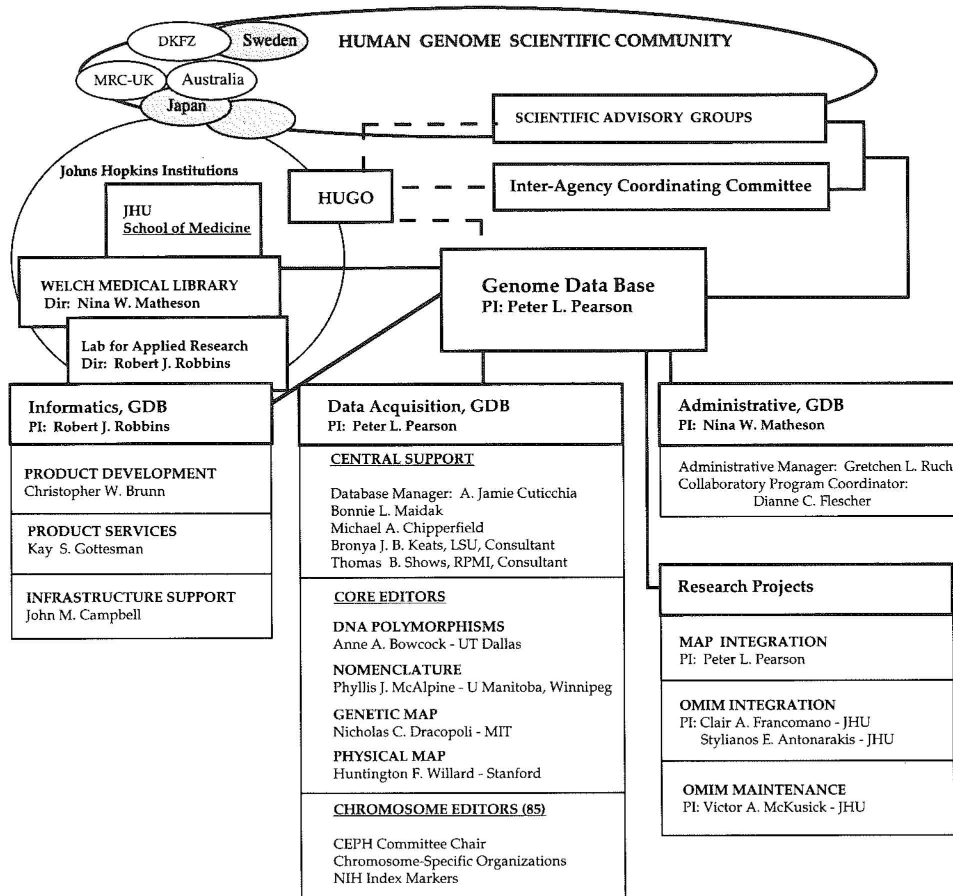
Databases that access more than the published literature are also important tools. The Genome Data Base (GDB) at Johns Hopkins University went on-line in September of 1990. It collects, organizes, stores, and distributes gene-mapping information provided by investigators, including data from the Human Gene Mapping Workshops (see Conferences, below). GDB also serves as a repository for genetic disease information applicable to patient care. It is free to scientists and the public and is accessible through SprintNet or the Internet. So far, GDB is working out well. Committee members at the 11th International Workshop on Human Gene Mapping (16–23 August 1991) used the system to enter and verify data, adding over 2,000 new entries to the database. This illustrates a new application for scientific databases: instead of building databases by reproducing published findings, investigators can use them as a means of publication, thus shortening the publication cycle.[43]



Computers and the HGP. *The Human Genome Project supports many databases. Here an investigator performs computer analysis of chromosome 19 map data at Lawrence Livermore National Laboratory. From* HGPR 1991–92, *p. 3.*

The DOE's GenBank, which is devoted to genome sequencing, has developed along similar lines. Begun in 1979 as the Los Alamos Sequence Library, it became GenBank in 1982. Its managers, spurred by the length of time it took to enter data published in journals, by journals' increased reluctance to publish nucleotide sequences, and by plans for the HGP, began work in the mid 1980s

39. "Publications," *Human Genome News* 3(3) (Sept. 1991), 19–20, on 20.
40. "Genome-Related Publications," *ibid.* 4(1) (May 1992), 10.
41. "Human Genome Abstracts," *ibid.* 2(2) (July 1990), 16.
42. "Genome-Related Publications" (see note 40).
43. "NIH–DOE Award Supports Human Genome Data Base at Hopkins," *Human Genome News* 3(4) (Nov. 1991), 5–6; see also "GDB Proves Itself at HGM 11," *ibid.*, 1–4, for a detailed description of how GDB works.



Organizational chart of the Genome Data Base (GDB). *From* HGPR 1991–92, *p. 17.*

on allowing researchers to publish sequence data directly in the database. As with GDB, this meant developing a large software system, here including a relational database, an interactive interface, and a program to check the quality and integrity of data submitted for automatic processing. GenBank now receives 95 percent of its data through its Electronic Data Publishing. In 1992 it processed over twenty times the data—30 million base pairs—it had in 1984, at a hundredth the cost (10¢ a base pair, as opposed to $10), with a turnaround time of two weeks instead of a year.[44]

The Directory of Biotechnology Information Resources (DBIR) is an on-line database at the National Library of Medicine. It includes information on other resources such as publications, committees, biological culture collections, and biotechnology centers. Most important, it covers other computerized databases, data networks, electronic bulletin boards, and other resources for collecting and distributing biotechnology information.[45]

Several electronic bulletin boards or *newsgroups* are available to the genome research community. The Human Genome Program Newsgroup began operating on the BioSci network in early 1990, to provide another way to distribute information about genome programs, grant applications, and meetings reports. It was also intended to serve as a forum for issues connected to genome research and is open to the public.[46] According to David Benton at the National Center for Human Genome Research, usage of the newsgroup has so far been "very light" but regular. Scientists tend to use other, more specialized bulletin boards on the BioSci network, such as "Methods and Reagents," to exchange information about techniques and methods.

Finally, the Human Genome Information Management System at Oak Ridge National Laboratory—the same group that puts out the *Human Genome News*—has developed the Human Genome Information Database as "a text management and user-conferencing mechanism." Due to go on line by the end of 1992, the HGID will provide users with text from meeting reports, newsletters, program and technical reports, and archived announcements from the BioSci Human Genome Program Newsgroup as well as bibliographic and abstract information related to the HGP.[47]

Besides these electronic newsgroups are the newsgroups for researchers working on particular chromosomes. Although conceived as a means of keeping colleagues up to date on the latest developments, these have proved somewhat disappointing. Researchers are reluctant to "publish" information immediately this way for various reasons, so entries tend to be about six months old before they are entered on the net.

Conferences

Meetings, conferences, and workshops on genome mapping and sequencing are another important—perhaps the main—means of communication for genome scientists. There are far too many to list here, but the International Human Gene Mapping Workshops deserve mention since they have long been successful in bringing gene mappers together to discuss their data and build improved maps.

These biannual workshops began meeting in 1973 and are now complemented by chromosome-specific workshops that meet more frequently and between the main meetings. Prior to each workshop, separate committees for each chromosome evaluate the research done on the chromosome, soliciting papers from international researchers to be presented. At the conference each committee works toward a consensus on which mapping data will be accepted as standard, as well as determining official nomenclature for map sites and probes. Data accepted at the conference are submitted to the Human Gene Mapping Library in New Haven and entered into that database. The conference proceedings are published in *Cytogenetics and Cell Genetics*.[48]

---

44. "Electronic Data Publishing in GenBank," *The Human Genome Project*, no. 20 of *Los Alamos Science* (1992), 270–273.

45. "Databases Organize Information Resources," *Human Genome News* 2(2) (July 1990), 17.

46. "Newsgroup and Information Database Services Offered," *ibid.* 2(1) (May 1990), 10.

47. *Ibid.*

48. OTA, *Mapping Our Genes* (see note 6), 157.

ETHICAL, LEGAL, AND SOCIAL IMPLICATIONS

In addition to its technical and administrative complexities, the Human Genome Project poses ethical, legal and social questions. Will the project accelerate the rise of a "new eugenics," technologically rather than socially driven? Will a complete map of the human genome change the interpretation of human rights? Will it shape reproductive choices in ethically problematic ways? How should information about genetic conditions which can be diagnosed, but not treated, be handled? And who should have access to genetic information about individuals?

A fairly convincing argument can be made that molecular genetics is already playing a role in a "new eugenics" grounded in concepts of health and disease and mediated by genetic counselors and physicians.[49] Many institutions could conceivably use new genetic information to control individuals. Schools, health insurers, and the criminal justice system have a stake in predicting and controlling behavior, and a knowledge of the genetic characteristics of individuals could be used to exclude those "at risk" from insurance coverage or to place students in different academic "tracks."[50] Some scientists involved in the genome project have suggested that the advantage of a complete genetic map is that it will permit the analysis of the inheritance of multifactorial traits such as intelligence, personality, and mental illness.[51] The Human Genome Project therefore promises to accelerate the date at which society must deal with the ethical implications of a model of human behavior that is grounded in biological determinism. As Troy Duster points out, "the gun is on the wall," and, to borrow from Anton Chekhov, if the gun is on the wall in the first act it must be used by the third.[52] How it will be used remains open, but that it will be used seems indisputable.

Genetic information has tremendous social power. Current scientific advances in molecular genetics have converged with work in behavioral genetics to encourage a widespread popular belief that all the basic characteristics of human beings are determined by their genes—that the individual is a molecular recipe present at the moment of conception. Rich in symbol and metaphor, genetic concepts are easily assimilated and interpreted in ways that reflect prevailing social anxieties—concerns about risks, differences, danger, and changing social relationships. As the results of contemporary research in genetics are reported in the media, they are used to justify popular stereotypes of gender (men are superior in mathematics, women in "nurturing behavior") and of race (black athletes have "gifted bodies"). They are also used to explain troublesome problems of crime and addiction (some people are predisposed to addiction, aggression, or destructive social behavior). Genetics is often used to explain behavior, physical illness, individual problems, and personal skills and to define what is normal, different, or deviant. The research of the Human Genome Project feeds into this cultural tendency in ways that may have implications for individual rights.[53]

In an effort to address such concerns, both the NIH and DOE have set aside 5 percent of their total genome project funding for research on the ethical, legal, and social implications of the project. At NIH alone the funding totaled $5 million in 1993. This program, ELSI, has brought together ethicists, scientists, sociologists, and historians to explore some of the more immediate problems. Georgetown University has begun an archival repository for materials relating to these issues, and Michael Yesley at Los Alamos National Laboratories has developed an on-line database on them. ELSI has also established a Joint Working Group on the Ethical, Legal, and Social Issues in Human Genome Research. The group has identified job-related and insurance-related discrimination as top priorities, concluding that Equal Employment Opportunity Commission rules should be expanded to limit genetic testing of job applicants and employees, that employees should have full access to medical records compiled by employers, and that safeguards should be instituted to guarantee the privacy of genetic information contained in medical records. Thus genetic information—data from an individual's genome—becomes a form of private and personal information that must be legally protected from those institutions that could benefit economically from use of the information.

The more fundamental question raised by the project involves whether there can or should be any knowledge identified as "forbidden." Some in the religious community have identified the genome as the location of personhood or the "genetic soul" and suggested that contemporary molecular genetics threatens our understanding of human rights. Others have interpreted the genetic prediction the project makes possible in positive terms as facilitating human rights and helping each individual to adopt habits that will prolong and enhance life. Certainly there are historical precedents for both destructive and positive social effects of increasing access to scientific information.

49. Troy Duster, *Backdoor to Eugenics* (New York: Routledge, 1990).

50. Dorothy Nelkin and Lawrence Tancredi, *Dangerous Diagnostics: The Social Power of Biological Information* (New York: Basic Books, 1989).

51. See, e.g., David Baltimore, in discussion in *Human Genome Project* (see note 44), 321.

52. Duster, *Backdoor to Eugenics* (see note 48).

53. Dorothy Nelkin and Susan Lindee (work in progress).

APPENDIXES

A. Key Personnel

Many individuals involved in the human genome project are scientists who function also as policymakers. The obvious example is James D. Watson, who as the first Director of the NIH National Center for Human Genome Research (NCHGR) played an active role in both scientific developments and administrative policy. Most of those who direct genome centers also serve on committees for NIH or DOE, reviewing applications, for example, or assessing technologies. While the cast of characters is massive—particularly when the international effort is considered—the community is intellectually close and, particularly in the United States, dominated by individuals who are both scientifically eminent and politically powerful.

This appendix contains brief profiles of a few key figures from the NIH program; a few from the DOE program, notably several members of the Human Genome Coordinating (originally Steering) Committee (HGCC); the directors of all designated genome centers; and a small group of individuals who played an important role in the historical development of the project. The list here does not by any means include all the important players, but it is an introduction to some of those with high profiles in the American effort.[54]

Paul Berg (B.S. Pennsylvania State University; Ph.D. Western Reserve University) is Director of the Beckman Center for Molecular and Genetic Medicine at Stanford University, where he has taught biochemistry since 1970. He shared a 1980 Nobel Prize in chemistry with Walter Gilbert and Frederick Sanger for studies of the biochemistry of recombinant DNA. He is the author, with Maxine Singer, of *Genes and Genomes* and has served on several panels connected with the Genome Project; currently he chairs the NCHGR Program Advisory Committee on the Human Genome and cochairs the DOE–NIH Subcommittee on the Human Genome.

David Botstein (A.B. Harvard University; Ph.D. in human genetics, University of Michigan) is Chairman of the Department of Genetics at the Stanford University School of Medicine. He led the research group that in 1980 first proposed mapping human genes with the markers known as restriction fragment length polymorphisms (RFLPs).

Charles R. Cantor (A.B. Columbia University; Ph.D. in chemistry, University of California at Berkeley), formerly Director of the Human Genome Center at Lawrence Berkeley Laboratory, is now Principal Scientist for the DOE HGP. He is the Higgins Professor of Genetics and Development at the Columbia University College of Physicians and Surgeons and chair of that department. Cantor is an internationally recognized expert in the fields of genetic mapping and sequencing; his group developed the pulsed-field gel electrophoresis technique to separate DNA fragments. He has been a member of DOE's HGCC since October 1988.

Anthony V. Carrano (B.S. in chemistry, Rensselaer Polytechnic Institute; Ph.D. in biophysics, University of California at Berkeley) is Genetics Section leader of the Biomedical Sciences Division and Director of the Human Genome Center at Lawrence Livermore National Laboratory, as well as a member of the DOE's HGCC. Carrano's research has centered on cytogenetics, molecular cytogenetics, mechanisms of mutagenic damage and repair, and genetic consequences of mutagen exposure. He has published methods for fluorescence-based, high-resolution, semiautomated methods for DNA fingerprinting. He is also closely involved in the National Gene Library Project.

C. Thomas Caskey (B.S. University of South Carolina; M.D. Duke University School of Medicine) is Director of the Human Genome Center at Baylor College of Medicine and a member of the DOE's HGCC. At Baylor he holds the Henry and Emma Meyer Chair in Molecular Genetics; at the Howard Hughes Medical Institute in Utah he is chief of the Medical Genetics Section and an investigator. Caskey's research interests include inherited disease and mammalian genetics. He has served on review panels for the General Medical Sciences Council of the NIH Department of Health and Human Services and for the Office of Technology Assessment.

Francis S. Collins (B.S. University of Virginia; Ph.D. Yale University; M.D. University of North Carolina), until recently Director of the Genome Center at the University of Michigan, Chief of Medical Genetics at the University of Michigan Medical Center, and an investigator at the Howard Hughes Medical Institute in Utah, is the new director of NCHGR, succeeding James Watson. His group at Michigan succeeded in identifying the genes that cause cystic fibrosis and neurofibromatosis and is searching for the Huntington's disease gene on chromosome 4.[55]

David Cox (M.D., Ph.D. in genetics, University of Washington) is Professor of Psychiatry, Biochemistry, and Genetics at the University of California at San Francisco, where he is codirector of the Human Genome Mapping Center.

Charles DeLisi (B.A. City College of New York; Ph.D. in physics, New York University) has been a professor in and chair of the Department of Biomathematical Science at Mt. Sinai Medical School in New York since 1987. He conceived the idea of a federally funded genome sequencing project in 1985, while he was Director of the Office of Health and Environmental Research at DOE. He was Senior Investigator in Biophysics at the National Cancer Institute between 1977 and 1985.

Helen Donis-Keller (B.S. Lakehead University, Ontario; Ph.D. Harvard University) is Professor of Genetics in Surgery and Director of the Division of Human Molecular Genetics in the Department of Surgery at Washington University School of Medicine in St. Louis. When at Harvard she worked for Walter Gilbert. In the 1980s she worked at Collaborative Research, leading the team that produced the first complete genetic linkage map in 1987. Her current research interests include high-resolution linkage mapping of the human genome and construction of a clone-based map of human chromosome 7.

Beverly Emanuel (Ph.D., postdoctoral work in human genetics, University of Pennsylvania) is Director of the NIH Genome Center for Chromosome 22 at the Children's Hospital of Philadelphia and the University of Pennsylvania. She is a member of the Genome Research Review Committee of NCHGR, on the board of directors of the American Society of Human Genetics, and a member of HUGO. She has published on the molecular cytogenetics of cancer, DiGeorge syndrome, Miller-Dieker syndrome and the mapping of chromosome 22. She is currently professor of pediatrics and human genetics at and director of the Section of Genome Analysis at CHOP.

54. Information for these biographies comes from *American Men and Women of Science*, 18th ed., 1992–1993; from solicited curricula vitae; items in *Human Genome News* and its predecessor *Human Genome Quarterly;* and from the various chapters of Robert Cook-Deegan, "Gene Quest: Science, Politics and the Human Genome Project" (prepublication draft, Oct. 1991). Undergraduate degrees and graduate fields were supplied when possible.

55. See Thompson, "Healy and Collins" (see note 27).

Glen Evans (Ph.D. in chemistry, M.D. University of California at San Diego) heads the DOE–NIH Genome Center at the Salk Institute for Biological Studies. He is also an associate adjunct professor of biology, neuroscience, and pharmacology at UC San Diego, and on the Science Advisory Committee for the Life Sciences Division at LANL. He is a member of the Genome Research Review Committee of NIH's NCHGR.

David J. Galas (A.B. University of California at Berkeley; Ph.D. in physics, University of California at Davis) is Associate Director of the DOE Office of Health and Environmental Research, which administers the DOE Human Genome program. Galas was previously Director of the Department of Molecular Biology at the University of Southern California.

Raymond Gesteland (B.S. University of Wisconsin; Ph.D. in biochemistry, Harvard University) heads the NIH Genome Center at the University of Utah with Ray White; he is a member of the DOE HGCC. He has been an investigator and Professor of Human Genetics at the Howard Hughes Medical Institute at Utah since 1978. His research has focused on the regulation of gene expression and translation.

Walter Gilbert (A.B. Harvard University; Ph.D. in mathematics, Cambridge University) has been part of several major developments in molecular biology, including the discovery of mRNA and the isolation of the *lac* repressor and its mechanism. In 1980 he shared a Nobel prize with Allan Maxam for inventing the Maxam-Gilbert sequencing technique (later automated by Leroy Hood's group at Caltech). He has held several faculty positions at Harvard, in between forays into the biotechnology industry. He helped found the Swiss-American firm Biogen, later serving as chief executive officer. Later he formed the short-lived Genome Corporation. Gilbert has been one of the strongest proponents of the effort to sequence the entire genome; he was responsible for spreading the word to colleagues and the public during the late 1980s.

Leroy Hood (M.D. Johns Hopkins University; B.S., Ph.D. in biochemistry, California Institute of Technology) is Director of the NSF Science and Technology Center for Molecular Biotechnology at the University of Washington School of Medicine in Seattle. Before that he was Director of the NSF Science and Technology Center for Integrated Protein and Nucleic Acid Chemistry and Biological Computation at Caltech. Hood's laboratory at Caltech played a major role in developing automated instruments that analyze the sequences of proteins and DNA and synthesize peptides and gene fragments. Hood serves on the DOE HGCC.

Elke Jordan (B.A. Goucher College; Ph.D. in biochemistry, Johns Hopkins University; postdoctoral work in molecular biology, Harvard University) is Deputy Director of the NCHGR. She worked at NIH's National Institute of General Medical Sciences from 1976 to 1988 as a program administrator, deputy director of the genetics program, and Associate Director of Program Activities.

Eric T. Juengst (B.S. in biology, University of the South; Ph.D. in philosophy, esp. bioethics, Georgetown University) has been Manager of the NCHGR Program on Ethical, Legal, and Social Issues since May 1990. He has held academic positions at the Division of Medical Ethics at the Medical School of the University of California at San Francisco, where he researched ethical issues in prospective "gene therapy" of germ line cells, and at Pennsylvania State University College of Medicine. He has also served on review panels of the Office of Technology Assessment and the NIH–NSF Program on Ethics in Science and Technology.

Eric Lander (B.A. Princeton University; Ph.D. in mathematics, Oxford University) has been Director of the Center for Genome Research at MIT since 1990. He is Associate Professor of Biology at MIT and a member of the Whitehead Institute for Biomedical Research, where he directs projects to construct physical maps of the mouse and human genomes. He received a MacArthur Fellowship in 1987.

Victor A. McKusick (M.D. Johns Hopkins University, 1946) has been involved in the study of genetics for more than forty years. At Hopkins he was Professor of Medicine from 1952 to 1985; he has been Professor of Medical Genetics since 1985. McKusick began keeping track of genetic disorders and variants in 1960 and published the first catalogue of human genes (a catalogue of data on the X chromosome) in 1962. In 1966 he published an expanded list of disorders and genes as *Mendelian Inheritance in Man,* which became a crucial reference work in the field. The work went through seven printed editions and became available on line in 1987. McKusick was the founding president of HUGO and now chairs its ethics committee. He also served on the National Academy of Sciences Committee on Mapping and Sequencing the Human Genome from 1986 to 1988.

Robert K. Moyzis (B.S. in biology and chemistry, Northwestern Illinois University; Ph.D. Johns Hopkins University, 1978) is Director of the Center for Human Genome Studies at Los Alamos National Laboratory (LANL) and a member of the DOE's HGCC. As head of the LANL Genetics Group for five years, Moyzis led the mapping effort on chromosome 16 and received a Distinguished Performance Award from LANL for identifying and isolating the highly conserved functional telomere, the region of DNA located at the ends of each human chromosome. This discovery enables biologists to determine how and where chromosomes end and will provide physical orientation in constructing maps of the human genome.

Jeffrey Murray (M.D. Tufts Medical School) heads the NIH Genome Center at the University of Iowa, where he is Associate Professor of Pediatrics. His research has focused on mapping human chromosome 4.

Richard Myers (B.S. University of Alabama; Ph.D. in biochemistry, University of California at Berkeley) is Director of the Human Genome Mapping Center at the University of California at San Francisco, where he is Associate Professor of Physiology and of Biochemistry and Biophysics.

Susan L. Naylor (Ph.D. in human genetics, University of Medical Branch, Galveston) is Director of the NIH Genome Center at the University of Texas Health Science Center at San Antonio and a full professor in the Department of Cellular and Structural Biology.

Maynard Olson (B.S. California Institute of Technology; Ph.D. in inorganic chemistry, Stanford University) is Professor of Molecular Biotechnology and Medicine at the University of Washington in Seattle. Until September 1992 he headed the NIH Genome Center at Washington University in St. Louis with David Schlessinger. Olson's lab in St. Louis created the first yeast artificial chromosomes, or YACs (see Technologies and Techniques, above), for physical mapping. He was involved in the early planning of the HGP as a member of the National Research Council Committee on Mapping and Sequencing the Human Genome. Olson currently serves on the NIH Program Advisory Committee on the Human Genome.

Jane L. Peterson (B.A. Western College; Ph.D. University of Colorado; postdoctoral work, Yale University) is Chief of Research Centers Branch of NCHGR, where she oversees funding and operation of multidisciplinary research centers. She was Assistant Program Director of the Developmental Biology Program at NSF from 1981 to 1985, and then served as Program Administrator of the Genetics Program at the National Institute of General Medical Sciences from 1985 to 1989.

Jasper D. Rine (Ph.D. University of Oregon; postdoctoral work, Stanford University School of Medicine) heads the Human Genome Center at Lawrence Berkeley Laboratory. He is also Professor of Genetics at the University of California at Berkeley, and serves on the DOE's HGCC.

Gerald M. Rubin (Ph.D. in biology, Cambridge University) heads the NIH Genome Center at the University of California at Berkeley, which focuses on *Drosophila* genetics. He is also Professor of Genetics in both the Department of Molecular and Cell Biology and the Howard Hughes Medical Institute at Berkeley.

David Schlessinger (B.A., B.S. University of Chicago; Ph.D. in biochemistry, Harvard University) is Director of the Center for Genetics in Medicine at Washington University in St. Louis. He has been Professor of Microbiology at Washington University School of Medicine since 1972. His research interests are in human genome mapping, using yeast artificial chromosomes (YACs).

James D. Watson (B.S. University of Chicago; Ph.D. in genetics, Indiana University) was, until 10 April 1992, Executive Director of the NCHGR. He is also one of the most famous living scientists. With Francis Crick he discovered the double-helix structure of DNA in 1953 and received a Nobel Prize for the work in 1962. He taught at Harvard from 1956 to 1976 and has directed the Cold Spring Harbor Laboratory since 1968. His autobiography, *The Double Helix,* inspired a generation of young people to pursue careers in molecular biology and his classic molecular genetics text, *The Molecular Biology of the Gene,* continues to be widely used and respected. If any one individual has been crucial to the Human Genome Project and its success so far, it has been James Watson.

Nancy S. Wexler (A.B. Radcliffe College; Ph.D. in psychology, University of Michigan) is chair of the Joint NIH–DOE ELSI Working Group and serves on the NIH Program Advisory Committee on the Human Genome. She is a clinical psychologist in the Department of Neurology and Psychiatry at the College of Physicians and Surgeons of Columbia University and president of the Hereditary Disease Foundation. In the 1970s she was a member of the Huntington's Disease Commission and conducted research at the National Institute of Neurological and Communicative Disorders and Stroke.

Ray White (Ph.D. in microbiology, Massachusetts Institute of Technology) is codirector, with Raymond Gesteland, of the NIH Genome Center at the University of Utah. He conducted studies of the DNA of large, multigeneration Mormon families to find RFLPs and construct genetic linkage maps. This work led to the mapping of the gene causing cystic fibrosis.[56]

Norton Zinder (B.A. Columbia University; Ph.D. in medical microbiology, University of Wisconsin) was Chair of the NCHGR Program Advisory Committee on the Human Genome until 1991. He has taught at Rockefeller University since 1958 and has been John D. Rockefeller, Jr., Professor of Microbial Genetics since 1976. His research has focused on virology, protein biosynthesis, and genetics.

56. Wills, *Exons, Introns, and Talking Genes* (see note 29), 201–205.

B. Chromosomes and Their Associated Diseases

Over 1,200 diseases and disabilities have been definitely or tentatively mapped to chromosomes. This list includes diseases (and a few significant biological functions) whose location has been verified by at least three different mapping teams and reported in the proceedings of the Eleventh International Human Gene Mapping Workshop, published in *Cytogenics and Cell Genetics* 58(1–4), 1991 (see under Communications, above), as *Human Gene Mapping 11.*

1 Progressive encephalopathy; neuronal lipofuscinosis; nemaline myopathy; Gaucher's disease (glucocerebrosidase deficiencies)

2 Cleft lip and palate

3 Von Hippel–Linden disease; retinitis pigmentosa; site for gene deletion in small-cell lung cancer and renal-cell carcinoma

4 Huntington's disease; Hurler's or Scheie's syndrome

5 Polyposis; Treacher Collins–Franceschetti syndrome; congenital contractural arachnodactyly; adenomatous polyposis; spinal muscular atrophy

6 One form of spinocerebellar ataxia; drug receptors for cannabinoids and diazepam

7 Split hand and foot deformities; holoprosencephaly; Grieg cephalopolysyndactyly syndrome; cystic fibrosis transmembrane conductance regulator

8 Hereditary spherocytosis; lipoprotein lipase deficiency; Langer-Giedion syndrome; autosomal-dominant retinitis pigmentosa

9 Xeroderma pigmentosum; Finnish hereditary amylocytosis; acute hepatic porphyria; idiopathic torsion dystonia; Friedreich's ataxia; tuberous sclerosis; some carcinomas

10 Medullary thyroid carcinoma; congenital erythropoietic porphyria

11 Wilms' tumor; ataxia telangiectasia, complementary groups C and D; insulin-dependent diabetes; long Q-T syndrome (ventricular arrhythmia); bipolar affective disorder; hypoparathyroidism; Niemann-Pick disease; familial combined hyperlipidism.

12 Stickler syndrome (progressive arthro-opthalmopathy); epidermolysis bullosa; congenital spondyloepiphyseal dysplasia; tuberous sclerosis 3; vitamin D receptor

13 Retinoblastoma; breast, lung, and bone tumors; Wilson's disease; Hirschsprung disease; Moebius syndrome

14 Hers disease; Krabbe's disease; Usher syndrome

15 Marfan's syndrome; Prader-Willi syndrome; Angelman's disease; limb-girdle muscular dystrophy, recessive; xeroderma pigmentosum, complementary group F

16 Hemoglobin synthesis; polycystic kidney disease; Spiegelmeyer-Vogt disease

17 Charcot-Marie-Tooth neuropathy; early-onset familial breast cancer; Miller-Dicker syndrome; Li-Fraumeni syndrome; neurofibromatosis 1; colon cancers

18 Colon cancers; erythropoietic protoporphyria

19 Malignant hyperthermia; Alzheimer's disease, 2 (late onset); central core disease of muscle; maple-syrup urine disease; dystrophia myotonia

20 Neurohypophyseal diabetes insipidus; Gerstmann-Straussler-Schenker syndrome; Creutzfeldt-Jakob disease; Alagille syndrome; non-insulin-dependent diabetes mellitus; adenosine deaminase (ADA) deficient form of severe combined immune deficiency (SCID); hemolytic anemia from ADA excess; Albright hereditary osteodystrophy; galactosialidosis; Fanconi's anemia; epilepsy (benign neonatal)

21 Alzheimer's disease, 1 (early onset); amyotrophic lateral sclerosis; Down syndrome; progressive myoclonus epilepsy

22 DiGeorge syndrome; neurofibromatosis 2

X Fragile X syndrome; night blindness; color blindness; Duchenne and Becker types of muscular dystrophy; Menke's syndrome; Kallmann syndrome; Wiskett-Aldrich syndrome; severe combined immunodeficiency; Hunter's syndrome; retinoschisis; Aland island eye disease; many more conditions

Y Many male-specific characteristics; no disease associations

SELECTED BIBLIOGRAPHY

Histories of Genetics and Guides to the Human Genome Project

Angier, Natalie. *Natural Obsessions: The Search for the Oncogene.* Boston: Houghton Mifflin, 1988.

Bishop, Jerry E.; Waldholz, Michael. *Genome: The Story of the Most Astonishing Scientific Adventure of Our Time—The Attempt to Map All the Genes in the Human Body.* New York: Simon & Schuster, 1990.

Carlson, Elof Axel. *The Gene: A Critical History.* Philadelphia: W. B. Saunders, 1966.

Cook-Deegan, Robert Mullan. "Gene Quest: Science, Politics, and the Human Genome Project." 1991. Draft on deposit at Human Genome Archive, Georgetown University, and elsewhere.

Cook-Deegan, Robert Mullan; Watson, James D. "Origins of the Human Genome Project." *FASEB Journal* 5 (1991), 8–11.

Davis, Bernard, editor. *The Genetic Revolution: Scientific Prospects and Public Perceptions.* Baltimore: Johns Hopkins University Press, 1991.

Hall, Stephen S. *Invisible Frontiers: The Race to Synthesize a Human Gene.* New York: Atlantic Monthly Press, 1987.

The Human Genome Project. Number 20 of *Los Alamos Science* (1992).

Judson, Horace Freeland. *The Eighth Day of Creation: Makers of the Revolution in Biology.* New York: Simon & Schuster, 1979.

Kevles, Daniel J.; Hood, Leroy, editors. *The Code of Codes: Scientific and Social Issues in the Human Genome Project.* Cambridge, Mass.: Harvard University Press, 1992.

Krimsky, Sheldon. *Genetic Alchemy: The Social History of the Recombinant DNA Controversy.* Cambridge, Mass.: MIT Press, 1982.

Lear, John. *Recombinant DNA: The Untold Story.* New York: Crown, 1978.

McCarty, Maclyn. *The Transforming Principle: Discovering That Genes Are Made of DNA.* New York: Norton, 1985.

Office of Technology Assessment, U.S. Congress. *Mapping Our Genes: Genome Projects—How Big, How Fast?* Washington, D.C.: U.S. Government Printing Office, 1988.

Olby, Robert. *The Path to the Double Helix.* Seattle: University of Washington Press, 1974.

Portugal, Franklin H.; Cohen, Jack S. *A Century of DNA.* Cambridge, Mass.: MIT Press, 1977.

Watson, James D. *The Double Helix: A Personal Account of the Discovery of the Structure of DNA.* Norton Critical Edition, edited by Gunther S. Stent. New York: Norton, 1980.

Wills, Christopher. *Exons, Introns, and Talking Genes.* New York: Basic Books, 1991.

Wright, Susan. "Recombinant DNA Technology and its Social Transformation, 1972–1982." *Osiris* n.s., 2 (1986), 303–360.

Techniques and Technologies

Botstein D.; White R. L.; Skolnick, M.; Davis, R. W. "Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms." *American Journal of Human Genetics* 32 (1980), 314–331.

Burke, D. T.; Carle, G. F.; Olson, M. V. "Cloning of Large Segments of Exogenous DNA Into Yeast Using Artificial Chromosome Vectors." *Science* 236 (1987), 806–812.

Carle, G. F.; Frank, F.; Olson, M. V. "Electrophoretic Separations of Large DNA Molecules by Periodic Inversion of the Electric Field." *Science* 232 (1986), 665–668.

Church, G. M.; Gilbert, W. "Genomic Sequencing." *Proceedings of the National Academy of Sciences USA* 81 (1984), 1991–1995.

Connell, C.; Fung, C.; Heiner, J.; et al. "Automated DNA Sequence Analysis." *BioTechniques* 5 (1987), 342–348.

Crick, F .H. C.; Watson, J. D. "The Complementary Structure of Deoxyribonucleic Acid." *Proceedings of the Royal Society (A)* 223 (1954), 80–96.

Donis-Keller, H.; Green, P.; Helms, C.; et al. "A Genetic Linkage Map of the Human Genome." *Cell* 51 (1987), 319–337.

Evans, G. A.; Wahl, G. M. "Cosmid Vectors for Genomic Walking and Restriction Mapping." In *A Guide to Molecular Cloning Techniques,* edited by S. L. Berger and A. R. Kimmel. Methods in Enzymology, 152. New York: Academic Press, 1988.

Gerhard, D. S.; Kawasaki, E. S.; Bancroft, F. C.; et al. "Localization of a Unique Gene by Direct Hybridization." *Proceedings of the National Academy of Sciences USA* 78 (1981), 3755–3759.

Gusella, J. F.; Wexler, N. S.; Conneally, P. M.; et al. "A Polymorphic DNA Marker Genetically Linked to Huntington's Disease." *Nature* 306 (1983), 234–238.

Jones, S. S.; Brown, J. E.; Vanstone, D. A.; et al. "Automating the Purification and Isolation of Synthetic DNA." *BioTechnology* 5 (1987), 67–70.

Knobeloch, D. W.; Hildebrand, C. E.; Moyzis, R. K.; et al. "Robotics in the Human Genome Project." *BioTechnology* 5 (1987), 1284–1287.

Lander, E. S.; Botstein, D. "Mapping Complex Genetic Traits in Humans: New Strategies Using a Complete RFLP Linkage Map." *Cold Spring Harbor Symposia on Quantitative Biology* 51 (1986), 49–62.

Maxam, A. M.; Gilbert W. "A New Method for Sequencing DNA." *Proceedings of the National Academy of Sciences USA* 74 (1977), 560–564.

———. "Sequencing End-Labeled DNA with Base-Specific Chemical Cleavage." In *Nucleic Acids,* edited by L. Grossman and K. Moldave. Methods in Enzymology, 65. New York: Academic Press, 1980. Pages 499–560.

McKusick, V. A.; Ruddle, F. H. "Toward a Complete Map of the Human Genome." *Genomics* 1 (1987), 103–106.

Sanger, F.; Nicklen, S.; Coulson, A. R. "DNA Sequencing with Chain-Terminating Inhibitors." *Proceedings of the National Academy of Sciences USA* 74 (1980), 5463–5468.

Schwartz, D. C.; Cantor, C. R. "Separation of Yeast Chromosome-Sized DNAs by Pulsed-Field Gel Electrophoresis." *Cell* 37 (1984), 67–75.

Southern, E. M. "Detection of Specific Sequences among DNA Fragments Separated by Gel Electrophoresis." *Journal of Molecular Biology* 98 (1975), 503–517.

Watson, J. D.; Crick F. H. C. "Genetic Implications of the Structure of Deoxyribonucleic Acid." *Nature* 171 (1953), 964–967.

———. "Molecular Structure of Nucleic Acids: A Structure of Deoxyribose Nucleic Acid." *Nature* 171 (1953), 737–738.

Watson, J. D.; Hopkins, N. H.; Roberts, J. W. *The Molecular Biology of the Gene.* Menlo Park, Calif.: Benjamin/Cummings Publishing Co., 1987.

White, R.; Lalouel, J.–M. "Chromosome Mapping With DNA Markers." *Scientific American* 258 (1988), 40–48.

Corporate Involvement

Beam, Alex. "A Grand Plan to Map the Gene Code: Nobelist Walter Gilbert Says His Genome Corp. Can Do It." *Business Week,* 27 April 1987, 116–117.

Gannes, Stuart. "Striking it Rich in Biotech." *Fortune* 116 (9 Nov. 1987), 131–132.

Teitelman, Robert. *Gene Dreams: Wall Street, Academia, and the Rise of Biotechnology.* New York: Basic Books, 1989.

Wyke, Alexandra. "The Genetic Alternative—A Survey of Biotechnology." *The Economist* 307 (30 April 1988), 1–18.

Yoxen, Edward. *The Gene Business: Who Should Control Biotechnology?* New York: Harper & Row, 1984.

"Biotechnology: Inherited Wealth." *The Economist,* 3 April 1988, pp. S3–S19.

"Gene Rush." *Scientific American,* January 1991, pp. 112–113.

Pros and Cons

Alberts, Bruce M. "Limits to Growth: In Biology, Small Science is Good Science." *Cell* 41 (1985), 337–338.

Davis, Bernard. "Some Problems With a Crash Program." *FASEB Journal* 5 (1991), 76.

Davis, Bernard, et al. "The Human Genome and Other Initiatives." *Science* 249 (1990), 342–343.

Fox, Jeffrey. "Critics Converge on the Genome Project." *BioTechnology* 6 (1988), 643.

Korn, D. "Patent and Trade Secret Protection in University-Industry Research Relationships in Biotechnology." *Harvard Journal on Legislation* 24 (1987), 191–238.

Lewin, Roger. "Proposal to Sequence the Human Genome Stirs Debate." *Science* 232 (1986), 1598–1600.

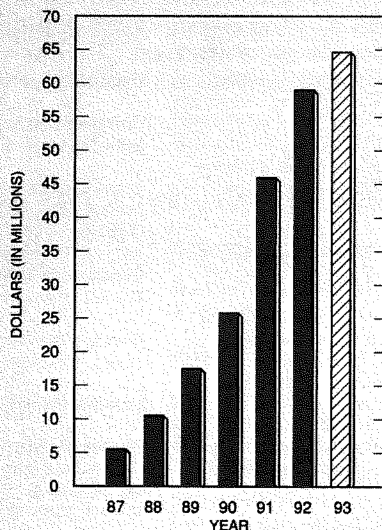
Many of the works in the first section contain discussions relevant to the other three topics covered in this bibliography. The discussions of techniques and technologies in particular are more accessible to the general reader than the primary sources listed here. See especially Kevles and Hood, eds., *Code of Codes;* and OTA, *Mapping Our Genes.*

Lewin. "Shifting Sentiments Over Sequencing the Human Genome." *Science* 233 (1986), 620–621.
———. "Politics of the Genome," *Science* 235 (1987), 1453.
McKusick, Victor A. "Current Trends in Mapping Human Genes." *FASEB Journal* 5 (1991), 12–20.
Rechsteiner, Martin. "The Folly of the Human Genome Project." *New Scientist* 127 (15 Sept. 1990), 20.
Roberts, Leslie. "Genome Backlash Going Full Force." *Science* 248 (1990), 804.
———. "Genome Patent Fight Erupts." *Science* 254 (1991), 184–186.
———. "A Meeting of the Minds on the Genome Project?" *Science* 250 (1990), 756–757.
———. "Plan for Genome Centers Sparks Controversy." *Science* 246 (1989), 204–205.
———. "Tough Times Ahead for the Genome Project." *Science* 248 (1990), 1600–1601.

Ethical, Legal, and Social Issues

Annas, George. "Who's Afraid of the Human Genome?" *Hastings Center Report* 19(4) (1989), 19–21.
Carmen, Ira H. *Cloning and the Constitution: An Inquiry into Governmental Policymaking and Genetic Experimentation.* Madison, Wis.: University of Wisconsin Press, 1985.
Duster, Tony. *Backdoor to Eugenics.* New York: Routledge, 1990.
Lappé, Marc. "The Predictive Power of the New Genetics." *Hastings Center Report* 14(2) (1984), 18–21.
Macklin, R. "Mapping the Human Genome: Problems of Privacy and Free Choice." Pp. 107–114 in *Genetics and the Law III,* edited by A. Milunsky and G. J. Annas. New York: Plenum Press, 1985.
National Center for Human Genome Research, National Institutes of Health. *Report of the Working Group on Ethical, Legal and Social Issues Related to Mapping and Sequencing the Human Genome.* Bethesda, Md., 1989.
Nelkin, Dorothy; Tancredi, Laurence. *Dangerous Diagnostics: The Social Power of Biological Information.* New York: Basic Books, 1989.
Nolan, Kathleen; Swenson, Sarah. "New Tools, New Dilemmas: Genetic Frontiers." *Hastings Center Report* 18(5) (1988), 40–46.
Office of Technology Assessment, U.S. Congress. *New Developments in Biotechnology: Ownership of Human Tissues and Cells.* Washington, D.C.: U.S. Government Printing Office, 1987.
President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research. *Screening and Counseling for Genetic Conditions: A Report on the Ethical, Social, and Legal Implications for Genetics Screening, Counseling and Education Programs.* Washington, D.C.: U.S. Government Printing Office, 1983.
Roberts, Leslie. "Who Owns the Human Genome?" *Science* 237 (1987), 358–361.
Westoff, C. F. "Sex Preselection in the United States: Some Implications." *Science* 184 (1974), 633–636.
Woodhead, Avril D.; Barnhart, Benjamin, editors. *Biotechnology and the Human Genome: Innovations and Impact.* New York: Plenum Press, 1988.

Expenditures and FY 1993 Presidential Budget for the DOE Human Genome Program



Types of Institutions Conducting DOE-Sponsored Genome Research

8	National laboratories
3	Other federal organizations
41	Academic institutions
10	Private-sector institutions
12	Nonacademic, commercial organizations

Human Genome Program Funds Distribution in FY 1992 (in $K)
(Commitments as of May 1, 1992)

	Project Type					
Organization Type	**Mapping & Sequencing**	**Instrumentation Development**	**Informatics**	**ELSI**	**Totals**	**Percent of 56800[1]**
DOE laboratories	23671	7559	5122	236	36588	64.4
Academic sites	5462	3341	4528	736	14067	24.8
Institutions (nonprofit)	2173	0	602	847	3622	6.4
NIH laboratories	680	0	0	0	680	1.2
Companies and SBIR[2]	1550	0	314	392	2256	3.9
All organizations	33536	10900	10566	2211	57213	
[Percent of 56800]	[59.0]	[19.2]	[18.6]	[3.9]	[100.7][3]	

[1]Total allocation of $59 million less capital equipment funds of $2.2 million.
[2]Small Business Innovation Research grants.
[3]Excess occurs because funding for genome SBIR projects is received from the DOE-wide SBIR program, to which OHER contributes.

From HGPR 1991–92, *p. 62.*

GLOSSARY AND ACRONYMS

Adenine (A): A nitrogenous base, one member of the *base pair* A-T (adenine-*thymine*).

Alleles: Alternative forms of a genetic *locus*; a single allele for each locus is inherited separately from each parent (e.g., at a locus for eye color the allele might result in blue or brown eyes).

Amino acid: Any of a class of 20 molecules that are combined to form *proteins* in living things. The sequence of amino acids in a protein and hence protein function are determined by the *genetic code.*

Amplification: An increase in the number of copies of a specific DNA fragment; can be in vivo or in vitro. See *cloning, polymerase chain reaction.*

Arrayed library: Individual primary recombinant clones (hosted in *phage, cosmid, YAC,* or other *vector)* that are placed in two-dimensional arrays in microtiter dishes. Each primary clone can be identified by the identity of the plate and the clone location (row and column) on that plate. Arrayed libraries of clones can be used for many applications, including screening for a specific *gene* or genomic region of interest as well as for *physical mapping.* Information gathered on individual clones from various genetic *linkage* and *physical map* analyses is entered into a relational database and used to construct physical and genetic *linkage maps* simultaneously; clone identifiers serve to interrelate the multilevel maps. Compare *library, genomic library.*

Autoradiography: A technique that uses X-ray film to visualize radioactively labeled molecules or fragments of molecules; used in analyzing length and number of DNA fragments after they are separated by gel *electrophoresis.*

Autosome: A *chromosome* not involved in sex determination. The *diploid* human *genome* consists of 46 chromosomes, 22 pairs of autosomes, and 1 pair of *sex chromosomes* (the X and Y chromosomes).

Bacteriophage: See *phage.*

Base pair (bp): Two nitrogenous bases (*adenine* and *thymine* or *guanine* and *cytosine*) held together by weak bonds. Two strands of DNA are held together in the shape of a double helix by the bonds between base pairs.

Base sequence: The order of *nucleotide* bases in a DNA molecule.

Base sequence analysis: A method, sometimes automated, for determining the *base sequence.*

Biotechnology: A set of biological techniques developed through basic research and now applied to research and product development. In particular, the use by industry of *recombinant DNA,* cell fusion, and new bioprocessing techniques.

bp: See *base pair.*

cDNA: See *complementary DNA.*

Centimorgan (cM): A unit of measure of *recombination* frequency. One centimorgan is equal to a 1% chance that a marker at one genetic *locus* will be separated from a marker at a second locus due to *crossing over* in a single generation. In human beings, 1 centimorgan is equivalent, on average, to 1 million *base pairs.*

Centromere: A specialized *chromosome* region to which spindle fibers attach during cell division.

Chromosomes: The self-replicating genetic structures of cells containing the cellular DNA that bears in its *nucleotide* sequence the linear array of *genes.* In *prokaryotes,* chromosomal DNA is circular, and the entire genome is carried on one chromosome. *Eukaryotic* genomes consist of a number of chromosomes whose DNA is associated with different kinds of *proteins.*

Clone bank: See *genomic library.*

Clones: A group of cells derived from a single ancestor.

Cloning: The process of asexually producing a group of cells (clones), all genetically identical, from a single ancestor. In *recombinant DNA technology,* the use of DNA manipulation procedures to produce multiple copies of a single *gene* or segment of DNA is referred to as cloning DNA.

Cloning vector: DNA molecule originating from a *virus,* a *plasmid,* or the cell of a higher organism into which another DNA fragment of appropriate size can be integrated without loss of the vector's capacity for self-replication; vectors introduce foreign DNA into host cells, where it can be reproduced in large quantities. Examples are *plasmids, cosmids,* and *yeast artificial chromosomes*; vectors are often *recombinant* molecules containing DNA sequences from several sources.

cM: See *centimorgan.*

Code: See *genetic code.*

Codon: See *genetic code.*

Complementary DNA (cDNA): DNA that is synthesized from a *messenger RNA* template; the single-stranded form is often used as a *probe* in *physical mapping.*

Complementary sequences: *Nucleic acid base sequences* that can form a double-stranded structure by matching *base pairs*; the complementary sequence to G-T-A-C is C-A-T-G.

Conserved sequence: A *base sequence* in a DNA molecule (or an *amino acid* sequence in a *protein*) that has remained essentially unchanged throughout evolution.

Contig map: A map depicting the relative order of a linked *library* of small overlapping clones representing a complete chromosomal segment.

Contigs: Groups of *clones* representing overlapping regions of a *genome.*

Cosmid: Artificially constructed *cloning vector* containing the *cos* gene of *phage* lambda. Cosmids can be packaged in lambda phage particles for infection into *E. coli*; this permits cloning of larger DNA fragments (up to 45 kb) than can be introduced into bacterial hosts in *plasmid* vectors.

Crossing over: The breaking during *meiosis* of one maternal and one paternal *chromosome,* the exchange of corresponding sections of DNA, and the rejoining of the chromosomes. This process can result in an exchange of *alleles* between chromosomes. Compare *recombination.*

Cytosine (C): A *nitrogenous base,* one member of the *base pair* G-C (*guanine* and cytosine).

Deoxyribonucleotide: See *nucleotide.*

Diploid: A full set of genetic material, consisting of paired *chromosomes*—one chromosome from each parental set. Most animal cells except the *gametes* have a diploid set of chromosomes. The diploid human *genome* has 46 chromosomes. Compare *haploid.*

DNA (deoxyribonucleic acid): The molecule that encodes genetic information. DNA is a double-stranded molecule held together by weak bonds between *base pairs* of *nucleotides.* The four nucleotides in DNA contain the bases: *adenine* (A), *guanine* (G), *cytosine* (C), and *thymine* (T). In nature, *base pairs* form only between A and T and between G and C; thus the *base sequence* of each single strand can be deduced from that of its partner.

DNA probes: See *probe.*

DNA replication: The use of existing DNA as a template for the synthesis of new DNA strands. In humans and other *eukaryotes,* replication occurs in the cell *nucleus.*

DNA sequence: The relative order of *base pairs,* whether in a fragment of DNA, a *gene,* a *chromosome,* or an entire *genome.* See *base sequence analysis.*

Domain: A discrete portion of a *protein* with its own function. The combination of domains in a single protein determines its overall function.

Double helix: The shape that two linear strands of DNA assume when bonded together.

***E. coli*:** Common bacterium that has been studied intensively by geneticists because of its small genome size, normal lack of pathogenicity, and ease of growth in the laboratory.

Electrophoresis: A method of separating large molecules (such as DNA fragments or *proteins*) from a mixture of similar molecules. An electric current is passed through a medium containing the mixture, and each kind of molecule travels through the medium at a different rate, depending on its electrical charge and size. Separation is based on these differences. Agarose and acrylamide gels are the media commonly used for electrophoresis of proteins and nucleic acids.

Endonuclease: An *enzyme* that cleaves its nucleic acid substrate at internal sites in the *nucleotide* sequence.

Enzyme: A *protein* that acts as a catalyst, speeding the rate at which a biochemical reaction proceeds but not altering the direction or nature of the reaction.

EST: Expressed sequence tag. See *sequence tagged site.*

Eukaryote: Cell or organism with membrane-bound, structurally discrete *nucleus* and other well-developed subcellular compartments. Eukaryotes include all organisms except *viruses,* bacteria, and blue-green algae. Compare *prokaryote.* See *chromosomes.*

Acronyms for organizations appear on page 38.

Exogenous DNA: DNA originating outside an organism.

Exons: The *protein*-coding DNA sequences of a *gene.* Compare *introns.*

Exonuclease: An *enzyme* that cleaves *nucleotides* sequentially from free ends of a linear nucleic acid substrate.

Expressed gene: See *gene expression.*

FISH (fluorescence in situ hybridization): A *physical mapping* approach that uses fluorescein tags to detect *hybridization* of *probes* with *metaphase chromosomes* and with the less-condensed *somatic interphase* chromatin.

Flow cytometry: Analysis of biological material by detection of the light-absorbing or fluorescing properties of cells or subcellular fractions (i.e., *chromosomes*) passing in a narrow stream through a laser beam. An absorbance or fluorescence profile of the sample is produced. Automated sorting devices, used to fractionate samples, sort successive droplets of the analyzed stream into different fractions depending on the fluorescence emitted by each droplet.

Flow karyotyping: Use of flow cytometry to analyze and/or separate *chromosomes* on the basis of their DNA content.

Gamete: Mature male or female reproductive cell (sperm or ovum) with a *haploid* set of *chromosomes* (23 for humans).

Gene: The fundamental physical and functional unit of heredity. A *gene* is an ordered sequence of *nucleotides* located in a particular position on a particular *chromosome* that encodes a specific functional product (i.e., a *protein* or *RNA* molecule). See *gene expression.*

Gene expression: The process by which a *gene's* coded information is converted into the structures present and operating in the cell. Expressed genes include those that are transcribed into *mRNA* and then translated into *protein* and those that are transcribed into *RNA* but not translated into protein (e.g., *transfer* and *ribosomal RNAs*).

Gene families: Groups of closely related *genes* that make similar products.

Gene library: See *genomic library.*

Gene mapping: Determination of the relative positions of *genes* on a DNA molecule (*chromosome* or *plasmid*) and of the distance, in *linkage* units or physical units, between them.

Gene product: The biochemical material, either *RNA* or *protein*, resulting from expression of a gene. The amount of gene product is used to measure how active a gene is; abnormal amounts can be correlated with disease-causing alleles.

Genetic code: The sequence of *nucleotides*, coded in triplets (*codons*) along the *mRNA*, that determines the sequence of *amino acids* in *protein* synthesis. The DNA sequence of a *gene* can be used to predict the mRNA sequence, and the genetic code can in turn be used to predict the *amino acid* sequence.

Genetic engineering technologies: See *recombinant DNA technologies.*

Genetic map: See *linkage map.*

Genetic material: See *genome.*

Genetics: The study of the patterns of inheritance of specific traits.

Genome: All the genetic material in the *chromosomes* of a particular organism; its size is generally given as its total number of *base pairs.*

Genome projects: Research and technology development efforts aimed at *mapping* and *sequencing* some or all of the *genome* of human beings and other organisms.

Genomic library: A collection of *clones* made from a set of randomly generated overlapping DNA fragments representing the entire *genome* of an organism. Compare *library, arrayed library.*

Guanine (G): A nitrogenous base, one member of the *base pair* G-C (guanine and *cytosine*).

Haploid: A single set of *chromosomes* (half the full set of genetic material), present in the egg and sperm cells of animals and in the egg and pollen cells of plants. Human beings have 23 chromosomes in their reproductive cells. Compare *diploid.*

Heterozygosity: The presence of different *alleles* at one or more *loci* on *homologous chromosomes.*

Homeobox: A short stretch of *nucleotides* whose *base sequence* is virtually identical in all the *genes* that contain it. It has been found in many organisms from fruit flies to human beings. In the fruit fly, a homeobox appears to determine when particular groups of genes are expressed during development.

Homologies: Similarities in DNA or *protein* sequences between individuals of the same species or among different species.

Homologous chromosomes: A pair of *chromosomes* containing the same linear *gene* sequences, each derived from one parent.

Human gene therapy: Insertion of normal DNA directly into cells to correct a genetic defect.

Human Genome Initiative: Collective name for several projects begun in 1986 by DOE to (1) create an ordered set of DNA segments from known chromosomal locations, (2) develop new computational methods for analyzing genetic map and DNA sequence data, and (3) develop new techniques and instruments for detecting and analyzing DNA. This DOE initiative is now known as the Human Genome Program. The national effort, led by DOE and NIH, is known as the Human Genome Project.

Hybridization: The process of joining two *complementary* strands of DNA or one each of DNA and RNA to form a double-stranded molecule.

Informatics: The study of the application of computer and statistical techniques to the management of information. In *genome* projects, informatics includes the development of methods to search databases quickly, to analyze DNA sequence information, and to predict *protein* sequence and structure from DNA sequence data.

In situ hybridization: Use of a DNA or RNA probe to detect the presence of the *complementary DNA* sequence in cloned bacterial or cultured *eukaryotic* cells.

Interphase: The period in the cell cycle when DNA is replicated in the nucleus; followed by *mitosis.*

Introns: The DNA *base sequences* interrupting the *protein*-coding sequences of a *gene*; these sequences are *transcribed* into *RNA* but are cut out of the message before it is *translated* into protein. Compare *exons.*

In vitro: Outside a living organism.

Karyotype: A photomicrograph of an individual's *chromosomes* arranged in a standard format showing the number, size, and shape of each chromosome type; used in low-resolution *physical mapping* to correlate gross chromosomal abnormalities with the characteristics of specific diseases.

kb: See *kilobase.*

Kilobase (kb): Unit of length for DNA fragments equal to 1000 *nucleotides.*

Library: An unordered collection of *clones* (i.e., cloned DNA from a particular organism), whose relationship to each other can be established by *physical mapping.* Compare *genomic library, arrayed library.*

Linkage: The proximity of two or more *markers* (e.g., *genes, RFLP* markers) on a *chromosome*; the closer together the markers are, the lower the probability that they will be separated during DNA repair or replication processes (binary fission in *prokaryotes, mitosis* or *meiosis* in *eukaryotes*), and hence the greater the probability that they will be inherited together.

Linkage map: A map of the relative positions of genetic *loci* on a *chromosome*, determined on the basis of how often the loci are inherited together. Distance is measured in *centimorgans* (*cM*).

Localize: Determination of the original position (*locus*) of a *gene* or other *marker* on a chromosome.

Locus (pl. loci): The position on a *chromosome* of a *gene* or other chromosome *marker*; also, the DNA at that position. The use of *locus* is sometimes restricted to mean regions of DNA that are *expressed.* See *gene expression.*

Macrorestriction map: Map depicting the order of and distance between sites at which *restriction enzymes* cleave *chromosomes.*

Mapping: See *gene mapping, linkage map, physical map.*

Marker: An identifiable physical location on a *chromosome* (e.g., *restriction enzyme cutting site, gene*) whose inheritance can be monitored. Markers can be expressed regions of DNA (genes) or some segment of DNA with no known coding function but whose pattern of inheritance can be determined. See *RFLP, restriction fragment length polymorphism.*

Mb: See *megabase.*

Megabase (Mb): Unit of length for DNA fragments equal to 1 million *nucleotides* and roughly equal to 1 *cM.*

Meiosis: The process of two consecutive cell divisions in the *diploid* progenitors of sex cells. Meiosis results in four rather than two daughter cells, each with a *haploid* set of *chromosomes.*

Messenger RNA (mRNA): RNA that serves as a template for *protein* synthesis. See *genetic code.*

Metaphase: A stage in *mitosis* or *meiosis* during which the *chromosomes* are aligned along the equatorial plane of the cell.

Mitosis: The process of nuclear division in cells that produces daughter cells that are genetically identical to each other and to the parent cell.

mRNA: See *messenger RNA.*

Multifactorial or multigenic disorders: See *polygenic disorders.*

Multiplexing: A *sequencing* approach that uses several pooled samples simultaneously, greatly increasing sequencing speed.

Mutation: Any heritable change in DNA *sequence*. Compare *polymorphism.*

Nitrogenous base: A nitrogen-containing molecule having the chemical properties of a base.

Nucleic acid: A large molecule composed of *nucleotide* subunits.

Nucleotide: A subunit of DNA or *RNA* consisting of a nitrogenous base (*adenine, guanine, thymine*, or *cytosine* in DNA; adenine, guanine, *uracil*, or cytosine in RNA), a phosphate molecule, and a sugar molecule (deoxyribose in DNA and ribose in RNA). Thousands of *nucleotides* are linked to form a DNA or RNA molecule. See *DNA, base pair, RNA.*

Nucleus: The cellular organelle in *eukaryotes* that contains the genetic material.

Oncogene: A *gene*, one or more forms of which is associated with cancer. Many oncogenes are involved, directly or indirectly, in controlling the rate of cell growth.

Overlapping clones: See *genomic library.*

PCR: See *polymerase chain reaction.*

Phage: A *virus* for which the natural host is a bacterial cell.

Physical map: A map of the locations of identifiable landmarks on DNA (e.g., *restriction enzyme cutting sites, genes*), regardless of inheritance. Distance is measured in *base pairs*. For the human *genome*, the lowest-resolution *physical map* is the banding patterns on the 24 different *chromosomes*; the highest-resolution map would be the complete *nucleotide* sequence of the chromosomes.

Plasmid: Autonomously replicating, extrachromosomal circular DNA molecules, distinct from the normal bacterial *genome* and nonessential for cell survival under nonselective conditions. Some plasmids are capable of integrating into the host genome. A number of artificially constructed plasmids are used as *cloning vectors.*

Polygenic disorders: Genetic disorders resulting from the combined action of *alleles* of more than one *gene* (e.g., heart disease, diabetes, and some cancers). Although such disorders are inherited, they depend on the simultaneous presence of several alleles; thus the hereditary patterns are usually more complex than those of single-gene disorders. Compare *single-gene disorders.*

Polymerase chain reaction (PCR): A method for amplifying a DNA *base sequence* using a heat-stable *polymerase* and two 20-base *primers*, one *complementary* to the (+)-strand at one end of the sequence to be amplified and the other complementary to the (–)-strand at the other end. Because the newly synthesized DNA strands can subsequently serve as additional templates for the same primer sequences, successive rounds of primer annealing, strand elongation, and dissociation produce rapid and highly specific amplification of the desired sequence. PCR also can be used to detect the existence of the defined sequence in a DNA sample.

Polymerase, DNA or RNA: *Enzymes* that catalyze the synthesis of *nucleic acids* on preexisting nucleic acid templates, assembling RNA from ribonucleotides or DNA from deoxyribonucleotides.

Polymorphism: Difference in DNA sequence among individuals. Genetic variations occurring in more than 1% of a population would be considered useful polymorphisms for genetic *linkage* analysis. Compare *mutation.*

Primer: Short preexisting polynucleotide chain to which new deoxyribonucleotides can be added by DNA *polymerase.*

Probe: Single-stranded DNA or *RNA* molecules of specific base *sequence*, labeled either radioactively or immunologically, that are used to detect the *complementary* base sequence by *hybridization.*

Prokaryote: Cell or organism lacking a membrane-bound, structurally discrete *nucleus* and other subcellular compartments. Bacteria are prokaryotes. Compare *eukaryote*. See *chromosomes.*

Promoter: A site on DNA to which *RNA polymerase* will bind and initiate *transcription.*

Protein: A large molecule composed of one or more chains of *amino acids* in a specific order; the order is determined by the *base sequence* of *nucleotides* in the *gene* coding for the protein. Proteins are required for the structure, function, and regulation of the body's cells, tissues, and organs, and each protein has unique functions. Examples are hormones, *enzymes*, and antibodies.

Purine: A nitrogen-containing, single-ring, basic compound that occurs in nucleic acids. The purines in DNA and RNA are adenine and guanine.

Pyrimidine: A nitrogen-containing, double-ring, basic compound that occurs in nucleic acids. The pyrimidines in DNA are cytosine and thymine; in RNA, cytosine and uracil.

Rare-cutter enzyme: See *restriction enzyme cutting site.*

Recombinant clones: *Clones* containing *recombinant DNA molecules.* See *recombinant DNA technologies.*

Recombinant DNA molecules: A combination of DNA molecules of different origin that are joined using *recombinant DNA technologies.*

Recombinant DNA technologies: Procedures used to join together DNA segments in a cell-free system (an environment outside a cell or organism). Under appropriate conditions, a recombinant DNA molecule can enter a cell and replicate there, either autonomously or after it has become integrated into a cellular *chromosome.*

Recombination: The process by which progeny derive a combination of *genes* different from that of either parent. In higher organisms, this can occur by *crossing over.*

Regulatory regions or sequences: A DNA *base sequence* that controls *gene expression.*

Resolution: Degree of molecular detail on a *physical map* of DNA, ranging from low to high.

Restriction enzyme, endonuclease: A *protein* that recognizes specific, short *nucleotide sequences* and cuts DNA at those sites. Bacteria contain over 400 such *enzymes* that recognize and cut over 100 different DNA sequences. See *restriction enzyme cutting site.*

Restriction enzyme cutting site: A specific *nucleotide sequence* of DNA at which a particular *restriction enzyme* cuts the DNA. Some sites occur frequently in DNA (e.g., every several hundred *base pairs*), others much less frequently (*rare-cutter*, e.g., every 10,000 base pairs).

Restriction fragment length polymorphism (RFLP): Variation between individuals in DNA fragment sizes cut by specific *restriction enzymes*; *polymorphic sequences* that result in RFLPs are used as *markers* on both *physical maps* and genetic *linkage maps.* RFLPs are usually caused by *mutation* at a cutting site. See *marker.*

RFLP: See *restriction fragment length polymorphism.*

Ribonucleic acid (RNA): A chemical found in the *nucleus* and cytoplasm of cells; it plays an important role in *protein* synthesis and other chemical activities of the cell. The structure of RNA is similar to that of DNA. There are several classes of RNA molecules, including *messenger RNA, transfer RNA, ribosomal RNA*, and other small RNAs, each serving a different purpose.

Ribonucleotides: See *nucleotide.*

Ribosomal RNA (rRNA): A class of RNA found in the ribosomes of cells.

Ribosomes: Small cellular components composed of specialized ribosomal RNA and protein; site of protein synthesis. See *ribonucleic acid (RNA).*

RNA: See *ribonucleic acid.*

Sequence: See *base sequence.*

Sequence tagged site (STS): Short (200 to 500 *base pairs*) DNA sequence that has a single occurrence in the human *genome* and whose location and base sequence are known. Detectable by *polymerase chain reaction*, STSs are useful for localizing and orienting the mapping and sequence data reported from many different laboratories and serve as landmarks on the developing *physical map* of the human genome. Expressed sequence tags (ESTs) are STSs derived from cDNAs.

Sequencing: Determination of the order of *nucleotides* (*base sequences*) in a DNA or *RNA* molecule or the order of *amino acids* in a *protein.*

Sex chromosomes: The X and Y *chromosomes* in human beings that determine the sex of an individual. Females have two X chromosomes in diploid cells; males have an X and a Y chromosome. The sex chromosomes comprise the 23rd chromosome pair in a *karyotype*. Compare *autosome.*

Shotgun method: *Cloning* of DNA fragments randomly generated from a *genome*. See *library, genomic library.*

Single-gene disorder: Hereditary disorder caused by a *mutant* allele of a single *gene* (e.g., Duchenne muscular dystrophy, retinoblastoma, sickle cell disease). Compare *polygenic disorders.*

Somatic cells: Any cell in the body except *gametes* and their precursors.

Southern blotting: Transfer by absorption of DNA fragments separated in electrophoretic gels to membrane filters for detection of specific *base sequences* by radiolabeled complementary probes.

STS: See *sequence tagged site.*

Tandem repeat sequences: Multiple copies of the same *base sequence* on a *chromosome*; used as a marker in *physical mapping.*

Technology transfer: The process of converting scientific findings from research laboratories into useful products by the commercial sector.

Telomere: The ends of *chromosomes.* These specialized structures are involved in the replication and stability of linear DNA molecules. See *DNA replication.*

Thymine (T): A nitrogenous base, one member of the *base pair* A-T (*adenine*-thymine).

Transcription: The synthesis of an *RNA* copy from a *sequence* of DNA (a *gene*); the first step in *gene expression.* Compare *translation.*

Transfer RNA (tRNA): A class of *RNA* having structures with triplet *nucleotide* sequences that are *complementary* to the triplet nucleotide coding sequences of *mRNA.* The role of tRNAs in protein synthesis is to bond with *amino acids* and transfer them to the ribosomes, where proteins are assembled according to the genetic code carried by mRNA.

Transformation: A process by which the genetic material carried by an individual cell is altered by incorporation of exogenous DNA into its *genome.*

Translation: The process in which the genetic code carried by mRNA directs the synthesis of *proteins* from *amino acids.* Compare *transcription.*

tRNA: See *transfer RNA.*

Uracil: A nitrogenous base normally found in RNA but not DNA; uracil is capable of forming a *base pair* with *adenine.*

Vector: See *cloning vector.*

Virus: A noncellular biological entity that can reproduce only within a host cell. Viruses consist of *nucleic acid* covered by *protein*; some animal viruses are also surrounded by membrane. Inside the infected cell, the virus uses the synthetic capability of the host to produce progeny virus.

VLSI: Very large-scale integration allowing over 100,000 transistors on a chip.

YAC: See *yeast artificial chromosome.*

Yeast artificial chromosome (YAC): A vector used to clone DNA fragments (up to 400 kb); it is constructed from the telomeric, centromeric, and replication origin sequences needed for replication in yeast cells. Compare *cloning vector, cosmid.*

AEC Atomic Energy Commission
ANL* Argonne National Laboratory, Argonne, IL
ATCC American *Type Culture* Collection, Rockville, MD
BNL* Brookhaven National Laboratory, Upton, NY
CEPH Centre d'Etude du Polymorphisme Humain
CRADA Cooperative Research and Development Agreement
DKFZ German Cancer Research Center
DOE Department of Energy
ERDA Energy Research and Development Administration
FCCSET Federal Coordinating Council on Science, Engineering, and Technology
GDB*† Genome Data Base
HERAC* Health and Environmental Research Advisory Committee
HGCC* Human Genome Coordinating Committee
HGMIS* Human Genome Management Information System (ORNL)
HUGO Human Genome Organization (international)
JHU Johns Hopkins University
JITF*† Joint Informatics Task Force
LANL* Los Alamos National Laboratory, Los Alamos, NM
LBL* Lawrence Berkeley Laboratory, Berkeley, CA
LLNL* Lawrence Livermore National Laboratory, Livermore, CA
MRC Medical Research Council (U.K.)
NAS National Academy of Sciences (U.S.)
NCHGR† National Center for Human Genome Research
NIH† National Institutes of Health, Bethesda, MD
NLGLP* National Laboratory Gene Library Project (LANL, LLNL)
NRC National Research Council (NAS)
NSF National Science Foundation
OHER* Office of Health and Environmental Research
ORNL* Oak Ridge National Laboratory, Oak Ridge, TN
OSTP Office of Scientific and Technology Policy (White House)
OTA Office of Technology Assessment (U.S. Congress)
PACHG† Program Advisory Committee on the Human Genome
PNL* Pacific Northwest Laboratory, Richland, WA
SBIR Small Business Innovation Research
SCC Scientific Coordinating Committee
TWAS Third World Academy of Sciences
UNESCO United Nations Educational, Scientific, and Cultural Organization
USDA U.S. Department of Agriculture

*Denotes U.S. Department of Energy organizations.
†Denotes U.S. Department of Health and Human Services organizations.

The glossary is reproduced from HGPR 1991–92, pp. 230–241; the acronyms are from the inside back cover.

A HUMAN GENOME PROJECT TIMELINE

1940–41

George W. Beadle and **Edward L. Tatum** (Stanford), on discovering that genes are associated with biochemical defects in *Neurospora*, propose the "one gene, one enzyme" hypothesis.

1944

Oswald T. Avery, Colin MacLeod, and **Maclyn J. McCarty** (Rockefeller Institute) find that DNA is the transforming principle in *Pneumococcus.*

1951–52

Norton Zinder and **Joshua Lederberg** (Wisconsin) discover transduction in *Salmonella:* viruses mediate genetic exchange between other organisms.

1953

James D. Watson and **Francis H. C. Crick** (Cambridge) propose the double helix structure for DNA.

1956

Joe-Hin Tijo and **Albert Levan** (Lund) discover the correct number of chromosomes in humans.

1957–58

Matthew Meselson and **Franklin W. Stahl** (Caltech), using a density gradient experiment, show semiconservative replication of DNA, as the Watson-Crick model predicts.

1959

Jérôme Lejeune (Paris) and two Medical Research Council (U.K.) teams—with **Charles E. Ford** and **Patricia A. Jacobs**—establish that an extra chromosome causes Down syndrome: the first *trisomy.*

MRC teams also map male characteristics to the Y chromosome, on the basis of sex chromosome anomalies (XO, XXY).

1960

The Denver conference for standardization of cytogenetic nomenclature convenes; numbers autosomes and divides them into 7 groups by size.

1962

Victor McKusick (Johns Hopkins) produces a catalogue of X-linked traits, "On the X Chromosome of Man."

1964

Michael Lesch and **William Nyhan** (Johns Hopkins) describe the Lesch-Nyhan syndrome.

1965

The genetic code is completely elucidated at the Cold Spring Harbor Symposium (see figure).

1966

Chicago conference adopts *p* for short arms, *q* for long arm, of chromosome bands.

Victor McKusick publishes *Mendelian Inheritance in Man.*

1966–67

Walter Gilbert and **Benno Muller-Hill** (Harvard) isolate the *lac* repressor and prove that the operator is DNA.

1967

Mary Weiss and **Howard Green** (NYU) successfully culture mouse-human cells: somatic-cell hybridization.

The genetic code. *From OTA,* Mapping Our Genes, *p. 23.*

Codon	Amino Acid	Codon	Amino Acid	Codon	Amino Acid	Codon	Amino Acid
UUU	Phenylalanine	UCU	Serine	UAU	Tyrosine	UGU	Cysteine
UUC	Phenylalanine	UCC	Serine	UAC	Tyrosine	UGC	Cysteine
UUA	Leucine	UCA	Serine	UAA	stop	UGA	stop
UUG	Leucine	UCG	Serine	UAG	stop	UGG	Tryptophan
CUU	Leucine	CCU	Proline	CAU	Histidine	CGU	Arginine
CUC	Leucine	CCC	Proline	CAC	Histidine	CGC	Arginine
CUA	Leucine	CCA	Proline	CAA	Glutamine	CGA	Arginine
CUG	Leucine	CCG	Proline	CAG	Glutamine	CGG	Arginine
AUU	Isoleucine	ACU	Threonine	AAU	Asparagine	AGU	Serine
AUC	Isoleucine	ACC	Threonine	AAC	Asparagine	AGC	Serine
AUA	Isoleucine	ACA	Threonine	AAA	Lysine	AGA	Arginine
AUG	Methionine (start)	ACG	Threonine	AAG	Lysine	AGG	Arginine
GUU	Valine	GCU	Valine	GAU	Aspartic acid	GGU	Glycine
GUC	Valine	GCC	Alanine	GAC	Aspartic acid	GGC	Glycine
GUA	Valine	GCA	Alanine	GAA	Glutamic acid	GGA	Glycine
GUG	Valine	GCG	Alanine	GAG	Glutamic acid	GGG	Glycine

Each codon, or triplet of nucleotides in RNA, codes for an amino acid. Twenty different amino acids are produced from a total of 64 different RNA codons, but some amino acids are specified by more than one codon (e.g., phenylalanine is specified by UUU and by UUC). In addition, one codon (AUG) specifies the start of a protein, and three codons (UAA, UAG, and UGA) specify the termination of a protein. Mutations in the nucleotide sequence can change the resulting protein structure if the mutation alters the amino acid specified by a codon or if it alters the reading frame by deleting or adding a nucleotide.

U=uracil (thymine) A=adenine
C=cytosine G=guanine

SOURCES: Office of Technology Assessment and National Institute of General Medical Sciences, 1988.

1969

Torbjörn Caspersson and colleagues (Stockholm) succeed in staining ordinary chromosomes so that the bands show, using quinacrine fluorescence (banding).

Herbert A. Lubs (Yale) first describes "marker X," later called *fragile X*.

1970

Mary Lou Pardue and **Joseph Gall** (Yale) develop *in situ* hybridization using radioactive DNA probes.

1971

The Paris conference on standardization of nomenclature in human cytogenetics adopts numbers for chromosome bands.

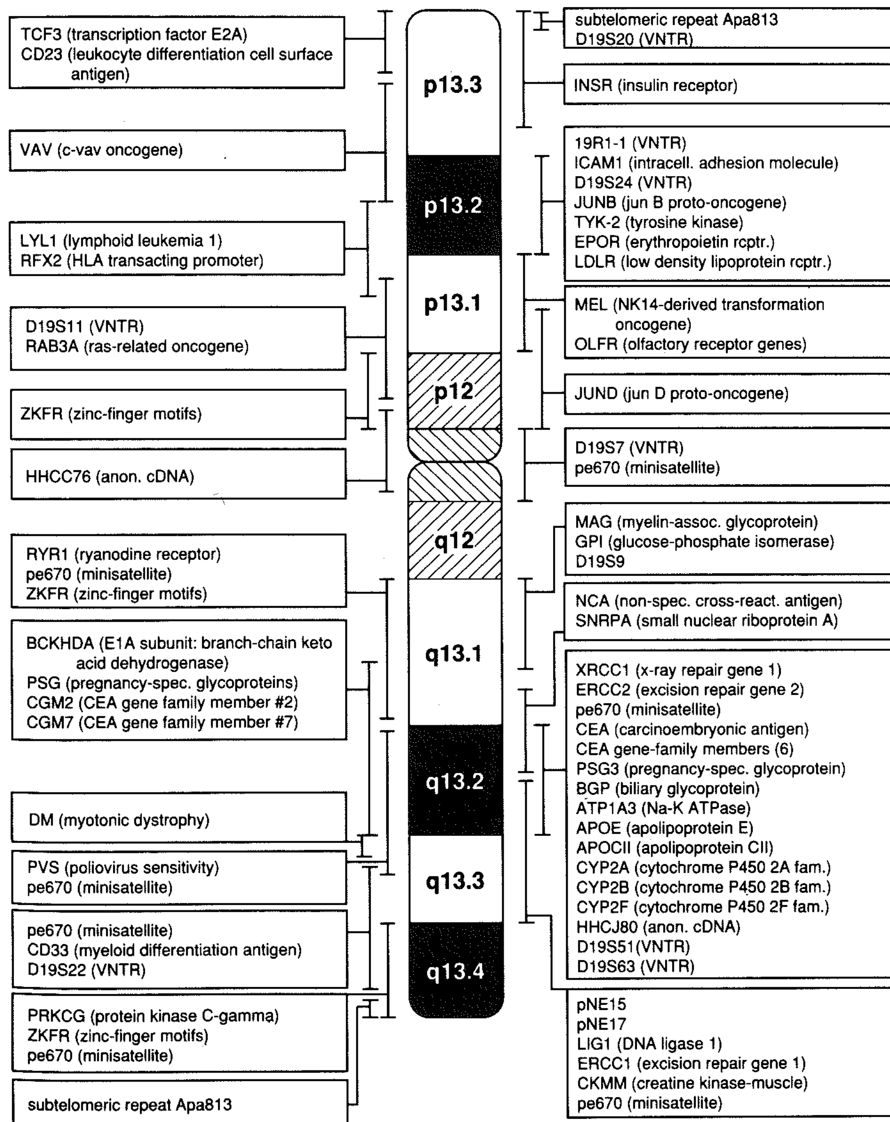
1971–72

David A. Jackson, Robert H. Symons, and **Paul Berg** (Stanford) use restriction enzymes to make recombinant DNA.

1973

The First Human Gene Mapping Workshop convenes at New Haven.

Known genes and genetic markers on chromosome 19. Note the use of p *for the short arm,* q *for the long, and the numbering of the different-colored bands. From* HGPR 1991–92, *p. 43 (figure by Barbara Trask, Lawrence Livermore National Laboratory).*



1975

Edwin M. Southern (Edinburgh) develops the Southern blot technique, for identifying DNA fragments from restriction enzyme digests.

Asilomar conference convened to draft guidelines for research on recombinant DNA.

1976

Each human chromosome has at least one gene assigned to it.

1975–77

Frederick Sanger (Cambridge) and **Allan M. Maxam** and **Walter Gilbert** (Harvard) develop basic DNA sequencing techniques.

1977

Susan M. Berget, Phillip A. Sharp, and colleagues (MIT) and **Louise T. Chow, Richard E. Gelinas, Thomas R. Broker,** and **Richard J. Roberts** (Cold Spring Harbor) discover introns and RNA splicing.

1979

Charles Verellen-Dumoulin (Louvain) locates Duchenne muscular dystrophy on band p21 of chromosome X by studying cases with X-autosome translocation.

1980

David Botstein and colleagues (Stanford) publish prospectus for genetic mapping using restriction fragment length polymorphisms (RFLPs).

1981

Mary Harper and **Grady F. Saunders** (Houston) modify *in situ* hybridization to locate infrequent markers.

3 1151 00940 2623
THE EISENHOWER LIBRARY

1983

James F. Gusella, Nancy Wexler, Michael Conneally, and colleagues (Harvard, Hereditary Disease Foundation) assign Huntington's disease gene to the short end of chromosome 4, by linkage to RFLP.

1984

Alta Conference discusses large-scale effort to map human genome.

1985

Robert Sinsheimer (UC Santa Cruz) organizes a workshop on human genome project.

1986

Charles DeLisi (DOE) launches international workshop in Santa Fe, organized by **Mark Bitensky** (Los Alamos). Further meetings sponsored by Cold Spring Harbor, Howard Hughes Medical Institute, and NIH to sketch plans for HGP.

OTA and NAS study proposals for a human genome project.

Anthony P. Monaco, Louis Kunkel, and colleagues (Harvard) use "reverse genetics" to identify basic defect from map location in Becker and Duchenne muscular dystrophy.

Victor McKusick reports that about 950 genes have been assigned to specific chromosomes in man.

1987

Applied Biosystems and Du Pont put first automated sequencers on market.

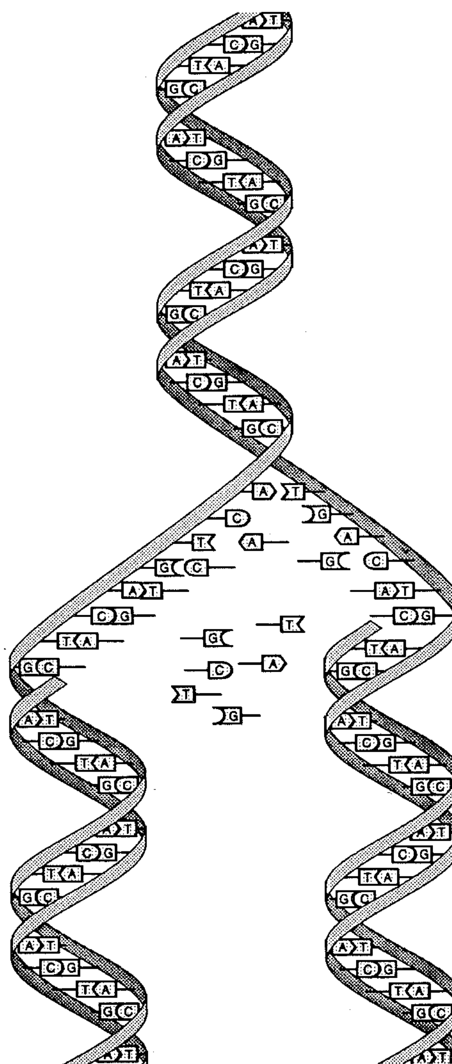
Helen Donis-Keller and colleagues (Collaborative Research) produce the first genetic linkage map.

1988

Human Genome Project is funded.

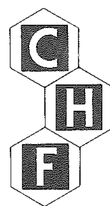
Kary Mullis (Cetus Corp.) develops polymerase chain reaction (PCR) for DNA amplification.

Replication of DNA. *From OTA,* Mapping Our Genes, *p. 22.*



When DNA replicates, the original strands unwind and serve as templates for the building of new, complementary strands. The daughter molecules are exact copies of the parent, each daughter having one of the parent strands.
SOURCE: Office of Technology Assessment, 1988.

The timeline is based in part on Victor A. McKusick, Thomas H. Roderick, Joe Mori, and Natalie W. Paul, eds., *Medical and Experimental Mammalian Genetics: A Perspective* (New York: Alan R. Liss, 1987), Table 1.



CHEMICAL
HERITAGE
FOUNDATION

The Chemical Heritage Foundation (CHF) was established by joint action of the American Chemical Society and the American Institute of Chemical Engineers. The purpose of CHF is to advance the heritage of the chemical sciences. Affiliated organizations include Alpha Chi Sigma, The American Association for Clinical Chemistry, Inc., the American Association of Textile Chemists and Colorists, the American Institute of Chemists, Inc., the American Society for Biochemistry and Molecular Biology, the American Society for Mass Spectrometry, The Chemists' Club, the Electrochemical Society, and the Société de Chimie Industrielle.

CHF seeks to put our chemical heritage to work by discovering and disseminating information about historical resources; encouraging research, scholarship, and popular writing; publishing resource guides and historical materials; conducting oral histories; creating traveling exhibits; and taking other appropriate steps to make known the achievements of chemical scientists and the chemical process industries.

CHF operates through public outreach programs and through the Arnold and Mabel Beckman Center for the History of Chemistry and the Donald F. and Mildred Topp Othmer Library of Chemical History.